

PODEN PENSAR LES MÀQUINES? QUATRE TESIS SOBRE LA INTEL·LIGÈNCIA ARTIFICIAL

Joan Albert VICENS

Resum

Aquest article exposa quatre tesis sobre la naturalesa intel·ligent de la intel·ligència artificial en comparació amb la intel·ligència humana: 1) Un sistema equivalent a la intel·ligència humana que no se'n distingeix pròpiament; 2) Un poder destinat a desbordar-la i fins i tot a dominar-la (superintel·ligència); 3) Quelcom d'essencialment diferent de la nostra intel·ligència, o bé 4) capaç d'imitar-la i obtenir almenys els mateixos resultats que ella. Més enllà del debat sobre el que és o pot arribar a ser la intel·ligència artificial, es mostra també que en qualsevol cas ja emergeix invasivament com una tecnologia que substitueix les persones en el treball i en les seves capacitats, els imposa les seves maneres d'accedir a la informació i al coneixement i les reemplaça finalment en les seves decisions. La col·laboració entre ambdues intel·ligències s'ha de gestionar garantint la preservació de la dignitat, la llibertat i la responsabilitat humanes.

Paraules clau: intel·ligència artificial, màquines, intel·ligència humana, superintel·ligència, naturalesa.

Can the Machines Think? Four thesis on Artificial Intelligence

Abstract

This article presents four theses on the intelligent nature of artificial intelligence in comparison with human intelligence: 1) A system equivalent to human intelligence that is not properly distinguished from it; 2) A power destined to overwhelm it and even dominate it (superintelligence); 3) Something essentially different from our intelligence; or 4) being able to imitate it and obtaining at least the same results as it. Beyond the debate on what artificial intelligence is or could become, it is also shown that in any case, it already emerges invasive as a technology that replaces the person in their work and their capacities. It also imposes its ways of accessing information and knowledge and ultimately replaces the person in their decisions. Collaboration betwe-

en both intelligences must be managed to ensure the preservation of human dignity, as well as their freedom and responsibility.

Key Words: Artificial Intelligence, Machines, Human Intelligence, Superintelligence, Nature.

Compondre una peça musical que sembla l'obra d'un clàssic, redactar un article que resisteix l'examen *peer review* d'una revista científica, seleccionar el millor candidat per a un lloc de treball, diagnosticar un melanoma, decidir la llibertat condicional d'un presoner, dirigir una orquestra simfònica... Són exemples d'activitats que sempre hauríem considerat exclusives de persones intel·ligents i que poden ser realitzades avui per intel·ligències artificials (en endavant, IA).

No és estrany que això ens confongui. Són realment intel·ligents, les IA? Ens tocarà cedir el lloc a IA que, cada cop més, s'apropriaran de tasques humanes? O els humans i les IA col·laborarem pel progrés i el benestar de la humanitat? Hem de témer que les IA superin les persones i que fins i tot prenguin el control de la societat?

Els debats sobre aquestes qüestions inunden la bibliografia sobre la IA, el transhumanisme i el posthumanisme. Sovint el discurs abandona la terra ferma dels fets i s'enfila, a l'estil del colom kantià, en una atmosfera ultrallegera de ciència-ficció on els arguments i les conclusions esdevenen agosarats i inquietants. No se sap quan passem la línia que separa el saber real de la simple fantasia amb reminiscències científiques. Margaret Boden, especialista en ciències cognitives, diu que «la IA ha posat en entredit el nostre concepte de la humanitat i el seu futur».¹

Ara bé, què entenem per IA?² Boden diu que «la IA té com a objecte que els ordinadors facin la mateixa mena de coses que pot fer la ment».³ Es tracta de tecnologies que fan tasques complexes que l'home només pot fer en virtut de la seva intel·ligència i que tenen una dimensió formal (llenguatges de programació, algorismes, etc.) i una base material (processadors, unitats de memòria, targetes gràfiques, servidors, robots, escàners, etc.). D'ençà del seu naixement oficial el 1957, la IA ha evolucionat a batzegades, se n'han experimentat uns quants models (simbòlic, connexionista, evolutiu...) i s'ha basat en diferents lògiques i llenguatges de programació. Fins i tot ha passat algun període d'«hivern», de paràlisi, fins a l'actual present «primaveral», on experimenta una implantació espectacular a tot arreu i ja forma part de la quotidianitat de les nostres vides.

¹ M. A. BODEN, *Inteligencia artificial*. Madrid: Turner, 2017, pàg. 13.

² R. LÓPEZ DE MANTARAS i P. MESEGUER, *Inteligencia artificial*. Madrid, CSIC, 2017. En aquest llibre hi trobem una exposició assequible de la història i el present de la IA.

³ M. A. BODEN, *op. cit.*, pàg. 11.

Els sistemes de què disposem avui són IA de tipus *específic*, sistemes experts en relació amb una qüestió concreta però incapaços de fer altres coses, i *dèbil*, és a dir, que fan coses semblants a nosaltres sense pensar com nosaltres. No disposem encara de cap IA *general*, que s'apliqui a qualsevol tasca, ni *forta*, que no tan sols imiti la ment sinó que sigui capaç de pensar com un ésser humà.⁴

La IA ha evolucionat des del processament de símbols segons regles a l'aprenentatge automàtic (*machine learning*). Tenim sistemes capaços de descobrir correlacions en immenses quantitats de dades relatives a qualsevol cosa: la IA és avui sobretot un «motor de classificació», un «poder de comparació»⁵ que usa les correlacions que descobreix per inferir hipòtesis o substituir operacions que fem les persones: identificar objectes, entendre discursos i parlar, predir fets, planificar tasques, resoldre problemes, assolir noves capacitats, etc. En el màxim nivell de la IA trobem el *deep learning*, aprenentatge profund: sistemes capaços de descobrir patrons en muntanyes de dades segons diversos nivells d'abstracció, màquines que aprenen soles i que, en aquest sentit, s'autoprogramen. Més que per la lògica i la deducció, la IA està avui dominada per l'estadística, el càlcul de probabilitats, la teoria de jocs, la inducció. P. Cardon ho resumeix dient que «els dissenyadors han abandonat l'ambició de fer màquines “intel·ligents” [= deductives]. Prefereixen fer-les “estadístiques”».⁶

Els sistemes d'IA poden combinar diferents recursos: càlcul, anàlisi de dades, aprenentatge, visió artificial, navegació, processament del llenguatge, robòtica, etc. Un «sistema integrat d'IA»⁷ podria associar tots aquests recursos, tal com passa en l'ésser humà. A la llarga, es vol crear una IA general i forta: «la intel·ligència, allò que permet comprendre el món que ens envolta, adaptar-nos als canvis, anticipar i predir, és l'objectiu a assolir».⁸

Hom discuteix, però, sobre la semblança o no entre IA i la intel·ligència humana (en endavant, IH) i sobre la possibilitat que arribi a existir una IA general i forta. Tanmateix, el debat sobre IA i IH es mou entre *dues ignoràncies* que el condicionen essencialment. En primer lloc, no sabem què podrà donar de si la IA en el futur. Els canvis que experimenta aquesta tecnologia són vertiginosos. La computació quàntica, que es troba a les beceroles, significarà un salt abismal. Joan Ignasi Cirac, director de l'Institut Max Planck de Munic, assegura que els futurs computadors quàntics (ara en tenim només alguns prototips) podran realitzar en pocs minuts càlculs que amb un ordinador actual

⁴ La distinció d'IA dèbil o forta procedeix de J. SEARLE: *Minds, Brains, and Programs*, pàg. 2 de l'edició digital a <http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.searle2.html>.

⁵ E. SADIN, *L'intelligence artificielle ou l'enjeu du siècle*. París: Ed. L'échappé, pàg. 158.

⁶ P. CARDON, *Con qué sueñan los algoritmos*. Madrid: Dado Ed., pàg. 78.

⁷ R. LÓPEZ DE MANTARAS i P. MESEGUER, *op. cit.*, posició a Kindle: 1703.

⁸ W.AA. AUTORS, *Intelligence artificielle*. París: Champs actuels, 2017, pàg. 14.

requeririen tant de temps com l'edat de l'univers.⁹ No sabem què significarà aquesta potència descomunal, si s'arriba a concretar, ni què donaran de si les tecnologies del futur.

D'altra banda, també tenim un coneixement molt limitat del funcionament del nostre cervell. En realitat, sabem poca cosa dels mecanismes biològics i fisiològics del pensament complex i la consciència humana. Ho reconeix Juan Lerma, president de la Societat Espanyola de Neurociència: «El problema de la cognició, d'entendre com es genera la consciència, no està resolt. Cal treballar més sobre aquest i altres temes fonamentals del funcionament del sistema nerviós. Del cervell en sabem molt, però ens queda més encara per descobrir i descriure».¹⁰ La imitació tecnològica del cervell i la comparació de la IA amb la ment lligada al nostre cervell avancen a peu coix.

