

BARCELONA
METS
POLIS

N. 127-JUL. 23

Les noves eines d'intel·ligència artificial estan colonitzant totes les activitats, sense una regulació que permeti optimitzar-ne els beneficis i minimitzar-ne els riscos.

Impactes de la intel·ligència artificial



Controlar la intel·ligència artificial

Els avenços en intel·ligència artificial (IA) desperten tanta expectació com recels. La irrupció del ChatGPT mostra que aquesta tecnologia avança

més ràpidament que la capacitat de la societat per entomar-ne les conseqüències. La dificultat de comprendre'n l'abast i la manca de control sobre les seves aplicacions engendren tota mena de temors, fins al punt que més d'un miler de científics han demanat una moratòria per debatre'n la regulació abans no sigui massa tard. Veus molt reconegudes han advertit sobre el perill que, algun dia, la IA sigui tan poderosa que pugui sotmetre la intel·ligència humana, en la línia del que va dir Claude Shannon, precursor de la teoria de la informació: "Visualitzo un moment en què serem per als robots allò que els gossos són ara per als humans".

En realitat, la IA està lluny de poder superar la intel·ligència humana corpòria, entre altres coses perquè no té la capacitat d'interactuar amb l'entorn, que és la manera com el cervell humà és capaç de crear un tipus de coneixement que cap màquina ha pogut imitar. Una cosa és guanyar als escacs i una altra de molt diferent és respondre davant d'una situació inesperada, com ho fa l'engranatge emocional i racional del nostre cervell. Però mentre ens distraiem imaginant futurs distòpics, deixem sense control les aplicacions que ja existeixen. Més que el que pot fer la intel·ligència artificial en si mateixa, ens ha de preocupar el que alguna gent pot arribar a fer-ne, de la intel·ligència artificial.

Com passa amb totes les tecnologies fortament disruptives, el potencial de la IA és enorme. És present en moltes de les eines que fem servir, des dels navegadors fins a les aplicacions d'internet; està aportant avenços en sectors com la medicina, i, també, pot ajudar les ciutats a gestionar processos complexos com ara la mobilitat o la logística. Però, de la mateixa manera que pot ser molt útil per processar una gran quantitat de dades i simplificar tasques, també pot servir per crear sistemes de control i vigilància de la ciutadania. Ens hem de plantejar, per tant, amb quines regles volem que la IA operi. No pot ser que una eina tan poderosa estigui, a la pràctica, en mans d'unes poques corporacions privades que operen sense cap mena de control. — Milagros Pérez Oliva

Dossier

Impactes de la intel·ligència artificial

4-9

Ramon López de Mántaras

Mites i realitats de la intel·ligència artificial

10-14

Pablo Martínez i Mar Santamaria

Que la ciutat ens incomodi

16-21

Pau Garcia

La ignorància dels algoritmes i les noves realitats sintètiques

22-25

José Luis de Vicente

Quatre vectors cap a una cultura de la IA

26-30

Joan Manuel del Pozo

Tecnologia i humanitats: un diàleg necessari

32-36

Beatriz López

Els robots i els sistemes intel·ligents revolucionen la medicina

38-42

Filippo Bistaffa, Pedro Meseguer i Juan Antonio Rodríguez Aguilar

Eines d'IA per a una mobilitat compartida i sostenible

44-48

Nuria Oliver

Intel·ligència artificial ètica, responsable i per al bé social





Mites i realitats de la intel·ligència artificial

Ramon López de Mántaras

Les intel·ligències artificials desenvolupades fins ara són específiques per a tasques molt concretes, però no tenen capacitat per entendre

què fan ni poden interactuar amb l'entorn o generar certs tipus de coneixement, com és el sentit comú humà, fruit de les experiències viscudes. Mentre especulem si la IA substituirà la intel·ligència humana, ens distraiem dels veritables problemes: qui controla les seves aplicacions i per a què es fan servir.

Arthur Clarke, als anys seixanta, va dir: "Qualsevol tecnologia prou sofisticada és indistingible de la màgia". La intel·ligència artificial (IA) n'és un clar exemple. Sembla que no hi hagi límits en les seves capacitats, però, en realitat, la IA roman encallada des de fa més de cinquanta anys en una de les qüestions més fonamentals: com dotar les màquines de coneixements de sentit comú.

Els coneixements de sentit comú són fruit de les nostres vivències i experiències. Hi ha milions de coneixements de sentit comú que les persones fem servir contínuament i que ens permeten entendre el món en què vivim. Malgrat que alguns afirmen que el sentit comú ja està "emergint" en els grans models de llenguatge recents de la IA generativa, de fet és impossible assegurar-ho, ja que la complexitat i opacitat d'aquests models fa pràcticament impossible analitzar com funcionen internament. Només sabem que aprenen



© Curro Palacios Taberner

Ramon López de Mántaras

És professor emèrit d'Investigació del Consell Superior d'Investigacions Científiques (CSIC) a l'Institut d'Investigació en Intel·ligència Artificial (IIIA), del qual fou fundador i director. Ha estat professor visitant de la University of Technology Sydney i de la Western Sydney University. Ha rebut el Premi Ciutat de Barcelona d'Investigació (1982), el Robert S. Engelmore Memorial Award de l'Associació Americana d'Intel·ligència Artificial (2011), el Distinguished Award de l'Associació Europea d'Intel·ligència Artificial (EurAI, 2016) i el Premio Nacional de Investigación Julio Rey Pastor (2018), entre d'altres. És membre d'honor de l'EurAI, membre numerari de l'Institut d'Estudis Catalans i de diversos comitès assessors.

Notes

1. Jumper, J., Evans, R., Pritzel, A. *et al.* "Highly accurate protein structure prediction with AlphaFold". *Nature*, 596, 583-589. 2021.
<http://ow.ly/XiVW50OnOjm>

patrons lingüístics per tal de predir quina és la següent paraula en un text. Segons la meua opinió, no crec que aquesta tècnica permeti dotar de sentit comú la IA per molt que escalem la complexitat (en termes de nombre de paràmetres) d'aquests models de llenguatge. El problema de dotar de sentit comú les màquines és massa complicat per creure que es pot resoldre a partir de calcular quina és la següent paraula més probable. De fet, l'Agència de Projectes de Recerca Avançats de Defensa (DARPA), la institució que més inverteix en programes de recerca als Estats Units, a finals del 2018, va endegar un programa plurianual de llarg termini que finança, amb milers de dòlars, projectes de recerca sobre com dotar les màquines de coneixements de sentit comú.

Una línia de recerca molt interessant per dotar les màquines de coneixements per entendre el món és la que té com a objectiu la modelització matemàtica i l'aprenentatge de relacions causa-efecte. Els sistemes actuals basats en aprenentatge profund simplement poden aprendre funcions matemàtiques simètriques, en particular correlacions, però no poden aprendre relacions asimètriques i, per tant, no són capaços de diferenciar entre causes i efectes, com ara que la sortida del sol és la causa del cant del gall i no a l'inrevés.

Situació real de la IA

Quina és, doncs, la situació real de la IA? La realitat és que tenim intel·ligències artificials específiques, en el sentit que cadascuna sap fer bé una única tasca, tot i no entendre realment res del que fan. Focalitzem-nos en una tècnica d'IA, coneguda com a *aprenentatge profund*, que, durant els darrers deu anys, ha aconseguit uns resultats espectaculars d'acord amb la seva extraordinària capacitat per detectar patrons estadístics analitzant grans quantitats de dades. Segons aquesta tècnica, hi ha un programari anomenat AlphaFold¹ que és capaç de predir l'estructura

3D de les proteïnes, és a dir, com es plega una proteïna a l'espai tridimensional. Resoldre el problema del plegament de proteïnes és important per descobrir i dissenyar nous fàrmacs, perquè la funció de les proteïnes depèn de la seva estructura tridimensional. Un altre

Els sistemes actuals basats en aprenentatge profund no són capaços de diferenciar entre causes i efectes, com ara que la sortida del sol és la causa del cant del gall i no a l'inrevés.

exemple molt recent d'èxit de l'aprenentatge profund, que combina xarxes neuronals i aprenentatge per reforç, és la ja esmentada IA generativa, que permet processar llenguatge natural i també música, imatges o vídeos a partir d'una descripció inicial. Malgrat la seva aparent intel·ligència, aquests sistemes són molt limitats. De fet, no aprenen res en el sentit humà que entenem per aprendre, ja que, contràriament a nosaltres, no aprenen de manera incremental, no poden relacionar el que aprenen novament amb el que havien après anteriorment i gairebé no tenen capacitat de generalitzar ni de raonar.

Notes

2. Dennet, D. C. *From Bacteria to Bach and Back: The Evolution of Minds*. Penguin Books, 2017.

3. Bender, E. M., Gebru, T., McMillan-Major, A. i Shmitchell, S. "On the Danger of Stochastic Parrots: Can Language Models Be Too Big?". ACM Conference on Fairness, Accountability, and Transparency, 610-623. 2021.

Quin és, doncs, el motiu pel qual molts creuen que la IA està a punt d'igualar la intel·ligència humana i, a partir d'aquesta premissa, fan prediccions sobre hipotètiques intel·ligències artificials superiors a la intel·ligència humana? Al meu entendre, l'excessiu antropomorfisme és el motiu principal que fa que la societat tingui una percepció errònia de l'estat real de la intel·ligència artificial. Quan ens informen d'èxits espectaculars d'una IA que resol una tasca molt concreta, tendim a generalitzar i li atribuïm la capacitat de fer pràcticament qualsevol cosa que fem els éssers humans i, fins i tot, de fer-la molt millor. En realitat, el que tenen els actuals sistemes d'IA no és intel·ligència sinó "habilitats sense comprensió" en el sentit que apunta Daniel Dennett en el seu llibre *From Bacteria to Bach and Back: The Evolution of Minds*². És a dir, sistemes que poden arribar a ser molt hàbils duent a terme tasques específiques, com reconèixer un seguit d'elements en una imatge o generar text, però sense comprendre absolutament res sobre la naturalesa d'aquests elements ni de les propietats i relacions entre ells.

El problema és que els sistemes actuals d'IA no tenen accés als referents del món real que donen contingut a les paraules perquè, en ser intel·ligències artificials no corpòries, no poden interactuar amb el món.

L'excessiu antropomorfisme és el motiu principal que fa que la societat tingui una percepció errònia de l'estat real de la intel·ligència artificial.

Aquesta incapacitat per entendre realment el món impossibilita que grans models de llenguatge com GPT-4 entenguin el significat dels textos que generen. De fet, als dissenyadors d'aquests sistemes l'absència de referents del món real no els importa. Només els importa el

poder retòric, enganyant els oients i els lectors fent-los creure que entenen el llenguatge com els humans, encara que generin multitud de falsedats.

L'any 2021, les investigadores Timnit Gebru i Margaret Mitchell, aleshores col·laboradors de l'equip d'ètica de Google, ja van advertir del risc que les persones imputéssim intenció comunicativa i comprensió del llenguatge a artefactes que semblen humans³. Irònicament, a causa d'haver fet públic aquest important aspecte ètic, Google les va acomiadar. Emily M. Bender, experta en lingüística computacional, encertadament afirma que crear tecnologia que imiti els humans fent-se passar per quelcom d'humà requereix que tinguem molt clar què vol dir ser humà, ja que, en cas contrari, correm el risc de deshumanitzar-nos.

Altres, com Daniel Dennett, són encara més contundents. Dennett afirma que no podem viure en un món amb *counterfeit people* ('persones falsificades'), ja que una societat amb persones falses que no podem diferenciar de les reals deixaria de ser una societat. Ens recorda que fabricar diners falsificats és un acte delictiu i afirma que falsificar persones és tant o més greu. A més, a les persones artificials no se'ls poden demanar responsabilitats, i això les converteix en actors amorals, amb una gran capacitat de generar multitud de falsedats, és a dir, amb la capacitat



de desinformar. Els responsables són els creadors d'aquestes tecnologies,afegeix Dennett. Si no comencem a regular-les, la democràcia es pot veure aviat aclaparada per la desinformació i la consegüent polarització. Estan en joc l'estabilitat i la seguretat de la societat.

Cap a intel·ligències artificials més generals

Les tasques més complicades de dur a terme són aquelles que requereixen interaccionar amb entorns dinàmics, incerts, imprevisibles i no restringits. Dissenyar sistemes que tinguin aquestes capacitats requereix integrar desenvolupaments de moltes àrees de la IA. En particular, necessitem llenguatges de representació de coneixements que codifiquin informació sobre molts tipus diferents d'objectes, situacions, accions..., així com sobre les seves propietats i les relacions entre ells, en particular relacions causa-efecte, per tal que les intel·ligències artificials entenguin el món. També necessitem nous algorismes que, a partir d'aquestes representacions, puguin, de manera robusta i eficient, resoldre problemes sobre pràcticament qualsevol tema.

Finalment, atès que necessitaran adquirir un nombre pràcticament il·limitat de coneixements, aquests sistemes hauran de ser capaços d'aprendre de manera contínua al llarg de tota la seva existència. En definitiva, és imprescindible que les arquitectures cognitives incorporin de manera adequada percepció, representació, raonament, acció i aprenentatge. Aquest és un problema molt important, ja que encara no sabem com integrar tots aquests components de la intel·ligència. Resoldre aquesta qüestió és un pas previ cap a intel·ligències artificials generals.

Reflexions finals

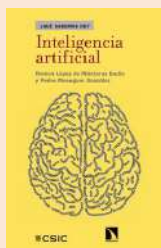
La intel·ligència humana és el referent principal de cara a assolir l'objectiu últim de la IA general, però, per molt sofisticada que arribi a ser aquesta futura IA, sempre serà diferent de la humana. El motiu és que el desenvolupament mental que requereix tota intel·ligència complexa depèn de les interaccions amb l'entorn, i aquestes depenen alhora del cos, en particular dels sistemes perceptiu i motor. Es tracta de sistemes que no existeixen en el cas de les intel·ligències artificials no corpòries i que són molt diferents dels nostres en el cas de les intel·ligències artificials corpòries.

El perill de la IA no és la suposada singularitat tecnològica conseqüència de futures hipotètiques superintel·ligències artificials. Possiblement, els discursos apocalíptics sobre els perills a llarg termini s'han posat sobre la taula de manera interessada, per tal de distreure'ns dels veritables problemes actuals que presenta. És a dir, la privadesa, la manipulació, la vigilància i el control massiu de la ciutadania,

l'autonomia dels sistemes (en particular, les armes autònomes), la confiança excessiva sobre les capacitats de la IA, el biaix dels algorismes, la impossibilitat de retre comptes justificant les seves decisions, i l'excessiu poder d'unes quantes grans empreses tecnològiques.

Necessitem futurs ciutadans més i millor informats, amb més capacitat per avaluar els riscos tecnològics, amb molt més sentit crític i capaços de fer valer els seus drets.

En general, com més sofisticats siguin els sistemes d'IA, més responsabilitat hauríem d'exigir als seus dissenyadors i programadors, per assegurar que compleixin principis legals i ètics. Les intel·ligències artificials no tenen, ni tindran mai, intencionalitat; són el reflex de les intencions i els biaixos dels equips de programadors i entitats implicats en la seva implementació i el seu desplegament. Els agents morals som nosaltres i no pas les màquines. És per això que molts experts assenyalen la necessitat de regular-ne el desenvolupament i, fins i tot, prohibir-ne l'ús en casos de risc extrem, com ara les armes letals. Però, a més de regular, és imprescindible educar els ciutadans, en particular els polítics, sobre els beneficis i riscos de les tecnologies intel·ligents. Necessitem futurs ciutadans més i millor informats, amb més capacitat per avaluar els riscos tecnològics, amb molt més sentit crític i capaços de fer valer els seus drets. Aquest procés de formació ha de començar a l'escola i tenir continuïtat a la universitat. En particular, cal que els estudiants de ciència i enginyeria rebin una formació ètica que els permeti comprendre millor les implicacions socials de les tecnologies que desenvoluparan. Només si invertim en educació aconseguirem una societat que pugui aprofitar els avantatges de les tecnologies intel·ligents minimitzant-ne els riscos. Així, la intel·ligència artificial servirà per fer un gran pas en el progrés de la humanitat.