En aquest article, assumint aquesta doble ignorància, presentarem quatre tesis sobre la naturalesa de la IA, sobre allò que és o arribarà a ser (1-4), i una reflexió afegida (5) sobre el seu impacte en la nostra vida personal i social.

1. La IA equival a la IH

Poden pensar les màquines? Alain Turing deia que per respondre aquesta pregunta, que ell es feia el 1947, cal primerament aclarir els significats de «màquina» i de «pensar», però reconeixia la dificultat de fer-ho de manera unànime.¹¹

Més expeditius, altres autors aposten per la tesi que les màquines pensen com les persones perquè les persones també pensen com les màquines. La IA reproduïx bàsicament el pensament humà, perquè pensar, segons ells, és *computar* informació; quan pensem, «calculem» un resultat a partir de dades prèvies; el cervell és un ordinador supercomplex que rep dades i hi respon amb una reacció adaptativa. No hi hauria diferència *qualitativa* entre el sistema biològic i el sistema informàtic, entre un procediment mental i un algorisme, entre el coneixement i la informació. Aquestes equivalències són la base de l'anomenat «paradigma cognitivista», on s'inclou la «metàfora computacional» de la ment.¹²

Allen Newell i Herbert Simon van veure en el concepte de «sistema de símbols físic» el nexa entre IA i IH. Un sistema de símbols es pot transformar d'acord amb regles i relacions en un altre sistema de símbols. Aquest procés es realitza sobre un substrat físic, electrònic o biològic. La naturalesa del substrat no és rellevant si facilita la trans-

⁹ Entrevista a Joan Ignasi Cirac, «Tendremos tecnologías cuánticas en nuestros móviles», *El País*, 21-10-2018: https://retina.elpais.com/retina/2018/10/20/tendencias/1540022033_896762.html

¹⁰ J. LERMA, «El cerebro, ese gran reto», *SEBB, Revista de la Sociedad española de bioquímica y biología molecular*. Madrid, 183, Març 2013, pàg. 4.

¹¹ A. M. TURING, «Computing Machinery and Intelligence». *Mind*, vol. LIX, núm. 236, oct. 1950, pàg. 433.

¹² M. CARABANTES, *Inteligencia artificial: una perspectiva filosófica*. Salamanca: Escolar y Mayo Ed., 2016, pàg. 100.

formació del sistema de símbols. En un ordinador aquest procés s'anomena computació, i en un cervell, pensament.¹³ Deixem de banda si el seu curs és deductiu, com en la «IA al vell estil», o inductiu (estadístic-probabilístic), com és el cas del *machine learning*: també la ment fa inferències deductives o inductives, a conveniència. La IH i la IA són formalment equivalents encara que difereixin materialment.

Max Tegmark justifica la tesi de l'equivalència sobre la base d'una definició simple d'intel·ligència predicable igualment de la IH i la IA: «la intel·ligència és la capacitat d'assolir objectius complexos». Construït segons el model de la «màquina universal de Turing»,¹⁴ un ordinador és capaç de transformar qualsevol conjunt de símbols en un altre conjunt de símbols d'acord amb un procediment (algorisme) que segueix un nombre finit de passos fins a obtenir-ne un resultat. Com que això també ho podem dir del cervell, és indiferent que la computació es faci en una màquina o en un cervell: tots dos transformen informació d'un estat en un altre, reben *inputs* i emeten *outputs*. La intel·ligència és pròpiament quelcom «intangible, abstracte i eteri» –diu Tegmark–, «resulta tan poc físic perquè és independent del substrat i adquireix una vida pròpia que no reflecteix els detalls físics ni en depèn».¹⁵ La intel·ligència és programari i és independent del seu maquinari: una xarxa neuronal artificial o les neurones d'un cervell. Així com podem estudiar la transmissió i les propietats de les ones en diferents substrats (aigua, aire, etc.), també podem autonomitzar els processos intel·ligents del seu suport material i considerar-los en la seva estricta formalitat computacional: «la intel·ligència es redueix a informació i computació, no pas a carn, sang i àtoms de carboni».¹⁶

L'equivalència entre IA i IH es manté també, seguint Tegmark, si baixem al detall. Les funcions intel·ligents bàsiques (registre d'informació, memòria, computació, aprenentatge, intencionalitat, consciència, etc.) són independents del seu substrat material i ja les fan o les faran aviat els sistemes artificials. Es pot dir que la IA té intencionalitat, com la IH, perquè actua en funció d'objectius i optimitza els recursos disponibles. La IA també podrà «sentir-se» lliure, com els humans, si entenem la llibertat com l'experiència subjectiva (que no implica «llibertat real») que tenim que un procés intel·ligent resta obert abans de completar-lo. Si algú li nega caràcter intel·ligent a la IA perquè no és conscient del que fa, respondrem que l'activitat intel·ligent no ha d'anar necessàriament lligada a l'autoconsciència; de fet, només algunes activitats intel·ligents humanes són conscients, les més clarament executives. Com que la consciència és «l'experiència

¹³ R. LÓPEZ DE MANTARAS i P. MESEGUER, posició a Kindle: 60-67.

¹⁴ A. M. Turing va presentar la noció de «màquina universal» en l'article fundacional de la informàtica: A. M. TURING, «On Computable Numbers, with an Application to the *Entscheidungsproblem*». *Proceedings of the London Mathematical Society*, 2 (1937) 42, pàg. 230-265.

¹⁵ M. TEGMARK, *Vida 3.0*. Madrid: Taurus, 2018, pàg. 87-89.

¹⁶ *Idíd.*, pàg. 75.

subjectiva d'allò que fem», no és descartable la creació de màquines conscients quan compreguem de quins processos complexos de computació de la informació emergeix la consciència i sapiguem replicar-los en substrats no biològics. La inconsciència permanent de la IA és, per tant, provisional.¹⁷

En definitiva, IA i IH són bàsicament el mateix i entre elles només hi ha, de moment, una diferència de capacitats, no pas de «naturalesa». Té tota la lògica del món esperar que IA i IH s'equiparin *algun dia* perquè *ja són essencialment el mateix*. En general, assegura Tegmark: «No és gaire interessant traçar un límit artificial entre la intel·ligència i la no-intel·ligència i resulta més útil quantificar el grau de capacitat per aconseguir diferents objectius».¹⁸ La intel·ligència entesa com a poder de computació per resoldre problemes ha de ser susceptible de gairebé tots els graus imaginables. Només cal veure les possibilitats actuals dels ordinadors per adonar-se que, en alguns aspectes, desborden les capacitats humanes. Un microprocessador bàsic de 2GHz computa informació set vegades més de pressa que el nostre cervell; la informació es mou per canals artificials a la velocitat de la llum, i al nostre cervell, a 120 m/s aproximadament; el cervell humà té 100.000 milions de neurones, però no hi ha límits per a un sistema intel·ligent artificial, que no s'ha encabir dins d'un crani.¹⁹ J. I. Cirac aposta que un computador quàntic de 300 qbits tindrà més poder de computar que un sistema equivalent a la totalitat dels àtoms de l'univers.²⁰ La memòria humana tampoc no pot «competir» amb l'artificial pel que fa a la «quantitat» d'informació disponible i a la velocitat d'accés a aquesta informació: la IA té tot el que hi ha a internet immediatament a l'abast. A més, el cervell humà es fatiga i es deteriora, mentre que un sistema d'IA pot funcionar indefinidament i en podem canviar el maquinari quan vulguem.

Tot plegat ens obliga a estimar el pronòstic que, malgrat el seu parentiu inicial, la IA i la IH estan condemnades a separar-se amb l'esdeveniment de la «superintel·ligència», una nova realitat que consumaria tot el potencial de la IA.

2 La IA desborda la IH: la superintel·ligència (SI)

L'informàtic visionari Ray Kurzweil i el professor d'Oxford Nick Bostrom auguren la creació, tard o d'hora, d'una «superintel·ligència» (en endavant SI), és a dir, una IA *general*, capaç de desenvolupar habilitats de tota mena, i *forta*, capaç de fer-ho a un

¹⁷ M. Tegmark desplega aquests arguments al llarg del llibre *Vida 3.0*, *op. cit.*

¹⁸ M. TEGMARK, *op. cit.*, pàg. 71.