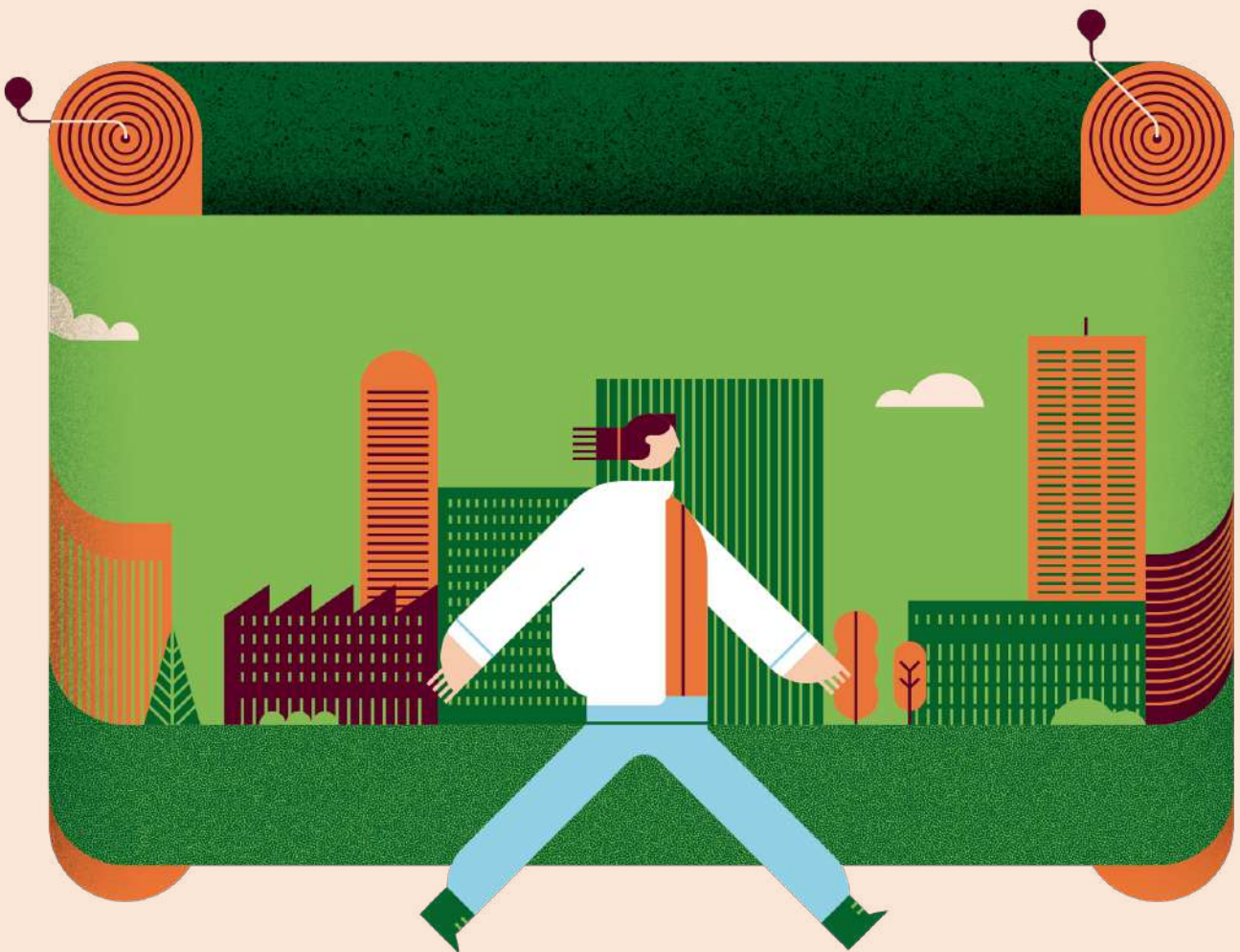
Inteligencia artificial

Ramon López de Mántaras Badia
i Pedro Meseguer González
Col·lecció "¿Qué sabemos de?",
Los Libros de la Catarata, 2017

**Que la ciutat
ens incomodi**
Pablo Martínez
Mar Santamaria

La polarització i la segregació que fomenten els algoritmes poden acabar donant forma a la ciutat. La IA

s'alimenta de si mateixa i, com si es tractés d'un mirall infinit, s'entesta a mostrar-nos el que ja coneix, però no proposa models alternatius. Com a ciutadans, hem d'aprendre a fer-la servir per evitar que s'apoderi de les nostres decisions.



A la ciutat del present, els cotxes no volen ni els hologrames envaeixen els carrers. Les nostres cases són majoritàriament de maons, cartó, guix i, en el millor dels casos, de fusta. Això no impedeix que les tecnologies avançades ja estiguin imbricades en els nostres teixits urbans, encara que no les percebem ni hi estiguin presents de manera evident. La ciutat que veiem i toquem és producte d'un llarg procés històric en què la pedra infereix la velocitat de la lenta transformació física, mentre la ciutat que practiquem (feta de relacions i filies) es transforma ràpidament al ritme de la tecnologia. En aquesta ciutat hi transcorre, avui, l'anàlisi massiva de dades generades amb eines d'intel·ligència artificial (IA), que es nodreixen d'un món paral·lel anomenat internet, en el qual no existeixen els pactes socials de l'espai públic i els consensos del qual són definits per les grans corporacions.

En aquesta ciutat, enamorar-se al pas de vianants ja no és una casualitat del destí atzarós. És conseqüència inevitable del *microtargeting* que orienta les nostres decisions, una estratègia que disposa de manera pròxima els comerços i altres llocs afins a públics similars, als quals també assisteixen altres ciutadans (que han rebut anuncis o recomanacions de *followers* de xarxes socials), com a resultat de l'anàlisi del rastre de dades que deixem constantment i que altres van capturar sense que nosaltres ho notéssim. Així és com la dansa urbana de vianants que circulen pels carrers no només és predictable, sinó que també és, alhora, induïda pel pla digital que precedeix les nostres decisions. En aquesta ciutat, l'única deriva urbana possible és la d'un turista sense *roaming* que mira de trobar el camí a Google Maps.

Si Richard Sennett¹ explica la ciutat com el lloc on és més probable trobar-se amb estranys, podem afirmar que la ciutat tal com ell l'entén està en crisi. El que és casual, espontani, improbable i desendreçat desapareix a poc a poc, a mesura que el pla digital que facilita i orienta les nostres decisions es fusiona amb les nostres vides (el que alguns ja anomenen *metavers*). Així és com la polarització i la segregació existents a les xarxes socials que alimenten tants algorismes d'IA poden acabar donant forma a la ciutat, organitzant l'espai públic, reduint la complexitat i la mixtura de diversos i contraris, minvant la mobilitat social i, finalment, reduint la capacitat de la mateixa ciutat com a instrument de redistribució de la riquesa.

Estem convençuts que descriure amb precisió les nostres ciutats ens ha de facilitar i empènyer cap a una millor presa de decisions. Aquesta convicció ve acompanyada de l'increment de la capacitat d'emmagatzematge de la informació, de la facilitat per implantar sensors a qualsevol procés i de l'omnipresència de xarxes de telecomunicacions capaces de transferir tots aquests fluxos de dades. Això no obstant, davant del gran volum d'informació adquirit (i que ens hauria d'illuminar), descobrim que aquest excés s'escapa de la nostra comprensió i que necessitem tècniques algorítmiques que ens ajudin a operar-hi.



© Gunnar Knechtel

Pablo Martínez Díez

Arquitecte urbanista (Universitat Politècnica de Catalunya, 2006), especialitzat en ciència de dades i anàlisi espacial. Té més de 15 anys d'experiència en el camp del diagnòstic urbà i en la creació d'eines digitals per millorar la presa de decisions en polítiques públiques a través de dades.



© Gunnar Knechtel

Mar Santamaria Varas

Arquitecta urbanista (Universitat Politècnica de Catalunya, 2006) amb més de 15 anys d'experiència en anàlisi espacial, governança i planificació, i projectes urbans estratègics.

Tots dos són cofundadors de 300.000 Km/s, un *think tank* urbà amb seu a Barcelona, que explora el potencial de les dades massives i els nous paradigmes de computació per millorar l'anàlisi urbana, el planejament estratègic i la presa de decisions.

Notes

1. Sennett, R. *Construir i habitar. Ètica per a la ciutat*. Editorial Arcàdia, Barcelona, 2019.
2. <https://impetus4change.eu/>
3. <http://merce.300000.eu/>

La IA com a mirall infinit

Les tècniques basades en IA no desxifren les lògiques del fenomen urbà, sinó que s'alimenten del que se'ls ha mostrat i són capaces de reproduir-ho amb cert virtuosisme, la qual cosa crea la fantàstica il·lusió de prendre decisions amb un encert ajustat sobre sistemes altament complexos que no poden entendre. Qui se senti incòmode en capbussar-se en les aparents incoherències del caràcter urbà o busqui millorar de manera aïllada una part dels processos que el formen, sense mirar-ne la totalitat, pot trobar en aquestes eines un gran confort i una gran satisfacció.

Sota aquest marc operacional, la IA, capaç de reproduir i recombinar informacions sobre situacions precedents, presenta una clara habilitat per gestionar processos: optimitza rutes, redueix consums o minimitza impactes negatius. Tanmateix, malgrat que tot això és necessari, no proposa models alternatius. Per exemple, en un sistema de mobilitat gestionat mitjançant IA, es redueixen els embussos i es milloren els temps de trajecte i la satisfacció dels conductors. Això no obstant, no es redueix el trànsit, sinó que s'aconsegueix que, als carrers, hi càpiguen més cotxes. Per tant, no es transforma el sistema de mobilitat actual que té impactes tan clars sobre el canvi climàtic.

Davant la impossibilitat de definir un model alternatiu, la IA accentua el present. Moltes de les IA que ja operen amb regularitat s'alimenten dels resultats de les seves decisions. Com si es tractés d'una cambra d'eco o un mirall infinit, s'entesten a mostrar-nos el que saben, una vegada i una altra. Un exemple quotidià són els algorismes de recomanació de restaurants, que fan que els més ben posicionats en el rànquing siguin els que obtinguin més visites i que així, dia rere dia, mantinguin el posicionament dominant, la qual cosa fa gairebé impossible alterar-lo si no és per una intervenció externa. I això, que pot semblar una anècdota, configura el carrer en què se situa aquest establiment, influeix en els preus dels locals adjacents i relega a l'oblit altres racons de la ciutat.

També hi ha altres IA més enllà dels algorismes de recomanació de compra o dels xats que poden ajudar els més entremaliats a fer els

deures de l'escola. Posem la mirada en investigacions com les que ara desenvolupa el Barcelona Supercomputing Center en el marc del projecte Impetus4Change² per millorar les prediccions d'episodis climàtics extrems associats al canvi climàtic. Gràcies a l'ús d'aquestes noves

metodologies, és possible anticipar-se a onades de calor o inundacions, amb una antelació de setmanes i mesos, i amb una precisió més alta que l'obtinguda a través de les metodologies actuals (tant en la dimensió temporal com en la resolució espacial), que ens podran ajudar a adaptar-nos millor al canvi climàtic.

Un altre cas és el de l'experiment Mercè, emmarcat en el projecte Urban Citizen Learning,³ en el qual els ciutadans entrenen conscientment un

Els algorismes de recomanació de restaurants configuren els carrers en què se situen, influeixen en els preus dels locals i releguen a l'oblit altres racons de la ciutat.

Notes

4. Amantegui, A. "Por qué las streamers chinas se juntan bajo el puente de un barrio rico para emitir sus directos". *La Vanguardia*, 2023. <https://ow.ly/1KxN50OQjgr>

5. Weckert, S. "Google Maps Hacks". 2020. <http://ow.ly/F4me50OoGfM>

6. Tilbe, A. "Complete ChatGPT Guide for Lawyers: Top 20 Essential Prompts". Medium, 2023. <http://ow.ly/WKkn50OoGgX>

7. Portal jurídic de l'Ajuntament de Barcelona. <http://ow.ly/pWtT50OoGib>

algoritme d'IA perquè ens ajudi a entendre com es combinen els paràmetres que descriuen el medi urbà per produir entorns habitables. L'objectiu del projecte és la quantificació de certes dimensions del món urbà, que fins ara s'escapaven d'una descripció objectiva però que, gràcies a la IA, podem mesurar amb vista a regular i avaluar en un futur.

Hackejar la IA

Al marge de la nostra capacitat per participar en aquests grans projectes, com a ciutadans hem d'aprendre a fer servir i a *hackejar* la IA de manera quotidiana. Hem d'aprendre a emprar-la perquè els beneficis que genera no siguin l'avantatge d'uns pocs i a *hackejar-la* per evitar que s'apoderi de les nostres decisions, per treure-li el poder de saber-ho tot i desactivar-la sense haver de prémer un botó.

Per exemple, uns joves *streamers*⁴ a la Xina retransmeten des de l'espai públic de barris benestants sobre els quals habitualment els algoritmes d'IA de posicionament en xarxes socials ofereixen més visibilitat. Un altre cas

és el de Simon Weckert⁵, que passeja lentament pels carrers un carretó de telèfons mòbils per simular, als ulls de Google Maps, un embús que alliberi de cotxes la ciutat, i això indueix la IA de planificació de rutes a cercar recorreguts menys congestionats. Una altra manera popular de "prendre el pèl" a la IA pot ser fer servir estampats de roba creats amb les

mateixes imatges que es van fer servir per entrenar-la. Això acaba provocant una confusió als algoritmes de reconeixement facial, que ens garanteix la invisibilitat davant les càmeres de videovigilància.

El bé comú davant el profit corporatiu

Hem creat un món difícil d'entendre, format per artefactes humans complexos (com és la ciutat mateixa), només operables per mitjà d'altres artefactes humans, com és la IA mateixa. Podem plantejar usos en què eines com el ChatGPT ens ajudin, per exemple, a interpretar⁶ el complex galimaties de documents legals⁷ necessaris per intervenir a la ciutat. Però també podem repensar tots aquests documents, simplificar-los, ajustar-los i cercar la manera de fer-los comprensibles sense intèrprets, sense cap més eina a part de les emrades per redactar-los.

Ni el futur està escrit ni és conseqüència inherent de la tecnologia. Per moltes decisions que deleguem en la IA, hem de ser nosaltres (els ciutadans) els qui definim el model de societat que volem que la tecnologia formi. És a les nostres mans decidir si volem que la IA converteixi la ciutat en el camp de profit d'uns quants o si volem que s'utilitzi per sobreviure al canvi climàtic en una societat més igualitària. I per a aquest darrer fi no n'hi haurà prou amb l'extenuació del model urbà actual fins a l'extrem, sinó que haurem de

Ni el futur està escrit ni és conseqüència inherent de la tecnologia. Per moltes decisions que deleguem en la IA, hem de ser nosaltres (els ciutadans) els qui definim el model de societat que volem que la tecnologia formi.