¹⁹ N. BOSTROM, *Superintelligence*. Oxford: Oxford Univ. Press, 2014. Citem l'edició castellana: *Superinteligencia*. TEELL, 2016, pàg. 54-61. Encara s'hi podrien afegir altres avantatges de la IA: estabilitat, capacitat de publicació o de compartir memòria, coordinació o ampliació indefinida.

²⁰ Així ho explica el físic J. I. Cirac, director de Institut Max Planck de Munic, en una entrevista amb Iñaki Gabilondo de la sèrie *Cuando ya no esté*: <https://www.youtube.com/watch?v=Tu7P7J3rXtY>

nivell humà o superior a l'humà en tots els aspectes. Aquesta SI serà, diu Bostrom, «qualsevol intel·lecte que excedeixi en gran mesura la capacitat cognitiva dels humans en pràcticament totes les àrees d'interès»; o més gràficament: «podria ser més adequat pensar que una IA és intel·ligent com diem que ho és un ésser humà mitjà en comparació amb un escarabat o un cuc de terra».²¹ La IA serà, promet Kurzweil, «uns mil milions de cops més potent que la reunió de tots els cervells humans».²² La SI serà, doncs, la IA en sentit absolut, en la plenitud il·limitada de la seva naturalesa intel·ligent. Dit d'una altra manera, la SI és el destí de la IA; en ella, si arriba a existir, es consumarà la naturalesa de la IA.

L'anomenada llei de Moore²³ i els presagis sobre la computació quàntica apunten una futura «explosió d'intel·ligència» o «singularitat» i, per tant, l'emergència d'un sistema superintel·ligent, probablement ubicat a internet, incontrolable i capaç de reconfigurar-se i millorar-se a si mateix contínuament. Aquesta SI seria l'última baula del que Tegmark anomena la Vida 3.0, una forma d'IA que encara no existeix i que es caracteritzaria per la seva capacitat de replicar-se a si mateixa, de redissenyar contínuament el seu programari i el seu maquinari. Els humans, en canvi, som capaços de dissenyar el nostre programari (la cultura), però no podem de moment alterar substancialment el nostre maquinari corporal, que depèn d'un llarg procés de selecció natural i de la transmissió genètica.²⁴

Tegmark troba lògica la conclusió que si arriba a existir una SI ens dominarà com nosaltres hem dominat els animals menys intel·ligents que nosaltres.²⁵ Bostrom ens posa en guàrdia davant d'una futura SI: «Una SI artificial podria ser en si mateixa un agent molt potent, que podria tenir èxit en afirmar-se davant del projecte que li va donar l'existència, i contra la resta del món».²⁶

En aquest punt, estem obligats a distingir dos escenaris. El primer seria la creació d'una SI a partir de la multiplicació bestial de les capacitats bàsiques de la IA actual: accés a dades i càlcul. La segona possibilitat és que algun dia es creï a partir de noves tecnologies una SI general i forta que repliqui perfectament i superi descomunament les capacitats humanes, que no es poden reduir a dades i càlcul.

En aquest article només podem tenir en compte la primera possibilitat, que es plantegen autors com Tegmark o Bostrom. La segona no es refereix a un impossible meta-

²¹ N. BOSTROM, *op. cit.*, pàg. 22, 93.

²² L. ALEXANDRE i J. M. BESNIER, *Les robots font-ils l'amour?* Malakoff: Dunod, 2016, posició a Kindle: 963.

²³ L'anomenada «llei de Moore» estableix que cada 18 mesos es duplica la capacitat de les unitats de memòria i la velocitat dels processadors.

²⁴ M. TEGMARK, *op. cit.*, pàg. 43-45.

²⁵ *Ibid.*, pàg. 172.

²⁶ N. BOSTROM, *op. cit.*, pàg. 95.

físic però ens situa en un terreny on qualsevol cosa es pot dir, com podem fantasiejar sobre *cyborgs* o intel·ligències extraterrestres. Les elucubracions sobre la SI són altament especulatives: Kurzweil, per exemple, imagina la SI del futur com una realitat omnipotent i incompreensible que podria créixer indefinidament fins a sortir de la Terra, ocupar la galàxia, «impregnar l'univers i decidir el seu destí».²⁷ Déu n'hi do...

Bostrom o Tegmark, més moderats que Kurzweil, anticipen els reptes i les amenaces que una SI representaria per a la humanitat i n'esbossen diverses hipòtesis: una SI sota control humà amb finalitats benèfiques o operada per estats o grups de poder amb finalitats malèfiques; o una IA autonomitzada del control humà, amb finalitats pròpies no per força coincidents amb els interessos humans i que, o bé podria tenir cura de nosaltres, o bé bandejar-nos com un destorb.²⁸

Altres especialistes rebutgen com a ciència-ficció la idea d'una IA general i forta que resulti exclusivament de l'ampliació del poder de càlcul, sigui com sigui que calculi. M. Boden pensa que «els escèptics de la singularitat tenen més arguments al seu favor» i afegeix que, amb les seves «boges especulacions», els profetes de la SI ignoren la diferència essencial d'IA i IH i especulen alegrement sobre els efectes de la llei de Moore.²⁹ J. G. Ganascia critica amb contundència el «mite de la singularitat», interpreta la llei de Moore com una simple projecció de certes regularitats fàctiques que no imposa cap necessitat, sosté que «la intel·ligència no equival ni a la freqüència d'execució d'operacions elementals, ni al nombre d'informacions acumulades en una memòria», i conclou que «ni el creixement de la potència de càlcul, ni de la capacitat d'emmagatzemar, no produeixen automàticament la intel·ligència».³⁰ Podem augmentar tant com vulguem les funcions elementals (càlcul, memòria, etc.) de la IA, podem multiplicar-ne la capacitat d'establir i aplicar correlacions: el que s'esdevindrà, creu Ganascia, no serà cap intel·ligència equiparable a la nostra i tampoc una SI que prengui el control de la humanitat.

El debat sobre la SI exposa en termes exagerats el que es pot dir en termes més limitats de tota IA: que emergeix com un nou poder davant l'ésser humà. No cal esperar que es realitzin les profecies de Kurzweil o les pors de Bostrom o Tegmark per adonar-nos que, si qualsevol tecnologia s'apodera poc o molt de les nostres capacitats, la IA ho farà encara més, com un contrapoder amb capacitat de suplantar-nos en molts terrenys.³¹

²⁷ R. KURZWEIL, pàg. 392-421.

²⁸ M. TEGMARK, *op. cit.*, pàg. 204 ss.; N. BOSTROM, *op. cit.*, pàg. 115 ss.

²⁹ M. BODEN, *op. cit.*, pàg. 150-152.

³⁰ J. G. GANASCIA, *Le mythe de la singularité*. París: Ed. du Seuil, 2017, pàg. 450-454, 558.

³¹ En aquest sentit, centenars de científics i investigadors en IA, entre els quals Gates, Hawking, Musk, Wozniak, Tegmark, Russell, Dennet, etc., van signar el manifest «*Autonomous Weapons: An Open Letter from AI & Robotics researchers*», publicat el 28-7-2015 i que es pot llegir a <https://futureoflife.org>, que adverteix dels perills del sistema d'armament basats en IA que operen al marge del control humà.

Més endavant parlarem de com la IA ja substitueix i controla l'ésser humà en alguns àmbits. Ara seguirem Ganascia quan exigeix que afinem la distinció entre IA i IH. La diferència essencial d'IA i IH és precisament la tercera tesi sobre la IA, que passem a explicar.

3. La IA difereix essencialment de la IH

L'equiparació de la IA i la IH és freqüent entre tecnòlegs, però la tesi que IA i IH són ben diferents domina el terreny filosòfic. N'esbossarem els principals arguments, sense pretensió de desenvolupar-los a fons ni de negar la importància d'altres raons. Ens centrarem en les dimensions evolutiva-cultural, corporal, sentent-intel·lectual i personal-pràctica de la intel·ligència humana. Volem posar en relleu que entre IA i IH no existeix avui només una diferència de grau, sinó que són realitats essencialment diverses.

a) Dimensió evolutiva-cultural

D'acord amb la metàfora computacional, la cultura d'un ésser humà equivaldria al programari i a les dades amb què treballa un ordinador. Tanmateix, la IH pertany a l'animal humà: està associada en els seus orígens a la feblesa instintiva dels nostres ancestres, a l'activitat caçadora-recol·lectora de grups socials cada cop més grans que creaven i es llegaven un bagatge cultural. La cultura és alhora el producte i el medi de l'activitat intel·lectual i el principal recurs evolutiu de l'espècie. Per això deia A. Gehlen: «Si la cultura és natural per a l'home, mai no copsarem la seva naturalesa tal i com és, sinó impregnada de compostos culturals molt precisos».³²

Una IA no és un producte evolutiu i no podem dir que li donem una cultura subministrant-li dades o connectant-la a internet. La cultura no és simple informació, sinó un complex bagatge de formes de vida, representacions, preferències, projectes, valors, hàbituds, maneres de sentir, d'expressar i d'interpretar el món, de les quals ens fem càrrec més o menys reflexivament. Parlem d'un complex de possibilitats i capacitats rebudes de generacions precedents, apropiades o rebutjades, transformades o bescanviades per unes altres de noves, i lliurades a la generació següent (tradicció). La transmissió cultural no és una transmissió d'informació amb els seus biaixos, sinó «el lliurament de maneres d'estar en la realitat» que ens ubiquen en un context social i en un temps històric determinats i que ens projecten vers el futur.³³

³² A. GEHLEN, *Antropología filosófica*. Barcelona: Paidós, 1993, «La naturalidad de lo cultural», pàg. 97 ss.

³³ X. ZUBIRI, *Tres dimensiones del ser humano*. Madrid: Fundación X. Zubiri-Alianza Ed., 2006, pàg. 138 i 155.

Amb uns quants clics puc descarregar, gravar o esborrar dades i programes dins d'un computador; la cultura, en canvi, ens va conformant mitjançant un lent i complex procés d'aprenentatge i socialització que dura tota la vida i que va deixant una empremta estrictament individual en cadascú, fins i tot física i corporal, i que no podem esborrar com ho fem amb les dades d'un disc dur. Puc clonar dos ordinadors, però no hi ha clònics culturals. L'home no té cultura com una IA disposa d'accés a internet o al *big data* perquè és en si mateix cultural i la seva intel·ligència opera sobre la base d'una herència biològica i cultural, social i històrica: rep, rebutja, adopta, transforma i transmet possibilitats i capacitats. La manca de societat, de cultura o d'història de la IA és allò que la fa replicable i aliena a qualsevol lloc, acrítica davant la informació esbiaixada de què s'alimenta, tan dèbil en contextos ambigus on els humans ens movem com el peix a l'aigua. La IA és maldestra en situacions connotades culturalment i socialment, marcades per pressupòsits i prejudicis, determinades per allò que Berger i Luckmann anomenaven «el món donat per descomptat», la vida interpretada, la realitat travessada per la cultura.³⁴

b) Dimensió corporal

L'home «no és una unitat additiva de dues substàncies [el cos i la ment], sinó una unitat sistemàtica de notes», és a dir, «una unitat psicoorgànica».³⁵ La intel·ligència forma part del *sub*-sistema psíquic inserit en un *únic* sistema psicocorporal. Intel·ligir és una activitat complexa que forma sistema amb activitats com respirar, digerir, dormir, veure, escoltar, tocar, agafar una cosa, girar el cap, etc. El psíquic i l'orgànic són dues cares de la mateixa moneda: «No es pot parlar de psique sense organisme —escriu Xavier Zubiri—. Per la mateixa raó, no es pot parlar d'organisme humà sense psique».³⁶ Si moc un braç per agafar una copa de vi, poso en marxa el cervell i el sistema nerviós, el sistema circulatori, l'esquelètic i muscular, la respiració, el metabolisme cel·lular..., i tot això es produeix *en funció d'intencions*, representacions, significats, valors, finalitats, esquemes habituals de conducta i segurament d'una decisió conscient. No és la intel·ligència qui pensa, sinó l'ésser humà complet, amb les capacitats i condicions individuals de cadascú. La intel·ligència humana no és independent del seu suport, com diuen molts informàtics que ho és la IA en relació amb el seu maquinari. Podem traslladar la informació i els programes d'un ordinador a un altre —per a la IA, el substracte particular és indiferent—, però no puc replicar en una màquina la ment d'una persona perquè aquesta forma un únic sistema real amb el seu cos particular, ni garantir,

³⁴ P.L. BERGER i T. LUCKMANN, *La construcció social de la realitat*. Barcelona: Herder, 1996.

³⁵ X. ZUBIRI, *Sobre el home*. Madrid: Fundación X. Zubiri-Alianza Ed., 1986, pàg. 47 ss.

³⁶ *Ibid.*, pàg. 59.

com pretén Kurzweil, la pervivència d'una persona (un sistema psicoorgànic i no pas quelcom que «resideix» en un cos) descarregant tota la informació del seu cervell (*backup*) i introduint-la en un robot.³⁷

L'home no resideix en un cos: és el seu cos; només amb ell «és aquí» ara mateix i no en un altre lloc; amb el seu cos constitueix el «lloc geomètric de la realitat».³⁸ La IH és una funció de l'home «de carn i ossos», i no pas la més rellevant per a la nostra felicitat, insistia Unamuno; deia Ortega també que pensar és una funció vital més, al costat d'altres coses que fem per a sobreviure en la nostra circumstància: «el pensament, la intel·ligència, són una de les reaccions a què la vida ens obliga, té les arrels i el sentit en el fet radical, previ i terrible de viure».³⁹ En cap sentit acceptable es pot dir que una IA «visqui». L'avió més intel·ligent vola sense pilot, però no fa nius ni pon ous.

c) Dimensió sentent-intel·lectiva

L'objecte de la intel·ligència no són les «dades» matemàtiques amb què opera la IA, els mapes de 0 i 1 als quals s'acaba reduint la complexitat d'un fenomen prèviament mesurat, quantificat, ni les correlacions o els patrons de dades. Intel·ligir no equival a processar dades mitjançant algorismes. Dir-ho és una ingent simplificació que deixa fora allò específic de la IH.

El famós exemple de «l'habitació xinesa», de John Searle, serveix per explicar que no existeix cap veritable intel·lecció darrere d'un mecanisme de computació: una persona es tanca dins d'una habitació i li fan arribar unes preguntes en xinès, que no entén, però aprèn a respondre-hi correctament aplicant mecànicament regles de correspondència entre paquets de símbols xinesos i anglesos. Això no vol dir que tingui la més remota idea del que contesta; només es limita a computar, a «manipular signes», sense entendre res.

Les dades codificades que rep un computador tampoc no tenen per a ell cap sentit, no signifiquen res del món en què vivim; a diferència del llenguatge humà, els símbols que processa un ordinador estan mancats d'intencionalitat. «No hi ha cap raó per suposar –conclou Searle– que la meua comprensió tingui res a veure amb programes de computadora, és a dir, amb operacions computacionals d'elements especificats formalment».⁴⁰ Només podem dir metafòricament que l'ordinador *entén* com una porta automàtica també *entén* les instruccions d'una cèl·lula fotoelèctrica, però, més enllà d'això, no hi trobarem cap mena de comprensió del que fa. Cathy O'Neil diu que «als

³⁷ R. KURWEIL, pàg. 222 ss.

³⁸ X. ZUBIRI, *Sobre el hombre*, op. cit., pàg. 79.

³⁹ J. ORTEGA Y GASSET, OC, V, 743. Madrid: Taurus, 2006.

⁴⁰ J. R. SEARLE, «Minds, brains, and programs», a: *Behavioral and Brain Sciences*, 3 (3), 1980, pàg. 417-457.

programes de programari els costa distingir una idea d'una simple seqüència de paraules».⁴¹

Però intel·ligir no és primàriament comprendre significats, com diu Searle, sinó, com mostra X. Zubiri, «aprehensió de realitat».⁴² D'aquesta tesi zubiriana només en destaquem tres aspectes.

Primerament, l'objecte formal de la IH és la «realitat»: l'home aprehèn les coses *com a reals* i no com a estímuls estructuralment lligats a determinades respostes (com fan els animals), ni com a dades que desencadenen determinats processos de computació, com fan els algorismes intel·ligents. No anomenem «realitat» a les coses existents més enllà dels nostres actes, sinó a la «formalitat» o la «manera de quedar» les coses en els nostres actes d'intel·lecció: les coses se'ns hi presenten amb una radical alteritat, com a independents dels actes mateixos en què sorgeixen; *sobretot*, remetent a si mateixes, com a essent «de si» («de suyo») o «pròpiament» les seves notes.

Les coses així apreheses desapareixen tots els dinamismes de la intel·lecció. A més de sentir-les de maneres ben diferents, les relacionem amb altres coses i *en parlem (logos)*, i *ens fan pensar* (raó): la seva realitat ens empeny, poderosa, a esbrinar què són realment, per què són com són, fins al punt, sovint, de posar en qüestió tot el que s'hagi dit i cregut sobre elles. Per contra, la IA, fins i tot en la versió més refinada de l'aprenentatge profund, no passa de ser un formidable motor de classificació de dades, de descoberta de patrons que van decantant automàticament uns determinats resultats, però sense que mai es produeixi en la màquina el miracle evolutiu que representa l'aprehensió de realitat. Avui la IA pot ser «lògica»: relaciona dades i n'infereix conclusions, però mai no aprehèn realitats ni arriba a ser «racional»: no pensa, no *inventa* noves possibilitats (com a màxim, troba relacions inadvertides), no qüestiona els seus propis resultats, no indaga el fonament real de les dades que relaciona o de les relacions que descobreix.