Nota

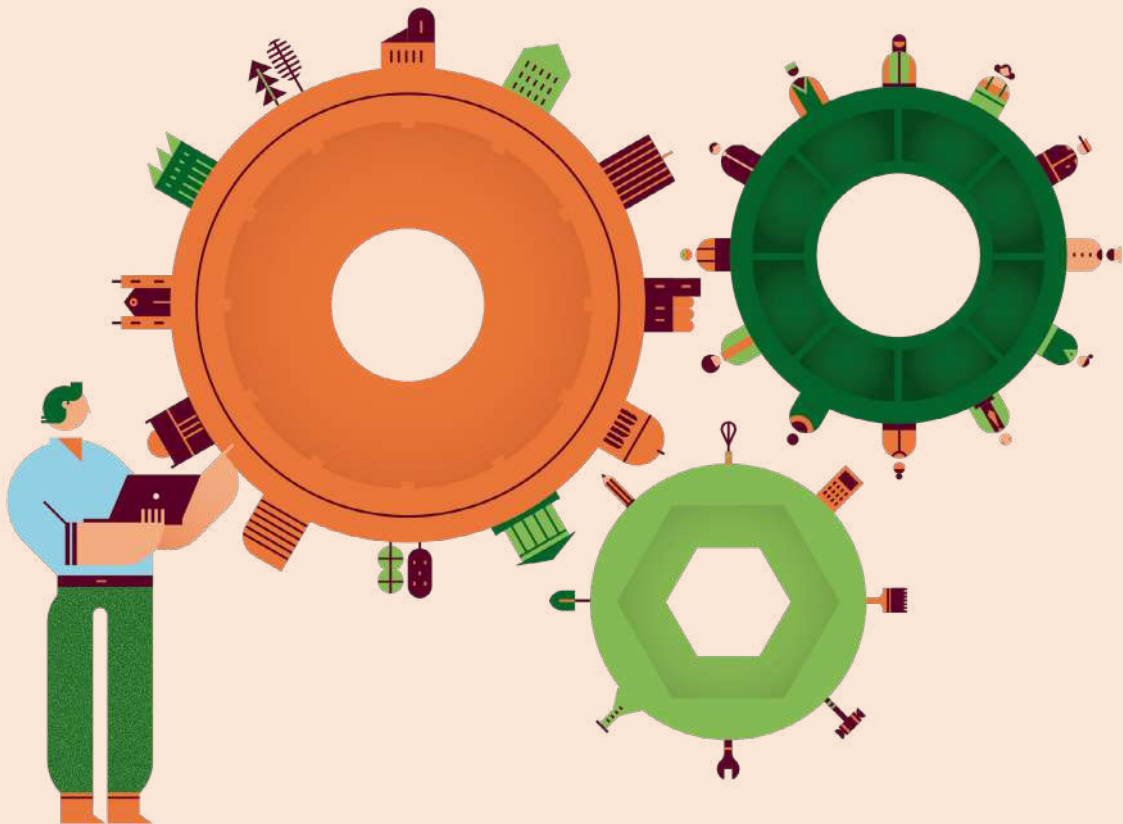
8. "Co-Opting AI: Public Conversations About Design, Inequality, and Technology". Institute for Public Knowledge. New York University, 2019. <http://ow.ly/ZvB150OoGjm>

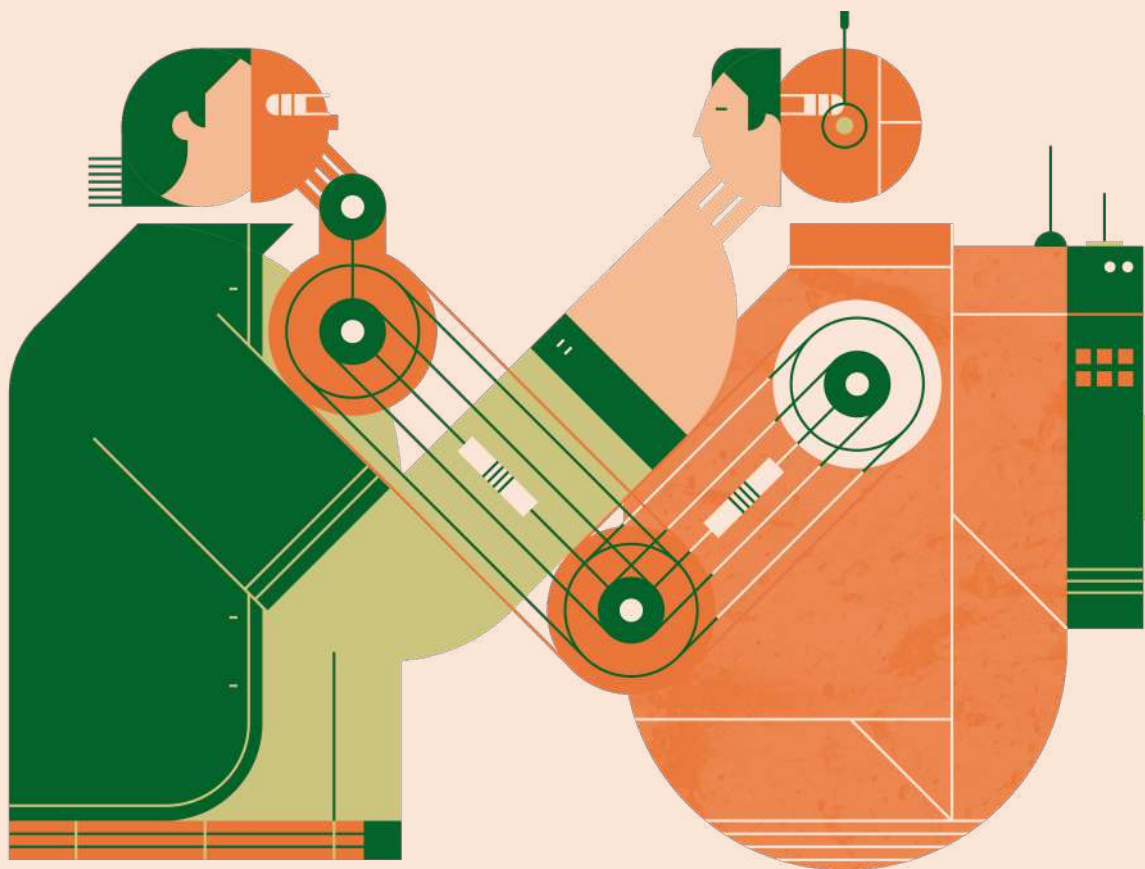
repensar les lògiques que la IA no arriba a desembellar, per poder assemblar i pactar un nou model amb el qual puguem orientar-nos cap al futur.

Per això hem d'aprendre a assumir el control de la IA, hem de *cooptar* la IA (no us podeu perdre la sèrie de vídeos dirigits per Mona Sloane sobre aquest tema⁸). Des de l'àmbit públic, cal vigilar i definir els límits, cal construir-la, dissenyar-la i experimentar-la, produir èxits positius al costat del bé comú davant la ingent quantitat d'algoritmes d'IA orientats avui al profit corporatiu.

Haurem de pensar quin paper podem representar en un escenari en què les grans corporacions inverteixen més en R+D que els mateixos governs (Amazon supera la despesa pública i privada destinada a innovació a Espanya). I plantejar-nos si té sentit donar suport, des dels ens públics, a les investigacions que fan algunes grans empreses en aquesta matèria per lucrar-se. O com podem entrellaçar aquestes tecnologies amb les polítiques públiques que vindran.

Si tornem a la ciutat, els ecos de la IA poden acabar configurant-la i donar lloc a una organització social més uniforme, en què desapareguin els errors i tot allò que sigui insòlit, casual, espontani i estrany. Que el model econòmic i social al qual ens resistim —i del funcionament del qual formen part la misèria i l'opulència— acabi sense ser qüestionat, sinó precisament el contrari, optimitzat. Que l'espai urbà sigui un entorn per al *microtargeting*. Tot això són riscos que haurem d'abordar i per als quals haurem de restar alerta. Resistim-nos al plaer d'una ciutat amable, clara i neta. Cerquem què hi ha darrere d'ella. Que la ciutat ens incomodi, tant allò que hi veiem com la dificultat per entendre-la.





La ignorància dels algoritmes i les noves realitats sintètiques **Pau Garcia**

Les tecnologies
d'intel·ligència
artificial són
una expressió
del que som.

Quan tot és possible, què decidim fer? Tenim a les mans eines perfectes tant per a la creació com per al control, que augmenten i perpetuen les desigualtats socials, però que, ahora, ens permeten enfrontar-nos al repte contemporani de revisar les maneres de mirar el món.

L'oportunitat d'un nou monstre

El 1818 es va publicar la novel·la *Frankenstein o el modern Prometeu* de Mary Shelley, que descriu la gestació d'una creació extraordinària, una promesa amb un potencial infinit que primer es viu des de l'esperança i, després, des de l'horror. El monstre es fa present en el moment en què el seu creador el veu despertar-se i se sent responsable d'allò a què acaba de donar vida.

Aquesta història és molt similar al que estem vivint actualment amb l'ús més estès i popular dels models generatius d'intel·ligència artificial (IA), on, davant la integració d'aquestes eines en espais tant laborals com domèstics, sorgeixen moltes preguntes: Què hem creat? Quin potencial i quin impacte té? És realment intel·ligent? I, si ho és, és realment intel·ligència humana?

Aquests interrogants omplen debats i espais en mitjans de comunicació i generen recerca en diverses universitats, la qual cosa crea espais d'opinió profundament polaritzats. L'investigador del Massachusetts Institute of Technology (MIT) Max Tegmark va publicar el mes de març passat una carta signada per personalitats clau del sector tecnològic (entre elles, Yoshua Bengio, guanyador del Turing Prize; Elon Musk, CEO de Tesla, o Steve Wozniak, cofundador d'Apple), en què demanaven a les principals empreses desenvolupadores d'IA que atuessin qualsevol desenvolupament relacionat amb els sistemes d'IA que fossin superiors, en dimensió, als existents. L'objectiu principal d'aquesta demanda era donar espai als legisladors, als polítics i a la societat en general per integrar aquestes tecnologies a una velocitat assumible. Avui en dia, la carta, que ja ha rebut més de 31.000 suports, no ha transcendit més enllà dels mitjans, i aquestes empreses no han frenat cap dels seus projectes de desenvolupament.

A més, les preocupacions de tota mena expressades per creadors —per exemple, que l'aparició de la IA generativa devaluï la seva feina— han estat perfectament exposades per periodistes i autors com Yuval Noah Harari o Dani Di Placido. Aquesta nova onada de models algorítmics generatius ha posat en qüestió espais i tècniques que, fins fa poc, pensàvem que eren segurs i molt difícils d'automatitzar, com són l'escriptura i els llenguatges visuals. Ara, més que mai, ens plantegem quin valor té allò que fem com a humans, quin impacte té i en què es diferencia la nostra intel·ligència de la d'una màquina.

Recentment, l'Oficina de Drets d'Autor dels Estats Units ha dictaminat que les imatges generades per IA “no són producte de l'autoria humana” i, per tant, no poden ser objecte de drets d'autor. Els autors estan desenvolupant eines per protegir el seu treball, però d'ara endavant serà una nova lluita constant en l'espai de la creació. Fa poc, la revista de ciència-ficció i fantasia *Clarkesworld Magazine* va haver de tancar temporalment les inscripcions després de rebre una allau d'històries escrites amb ChatGPT. Com sabrem si el 100% de les coses que llegim i veiem a les xarxes socials les ha escrit una persona? De moment, és impossible, perquè no tenim eines per esbrinar-ho, i aquí hi ha un gran perill i una gran oportunitat.



Pau Garcia

(Barcelona, 1989). És cofundador i director de Domestic Data Streamers, on, des del 2013, desenvolupa projectes de recerca i comunicació en dades per a institucions com ara les Nacions Unides, la California Academy of Sciences, la Tate Modern de Londres o el Citizen Lab de la Universitat de Toronto. Codirigeix el màster en Dades i Disseny d'Elisava, i ha impartit classes al Hong Kong Design Institute, al Royal College of Art de Londres, al Politecnico di Milano i a la UC Berkeley. És fundador de HeyHuman, un programa de residències artístiques que combina música, periodisme, recerca artística de dades i justícia social.

Notes

1. Crawford, K. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.
2. Katz, Y. "Noam Chomsky on Where Artificial Intelligence Went Wrong". *The Atlantic*. 2012. <http://ow.ly/58Qv50NW6vi>
3. Burgess, M., Schot, E. i Geiger, G. "This Algorithm Could Ruin Your Life". *Wired*. 2023. <http://ow.ly/zrkF50NW6Gu>

Avui, les eines polítiques i legals que tenim per defensar certs col·lectius i interessos socials necessiten camins més lents, perquè es puguin pensar i legislar de manera justa les capacitats de les empreses que hi ha darrere d'aquestes tecnologies. Aquest desequilibri entre la velocitat a la qual desenvolupem eines de manera privada i les maneres com entenem l'efecte que tenen sobre nosaltres, sobre la societat i el planeta és decisiu per esbrinar si el que estem creant és el monstre de Mary Shelley o una eina plena de potencial.

Sota la catifa artificial

Un aspecte d'aquesta tecnologia que cal tenir en compte és que, com en la majoria d'aquests avenços, hi ha un cert nivell d'alienació. La IA pot semblar una força etèria, que resideix als núvols, neta i un punt sobrenatural, però està formada per quantitats enormes de recursos naturals, combustibles fòssils, treball humà i infraestructures massives. Eines com ChatGPT poden fer l'efecte de ser lleugeres i estar desconnectades de qualsevol realitat material, però, de fet, requereixen quantitats massives de potència de càlcul i recursos extractius per funcionar¹.

Alguns personatges públics, com el Dr. Sean Holden, professor associat a la Universitat de Cambridge, o el reconegut lingüista Noam Chomsky² expliquen molt bé els límits de la intel·ligència artificial i les raons per les quals aquesta tecnologia queda lluny de ser perfecta. Per això aquí utilitzem el concepte d'*ignorància artificial* per referir-nos a tots aquests fenòmens generalment entesos com a errors o espais d'estúpidesa que venen determinats per la tecnologia algorítmica. Es poden posar dos exemples: l'octubre del 2021, la Imane, una mare migrant marroquina de 44 anys, divorciada, era interrogada a Rotterdam mentre es recuperava d'una cirurgia abdominal. Els ajuts públics que li permetien pagar el lloguer i comprar menjar per als seus tres fills estaven en risc, i no només això, hi havia la possibilitat que fos acusada de frau. Havia de demostrar la seva innocència en un procés burocràtic difícil i costós perquè un algoritme la considerava "d'alt risc" per cometre frau en l'obtenció d'ajuts socials.

Aquest no és un cas aïllat, sinó que forma part d'un patró global en el qual els governs i les empreses utilitzen algoritmes per reduir costos i augmentar l'eficiència. Com la Imane, a Rotterdam, més de 20.000 famílies han estat falsament acusades per un algoritme³ de rebre ajuts públics de manera fraudulenta. Sovint aquests sistemes no funcionen com es preveu i poden reproduir biaixos preocupants, i això, de vegades, afecta de manera irrecuperable comunitats que precisament són les que necessiten més ajuts.

Un altre exemple més popular és l'algoritme que Amazon

utilitzava el 2015 per fer la selecció de personal i realitzar recomanacions d'incrementos salarials. L'objectiu d'aquest sistema era que, per cada mil currículums introduïts, s'indiqués quins cinc s'havien de contractar o a qui calia augmentar el sou. El problema és

La IA pot semblar una força etèria, que resideix als núvols, però està formada per quantitats enormes de recursos naturals, combustibles fòssils, treball humà i infraestructures massives.

Notes

4. Larson, E. "Amazon Sued for Alleged Race, Gender Bias in Corporate Hires". *Bloomberg*. 2021. <http://ow.ly/mG3950NW6fT>
5. Haraway, D. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective". *Feminist Studies*, 14(3), 575-599. 1988. <http://ow.ly/Wyyyc50OoYwV>

que les dades que s'havien fet servir per entrenar l'algoritme procedien dels deu anys anteriors en una empresa que ja tenia, de base, un biaix que donava molta més preferència a homes que a dones. En conseqüència, el sistema va començar a descartar i a penalitzar tots els currículums dels quals es pogués extreure qualsevol inferència de gènere femení, reproduint exactament un model injust, que, si no hagués estat per l'evident desequilibri de gènere, hauria passat absolutament inadvertit⁴.

De nou, l'arrel del problema d'aquest algoritme és que basa totes les seves respostes en models estadístics. Aquestes tecnologies no distingeixen entre la veritat i la mentida; únicament busquen patrons. Els algoritmes predictius d'avui són màquines increïblement útils, però que poden generar resultats erronis profundament desconnectats de la realitat. Una IA tindrà un marge d'error molt més petit que una persona experta en meteorologia per preveure si demà plourà, però no entendre mai què significa que plougui, ni per què plou, ni el que això pot suposar per a una cultura o un poble. No existeix coneixement del *perquè*, només del *què*, i aquest caràcter desvinculat i descontextualitzat és especialment perillós quan s'actua en espais socials en els quals els prejudicis generalment es perpetuen i els individus i les institucions poden justificar certs comportaments amb la senzilla resposta de "l'algoritme ho recomana".

Quan la coral canta, jo ballo

Quan hom s'adona que la majoria d'usos actuals d'aquestes tecnologies en l'esfera pública estan vinculats al control, l'optimització i la supervisió de la ciutadania i els recursos, es fa palès que existeix una profunda falta d'imaginació en els espais d'ús de la intel·ligència artificial generativa. Si volem revertir-ho, hem de pensar en com utilitzar-la responant a preguntes diferents.

Un dels exercicis que feia sovint amb estudiants a l'aula era el de comprar quatre diaris de perspectives polítiques molt diferents. Amb els alumnes trobàvem la mateixa notícia i analitzàvem com l'explicava cada mitjà. En el reconeixement d'aquesta pluralitat de versions sempre emergien debats crítics molt més fonamentats i desapareixien les perspectives úniques i les posicions enrocades. Un dels futurs que veiem en l'ús de les IA el trobem, precisament, en la seva capacitat de ser plurals. Actualment, un dels projectes en desenvolupament, en el qual treballem, és una integració per a navegadors que permeti oferir una versió automatitzada de qualsevol notícia, en la qual es donin diverses lectures crítiques alternatives sobre els mateixos fets. Una nova manera d'entendre l'actualitat a través de molts altres ulls, tots ells amb el seu propi biaix.

El contrari d'un biaix no és l'eliminació d'aquest, no és la neutralitat absoluta; és la transparència, és aixecar la catifa i entendre com és el biaix i en què afecta. Un biaix pot ser positiu, pot donar una perspectiva concreta sobre una comunitat no representada, ja que la neutralitat mai existeix del tot, sempre hi ha una mirada, un coneixement situat⁵. El conflicte apareix quan s'utilitza una eina de la qual no s'és conscient que té una mirada predefinida amb què es representen els interessos d'uns pocs, sovint poderosos. Si som capaços d'invertir el biaix i fer-ne un ús transparent, aquesta

Notes

6. Cita original: "The test of a first-rate intelligence is the ability to hold two opposed ideas in mind at the same time and still retain the ability to function".

tecnologia ens pot donar un nou espai on podrem visualitzar la tensió entre idees diferents i estar més exposats a perspectives plurals.