En segon lloc, l'actualització d'una cosa com a real en la IH és l'«estar present la cosa des de si mateixa, des de la seva pròpia realitat», que es verifica en el més humil dels nostres actes de sentir. La intel·lecció humana és «*sentiente*», diu Zubiri. Sentir ja és «aprehendre el real en impressió»: el blau del cel en la visió o el so del violí en l'audició «queden» en les respectives percepcions, cadascun a la seva manera, en alteritat radical. Una IA no té aprehensió de realitat, sinó que processa i calcula símbols; necessita que li converteixin els fenòmens en dades matemàtiques computables digitalment, sense copsar-les mai com a «dades-de-realitat».

En tercer lloc, en qualsevol intel·lecció es co-actualitzen com a reals l'acte d'intel·lecció (visió, audició...) i allò que s'hi actualitza (la cosa vista, oïda...): «quan hi és

⁴¹ C. O'NEIL, *Armas de destrucción matemática*. Madrid: Capitán Swing, 2017.

⁴² Seguim bàsicament l'anàlisi de l'aprehensió de realitat de X. Zubiri en la seva trilogia *Inteligencia sentiente* (1980-1984).

present l'intel·ligit, per exemple, quan hi és present aquesta pedra, no tan sols veig la pedra, sinó que sento que estic veient aquesta pedra»; «així és com estic en mi», apunta Zubiri.⁴³ És la «*conciencia sentiente*», la co-presència en unitat indestruïble de l'acte i la cosa que obre camí a la reflexió o a la introspecció com a formes ulteriors d'«entrar en mi mateix».⁴⁴

Avui no hi ha IA autoconscients, però Tegmark diu que no hi ha cap impediment metafísic per crear formes complexes de processar la informació d'on emergeixi en el futur, com passa amb el nostre cervell, «l'experiència subjectiva» que s'anomena consciència: «Si la consciència és la manera en què la informació se sent quan es processa d'unes determinades maneres, aleshores ha de ser independent del substrat; l'únic que importa és l'estructura del processament de la informació»; la consciència seria el «fenomen emergent de més alt nivell» a què podria donar lloc una molt complexa forma de computació.⁴⁵

La tesi de Tegmark no suggereix un impossible, però resulta especulativa i en tot cas faria referència a una forma de consciència no humana, atès que la consciència «humana és específica del nostre sistema psicoorgànic, corporal, sensible, afectiu, volent... Avui ningú parla de màquines conscients ni es poden equiparar, més enllà d'una lleugera analogia, els processos pels quals la IA aprèn de si mateixa, avalua dades o practica la recursivitat, amb la intimitat, la consciència i la reflexió inherents a la IH.

d) Dimensió personal-pràctica

La IH és pròpia de «persones» humanes. Intel·ligir és una activitat personal, el subjecte humà és un *qui*, una persona, i no un *què*, un computador. Un *qui* que sent les coses com a reals (*de suyo*), i se'n fa càrrec, i que aprehèn la seva pròpia realitat com a *seva* (és la «*suidad*», diu Zubiri), i se'n fa càrrec. No som intel·ligents pel fet de ser persones, sinó que el nostre caràcter personal es fonamenta en la nostra intel·ligència i en la manera com aprehenem la nostra pròpia realitat, cadascú com a «seva», distensa, oberta.

Com que aprehenem coses reals, el nostre quefer sempre resta obert i no maquinament determinat, com passa amb un ordinador. Les coses intel·ligides «indeterminen» les nostres accions; no hi ha estrictes automatismes en la *conducta* humana, com sí que hi són en la computació. La realitat, en canvi, dona molt de si: una pedra a la mà pot ser un material de construcció, un objecte d'interès mineralògic, una arma homicida,

⁴³ X. ZUBIRI, *Inteligencia y realidad*. Madrid: Fundación X. Zubiri-Alianza Ed., 1980, pàg. 157.

⁴⁴ *Ibid.*, pàg. 161.

⁴⁵ M. TEGMARK, *op. cit.*, pàg. 373.

un petjapapers, el record d'un indret estimat, o l'humil pedra dels camins cantada en el poema de León Felipe, «Como tú».

D'altra banda, els humans aprehenem la nostra realitat psicoorgànica com a real, i cadascú l'aprehèn com a «seva»: «soc una realitat que, com a forma de realitat, no solament soc “de si” [“de suyo”] (en això coincideixo amb totes les altres realitats), sinó que a més soc “meu”. Tinc una realitat que és meva, cosa que no passa amb la pedra».⁴⁶ Aquest caràcter «meu» de la meua realitat és la «personèitat»; el fet d'intel·ligir em fa persona, algú que va modelant fins a la mort, a base d'actuar, la seva figura concreta o «personalitat».⁴⁷ Per mor de l'obertura que imposa la intel·ligència, podem parlar en les persones de moralitat, d'autonomia, de voluntat, d'intimitat, de vinculació («relligament») al poder del real, un poder que ens força a actuar i alhora «allibera» les nostres actuacions de qualsevol automatisme implacable.

L'opacitat i la imprevisibilitat dels resultats d'un procés de classificació (recerca de correlacions i de patrons) dut a terme per una IA no guarden cap analogia amb la indeterminació dels nostres actes resultant de l'obertura insondable de la realitat que s'hi actualitza: tan sols deriva de la incapacitat humana de seguir-li el pas a una IA i d'abastar el munt de dades amb què treballa. L'algorisme d'aprenentatge automàtic es limita a assenyalar «lleis que resumeixen observacions», per molt amagades que estiguin en les dades, però no copsa l'alteritat radical d'aquestes dades, no crea categories rupturistes, ni sotmet a crítica el seu propi llenguatge de programació,⁴⁸ com sí que ho pot fer la raó humana amb les seves categories i llenguatges. L'aparent «autonomia» d'una IA no suposa llibertat ni responsabilitat, cap mena d'autonomia –diu Ganascia– entesa com a «capacitat de donar-se a si mateix la pròpia llei», sinó que tan sols significa que el sistema informàtic desenvolupa determinades funcions sense que hi intervingui cap agent humà.⁴⁹ Automatisme no és veritable autonomia, per molt que l'automatisme doni lloc sovint a un procés opac i imprevisible de computació de dades. Equiparar màquina i home condueix abans a mecanitzar l'home que a legitimar la humanització de la màquina. Com diu Jordi Corominas: «els enginyers actuals diuen molt més sobre la seva concepció de l'home que sobre els propis robots, perquè l'home que prenen com a model, al qual els robots han de reemplaçar, és un home ja robotitzat».⁵⁰

Les raons exposades evidencien la distància ontològica avui existent entre la IA i la IH. No hi ha entre elles una simple diferència de grau ni tan sols de complexitat: són

⁴⁶ X. ZUBIRI, *El hombre y Dios*. Madrid: Fundación X. Zubiri-Alianza Ed., 2012, pàg. 48-49.

⁴⁷ Per això, diu Zubiri, «el hombre es siempre *lo mismo* pero nunca es *el mismo*». *Ibid.*, pàg. 51.

⁴⁸ J. G. GANASCIA, *op. cit.*, pàg. 783.

⁴⁹ *Ibid.*, pàg. 740.

⁵⁰ J. COROMINAS, «Crítica a la religión tecnocientífica». *Revista Perifèria CPG*, 4 (2019), pàg. 147.

realitats diferents, amb notes constitucionals (pròpies del sistema i no de les seves parts) i constitutives (essencials...) ben diverses.

Arribats aquí, però, encara hi hauria una manera d'esquivar parcialment aquesta conclusió i d'establir alguna semblança entre IA i IH. Es pot renunciar a l'equiparació entre la IA i la IH pel que fa a la seva naturalesa, però mantenir l'analogia entre elles pel que fa a les finalitats, els resultats i alguns procediments. És la quarta i última tesi sobre la naturalesa de la IA que volem examinar: la IA, almenys, imita la IH.

4. La IA imita la IH

Com que A. Turing no veia possible respondre *in recto* a la qüestió *Poden pensar les màquines?* a partir de definicions unànimes de «pensar» i de «màquina», va decidir replantejar-la proposant un experiment que permetria saber almenys si una màquina pot imitar una persona: «*The Imitation Game*».

L'experiment es faria així: situem una persona, l'interrogador, en una habitació parlant mitjançant un teclat amb un home i una dona ubicats en un altra habitació. Els pot preguntar el que vulgui per esbrinar qui és l'home i qui la dona. Ara imaginem que substituïm l'home per un computador que respon les preguntes de l'interrogador sense que aquest ho sàpiga. Si el joc continua sense que sospiti res, aleshores es pot dir que el computador ha sabut «imitar» prou bé el pensament humà, sense que sigui rellevant dir si efectivament pensa o no: «No poden realitzar les màquines alguna cosa que hauria de qualificar-se com a pensament, però que és molt diferent del que fa l'home?» –es pregunta Turing.⁵¹ Obviem si un computador pensa o no com un humà, el cas és que calcula, parla, respon, simula sentiments o fa música, que són coses que també fan les persones: obté els mateixos resultats que nosaltres però d'una altra manera. Com a mínim, doncs, és cert que la IA *imita* la IH.⁵²

Però caldrà detallar en què consisteix la imitació. La IA fa coses que també fa la IH *però les fa a la seva manera*, mitjançant llenguatges de programació, algorismes (programari), dades i unes bases materials (maquinari) que tenen poc a veure amb el nostre sistema psicoorgànic. Bostrom diu que «una IA no ha d'assemblar-se necessàriament a una ment humana», però no para de referir-se a les «IA de nivell humà», capaces de fer «igual o millor» les tasques que fem les persones.⁵³ N. Carr afirma que «els ordinadors són capaços de replicar les nostres finalitats sense replicar els nostres mitjans».⁵⁴ Una

⁵¹ A. M. TURING, *op. cit.*, pàg. 435.

⁵² Turing va anar molt més enllà del joc de la imitació i apostaria avui per una IA general i forta, capaç de replicar la ment humana i d'aprendre com ho fa ella i de relacionar-se amb les persones mitjançant interfícies humanitzades. R. LÓPEZ DE MANTARAS i P. MESEGUER, *op. cit.*, posició a Kindle: 183.

⁵³ N. BOSTROM, *op. cit.*, pàg. 29.

⁵⁴ N. CARR, *Atrapados. Cómo las máquinas se apoderan de nuestras vidas*. Madrid: Taurus, 2014, posició a Kindle: 186.

calculadora realitza ràpidament i infal·liblement operacions matemàtiques que les persones faríem molt penosament, però no direm que la calculadora *sap* calcular o que *sap que calcula*, i menys encara que ho faci *com* la nostra ment. Hi ha programari de reconeixement d'imatges, però necessita computar-ne desenes de milers per distingir bé una classe d'objectes (les persones, els gats, el foc...) en qualsevol context, mentre que una criatura veient-ne un parell de la mateixa classe ja en té prou. *Google Translator* ens podria oferir en menys d'un segon la versió anglesa d'aquest mateix article a base de fer el balanç de les equivalències més recurrents entre textos anglesos i les seves traduccions catalanes disponibles a internet, però, com Searle ha explicat, és evident que el programa no «parla» cap de les dues llengües.⁵⁵

A més de la coincidència en els resultats, alguns especialistes assenyalen entre IA i IH una semblança estructural, una semblança material i alguna analogia en la manera de procedir, similituds que no arriben a diluir les diferències essencials entre elles.

a) La *semblança estructural* existiria entre el complex estructural que fa possible l'activitat intel·ligent humana i l'estructura de l'anomenada «intel·ligència artificial integral». Aquesta associa capacitats d'afectació o registre (sensors 360°, càmeres, micròfons, lectors...), d'emmagatzemar dades (memòria, servidors...), de computar informació per resoldre problemes (processador, programari...) i aprendre de la pròpia experiència, de l'entorn o d'internet, i funcions efectores d'actuar sobre el medi o de compartir i comunicar resultats (robots, pantalles, màquines annexes...). Aquestes capacitats i funcions es poden atribuir també a molts animals per mor de la seva sensibilitat, però en els humans tenen totes un caràcter intel·lectiu. La «realitat» constitueix l'àmbit en què es mouen tots els actes humans: l'afectació és una afectació real de l'home i no una simple estimulació, el registre és aprehensió de la cosa en la seva «radical alteritat», la qual modifica la manera en què l'ésser humà se sent realment afectat (sentiment) i dispara una multitud de processos lògics i racionals per resoldre els nostres interrogants i ajustar, justificar o projectar les nostres conductes. L'activitat efectiva inclou en nosaltres la volició i no una mera resposta automàtica o una acció sobre el medi.⁵⁶ La semblança estructural entre IA i IH és una analogia que no suposa una correspondència real entre totes dues.

b) La *semblança material* entre IA i IH és un objectiu que es vol assolir mitjançant l'antropomorfització de la IA per convertir-la en una tecnologia «amable». La indústria s'esforça perquè la IA imiti l'ésser humà i estableixi amb ell una relació de «tu a tu». Cada dia podem parlar millor amb sistemes automàtics: «Arribem a una època on ens

⁵⁵ Luc Julià, un dels creadors de *Siri* (Apple) posa l'èmfasi en els «mètodes diferents» que fan servir persones i sistemes intel·ligents per parlar, reconèixer imatges, etc. La diferència és tan gran, diu ell, que caldria concloure que la tecnologia anomenada IA «no té res a veure amb la intel·ligència». Vg. L. JULIA, *L'intelligence artificiel n'existe pas*. París: Ed. First, 2019, pàg. 123-126.

⁵⁶ X. ZUBIRI, *El hombre y Dios*, op. cit., pàg. 44-46.

envoltaran paraules emeses per màquines», escriu Sadin.⁵⁷ Siri (Apple), Alexa (Amazon) o OK Google són exemples que la majoria hem experimentat. El seu èxit, explica Luc Julia, un dels creadors de Siri, no radica només en el fet que ens «entenem» i ens contesten, sinó que simulen aspectes tan humans com el sentit de l'humor, la picardia, els dubtes o el canviar de tema quan no saben què respondre. «En realitat, l'èxit de Siri –diu Julia– es deu a la introducció de l'estupidesa artificial!».⁵⁸ L'antropomorfització de la IA culmina amb la construcció de robots hominoides, com els famosos robots japonesos per a la cura personal o altres robots amb aspecte clarament humà, com la famosa Sophie o els robots de serveis sexuals.

La imitació de la IH per part d'alguns sistemes d'IA referma la sensació que som davant d'éssers «personals» amb qui podem mantenir una relació «humana» i a qui aviat haurem d'imposar deures o concedir drets. Tanmateix, la semblança material torna a ocultar la veritable naturalesa de la IA, que Julia emfatitza: tots aquests artefactes humanitzats «executen les tasques per a les quals han estat fets. No inventen res, només segueixen regles, exemples i codis utilitzant les dades que escollim per a aquests sistemes [...]. No són res més que el que nosaltres, humans, hem decidit que siguin».⁵⁹

c) L'*analogia procedimental* entre IA i IH es concreta en la capacitat comuna de relacionar dades, descobrir-hi correlacions i patrons, i d'actuar posteriorment sobre la base d'aquesta descoberta. No és estrany que Kurzweil ens recordi que «la capacitat de reconèixer patrons és un dels pilars de la nostra espècie».⁶⁰

David Hume va reduir els dinamismes mentals al «principi de l'associació d'idees» segons la semblança, la contigüïtat i l'associació constant: «D'aquí ve –deia– això que anomenem l'à *propos* del discurs: d'aquí la coherència dels escrits; i d'aquí ve també aquell fil o seqüència de pensaments que la persona sosté naturalment, fins i tot en la fantasia més desbocada».⁶¹ Per al filòsof escocès, les regularitats detectades en el passat són la base de les nostres creences sobre el futur: «Estem determinats a suposar que el futur és conforme al passat únicament pel costum».⁶² També «els algorismes de la IA –escriu D. Cardon– confien en la regularitat de les estructures de gust i interès del usuari» com a fonament de les seves prediccions, «parteixen de la suposició que existeix un caràcter regular i previsible de les pràctiques humanes», suposen que fins i tot

⁵⁷ E. SADIN, *Intelligence artificielle, op. cit.*, pàg. 74.

⁵⁸ L. JULIA, *op. cit.*, pàg. 63.

⁵⁹ *Ibid.*, pàg. 150.

⁶⁰ R. KURZWEIL, *op. cit.*, pàg. 165.

⁶¹ D. HUME, *Resum del Tractat de la naturalesa humana (Abstract)*. Trad. de Miquel Costa. Barcelona: La Busca Ed., pàg. 31.