Necessitem màquines que ens donin espai per a la reconsideració. Però és difícil: requereix valentia. Una màquina que ens ajudi a negar-nos la comoditat de ser sempre la mateixa persona, aquella que va arribar a una resposta fixa fa temps i mai ha tingut cap motiu per dubtar-ne. Una màquina que, en altres paraules, ens ajudi a mantenir-nos oberts.

Com va dir F. Scott Fitzgerald, "la prova d'una intel·ligència de primera categoria és la capacitat de mantenir dues idees oposades en ment al mateix temps i, així i tot, conservar la capacitat d'operar al món".⁶ La voluntat de canviar d'opinió és un superpoder en l'era moderna, i aquestes tecnologies ens poden ajudar, en un moment de màxima polarització, a variar el rumb i estar més oberts a idees diferents.

Les realitats sintètiques

Des del febrer del 2022, a Domestic Data Streamers hem estat desenvolupant un programa d'experimentació, formació i debat sobre les possibles aplicacions de les tecnologies d'intel·ligència artificial generativa. El projecte principal en què hem treballat ha estat el desenvolupament d'eines per a la recreació de memòries sintètiques a través de la generació d'imatges. La principal diferència amb altres *softwares* d'edició d'imatges és que la IA té la capacitat de crear imatges directament des de la descripció d'una escena, i amb cert entrenament es poden aconseguir resultats de molta qualitat a una velocitat de cinc a deu segons. Això fa que sigui accessible i que puguem iterar ràpidament, condicions fonamentals per treballar amb un gran nombre de col·lectius i individus.

Les aplicacions són molt extenses. Ja estem actuant amb treballadors socials, psicòlegs i experts mèdics en demència i psiconeurologia per entendre l'impacte positiu que poden tenir les memòries sintètiques durant l'acompanyament en el progrés de malalties degeneratives com l'alzheimer o la demència senil. Els primers resultats estan essent especialment emocionants i, pròximament, els publicarem documentats.

Aquesta tecnologia i aquests processos també els estem portant a altres espais, com ara la reconstrucció de la memòria històrica de comunitats de refugiats a Atenes o dels supervivents d'Hiroshima i Nagasaki, al Japó. Són comunitats que, per diverses raons, han perdut documentació visual sobre moments personals, sovint d'un gran interès històric; memòries esborrades de l'espai de la cultura visual que ara tenim la possibilitat d'evocar amb una tecnologia més assequible, des de les comunitats protagonistes, que generalment són excloses de l'accés a les últimes tecnologies, com són la gent gran, les persones migrants o els pacients de malalties mentals. Aquestes eines ens poden permetre entendre'ns tant a nosaltres com a les comunitats pertanyents a altres temps i realitats socials. Hem de ser capaços d'apropar-les de manera transparent, responsable i col·laborativa a tothom; la realitat sintètica pot ser un espai entre la ficció i la realitat, un espai intermedi en el qual puguem trobar-nos.

Reimaginar la ciutat

Aquesta mateixa tecnologia obre noves possibilitats en el procés de codisseny de l'espai públic. En els últims mesos, hem treballat en diverses sessions amb estudis d'arquitectura internacionals per entendre com es poden incorporar aquestes eines als processos de disseny i quin impacte poden arribar a tenir de cara a integrar veïns, comerços i altres agents en el desenvolupament de les propostes d'urbanisme i arquitectura. Amb aquestes tecnologies es facilita l'accés a la visualització d'alternatives urbanes de manera molt assequible: es poden repensar carrers, edificis i usos de l'espai públic en qüestió de segons. El límit més gran, generalment, és la imaginació d'un mateix.

Les imatges, per descomptat, tenen un aspecte estrany i poc detallat, no són rivals per al treball d'un arquitecte experimentat, però permeten visualitzar resultats amb prou detall com per fer-se una idea de les possibilitats de l'arquitectura alternativa. És una eina més per avançar molt ràpidament en la visualització del canvi i, conseqüentment, permet a veïns i habitants de la ciutat comprendre millor i donar suport a les alternatives urbanístiques que més els interessin. El perill resideix en el fet que els resultats d'aquestes eines, en ser tan visuals, atractius i assequibles, poden acabar passant per sobre d'altres factors importants, com poden ser les observacions d'ús, les dades demogràfiques o les diverses necessitats socials de cada espai. La visualització del canvi és una gran eina per a les comunitats, però s'ha d'utilitzar de manera responsable i informada.

Els punts de les preguntes

Quan observo com aquestes tecnologies es despleguen sobre la societat, es fa evident que encara som molt lluny de tenir una perspectiva plural i humanista. Sembla que les converses a tots els mitjans estan enfocades en la tecnologia en si mateixa, en el fenomen gairebé màgic que sembla contenir. Piques persones descriurien el funcionament d'un cotxe, un correu electrònic o un àudio de WhatsApp com una eina màgica, però la primera vegada que vam estar exposats a aquestes tecnologies ens ho semblaven. En el cas de la intel·ligència artificial generativa també succeeix un fenomen similar. I aquest fet ens distreu del que realment és important: Quin impacte volem que tinguin aquestes eines en la nostra societat? Per què les estem desenvolupant? Qui hi té accés? Qui se'n beneficia? A quins interessos responen i com ho fem perquè operin sota una sèrie de valors?

Si la intel·ligència artificial és la resposta, quina és la pregunta?

Referències bibliogràfiques

Lohr, S. *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. HarperBusiness, 2015.

Pasquinelli, M. i Joler, V. "The Noosphere manifested: AI as instrument of knowledge extractivism". *AI & Society*, 36(4), 1263-1280. 2020. <http://ow.ly/IRvb50NWaKS>

Quatre vectors cap a una cultura de la IA José Luis de Vicente

Les habilitats
creatives
ja es poden
repliar a través

d'aprenentatge profund (*deep learning*) i xarxes neuronals, però el seu veritable potencial resideix en la possibilitat de la col·laboració i no de la substitució. Dues subjectivitats diferents, una d'humana i una altra d'automatitzada, es poden trobar a mig camí i produir un espai simbòlic mutu i temptador, ple de possibilitats, però també de vies mortes. Aventurar quin impacte artístic tindrà la IA a mitjà termini és tan prematur com superb.



El mite de la creativitat

El fantasma de l'automatització i el seu impacte sobre el mercat laboral i la forma de la societat del demà ha generat in comptables anàlisis i pronòstics des de començaments del segle XXI. Durant un temps, en aquests debats s'hi aplicava una regla lineal: com més mecànica i rutinària fos una tasca, més senzill seria d'automatitzar-la. A partir d'aquesta lògica, les professions en què la creativitat tingués un paper central serien les que estarien menys amenaçades per la substitució. Aquesta mena d'“excepció creativa” reafirmava, en certa manera, un estatus especial dels artistes i creadors a la societat.

Si l'explosió —des de l'estiu del 2022— d'eines d'IA per a la creació en múltiples àmbits ha tingut algun efecte revelador, ha estat el de posar fi a aquest miratge. L'aplicació amb finalitats productives d'eines com el ChatGPT o Midjourney no cerca, en la majoria dels casos, l'excel·lència artística, sinó l'eficiència rendible. “Prou bo (i infinitament més barat)” és el factor decisiu en unes indústries creatives en què la gran majoria de tasques quotidianes no aspiren a l'originalitat extrema ni a la ruptura dels motlles establerts. Des d'aquesta posició, els dibuixants, músics, periodistes o realitzadors de vídeo es comencen a reconèixer en la mateixa posició que qualsevol altre col·lectiu subjecte a la imposició d'un nou ordre emergent, que no negocia les condicions.

Acumulació per desposseïció (digital)

Qualsevol aplicació creativa d'IA opera a partir d'un model entrenat amb quantitats de dades enormes. Stable Diffusion, un dels generadors d'imatges més populars, va desenvolupar el seu model a partir del conjunt de dades LAION-5B, publicat en obert a internet el març del 2022. LAION-5B consta de gairebé 6.000 milions d'arxius que combinen imatges amb etiquetes de text que n'identifiquen el contingut¹; imatges obtingudes a internet, producte de l'esforç col·lectiu quotidià dut a terme per tots nosaltres, els usuaris de la xarxa.

Al llarg dels darrers trenta anys, però sobretot durant l'era daurada del *social media* i l'emergència del telèfon intel·ligent a partir del 2007, la producció social de dades en grans volums ha estat part del pacte implícit en el model econòmic de la indústria d'internet. El tracte acceptat per tots era senzill: a canvi de fer servir de manera gratuïta tota mena de serveis digitals, se'ns anima a produir i compartir activament com més informació millor. Inicialment, la cadena de valor consistia a agregar a gran escala dades personals que permetessin la producció de perfils de consumidors, empaquetar-les i convertir-les en mercaderia.

La indústria de l'aprenentatge automàtic ha afegit una capa de valor addicional a aquestes muntanyes de contingut que la web atresora. És la matèria primera idònia per entrenar models amb capacitat de produir el que alguns anomenen “intel·ligència”, i d'altres, simplement, “capacitat de predicció estadística”. Un generador d'imatges automàtiques pot produir un gat de qualsevol raça i color amb fidelitat i precisió, perquè el model es pot alimentar d'una disponibilitat infinita d'imatges de gats procedents de la web.



José Luis de Vicente

Comissari i investigador cultural, des del 2023 és director del Museu del Disseny de Barcelona (DHUB). Ha estat fundador i director artístic del Sónar+D, el programa d'arts i idees del festival Sónar de Barcelona; cofundador de Model, el Festival d'Arquitectures de Barcelona, i comissari del festival Llum BCN. També ha comissariat múltiples exposicions, com ara Big Bang Data i Després de la fi del món (totes dues, al Centre de Cultura Contemporània de Barcelona, CCCB); Mirador Torre Glòries, i Curiosidad Radical: en la òrbita de Buckminster Fuller (Espacio Fundación Telefónica / ArtScience Museum Singapur), entre d'altres. El seu treball investiga els impactes de la innovació tecnològica i cultural a través d'artefactes, sistemes i narratives.

Nota

1. Qualsevol que tingui curiositat per esbrinar què hi ha a l'interior del model el pot descarregar a <https://laion.ai/blog/laion-5b/>

Darrere de cada una d'aquestes imatges pujades a una xarxa social, a un àlbum de fotos o a una notícia en un mitjà de comunicació, hi ha ciutadans, autors, creadors, artistes, fotògrafs professionals o aficionats, i qualsevol altra categoria de productor imaginable. Per descomptat, cap companyia de *deep learning* ha fet un esforç previ per obtenir els drets d'ús d'aquestes imatges. Tampoc s'ha limitat a utilitzar imatges lliures de drets o emparades per sistemes de llicències lliures, com les Creative Commons. Els creadors d'alguns dels serveis més populars, com ara Midjourney, s'han excusat al·ludint a la impossibilitat de produir una eina com aquesta si haguessin d'obtenir individualment els drets d'ús de milers de milions d'imatges. La responsabilitat s'ha invertit, i és l'autor qui ha de sol·licitar expressament que les imatges siguin excloses del model.

El terme *acumulació per desposseïció*, encunyat pel geògraf urbà i teòric social marxista David Harvey per designar la capacitat del capitalisme per mercantilitzar el que prèviament era compartit i comú, s'expressa a la perfecció en aquest procés. Fins i tot si considerem el potencial expressiu i creatiu d'aquestes noves eines com a socialment beneficiós i artísticament productiu, el capital simbòlic i cognitiu d'una gran tasca col·lectiva torna a quedar controlat i capturat per uns quants.

Centaures

En tots els diagnòstics inicials, l'emergència de la IA com una força transformadora es considera a partir de la lògica de la substitució: els robots i els sistemes automàtics ocuparan el nostre lloc. Però tan cert és que moltes habilitats es poden replicar, com que el veritable potencial de la IA com a

nou paradigma per a la creació procedeix de la possibilitat de la col·laboració, no de la substitució. Dues subjectivitats diferents, una d'humana i una altra d'automatitzada, es poden trobar a mig camí i produir un espai simbòlic mutu, amb resultats diferents dels que els creadors podrien obtenir per si sols.

L'evocadora imatge d'un centaure, meitat humà i meitat cavall, ofereix una visió simbiòtica de la col·laboració creativa entre humans i màquines.

En realitat, que escriptors, músics o arquitectes deleguin decisions fonamentalment creatives en sistemes de regles, processos evolutius autònoms o eines amb capacitat de produir respostes pròpies no és particularment nou. Hi ha precedents, des de la llarga tradició de l'escriptura automàtica i el *cut-up*, l'art basat en instruccions i l'art algorítmic dels anys seixanta del segle xx, fins a les múltiples eines generatives que productors de música electrònica o arquitectes parametricistes han fet servir de manera quotidiana les darreres dues dècades.

El dissenyador d'interacció Matt Jones utilitza la imatge evocadora del centaure, la criatura mitològica meitat humà i meitat cavall, per promoure una visió simbiòtica de la col·laboració creativa entre humans i màquines. Cada una de les dues parts d'aquest creador híbrid aporta els seus millors actius i capacitats i reposa sobre l'altra en els aspectes en què està en desavantatge.

El mitjà és el model

El juliol del 2015, el programador de Google Alexander Mordvintsev va llançar el projecte DeepDream, una eina que utilitza una xarxa neuronal per buscar el rastre d'imatges possibles dins d'altres de ja existents. Les imatges produïdes per DeepDream són calidoscopis visuals que és inevitable relacionar amb l'estètica de la psicodèlia, el referent més a mà per reconèixer-les com a part d'una tradició anterior. Els anys següents, els artistes que van explorar l'espai de possibilitat de les xarxes neuronals i l'aprenentatge profund es van abocar en l'ús de les xarxes neuronals generatives adversàries (GAN), una metodologia específica per a la producció d'imatges. En els projectes de Gene Kogan, Mario Klingemann o Anna Ridler, les GAN ens mostren totes les variacions possibles que podrien existir en el procés de transformar una imatge en una altra de diferent, en un flux visual potencialment infinit.

El juliol del 2022, OpenAI, un dels gegants de la indústria del *deep learning*, va permetre l'accés públic al seu servei DALL·E 2, la primera eina popular de generació d'imatges a utilitzar una tecnologia específica d'aprenentatge profund, anomenada *transformador* (*transformer*). Les imatges produïdes a través de DALL·E 2 han inundat els canals digitals ràpidament. Es caracteritzen per un fotorealisme imperfecte capaç de produir imatges a primera vista verídiques, però que quan s'exploren amb més detall revelen inconsistències, com rostres borrosos o deformats. És fàcil relacionar un retrat generat amb DALL·E amb un esbós del pintor Francis Bacon.

A DALL·E 2 l'han succeït ràpidament altres transformadors, com ara Stable Diffusion o Midjourney, potser el més popular de tots a l'hora d'escriure aquest article. Midjourney s'ha convertit en una mena d'atles implícit de la història de la cultura visual. Els seus usuaris aprenen aviat que utilitzar l'ordre "En l'estil de" permet produir imatges que reproduïxen l'estètica de qualsevol pintor, il·lustrador o cineasta. Però el model indubtablement té biaixos, com qualsevol tecnologia basada en l'anàlisi estadística. Per exemple, és sorprenentment fàcil produir imatges que clonen l'estil del director de cinema nord-americà Wes Anderson; els seus colors saturats, enquadraments plans i composicions simètriques resulten relativament senzills d'imitar per al model, per la qual cosa es generen sense parar imatges "en l'estil" d'Anderson.

Sense necessitat de cap formació especial, qualsevol podria diferenciar a primera vista una imatge produïda amb DeepDream d'altres de generades amb una GAN, amb DALL·E 2 o amb Midjourney. Però a més, els artistes més destacats en aquest mitjà (Refik Anadol, Holly Herndon i Mat Dryhurst) són els que produeixen els seus propis models, en comptes de limitar-se a utilitzar les mateixes eines genèriques que fan servir tots els altres.

En menys d'una dècada de pràctiques artístiques frenètiques al voltant de l'aprenentatge profund i les xarxes neuronals, els models grans del llenguatge, les GAN i els transformadors generatius, les estètiques possibles de la intel·ligència artificial s'estan revelant com un espai temptador, ple de possibilitats i de vies mortes, de llocs comuns, però també d'un estranyament profund. Qualsevol conclusió rotunda i categòrica sobre l'impacte artístic de les tecnologies d'IA a mitjà termini seria tan prematura com superba.



Tecnologia i humanitats: un diàleg necessari

Joan Manuel del Pozo

L'impacte de
la intel·ligència
artificial va
acompanyat
d'una preocupació

creixent pels riscos ètics, socials i democràtics d'un desplegament descontrolat. En aquest escenari, les humanitats, amb el seu component de pensament crític i de llibertats, haurien de ser-ne la base, per assegurar el respecte als principis ètics més bàsics. Seria incomprensible que la IA girés l'esquena a aquesta aspiració de la intel·ligència natural que l'ha inventada.

La presència social de la intel·ligència artificial (IA) és forta i creixent; la presència social de les humanitats és feble i minvant. Les dues realitats, però, no són directament homologables: la IA és una tecnologia informàtica i les humanitats són un corrent cultural secular. Però això, que relativitza l'impacte del contrast, fa rumiar en un sentit: si aquest corrent cultural, com és el cas, ha estat històricament un dels puntals de l'educació de la nostra intel·ligència natural (IN), és legítim plantejar-se algunes preguntes: aquesta tecnologia batejada com a *intel·ligència* —potser abusivament—, tan sovint confrontada a la IN humana per saber quina “guanyarà” en potència, velocitat i, finalment, domini, pot i ha de prescindir d'aquell puntal formatiu de la nostra IN?

Les humanitats, que han formatat —valgui aquest mot d'importació informàtica— la IN durant segles, fan nosa al desplegament de la IA? O, a l'inrevés, la IA podria beneficiar-se d'una influència humanística?, entenent les humanitats com el vessant educatiu de



Joan Manuel del Pozo

Doctor en Filosofia i professor emèrit de la Universitat de Girona, d'on fou vicerector i síndic. Estudiós de Ciceró i de Thomas More, ha publicat i fet conferències a Catalunya, la resta de Espanya i diversos països sobre educació, ètica i política. És coautor, amb Ulises Cortés, del llibre de diàlegs *La intel·ligència (artificial), a la butaca*. Fou diputat al Congrés i al Parlament de Catalunya, regidor de l'Ajuntament de Girona i conseller d'Educació i Universitats en el darrer govern de Pasqual Maragall.

Nota

1. Bellardi, G. *Introduzione a Cicerone. Le Orazioni III*, p. 35. Unione Tipografico-Editrice, Torí, 1975.

l'humanisme. O són preguntes banals i la IA ha de tirar pel dret sense més consideracions? Aquesta darrera pot ser la resposta “per defecte”, automàtica o inercial de persones irreflexives, però un mínim de consciència educativa, cívica, social o ètica —totes elles, ben mirat, vinculades entre si per la tradició humanística— ens convida a pensar-hi.

L'impacte creixent de la IA va acompanyat d'una també creixent preocupació pels riscos ètics, socials i democràtics d'un desplegament descontrolat. Una mostra en seria la “Carta de la moratòria”, en la qual un miler de científics ha demanat una pausa de, com a mínim, sis mesos en el seu desenvolupament, arran de l'aparatosos difusió del ChatGPT; però són moltes les veus, també des de dins mateix dels “fabricants” d'IA, que reclamen pausa, reflexió i regularització d'uns procediments que són invasius de tots els àmbits de la vida. En aquesta situació seria greu i irresponsable no aprofitar el cabal cultural de l'humanisme, o les humanitats en versió educativa, les quals són, fins ara, dipositàries del millor saber —entès en sentit plural i evolucionat fins avui i no limitat a la base del classicisme grecollatí— de caràcter ètic, social i democràtic. Prescindir-ne sense més consideració seria una opció radical de

Són moltes les veus, també des de dins mateix dels “fabricants” d'IA, que reclamen pausa, reflexió i regularització.

signe (anti)ètic, (anti)social i (anti)democràtic. Cal assumir el compromís d'aprofitar els sabers humanístics, de conrear-los —també amb ajuda de la IA— i de relacionar-hi el desplegament, l'adaptació i l'aplicació d'aquesta tecnologia.

L'humanisme, un saber complex

Aquesta reflexió pretén ajudar en l'intent. Potser cal, però, un aclariment conceptual sobre l'humanisme, per evitar confusions. L'humanisme, base cultural de les humanitats, no és ni una ideologia ni encara menys una religió: és un corrent de saviesa o cultura plena, iniciat a la Grècia i a la Roma clàssiques, reviscut al Renaixement, reimpulsat a la Il·lustració i en permanent renovació: Edgar Morin, Martha Nussbaum, Nuccio Ordine, Joan Fuster o George Steiner, per exemple, són humanistes contemporanis ben diferents. És un nervi cultural de les nostres societats, que està sobretot “fet de fe en la llibertat, en la intel·ligència, en la virtut de la paraula, en la moralitat i en el compromís cívic i social”: una síntesi d'un vell filòleg, òptima per comprendre el valor i la transcendència del perfil de la nostra “formatació” intel·lectual, i estímul per a un diàleg lliure i compromès amb una tecnologia que es diu intel·ligent¹.

El saber humanístic viu més en el territori dels símbols i els valors que en el dels fets i les funcionalitats aplicades. L'escriptor brasiler Mario Quintana va gosar dir que “els fets són un aspecte secundari de la realitat”, no pas per defensar prioritàriament els impresentables “fets alternatius” d'aquests temps de postveritat, sinó perquè considerava com a aspecte primari o bàsic de la realitat humana la comprensió, interpretació, contextualització, crítica i valoració de qualsevol fet que, es vulgui o no,

Notes

2. Lindo, E. "El algoritmo son los padres". *El País*. 2023. <http://ow.ly/noCh50OrIQO>

3. Harari, Y. N. *21 lliçons per al segle XXI*. Penguin Random House, Barcelona, 2018.

busca necessàriament un marc o format on pugui satisfer la necessitat que tenim de sentit.

Al fons, glateix la idea kantiana segons la qual l'ésser racional autònom —que pretenem ser— viu “en el regne de les finalitats”; i la IA, com a instrument, per potent, ràpid i eficaç que sigui, viu en el regne de les eines, incapaç —llevat de falsificacions— de produir per si mateixa una finalitat digna d'aquest nom. I és només en relació amb una finalitat o altra triada amb llibertat que es poden interpretar i valorar els nostres fets de cada dia. Els algoritmes de la IA no pensen, són pensats. Perquè, com ha dit Elvira Lindo², “l'algoritme són els pares”: no fa res per si mateix i, sobretot, no és màgic, sempre hi ha al darrere uns “pares i mares” que preparen els “regals” que li atribuïm.

El llenguatge natural, amenaçat

Com han coincidit a dir —en termes diferents, però coincidents de fons— dos pensadors tan allunyats entre si com Yuval Noah Harari³ i Slavoj Žižek, el risc de la nova onada d'IA —els xatbots com GPT— és que ens acabi “hackejant” el sistema operatiu de la humanitat, que no és altre que l'entrellat de sentit que ens proporciona cada llenguatge natural. Entrellat fet de claus interpretatives, marcs mentals, finalitats i valors, en forma de relats èpics, religions, mites, poesia, novel·les, teatralitzacions, ideologies i conceptes similars que ens permeten disposar de diverses cartografies per al trànsit vital.

Aquests xatbots, com a generadors de llenguatge sense més base que una potent, però freda, combinatòria d'infinat de dades no contrastades, no poden garantir la veritat de les seves afirmacions i consideren que la seva missió no és produir textos veritables o fiables sinó “plausibles”: és a dir, simulacions gratificants per a cada corresponsal que s'hi adreci. Es difumina la línia entre veritat i falsedat, i el llenguatge s'aplana cada cop més, alhora que afebleix i pot fer perdre la gran varietat de recursos del llenguatge natural. Com a exemples, dos de principals: la capacitat metafòrica i la irònica. El conreu i la cura del llenguatge natural es troben, des de sempre, a la base de les humanitats, que se senten interpel·lades per l'amenaça de degradació procedent de la IA.

Reclam d'ètica i democràcia

Cal reclamar una relació positiva de la IA amb les humanitats per algunes raons més que es desgranen tot seguit. Tota tecnologia, quan arriba, instaura un nou poder que sol estar en poques mans; més encara, en el cas de la IA i les poques grans tecnològiques. Els fabricants d'IA haurien d'aprendre de les humanitats el risc de la *hybris* o desmesura, capaç de portar els poders més ferrenys a l'autodestrucció, com a les tragèdies gregues. Sense tant de dramatisme, la IA necessita les humanitats com a punt de

El conreu i la cura del llenguatge natural es troben, des de sempre, a la base de les humanitats, que se senten interpel·lades per l'amenaça de degradació procedent de la IA.

Notes

4. Nowotny, H. *La fe en la inteligencia artificial. Los algoritmos predictivos y el futuro de la humanidad*.

Galaxia Gutenberg, Barcelona. 2022.

5. Tegmark, M. *Vida 3.0. Ser humano en la era de la inteligencia artificial*.

Penguin Random House, Barcelona. 2018.

Ja es coneixen alteracions greus en processos democràtics derivades de les manipulacions basades en IA, com les eleccions que van convertir Donald Trump en president o com el resultat del referèndum del Brexit.



La intel·ligència (artificial), a la butaca

Ulises Cortés i Joan Manuel del Pozo

Universitat de Girona, 2021

referència i gran repositori de figures, mites i relats de màxim interès humà. A l'altra banda, els usuaris d'IA necessitem les humanitats per una raó tan essencial com l'esmentada cura del llenguatge natural: la defensa de la llibertat, pròpia de tota la tradició humanística, atès el risc d'atac de la IA per la via de la invasió de la privacitat, per la de la manipulació a partir de l'acumulació de dades i per la via de la desestabilització democràtica.

Les humanitats, amb el seu foment del pensament crític i de l'esperit de llibertat, són la base necessària per satisfer la demanda molt generalitzada i creixent d'una IA respectuosa amb els principis ètics més bàsics; l'ètica, com a filosofia pràctica o "saber de la vida" des del classicisme grec, no és ni una moda ni una imposició de cap autoritat, és l'intent de resposta a la necessitat humana d'aprenentatge d'una vida en llibertat, d'una vida bona, plena i feliç.

Seria incompreensible que la IA —la nostra nova eina "intelligent", encara que sigui artificial— girés l'esquena a aquesta aspiració de la IN que l'ha inventada. Més encara si, com sabem, l'ètica és condició necessària no solament per a la vida personal, sinó també per a la col·lectiva; contra el que alguns pensen, el compromís social i polític de l'humanisme és irrenunciable. I avui, l'amenaça a la

democràcia —convivència en llibertat— per part de la IA és alguna cosa més que un malson: ja es coneixen alteracions greus en processos democràtics derivades de les manipulacions basades en IA, com les eleccions que van convertir Donald Trump en president o com el resultat del referèndum del Brexit.

Una tecnòloga d'IA molt respectada, Helga Nowotny⁴, defensa ni més ni menys que el cultiu de la saviesa —concepte de gran ressò humanístic— amb aquests mots: "Els algoritmes poden ajudar-nos en molts dels reptes que se'ns presenten. Però, si volem conservar la condició humana amb tot el que significa, haurem d'aprendre a usar-los sàviament i a conrear aquest tipus de saviesa que el futur necessita".

Un altre tecnòleg prestigiós, Max Tegmark⁵, rebla el clau: "I mirant lluny cap al futur, cap als reptes relacionats amb la IA general sobrehumana, ens convé adoptar com a mínim uns estàndards ètics bàsics abans de començar a inculcar aquests estàndards a màquines potents. [...] Ens interessa educar els nostres joves perquè facin que la tecnologia sigui sòlida i beneficiosa abans de cedir-li molt de poder".

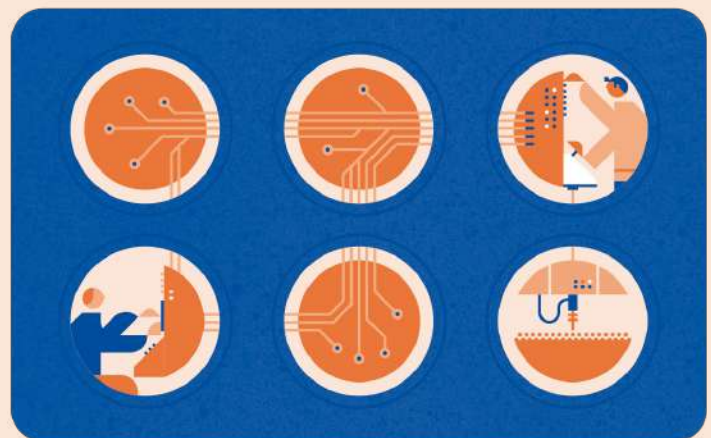
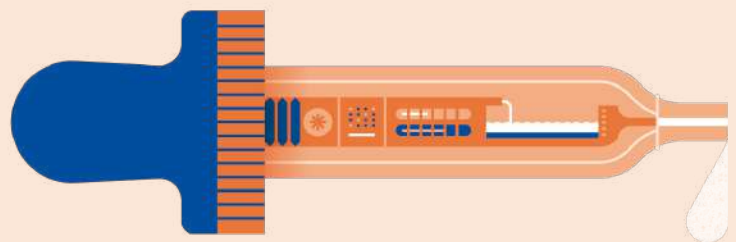
En podem dir "invitacions" al dit "nou humanisme tecnològic"? O simplement a l'humanisme? O al cultiu de les humanitats? El diàleg necessari entre les habilitats de la nova tecnologia d'IA i el pòsit de saber humà d'IN cap dins de qualsevol d'aquestes etiquetes o similars. No ens hi capfiquem, però treballem-ho seriosament. No és ni un repte tecnològic ni un repte humanístic. És un repte vital i de futur que necessita totes les formes d'intel·ligència.

Els robots i els sistemes intel·ligents revolucionen la medicina

Beatriz López

La intel·ligència artificial (IA) està revolucionant la medicina i l'atenció sanitària.

Ja hi ha desenvolupaments que poden identificar amb més precisió els tumors que els experts humans. Una intervenció complexa com la implantació d'una vàlvula aòrtica difícilment pot entendre's avui sense els sistemes robòtics que donen suport als clínics. I gràcies a les dades i els algorismes, s'està obrint el camí cap a una medicina personalitzada.



La IA ha donat suport, des del principi, a la medicina i a l'atenció sanitària. L'any 1972 ja existia una eina intel·ligent que permetia identificar els bacteris que causaven una infecció a la sang¹. Aquestes primeres eines s'anomenaven *sistemes experts*, perquè pretenien replicar el coneixement d'un expert humà en un camp d'aplicació —la medicina, en aquest cas—, amb la finalitat de fer accessible aquesta expertesa a un ampli col·lectiu, per exemple, hospitals amb manca d'un cert tipus d'especialistes. Ara bé, la pràctica clínica de la medicina consisteix en l'ús de la millor evidència científica disponible, normalment mitjançant assajos clínics, els resultats dels quals es publiquen en articles científics. És el que es coneix com *la medicina basada en l'evidència*. En aquest sentit, el coneixement d'un sistema necessita ser revisat de manera continuada, i els sistemes que donen suport a la presa de decisions —coneguts actualment com a *sistemes clínics d'ajut a la presa de decisions*— han esdevingut més complexos.

De fet, l'entorn clínic ha canviat molt des dels primers sistemes experts, i actualment els mètodes d'informació clínica recullen moltes dades sanitàries dels pacients (els coneguts com a *health care records*), així com informació genètica i d'altra d'obtinguda a través de nous dispositius, incloent-hi els sensors portables (*wearables*)². La disponibilitat de les dades ha permès que aquesta evidència mèdica no només es tingui mitjançant assajos clínics prospectius, que impliquen fer proves sobre persones físiques, sinó que es puguin provar hipòtesis amb assajos retrospectius, amb informació emmagatzemada als nostres sistemes de salut. Així mateix, sobre les dades dels registres sanitaris s'apliquen mètodes d'aprenentatge automàtic que permeten identificar, per exemple, quines persones estan subjectes a un risc cardiovascular important. Amb aquesta informació, els clínics poden recomanar actuacions per prevenir un possible deteriorament de la salut.

A més, les dades genètiques ens acosten al que coneixem com a *medicina de precisió*, que permet dissenyar fàrmacs específics per a persones amb condicions especials de salut, d'acord amb la seva genètica. De fet, la generació de noves proteïnes que permetin la síntesi de fàrmacs nous és una de les puntes de llança de l'ús de la IA, on s'utilitzen principalment tècniques anomenades d'*aprenentatge profund* (*deep learning*)³.

Pel que fa als nous dispositius, avui en dia disposem de sistemes de monitoratge complexos, d'imatge mèdica, de robòtica assistida... Si ens imaginem una UCI, veurem pacients connectats a molts aparells que els monitoren les constants vitals. Tota aquesta quantitat d'informació és difícil de controlar per la persona que està a càrrec del pacient, a la qual ajuden els sistemes d'IA, activant alertes si es reconeixen situacions de risc. Aquests sistemes de monitoratge estan desenvolupats a partir de dades històriques de pacients que han passat per situacions semblants, alguns dels quals s'han recuperat i d'altres, lamentablement, no. Per exemple, hi ha sistemes que permeten detectar ràpidament la septicèmia, un procés infecció greu que pot conduir amb rapidesa a la mort⁴.