⁶² *Ibid.*, pàg. 19.

els nostres comportaments més personals obeeixen a pautes i rutines.⁶³ Com deia Hume que fa la nostra ment, les IA assenyalen correlacions sense trobar-ne les causes. Més concretament, estableixen correlacions entre valors substitutius (viure en un barri determinat o els *like* a Facebook) de fenòmens complexos difícils d'estudiar (la conducta delictiva o l'amistat) i en dedueixen conclusions que aparenten ser ciència i veritable coneixement, quan l'elecció d'aquells valors substitutius respon sovint a biaixos ideològics gens innocus. Les prediccions dels algorismes d'IA pressuposen el manteniment de les pautes socials, però adverteix Carr que «en les ocasions peculiars on les condicions es desvien dels patrons establerts, els algorismes poden fer prediccions brutalment errònies».⁶⁴

Si l'activitat de la ment es reduís a l'associació d'idees, aleshores tindria sentit equiparar-la a una IA dedicada fonamentalment a la descoberta de patrons en muntanyes de dades. És cert que els humans també trobem patrons i regularitats en la nostra experiència, però la nostra ment no es limita a fer això, sinó que és una complexa *intelligència sentent* que es mou en la formalitat de realitat i apunta a les raons, les causes i els fonaments de les correlacions percebudes, es distancia del món conegut i imagina noves possibilitats, inventa projectes, crea llenguatges i els posa al servei de la comprensió del real, etc.

Daniel Innerarity sosté que la fortalesa dels algorismes d'IA «consisteix en el fet que extreuen les característiques que es repeteixen deixant al marge les propietats rares, variables i contingents. No tan sols és que es recolzin en estereotips sinó que estan calculats per maximitzar la conformitat als estereotips».⁶⁵ Per això la IA tendeix a justificar el sistema vigent més que no pas a millorar-lo, és presonera del passat, de les dades registrades i de les repeticions. La IH, en canvi, està sempre oberta a la novetat i al futur; només així ha pogut ser evolutivament útil a l'animal humà. La IH pot ser racional, la IA no.

Tampoc la nostra conducta racional encaixa en l'esquema conductista de la repetició dels comportaments reeixits, perquè, per molts condicionaments que pesin sobre nosaltres, la realitat «allibera» o indetermina sempre poc o molt la nostra praxi, i perquè la IH, en la seva dimensió racional, sotmet a crítica la nostra praxi, fins i tot la més exitosa, albira noves possibilitats i inventa nous projectes. Cardon assenjala que «com més espai es deixa per a una reflexió distant de l'usuari, els càlculs predictius [de la IA, sobre els nostres comportaments] se'n ressenten».⁶⁶

⁶³ D. CARDON, *Con qué sueñan los algoritmos*. Madrid: Dado Ed., pàg. 84-85.

⁶⁴ N. CARR, *Atrapados*, *op. cit.*, posició a Kindle: 2045.

⁶⁵ D. INNERARITY, «La inteligencia de la inteligencia artificial». *La Vanguardia*, 16-12-2019.

⁶⁶ D. CARDON, *op. cit.*, pàg. 92.

Ara bé, més semblant o més diferent, el cas és que la IA pot influir sobre la vida humana de manera decisiva. Deixant de banda el debat sobre la naturalesa de la IA, es pot dir que és una tecnologia de la substitució de l'home amb la finalitat de «facilitar-li» la vida però també d'administrar-lo i d'explotar comercialment fins al darrer detall de la seva vida. Ho expliquem en l'últim apartat d'aquest article.

5. L'impacte de la IA: la substitució de l'ésser humà

El debat sobre el parentiu entre IA i IH no posa sobre la taula la qüestió més humanament transcendent: el significat real de la IA i el seu impacte profund en la nostra vida personal i social. Mentre estem en aquella discussió i ens enorgullim de la nostra especificitat inimitable, resulta que la IA, a poc a poc, ens guanya la partida del control de la nostra vida mental, personal i social, i ens va substituïnt en tots els terrenys: ens dispensa bons serveis a base de deixar-nos de banda. No estem fent un judici de valor sinó enunciant un fet.

La tècnica no és simplement un «saber fer» (Aristòtil), ni «ciència aplicada» (Descartes), ni un mitjà de transformació del món per mor del nostre «benestar» (Ortega). Semblen ingènues les tesis de la «neutralitat» o la «complementarietat» de la tècnica, que posen l'accent en l'ús bo o dolent que l'home en pot fer. Tota tècnica és una «realitat» i se'n planta al davant amb radical alteritat (Zubiri), sigui un martell o un *smartphone*. «Les coses artificials –diu Antonio González–, com les coses naturals, sorgeixen davant nostre i s'imposen sobre nosaltres».⁶⁷ Això fa de qualsevol tecnologia quelcom no del tot disponible ni controlable: mai no acabem de saber què donarà de si, quins efectes tindrà sobre nosaltres, la societat o l'entorn natural. Cap tecnologia es deixa dirigir del tot; com a màxim es pot «co-regir»: «L'ésser humà, com a aprenent de buixot –conclou González–, queda dominat i abraçat per la tècnica que ell mateix ha creat».⁶⁸

La tècnica no és, per tant, un recurs que canvia el món sense alterar-nos a nosaltres mateixos, sinó que s'insereix en la nostra vida i la modifica, afecta les nostres relacions i ho fa en la mesura en què resulta útil i eficaç. Per dir-ho una mica bruscament, la tècnica té un poder «infecciós»: penetra la nostra vida i la canvia profundament. Més que un recurs per viure, constitueix el «medi» en què vivim: «Esdevinguda un Univers de mitjans, la Tècnica és de fet el *medi* de l'home» –diu Jacques Ellul–, és «un conjunt coherent que ens encotilla a tot arreu, s'introdueix en nosaltres i no ens en podem

⁶⁷ A. GONZÁLEZ, «Corregir la tècnica». *Cuadernos de Filosofía Latinoamericana*, Vol. 34, Núm. 109, 2013, Bogotá, pàg. 96.

⁶⁸ *Ibid.*, pàg. 98.

desfer».⁶⁹ La tècnica és una nota constitutiva o un «factor determinant» de l'actual sistema social (econòmic, polític, cultural) tecnoliberal i constitueix per si mateixa el «sistema tècnic», un complex de recursos i processos que s'impliquen recíprocament, s'alimenten els uns als altres i responen a inèrcies i dinamismes específics: sembla quelcom dotat de vida pròpia.

Ignasi Boada assenyala com a trets específics del sistema tècnic la simplificació matemàtica del real amb vistes a la seva dominació, l'esperit d'eficàcia, l'autonomització en relació amb el control humà, la permanent inestabilitat que introdueix en les nostres vides, sempre devaluades sota l'impuls de la «innovació» i el «progrés», l'acceleració que imposa a tot allò que fem, o el seu caràcter totalitari, perquè no deixa res fora del seu embat.⁷⁰ Doncs bé, si la IA és el vaixell insígnia de l'actual sistema tecnoliberal i la major promesa d'eficàcia i benestar que mai ha rebut la humanitat, és lògic que hi refluixin de manera eminent els altres trets del sistema tècnic i que ocasioni un trasbals insòlit en la nostra existència. La IA és una força dominant dins la societat global, informacional i capitalista «tecnoliberal» (Sadin) del segle XXI. Aquest poder desbordant de la IA apunta cap a una creixent substitució de l'ésser humà amb vistes, molt sovint, al seu control i a la seva explotació comercial. La IA és una tecnologia mai del tot disponible per part de l'home, i li va prenent el lloc. «Hem dissenyat un sistema que ens descarta», sentència Nicholas Carr.⁷¹

La substitució de què parlem és produeix almenys de tres maneres:

a) La IA substitueix el treball humà

Les previsions varien molt segons qui parli: pronostiquen en algun cas que podrien desaparèixer fins al 85% dels llocs de treball actuals per l'aplicació d'IA en tots els sectors productius.⁷² Els primers a ser substituïts seran els obrers que fan feines molt rutinàries, però aviat també els qui fan tasques de formació, disseny o direcció. Com és lògic, molts treballs que requereixen intel·ligència seran realitzats per la IA. Moltes persones esdevindran auxiliars de màquines intel·ligents, els seus entrenadors i supervisors,⁷³ o es limitaran a sancionar les decisions que ha pres un algorisme. Com més sofisticades són les màquines, menys feina queda per a nosaltres.

⁶⁹ J. ELLUL, *Le Systeme technicien*. París: Le cherche midi, pàg. 49 i 54.

⁷⁰ I. BOADA, *La tecnologia com a forma de vida*. Vic: ISCRV, 2001.