Beatriz López

Professora d'Intel·ligència Artificial al Departament d'Enginyeria Elèctrica, Electrònica i Automàtica de la Universitat de Girona, on lidera la investigació en IA en medicina i salut del grup de recerca eXIT. És sòcia cofundadora de l'empresa Newronia i membre del comitè assessor de MJN. Pertany a l'Associació Catalana d'Intel·ligència Artificial, de la qual fou presidenta del 2014 al 2016; a la xarxa espanyola IABiomed-Net, i a la xarxa internacional CLAIRE d'IA. Ha participat en els projectes PEPPER (tractament de persones amb diabetis tipus 1), NoaH (monitoratge de nadons prematurs a casa) i TDAp (recomanacions participatives i personalitzades a persones amb TDAH).

Notes

1. Shortliffe, E. H. "A rule-based computer program for advising physicians regarding antimicrobial therapy selection". ACM National Conference. 1974. <http://ow.ly/KUFh50OnBSz>
2. Rajpurkar, P., Chen, E., Banerjee, O. i Topol, E. J. "AI in health and medicine". *Nature Medicine*. 2022. <http://ow.ly/Tgbi50OnBTU>
3. Ferruz, N., i Höcker, B. "Controllable protein design with language models". *Nature Machine Intelligence*, 4, 521-532. 2022. <http://ow.ly/XXuL50OnBUy>
4. Goh, K. H., Wang, L., Yeow, A. Y. K., Poh, H., Li, K., Yeow, J. J. L. i Tan, G. Y. H. "Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare". *Nature Communications* 2021, 12:1, 1-10. 2021. <http://ow.ly/oURa50OnBVn>
5. Torrent-Fontbona, F. i López, B. "Personalized Adaptive CBR Bolus Recommender System for Type 1 Diabetes". *IEEE Journal of Biomedical and Health Informatics*, 23(1), 387-394. 2019. <http://ow.ly/ALPK50OnBXb>
6. Sánchez, E. "Un robot que ayuda a los niños con diabetes". *Euronews*. 2017. <http://ow.ly/iaOv50OqOVy>

Pel que fa als *wearables*, s'estan obrint noves vies d'obtenció de dades objectives per monitorar les persones amb malalties cròniques. Pensem, per exemple, en els pacients diabètics. Si no té cap dispositiu, la persona diabètica ha de comunicar al seu metge les dades sobre el nivell de sucre a la sang mitjançant diaris que anota manualment. Ara bé, si disposa d'un glucòmetre, un petit sensor que ho mesura contínuament, aquesta informació és molt més objectiva i precisa. Evidentment, tota la informació del sensor representa una gran quantitat de dades a visualitzar pel metge quan visita la persona a la consulta. Aquí és on la IA ajuda a identificar quins episodis crítics ha tingut el pacient i suggereix recomanacions perquè no es repeteixin. Es redueix, així, el temps que el metge necessita per processar la informació, temps que pot dedicar a atendre el pacient⁵.

Personalitzar els tractaments

L'ús de dades ha permès, en general, aplicar algoritmes d'IA per aconseguir una medicina personalitzada i per fer recomanacions particulars. Però els dispositius no només estan presents en la cura de les persones; la medicina fa temps que ha passat del paradigma de la "malaltia" al paradigma de la "salut", i els professionals se centren a donar-nos recomanacions per mantenir la qualitat de vida, el que es coneix com a *medicina preventiva*. En aquest context, els *wearables*, com ara els rellotges intel·ligents, ens ajuden a fer un seguiment del nostre exercici diari, per exemple. També disposem d'un conjunt d'aplicacions al mòbil que ens permeten seguir hàbits saludables i utilitzen la IA per fer-nos suggeriments segons la nostra evolució; d'entre totes, però, cal tenir en compte les que estan certificades pels organismes competents.

La IA també ha permès la participació activa de les persones en el seu procés de salut. Sense la IA, la medicina participativa es limita a informar el pacient sobre la malaltia que té; fins i tot hi ha jocs educatius que l'ajuden a entendre què li passa. Però la IA permet anar més enllà: un exemple d'això és el desenvolupament de robots que eduquen els nens petits amb diabetis⁶. Amb l'eina TDApp, els pacients que pateixen TDAH (trastorn per dèficit d'atenció i hiperactivitat) o les persones que en tenen cura poden expressar les seves preferències respecte dels efectes adversos que no voldrien patir. A partir d'aquesta informació i de les dades clíniques de la persona, l'aplicació



Notes

7. "AI 'outperforms' doctors diagnosing breast cancer". *BBC News*. 2020. <http://ow.ly/KjAg50OnBZ9>

8. Flores, M., Glusman, G., Brogaard, K., Price, N. D. i Hood, L. "P4 medicine: how systems medicine will transform the healthcare sector and society". *Personalized Medicine*, 10(6), 565. 2013. <http://ow.ly/Rrwx50OnC0f>

9. "Identificació d'algorismes d'IA". Observatori IA en salut. <http://ow.ly/gKUg50OnC1J>

10. RobSurgical. Bitrack System. <https://www.robsurgical.com/es/>

TDApp proposa una llista de tractaments possibles, d'acord amb l'evidència mèdica més actual. Si un pacient adolescent, per exemple, no vol engreixar-se a conseqüència del tractament, el sistema pot considerar fàrmacs alternatius. El psiquiatre és qui pren la decisió final, però, entre fàrmacs amb eficàcia similar, s'inclinarà per triar-ne un que satisfaci les preferències del pacient, amb l'objectiu d'assegurar l'adherència al tractament.

Un dels camps on la IA està revolucionant més la medicina i l'atenció sanitària és el de la predicció (diagnosi). En particular, s'ha avançat molt en sistemes de tractament d'imatges mèdiques utilitzant mètodes d'aprenentatge profund per identificar, per exemple, possibles tumors. Recentment, la IA ha demostrat que pot identificar amb més precisió els tumors que els experts humans⁷.

Les 4P de la medicina

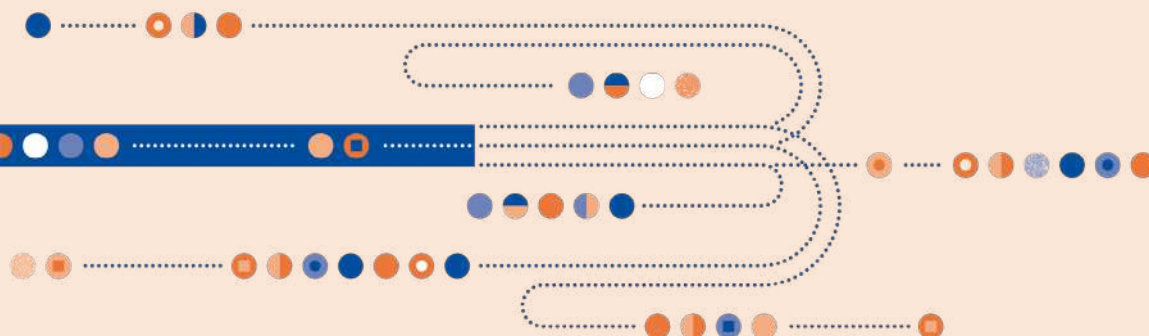
La IA, doncs, està present en les 4P de la medicina: personalització, prevenció, participació i predicció⁸. Hi ha moltes aplicacions disponibles, en tots els àmbits. D'acord amb l'Observatori d'IA en Salut a Catalunya⁹, el 47,68% dels algorismes desenvolupats i registrats estan destinats a l'atenció hospitalària; el 23,24%, a l'atenció primària, i, en molta menor mesura, el 8,61%, a les malalties mentals, i el 7,28%, a l'àmbit sociosanitari. De fet, un dels reptes actuals de la medicina és la salut mental, i aquí la IA pot aportar noves eines relacionades amb la detecció de l'estat emocional a partir de dades psicofisiològiques recollides per sensors (per exemple,

mesurant l'estrès a través de dades cardiovasculars i sudoració de la pell), que poden ser l'antesala de malalties neurològiques més greus com, per exemple, l'alzheimer.

Cal remarcar que només el 17% dels algorismes estan destinats a una millora de la gestió, mentre que el 60,53% es dediquen a la pràctica clínica

i la cirurgia. En aquest darrer àmbit, destaca el paper important de la robòtica. Robots com el Da Vinci o Bitrack¹⁰ estan suportats per sistemes d'imatge mèdica que mostren al clínic la informació més rellevant en cada pas. Una intervenció complexa com la implantació d'una vàlvula aòrtica (*transcatheter aortic valve*

La IA ajuda a identificar quins episodis crítics tenen els pacients amb malalties cròniques i suggereix recomanacions perquè no es repeteixin.



Notes

11. El-Fakdi, A., Gamero, F., Meléndez, J., Auffret, V. i Haigron, P. “A framework for a workflow-based CBR for interventional Clinical Decision Support Systems and its application to TAVI”. *Expert Systems with Applications*. 2014. <http://ow.ly/Vj0750OnC2R>
12. Zoë, C. “The future of elder care is here – and it’s artificial intelligence”. *The Guardian*. 2021. <http://ow.ly/X5fN50OnC4r>
13. Audio analytics for professionals. <https://www.soundintel.com/>
14. Zhavoronkov, A. “Caution with AI-generated content in biomedicine”. *Nature Medicine*. 2023. <http://ow.ly/SKy050OqPa1>
15. Barraud, E. “New implant offers promise for the paralyzed”. EPFL, 2022. <http://ow.ly/OEbe50OqPh7>

implantation, TAVI) difícilment podria entendre’s avui en dia sense l’ajuda de sistemes d’IA que donen suport als clínics en tot moment¹¹.

Tots aquests sistemes estan promoguts també per una realitat: tenim una població molt envellida. La medicina ha de donar respostes a persones que arriben a una edat avançada amb malalties cròniques (com les respiratòries o la hipertensió), amb comorbiditats (pateixen més d’una malaltia) i amb uns recursos sanitaris limitats per atendre-les. Tanmateix, a causa dels avenços en medicina, hi ha moltes persones que arriben a edats avançades amb un estat de salut que els permet viure amb autonomia a casa seva. En aquest context, apareixen les cases *smart*, dotades de sensors que permeten monitorar la vida d’una persona i detectar si, eventualment, cau a terra o no es lleva¹². La robòtica acompanya també aquestes persones grans proporcionant-los cuidadors que en tinguin cura en l’àmbit de les tasques rutinàries i deixant que l’espai de lleure sigui el que comparteixin amb els familiars, que, gràcies a aquesta ajuda, no estaran esgotats. Podem recordar aquí la pel·lícula *Un amic per a en Frank*, que il·lustra aquest futur que tenim al davant.

Les oportunitats per tenir cura de la gent gran són molt enginyoses. Per exemple, l’empresa Sound Intelligence ha desenvolupat

un programari per monitorar els sorolls en un geriàtric: si la seqüència observada correspon a un comportament normal (com la d’un moviment de llit, seguit d’un soroll d’una porta que s’obre, d’aigua que cau, d’una porta que es tanca i d’un altre moviment de llit), el sistema

no molesta el possible cuidador; però si la seqüència de sons és anormal (un moviment de llit i un cop fort), es dispara una alarma¹³. I el sistema ha de funcionar en un dia de tempesta!

Per concloure, podem dir que un dels activadors de la IA en medicina i atenció sanitària és la disponibilitat de les dades. Ara bé, hi ha reptes importants. En primer lloc, pel que fa a la qualitat de les dades, la presència de biaix i la privacitat. La IA generativa —un dels representants més populars de la qual és ChatGPT— pot obrir les portes a l’obtenció de nous medicaments i, eventualment, generar noves hipòtesis i tractaments, però presenta reptes importants respecte al corpus de dades sobre les quals es generen els resultats, que poden induir a biaixos significatius¹⁴. En segon lloc, en referència a la transparència dels sistemes: les solucions aportades pels algorismes d’IA han de ser explicables per merèixer la confiança dels professionals. Tenim un camí al davant que cal abordar també a nivell legislatiu, amb les normatives d’ús corresponents.

Quedem-nos, però, amb les oportunitats. I el millor exemple és el que es va publicar l’any 2022 sobre la recerca de l’Escola Politècnica Federal de Lausana (Suïssa), que demostra com una persona paraplàgica, postrada en una cadira de rodes, pot tornar a caminar gràcies a uns implants controlats per IA¹⁵.

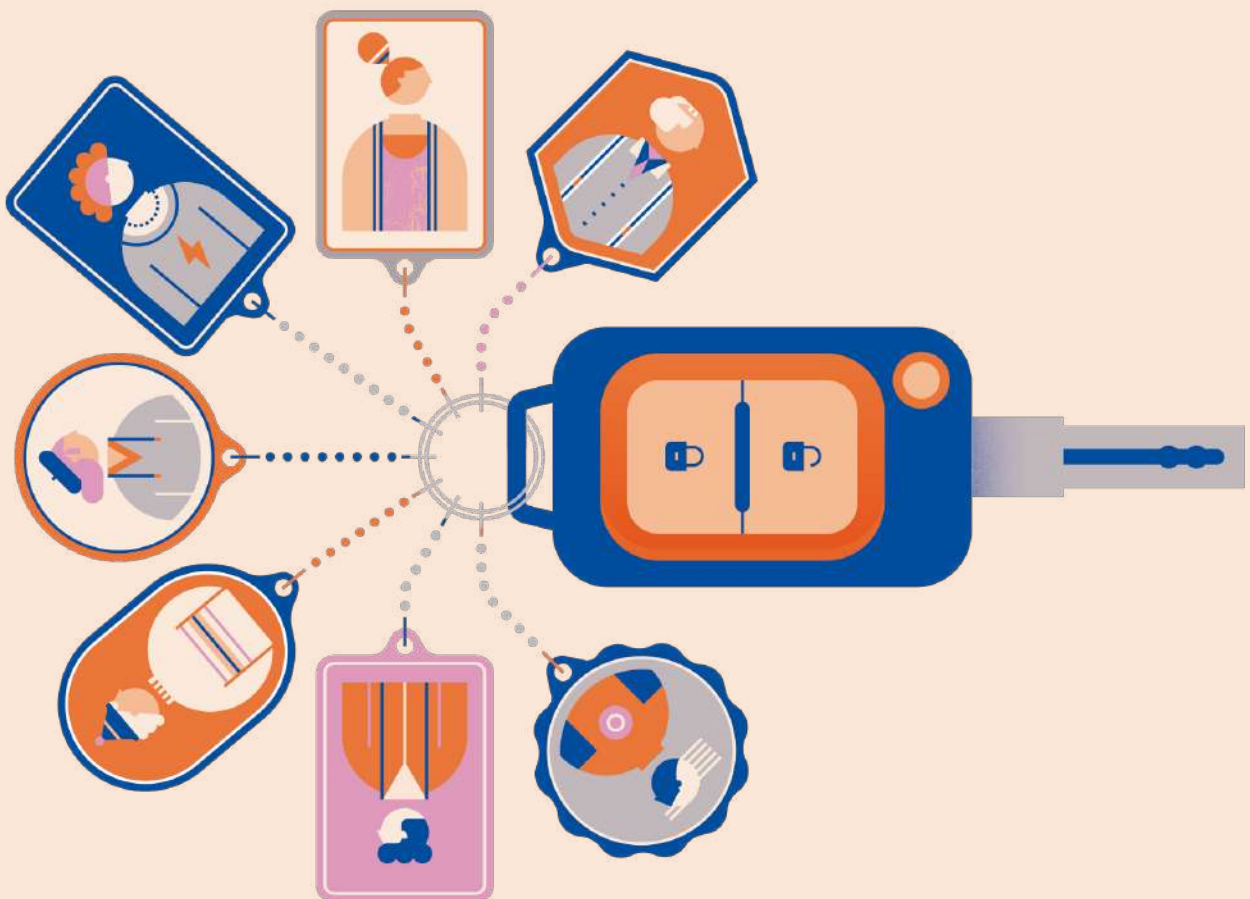
El millor exemple del que pot aportar la IA és el que es va publicar al 2022, quan una persona paraplàgica va tornar a caminar gràcies a uns implants.

Eines d'IA per a una mobilitat compartida i sostenible

**Filippo Bistaffa,
Pedro Meseguer
i Juan Antonio
Rodríguez Aguilar**

La IA pot ajudar a gestionar processos complexos de la vida urbana, com ara la mobilitat. Gràcies a simuladors

i servidors que posen en contacte persones que ofereixen el seu vehicle amb d'altres que necessiten desplaçar-se, es poden arribar a reduir un 80% el trànsit i un 70% les emissions de CO₂.



A començaments d'aquesta primavera havia d'anar, passada l'hora de dinar, a un lloc a la ciutat de Barcelona per fer-hi algunes gestions. La perspectiva d'una tarda llarga i assolellada em va fer concebre la idea d'una passejada amb bicicleta per la platja (visc a prop del mar). Com que la meua destinació no estava gaire ben comunicada amb transport públic, vaig decidir agafar el cotxe; a més, hi havia un pàrquing a prop i no era hora punta. Va ser un error. Quan vaig tornar de la meua ocupació fora de la ciutat, aquella tarda vaig continuar amb el cotxe fins a l'aparcament que havia triat. Amb trànsit fluid, els meus plans es complien de manera raonable. Però la tornada va ser molt diferent; en pocs minuts, i sense motiu aparent, em vaig trobar immers en un embús sense possibilitat d'escapar-me'n. Potser hi havia hagut un accident o hi havia carrers tallats per obres o manifestacions. No ho sé. Quan vaig arribar a casa m'havia quedat sense temps, però, sobretot, sense ganes de fer aquella passejada amb bicicleta. L'embús m'havia posat de mal humor, sense comptar la gasolina cremada, el temps malgastat i la contaminació, atmosfèrica i acústica, que havia generat.

Aquest episodi il·lustra la problemàtica del trànsit a les grans ciutats. El transport públic continua essent la millor solució per a la mobilitat, però és una resposta parcial: sempre hi haurà zones que no es comunicaran adequadament, ja que, quan les ciutats creixen, els barris nous triguen un cert temps — anys — a estar connectats de manera ràpida mitjançant el transport públic. Aquesta etapa sol ser prou llarga perquè sorgeixin barris nous en els quals es reproduïx el problema. És clar que la solució genèrica passa necessàriament per la mobilitat compartida; es tracta de treure vehicles que només transporten una persona de l'immens i frenètic formiguer humà en què es converteix una gran urbs en hora punta¹. Hi ha receptes, conegudes des de fa temps, a les quals avui la intel·ligència artificial està donant un nou impuls.

El viatge compartit, a gran escala

El concepte de viatge compartit —al món anglosaxó rep el nom de *carpooling* o *ridesharing*— va néixer alhora que els vehicles, i el seu ús es remunta a dècades enrere. Hi ha diverses versions d'aquesta idea. Aquí ens referim al cas en què un conductor ofereix el seu vehicle per portar altres persones, per aprofitar així el viatge (en la bibliografia especialitzada es coneix com a *peer-to-peer ridesharing*). Hi ha altres modalitats que fins i tot poden fer augmentar el nombre de vehicles.

Aquesta fórmula es fa servir des de la segona meitat del segle passat en àmbits restringits, com en el cas de companys de feina que comparteixen cotxe per anar a la mateixa destinació o membres de la mateixa família que es reparteixen per la ciutat. En tots dos casos, la comunicació entre els agents que comparteixen el viatge es produeix de manera directa, sense originar cap despesa extra, i el problema d'optimització que s'ha de resoldre és trivial. Però la seva aplicació a gran escala encara està pendent. Ara, els avenços de la intel·ligència artificial, juntament amb les facilitats de comunicació, permeten construir la



Filippo Bistaffa

Científic titular a l'Institut d'Investigació en Intel·ligència Artificial del CSIC, on va arribar el 2017 com a *fellow* Marie-Curie després d'obtenir el títol de doctor en Informàtica a la Universitat de Verona (Itàlia).



© Jordi Cortés

Pedro Meseguer

Investigador científic del CSIC a l'Institut d'Investigació en Intel·ligència Artificial (IIA), llicenciat en Ciències Físiques i en Informàtica i doctorat per la Universitat Politècnica de Catalunya. És membre de la European Association for Artificial Intelligence.



Juan Antonio Rodríguez Aguilar

Professor de recerca a l'Institut d'Investigació en Intel·ligència Artificial del CSIC. És membre de l'Associació Catalana d'Intel·ligència Artificial i de la European Association for Artificial Intelligence.

Notes

1. El director de l'Agència d'Ecologia Urbana de Barcelona afirmava en una entrevista: "El 65% de l'espai de les capitals és per aparcar o per circular i això comporta que no puguem fer servir llocs públics per a altres finalitats".

<http://ow.ly/LkuY50OyQtz>

2. En la resolució també s'inclouen dades de la zona urbana considerada, aconseguides mitjançant Google Maps. Una heurística aproxima la funció objectiu, cosa que permet accelerar-ne l'optimització.

El procediment complet es detalla a <http://ow.ly/j6OU50OyQwk>

3. Es va fer una anàlisi diferent per a cada dia i es van obtenir resultats molt similars. Les dades precises d'ambdós estudis es poden consultar a <http://ow.ly/72On50OyQwP>

tecnologia necessària per implantar aquesta fórmula en una gran ciutat. Hem desenvolupat els mètodes que demostren la viabilitat d'aquesta idea a gran escala, amb un simulador que anticipa els beneficis que se'n poden esperar.

L'operativa del sistema proposat consisteix en l'enviament d'una petició a un servidor per part de cada persona que ofereix traslladar o que vol ser traslladada. Aquest servidor proporciona contínuament un servei de mobilitat compartida: rep les ofertes o les peticions i, en un temps limitat —un minut—, combina les ofertes de trasllat que encara no s'han utilitzat amb les peticions pendents de servir, mitjançant un procés d'optimització estocàstica². A continuació, envia l'assignació corresponent a les ofertes i peticions considerades en aquest cicle, i continua amb el procés. En conjunt, una persona no ha d'esperar més de cinc minuts entre que envia la petició i rep la resposta.

L'estalvi que pot proporcionar depèn dels criteris que es prioritzin. N'hi ha diversos, i aquí ens centrarem en dos, que es tracten com a paràmetres: el temps extra necessari respecte d'un viatge individual i la ràtio d'adopció (el tant per cent de vehicles a tota la ciutat que fan viatges compartits). A partir de les dades reals de trànsit de dos dies³ a Nova York, un dia laborable —el dijous 1/12/2016— i un dia de cap de setmana —el dissabte 3/12/2016—, al nostre estudi experimental observem que la reducció de la contaminació atmosfèrica augmenta clarament amb la ràtio d'adopció, cosa que està directament correlacionada amb la disminució del nombre de viatges i de la distància total recorreguda. Aquestes dades milloren a mesura que permetem més temps extra per viatge compartit, acceptant retards superiors, de fins a deu minuts en el pitjor dels casos. En números rodons, amb el 100% de ràtio d'adopció, ocupant tots els seients

del vehicle i la màxima prioritat mediambiental, s'assoleix un estalvi del 70% en emissions de CO₂ i una disminució del 80% del trànsit.

En canvi, en el pitjor escenari, que és el 20% de ràtio d'adopció amb la mínima prioritat mediambiental i que representa un retard mitjà de

deu segons, s'obté una disminució del 8% en emissions de CO₂ i el trànsit decreix un 15%. Així doncs, fins i tot en nivells baixos d'implantació en una gran ciutat com Nova York, el mètode proposat permet assolir beneficis mediambientals significatius (un 1% de disminució en l'emissió de CO₂ significa un estalvi d'1,5 tones d'aquest contaminant). En escenaris intermedis, els avantatges obtinguts se situen entre aquests dos extrems.

Un simulador per ajudar en les decisions

Per disposar d'un sistema flexible que proporcioni resultats diferents en funció de la variació de prioritats en els viatges compartits, hem desenvolupat un simulador que implementa el procés de resolució, tractant com paràmetres les prioritats que se seleccionin. Aquest simulador és una eina

Els avenços de la intel·ligència artificial i les facilitats de comunicació permeten construir la tecnologia necessària per implantar la mobilitat compartida en una gran ciutat.

Notes

4. <https://ridesharing.iiia.csic.es>
5. Calo, S., Bistaffa, F., Jonsson, A. i Gómez, V. "A novel message passing approach to spatial air quality prediction in urban areas". *Deep Learning Barcelona Symposium*. 2022.
6. <https://opendata-ajuntament.barcelona.cat>
7. Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica (XVPCA). <http://ow.ly/gzNp50OyQxG>
8. Departament d'Acció Climàtica, Alimentació i Agenda Rural de la Generalitat de Catalunya. Qualitat de l'aire actual. <http://ow.ly/Aunt50OyQyp>

valuosa, ja que permet quantificar amb precisió els beneficis que donarà la implantació d'aquesta idea sobre una ciutat específica amb unes preferències concretes. A més, serveix d'ajuda a la decisió per als gestors públics que han de concretar polítiques de mobilitat urbana. El simulador s'ha d'alimentar amb les dades del plànol de la ciutat objectiu, i està disponible al web de l'Institut d'Investigació en Intel·ligència Artificial (IIIA)⁴ per al cas analitzat.

El simulador ofereix dues modalitats: simulació i anàlisi. En la primera, després de proporcionar valors per als diversos paràmetres del model, resol el problema proposat i mostra els resultats sobre l'àrea urbana considerada. En la segona, permet fer enginyeria inversa: s'estableixen els objectius que cal assolir en termes de beneficis mediambientals i de temps extra permès en viatges compartits, i es calcula la política de viatge compartit (*peer-to-peer ridesharing*) necessària per assolir-los.

El procediment és ràpid i sostenible, ja que s'executa en un temps límit (un minut per al procés d'optimització, cinc minuts en total) i no necessita infraestructures especials ni càlcul intensiu. Si la ciutadania hi coopera, la seva implantació sembrarà de forats les jungles de cotxes en què es converteixen sovint els nostres carrers. Representa una manera vàlida de repensar la mobilitat urbana i un element necessari per aprofitar les oportunitats que ofereixen les ciutats als habitants, i obtenir així beneficis mediambientals que són absolutament imperatius al segle XXI.

Altres aplicacions

Hem desenvolupat un altre projecte d'IA de mobilitat sostenible⁵ que recomana "rutes verdes" a ciutadans que es desplacen a peu o en bicicleta per Barcelona, tenint en compte la qualitat de l'entorn (nivells reduïts de pol·lució, presència d'espais verds...) i mirant de minimitzar el temps de circulació. Aquestes recomanacions es basen en les dades disponibles a l'Open Data BCN⁶ i les recopilades a la xarxa de sensors que monitoren el nivell d'emissions contaminants⁷ i la qualitat de l'aire⁸. Les nostres simulacions indiquen que aquestes rutes verdes permeten una reducció significativa de l'exposició dels vianants o ciclistes a la contaminació i al soroll (del 20% al 40%, a costa d'augmentar del 10% al 20% el temps de trànsit), comparat amb les rutes recomanades per serveis estàndard com Google Maps.

A més, hem identificat una alternativa a la distribució habitual de mercaderies a les ciutats, coneguda en anglès com a *co-loading*: quan diversos proveïdors logístics comparteixen el mateix vehicle per distribuir els seus productes. En termes pràctics, significaria una reducció dels vehicles de transport a cada barri, ja que

un mateix vehicle de repartiment distribuiria les comandes de diversos proveïdors. Aquesta idea s'ha mostrat molt útil en logística industrial, unida a la de *backhauling*, que consisteix a aprofitar el viatge de tornada dels

Si la ciutadania hi coopera, la implantació del viatge compartit sembrarà de forats les jungles de cotxes en què es converteixen sovint els nostres carrers.

Nota

9. Organització de les Nacions Unides (ONU). <https://sdgs.un.org/goals>

camions buits per transportar altres mercaderies. La seva aplicació resulta prometedora a les ciutats (el que s'anomena *last mile delivery* en el món del transport).

Hi ha moltes altres aplicacions d'IA per millorar la mobilitat a les ciutats: transport públic adaptat de manera dinàmica a les demandes dels usuaris, semàfors intel·ligents, gestió de l'aparcament... Un segon nivell inclou vehicles autònoms i internet de les coses. En conjunt, aquestes funcionalitats defineixen un nou concepte de ciutat *smart* o intel·ligent, on les necessitats de les persones estan intervingudes per la tecnologia, a fi de satisfer-les de manera eficient i saludable.

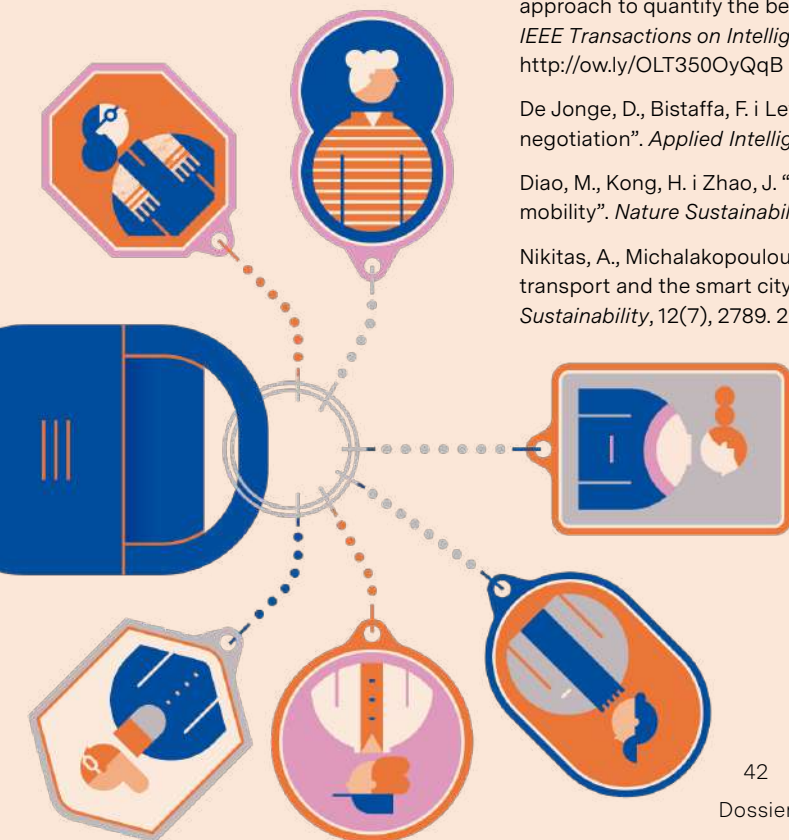
Les estratègies descrites en aquest article no són res més que granets de sorra en el que es coneix com a *intelligència artificial per al bé* (*AI for good*). Aconseguir ciutats més habitables, augmentar en quantitat i en qualitat els entorns naturals, reduir les desigualtats socials i millorar l'accés a la salut i al coneixement són alguns dels seus objectius bàsics. I aquestes dimensions de millora del món són les que donen un sentit més profund al nostre treball tècnic. Estan alineades amb els Objectius de

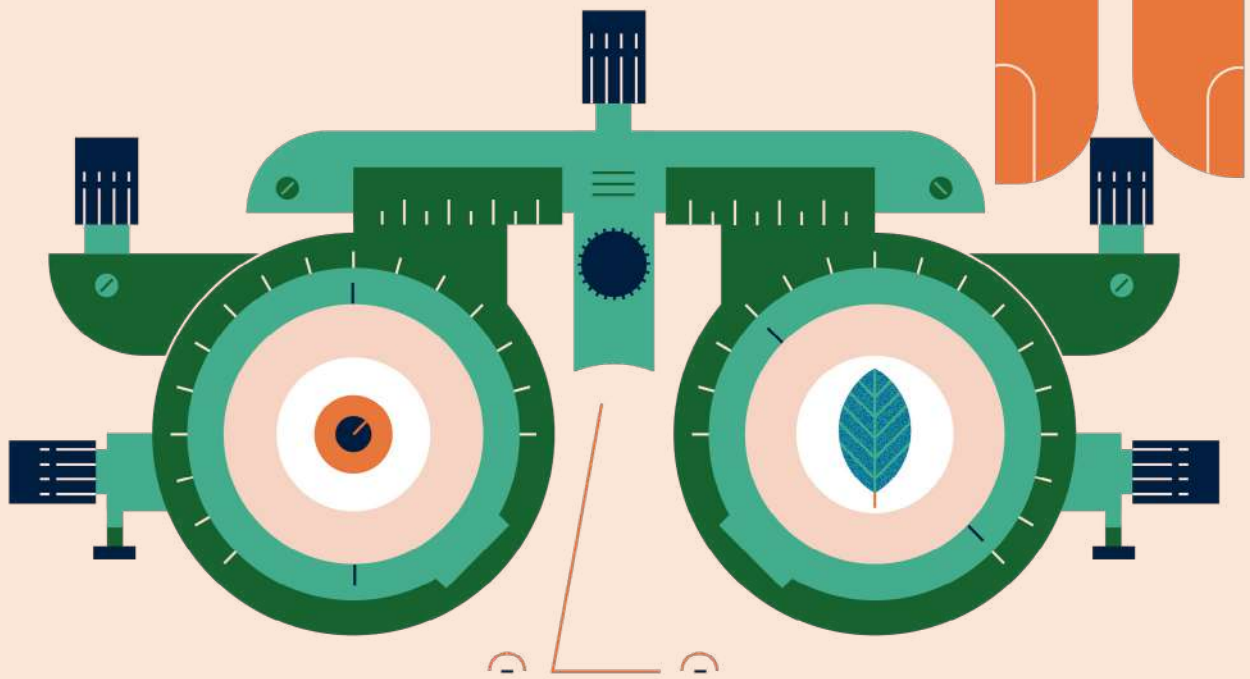
Desenvolupament Sostenible⁹, establerts per les Nacions Unides en la seva Agenda 2030. En aquesta iniciativa, la intel·ligència artificial hi té molt a dir.

La intel·ligència artificial per al bé (AI for good) pot aconseguir ciutats més habitables, augmentar en quantitat i en qualitat els entorns naturals, reduir les desigualtats socials i millorar l'accés a la salut i al coneixement.

Referències bibliogràfiques

- Bistaffa, F., Blum, C., Cerquides, J., Farinelli, A. i Rodríguez-Aguilar, J. A. "A computational approach to quantify the benefits of ridesharing for policy makers and travellers". *IEEE Transactions on Intelligent Transportation Systems*, 22(1), 119-130. 2021. <http://ow.ly/OLT350OyQqB>
- De Jonge, D., Bistaffa, F. i Levy, J. "Multi-objective vehicle routing with automated negotiation". *Applied Intelligence*, 52, 16916-16939. 2022. <http://ow.ly/Lfpl50OyQrc>
- Diao, M., Kong, H. i Zhao, J. "Impacts of transportation network companies on urban mobility". *Nature Sustainability*, 4, 494-500. 2021. <http://ow.ly/kZYh50OyQsx>
- Nikitas, A., Michalakopoulou, K., Njoya, E. T. i Karampatzakis, D. "Artificial intelligence, transport and the smart city: Definitions and dimensions of a new mobility era". *Sustainability*, 12(7), 2789. 2020. <http://ow.ly/HZJR50OyPKq>





Intel·ligència artificial ètica, responsable i per al bé social

Nuria Oliver

Quan la intel·ligència artificial es fa servir en escenaris que

afecten milions de persones, sorgeixen reptes de governança i dilemes ètics. Com podem estar segurs que les decisions algorítmiques s'apliquen sota el control social? Moltes de les propostes que es fan sobre principis ètics s'agrupen sota la sigla en anglès FATEN: justícia, autonomia, confiança, equitat i no maleficència.

Vivim en una època de prosperitat, però també ens enfrontem a uns reptes globals gegants que amenacen la nostra existència com a espècie. Sabem que necessitem intel·ligència artificial (IA) i, específicament, mètodes d'IA basats en l'aprenentatge a partir de les dades, que ens ajudin a abordar aquests reptes. El potencial és immens.

No obstant això, per fer realitat aquest potencial hem d'abordar nombroses barreres relacionades amb l'accés i l'anàlisi de dades valuoses —que, en molts casos, són de propietat privada—, amb una manca de principis ètics ben definits, amb limitacions legals i reguladores, amb reptes tècnics, amb uns clars interessos comercials que no estan necessàriament alineats amb el bé social i amb una petjada de carboni gens desdenyable dels sistemes actuals d'IA.

Quan els mètodes d'IA es fan servir en escenaris que tenen un impacte en la vida de milions de persones, com ara ciutats intel·ligents, sorgeixen desafiaments de governança i dilemes ètics importants. Com podem estar segurs que aplicar la IA de manera generalitzada no tindrà conseqüències negatives per a la societat? Les decisions algorítmiques quedaran fora del nostre control? Qui és responsable d'aquestes decisions? Aquests sistemes inclouran els mecanismes de seguretat necessaris per prevenir ciberatacs? Què passa amb l'ús maliciós de dades o la generació de contingut no veraç per servir interessos comercials, polítics o criminals? Aquestes preguntes són tan necessàries d'abordar com complexes, i fan que sigui imprescindible dotar-nos de principis ètics i normes de governança.

En l'última dècada s'han publicat nombroses propostes sobre principis ètics que cal respectar si es fa un ús generalitzat de la intel·ligència artificial, com ara l'*Informe Menlo*¹; els principis ètics inclosos en les estratègies nacionals d'IA de més de 50 països del món; l'informe de la Comissió Europea per al desenvolupament d'una intel·ligència artificial fiable²; els principis de l'OCDE per al desenvolupament de la intel·ligència artificial³; i diverses propostes professionals, com les de l'Institute of Electrical and Electronics Engineers (IEEE)⁴ i l'Association for Computing Machinery (ACM)⁵. La majoria dels principis proposats anteriorment es podrien agrupar en la sigla FATEN, que es detalla a continuació i que és una extensió dels quatre principis bàsics de l'ètica mèdica⁶.

F de justícia (fairness)

És a dir, sense discriminació. Els sistemes d'IA basats en l'aprenentatge a partir de dades poden discriminar per diversos motius, incloent-hi els biaixos en les dades utilitzades per entrenar els algorismes, una tria inadequada d'un algorisme o model, i una interpretació parcial dels resultats. En els darrers set anys s'han fet públics diversos casos de discriminació algorítmica d'allò més impactants en àrees com la justícia penal⁷, la concessió de crèdits⁸, els recursos humans i la contractació⁹, l'educació¹⁰ i l'atenció mèdica¹¹. La detecció i la mesura dels biaixos algorítmics i el desenvolupament d'algorismes justos d'aprenentatge automàtic són àrees fèrtils de recerca, tal com ho il·lustren la Conferència ACM FaccT sobre equitat, responsabilitat i transparència¹², creada recentment; el programa d'investigació ELLIS sobre aprenentatge automàtic



Nuria Oliver

Enginyera de Telecomunicació per la Universidad Politécnica de Madrid (UPM) i doctora en intel·ligència artificial pel Massachusetts Institute of Technology (MIT). És cofundadora i directora de la Fundació ELLIS Alicante —Instituto de Inteligencia Artificial centrada en la Humanidad—, vicepresidenta d'ELLIS Europa, Chief Data Scientist a la Data-Pop Alliance i Chief Scientific Advisor per a l'Instituto Vodafone. Ha signat més de 180 publicacions científiques i, segons Research.com, és la investigadora en informàtica dona amb un índex d'impacte més alt. És assessora de diverses universitats europees, de la Comissió Europea i de diversos governs en temes relacionats amb la IA. El seu treball ha estat reconegut internacionalment amb nombrosos premis.

Notes

1. Dittrich, D. i Kenneally, E. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. Departament de Seguretat Nacional dels Estats Units, 2012.
2. Comissió Europea. *Ethics guidelines for trustworthy AI*. Unió Europea, 2019.
3. OCDE. *Principios de la OCDE sobre IA*. OCDE, França, 2019.
4. Institute of Electrical and Electronics Engineers. *Diseño éticamente alineado*. Piscataway, NJ: IEEE, 2017.
5. Association for Computing Machinery. *Código de Ética y Conducta Profesional*. ACM, Nova York, 2018.
6. Gillon, R. "Medical ethics: four principles plus attention to scope". *British Medical Journal*. Regne Unit, 1994.
7. Angwin, J., Larson, J., Mattu, S. i Kirchner, L. *Machine Bias: There's software used across the country to predict future criminals and it is biased against blacks*. Pro Publica, Nova York. 2016.
8. Blattner, L. i Nelson, S. *How Costly is Noise? Data and Disparities in Consumer Credit*. 2021. <http://ow.ly/WpF250OkhuO>
9. Dastin, J. *Amazon scraps secret AI recruiting tool that showed bias against women*. 2018. <http://ow.ly/vRcB50Okhro>
10. Perry, A. M. i Turner, N. *AI is coming to schools, and if we're not careful, so will its biases*. Brookings, 2019. <http://ow.ly/wgXN50OkhsJ>
11. Ledford, H. *Millions of black people affected by racial bias in health-care algorithms*. 2019. <http://ow.ly/nQiH50Okhti>
12. <https://facctconference.org/>.
13. <https://ellis.eu/programs/human-centric-machine-learning>
<https://ellisalicante.org>

centrat en l'ésser humà, o l'Institut de Intel·ligència Artificial centrada en la Humanidad¹³, també de creació recent, que és una de les 39 unitats de la xarxa ELLIS d'excel·lència científica en intel·ligència artificial.

A d'autonomia, rendició de comptes i augment de la intel·ligència

El principi d'autonomia es troba en el cor de l'ètica occidental. D'acord amb aquest principi, cada persona hauria de ser capaç d'escollir lliurement els seus pensaments i les seves accions. No obstant això, podem construir models computacionals d'intel·ligència artificial sobre les nostres personalitats, interessos, gustos, necessitats, fortaleces/debilitats i comportaments que siguin —i probablement són— emprats per influir subliminarment en les nostres decisions, opcions i accions.

Per tant, hem de garantir que els sistemes d'IA que tenen un impacte directe o indirecte en la vida de les persones respectin sempre l'autonomia humana. La lletra *A* també és rendició de comptes (*accountability*, en anglès), és a dir, claredat en l'atribució de responsabilitat relacionada amb les conseqüències de l'ús de mètodes d'IA. Finalment, la *A* també significa augmentar la intel·ligència, en comptes de reemplaçar-la: els sistemes d'IA s'han de fer servir per donar suport i complementar la presa de decisions humanes i no pas per substituir els humans.

T de confiança (trust) i transparència

La confiança és un pilar fonamental en les nostres relacions i sorgeix quan es compleixen tres condicions: competència, que és la capacitat de dur a terme amb èxit la tasca compromesa; fiabilitat, entesa com la competència sostinguda en el temps, i honestedat i transparència. Per tant, la *T* a FATEN també hi és per a la transparència. Un sistema de presa de decisions basat en dades és transparent quan els no experts poden observar-lo i entendre'l amb facilitat. Els models d'IA transparents, interpretables i explicables són especialment necessaris en els casos d'ús amb impacte social.

E de bEneficiència i equitat

El principi de la *bEneficiència* es refereix a maximitzar l'impacte positiu de la IA amb sostenibilitat, diversitat i veracitat. No podem ignorar el cost ambiental del desenvolupament tecnològic, especialment quan es tracta d'algoritmes d'aprenentatge profund, atesa la necessitat de grans quantitats de computació i dades de les quals poden aprendre.

La diversitat també és essencial. En primer lloc, en la composició dels equips que desenvolupen sistemes d'IA, per maximitzar la probabilitat de

trobar solucions innovadores als reptes immensos als quals ens enfrontem¹⁴ i desenvolupar solucions inclusives que siguin rellevants a les comunitats on s'implementaran. En segon lloc, cal incorporar criteris de diversitat en els algoritmes per minimitzar la prevalença

Hem de garantir que els sistemes d'IA que tenen un impacte directe o indirecte en la vida de les persones respectin sempre l'autonomia humana.

Notes

14. Levine, S. R. *Diversity Confirmed To Boost Innovation And Financial Results*. Forbes, 2020.
<http://ow.ly/jMlp50OkhvS>

dels filtres bombolla i els efectes de la càmera de ressonància, que contribueixen, almenys parcialment, a polaritzar l'opinió pública.

També hem de garantir la “veracitat” de les dades. Actualment podem generar algorítmicament textos, àudios, fotos i vídeos falsos mitjançant models generatius de xarxes neuronals profundes (coneguts com *deep fakes*) que són indistingibles, per als humans, del contingut real. Aquests continguts falsos, compartits per milions de persones, poden tenir impacte en la formació de l'opinió pública sobre temes de gran rellevància social, com ara les vacunes o els processos electorals.

E també significa *equitat*. Els principis de l'accés universal al coneixement i la democratització de la tecnologia estan en perill a causa del domini extrem dels gegants tecnològics dels Estats Units (Apple, Amazon, Microsoft, Facebook i Alphabet/Google) i la Xina (Tencent, Alibaba o Baidu). Aquest domini del mercat condueix al domini de les dades. De fet, la majoria

L'accés universal al coneixement i la democratització de la tecnologia estan en perill a causa del domini extrem dels gegants tecnològics dels Estats Units i la Xina.

d'aquests gegants tecnològics són empreses de dades que guanyen milers de milions de dòlars analitzant i monetitzant la informació que recopilen sobre els seus usuaris. Una part significativa d'aquestes dades tan valuoses sobre el comportament humà, que es podrien fer servir per al bé social,

la generen i capturen aquestes empreses tecnològiques.

D'altra banda, en el segle XXI observem una polarització en la distribució de la riquesa. Segons el *Global Wealth Report* de Credit Suisse, les cent persones més riques del món acumulen més riquesa que els 4.000 milions de persones més pobres. Aquesta acumulació en mans de ben pocs s'ha atribuït, almenys de manera parcial, a la tecnologia i a la Quarta Revolució Industrial. Es podria argumentar que les dades, i especialment la capacitat d'aprofitar-les i donar-los sentit, són l'actiu que genera més riquesa (economia de dades). Per tant, si el nostre objectiu és maximitzar l'impacte positiu d'aquesta abundància de dades, hem de desenvolupar i promoure nous models de propietat, gestió, explotació i regulació d'aquestes.

N de no maleficiència

És a dir, minimitzar l'impacte negatiu que podria resultar de l'ús de mètodes d'IA. S'inclouen dins d'aquest principi la “prudència” en el desenvolupament de sistemes basats en IA; les garanties de fiabilitat, robustesa i reproductibilitat; la seguretat de les dades, i la preservació de la “privacitat”. És imprescindible verificar la seguretat i avaluar de manera exhaustiva els models d'IA abans de desplegar-los en el món real, particularment de xarxes neuronals profundes.

Com que els reptes són de naturalesa complexa, cal una combinació d'experts i expertes de diverses disciplines per tal que els projectes siguin un èxit. Aquesta naturalesa multidisciplinària ho fa encara més complex, però és necessària i particularment beneficiosa, ja que els equips inclourien especialistes en ètica.

Els òrgans de supervisió externa també són necessaris per garantir el compliment dels principis ètics. En els darrers anys, s'ha proposat crear la figura de l'administrador de dades (*data stewards*), responsable de la qualitat i governança de les dades en cada organització; comitès externs d'ètica de supervisió, i/o el nomenament d'un conseller/a amb responsabilitats de supervisió per garantir que els projectes amb impacte social estiguin alineats amb els principis de FATEN.

Una alternativa és exigir l'ús de processos, codis i sistemes oberts, a través del desplegament de regulacions que requereixin el seguiment de principis ètics i/o fomentant l'intercanvi de coneixement. A més, és imprescindible entendre les característiques culturals i socials de les societats en les quals es despleguen els projectes. Per tant, és absolutament necessari treballar amb les institucions locals i la societat civil dels països i les regions on es duran a terme els projectes.

És hora d'actuar com a societat, definint col·lectivament els horitzons ètics i polítics de la IA, perquè estem parlant de ciència i tecnologia, però també de drets, d'economia, de democràcia, d'igualtat, d'inclusió, de ciutadania, de pau i de poder.

És hora d'educar, de conèixer, de no caure en el sensacionalisme apocalíptic, de ser amos del nostre destí, de regular la IA amb intel·ligència i de centrar-nos a aturar les pràctiques abusives i els danys socials causats per les empreses que hi ha darrere dels avenços de la intel·ligència artificial, que en l'última dècada han acumulat un poder sense precedents i han contribuït a la desigualtat social.

És hora d'invertir en una intel·ligència artificial que contribueixi al progrés, sense deixar ningú enrere i sense destruir, durant el procés, el planeta. Com deia Theodore Roosevelt, “de vegades cal una revolució”. No tenim un planeta B, de manera que us convido a unir-vos a la revolució de la IA per i per a les persones i el planeta.

Referències bibliogràfiques

Oliver, N. “Artificial Intelligence for Social Good: The Way Forward”, capítol 11 del *SRIP Report 2022. SRIP Report of the European Commission*, 664-707. 2022.

Oliver, N. “Governance in the Era of Data-driven Decision-making Algorithms”. *Women Shaping Global Economic Governance*. Center for Economic Policy Research Press (CEPR), 2019.

Geschke, D., Lorenz, J., i Holtz, P. “The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers”. *British Journal of Social Psychology*, 58, 129-149. 2019.

Shorrocks, A. i Hechler-Fayd'herbe, N. *Global Wealth Report 2019: Global wealth rises by 2.6% driven by US & China, despite trade tensions*. Credit Suisse, Zúric, 2019.

Verhulst, S. G. “The Three Goals and Five Functions of Data Stewards. *Data Stewards: a new Role and Responsibility for an AI and Data Age*”. Medium i The Data Stewards Network, Nova York, 2018.