⁷¹ N. CARR, *Atrapados*, op. cit., posició a Kindle: 3315.

⁷² DIV. AUTORS, *Intelligence artificielle, 4. Les algorithmes vont-ils tuer l'emploi?* París: Flammarion, pàg. 65 ss.

⁷³ Fins i tot els qui parlen de col·laboració home-màquina imaginem un futur on els homes són simples facilitadors dels treballs de les màquines. Cf. H. DAUGHERTY i H. JAMES WILSON, *Human+Machine: Reimagining Work in the Age of AI*. Harvard Business Review Press, 2018.

Una IA no dubta, no es cansa, és «objectiva», no cobra un sou i no reivindica millores laborals. En canvi, els humans som fal·libles, poc fiables, «subjectius», reivindicatius, cal que ens paguin, ens posem malalts: «Sigui quina sigui la forma d'organització –diu Sadin–, hi ha un factor que, a més de generar costos, oposa una força d'inèrcia, comet errors, contesta les decisions: l'humà».⁷⁴ El 2017 un executiu de Goldman Sachs justificava l'acomiadament de 600 *traders* financers per deixar-ne només dos: «Ja saben vostès, 600 persones ocupen molt espai». La compravenda als mercats financers ja depèn en gran mesura de sistemes intel·ligents automàtics. No hi ha res més productiu per a moltes empreses que desfer-se de persones.

«La IA ens farà bascular vers una altra civilització on el treball i els diners podrien desaparèixer».⁷⁵ Hom parla obertament de la «societat post-treball» on se suposa que la gent viurà ociosa i es dedicarà a activitats culturals o solidàries. S'aventura que tot això es podrà finançar amb els rèdits generats per un sistema econòmic farcit d'IA. No ens ha d'estranyar que molts grans empresaris de la IA demanin una renda bàsica universal per subvencionar la vida de molts extreballadors que seran «consumidors» en estat d'insatisfacció perpètua, més que no pas artistes, filantrops o contemplatius. A mitjans del segle XX, encara sense IA, Hanna Arendt valorava la possibilitat d'una societat on els sistemes automàtics haurien eliminat el treball humà: «Es tracta d'una societat de treballadors que està a punt de ser alliberada de les traves de la feina, i aquesta societat desconeix aquestes altres activitats més elevades i significatives per a les quals mereixeria guanyar-se aquesta llibertat. [...] Ens enfrontem amb la perspectiva d'una societat de treballadors sense feina, és a dir, sense l'única activitat que els queda. Està clar que res podria ser pitjor».⁷⁶

Darrere del mite de l'alliberament del treball gràcies a la robotització, s'hi amaga el prejudici que el treball és *sempre* una desgràcia que cal evitar. Però també pot existir el treball com a ofici o professió realitzat en condicions dignes, salvat de l'explotació capitalista i de la mecanització del treballador. Si respon a les aptituds i l'interès de les persones, el treball «representa una part de la dignitat humana»,⁷⁷ estructura la vida, l'orienta, li dona un sentit, una funció social. El treball esdevé aleshores una forma d'estar en la realitat, una forma de coneixement i d'autoconeixement. Pot ser l'autoexpressió de l'ésser humà: treballant podem manifestar el que som, posar-nos a prova, assumir reptes, donar de si, col·laborar amb els altres, responsabilitzar-nos-en. «Massa sovint –lamenta Carr– l'automatització ens allibera d'allò que ens fa sentir-nos lliures».⁷⁸

⁷⁴ E. SADIN, *op. cit.* (IA), pàg. 147.

⁷⁵ L. ALEXANDRE i J.M. BESNIER, *op. cit.*, posició a Kindle: 856.

⁷⁶ H. ARENDT, *La condició humana*. Barcelona: Paidós, 2009, pàg. 15.

⁷⁷ E. SADIN, *op. cit.* (IA), pàg. 154.

⁷⁸ M. CARR, *op. cit.* (AT), posició a Kindle: 203.

b) La IA substitueix les capacitats humanes

Si la IA elimina feines i descarta treballadors, és perquè substitueix les nostres capacitats en tots els àmbits de la producció i l'administració. I ho fa també en la vida quotidiana, en les relacions personals, l'oci, la llar, la creació artística, la cura dels altres. De quines capacitats estem parlant? No pas de la nostra força física o de la capacitat de càlcul, com feien altres tecnologies: una excavadora o una calculadora, per exemple. La IA obté informació, la selecciona, la classifica i l'analitza, ens fa propostes, ens suggereix decisions o les pren directament en lloc nostre, executa tasques, es comunica amb nosaltres i ens suplanta en la comunicació amb els altres. Deixem d'usar moltes capacitats mentals perquè les exerceixi un sistema intel·ligent que iguala o millora els nostres resultats. A quin preu? L'apunta l'historiador de la tecnologia George Dyson: «I si el cost de tenir màquines que pensen és tenir gent que no pensa?».⁷⁹

Si la màquina fa les coses millor que l'home, aleshores és millor reduir al mínim la intervenció humana. La IA, per exemple, fa el diagnòstic i el metge el signa i el comunica al pacient. Haurà d'anar amb compte: si accepta el diagnòstic de la màquina i aquesta no l'encerta, ningú li retraurà res; com que la màquina encerta gairebé sempre, si el metge contradiu la màquina, se la juga. Eric Sadin parla de l'emergència d'un «nou poder *alétheic*», una espècie de «demiürg» que «personifica una forma de superjò dotat en tot moment de la intuïció de la veritat i que orienta el curs de les nostres accions individuals cap al millor dels mons possibles».⁸⁰ Un poder quasi omniscient i opac: en molts casos, veiem les conclusions a què arriba l'algorisme però no en podem conèixer les raons, amagades en el garbuix d'una computació fulgurant on entren en joc centenars de milions de dades i de possibilitats. No és estrany que, cada cop més, es reclami «transparència» a la IA.

c) La IA desposseeix els humans de les seves capacitats.

A més de substituir-les, la IA ens pren facultats i ens imposa les seves. El lema empresarial d'IBM, «Per un planeta més intel·ligent», ve a dir que si l'intel·ligent és el planeta, potser ja no cal que ho siguem cadascú de nosaltres. Hom parla d'*smartphones*, de ciutats intel·ligents, de cotxes intel·ligents, d'*Internet of things* o de la *www* (World Wide Web) com a *intel·ligència col·lectiva*.

La desposseïció de les nostres facultats es produeix de dues maneres. Primerament, perquè ens acostumem a no fer-les servir i les delegem en sistemes artificials: el GPS ens orienta en ruta, Tinder ens busca parella, Google ens facilita qualsevol cosa que cerquem, el pilot automàtic porta l'avió, el sistema EMR guia el tractament mèdic d'un

⁷⁹ Citat a N. CARR, *op. cit.* (AT), posició a Kindle: 1876.

⁸⁰ E. SADIN, *La siliconización del mundo*. Buenos Aires: Caja Negra, pàg. 37.

pacient. Lliurar-se a la multitud de serveis de la IA es paga en forma de desatenció, complaença acrítica, disminució de la memòria, desactivació del raonament profund, pèrdua dels hàbits que ens empoderen i dificultat per crear-ne de nous.

A més de perdre capacitats, en segon lloc, l'home adquireix paulatinament la *forma mentis* més adequada als sistemes d'IA. Carr ho resumeix dient que «programem els nostres ordinadors i, posteriorment, ells ens programen a nosaltres».⁸¹

La *virtualització* tradueix el real al format en què és computable, comunicat sense límits de lloc i de temps i exhibit en una pantalla. Quedem sotmesos, diu Sadin, a «una capa xifrada digital mitjancera en la nostra relació amb els fets i les coses [...]. El pas de la petjada analògica a la manipulació informatitzada indueix la pèrdua de visibilitat i de fricció sensible amb les coses, per introduir jocs d'interferència *imperceptibles* i *automatitzats*».⁸² El paisatge queda substituït pel mapa del GPS; l'amic, pel seu compte d'Instagram; el jo personal, pel jo públic a Facebook o Twitter; la conversa personal, per la relació de missatges de *WhatsApp*; el debat polític, per les polèmiques a la xarxa social; la visita real a la Sagrada Família, per una visita virtual. La realitat sentida és bescanviada per un mar de símbols; el món és l'artifici filtrat i personalitzat que la pantalla ens brinda.

La *dispersió* i la *distracció* constants que ens imposen els algorismes amb la finalitat d'impulsar el consum i facilitar la detecció de patrons en la nostra praxi digital dificulten el pensament lineal, concentrat, circumstanciat, pausat, complex, profund. La multitasca ens distreu, l'estimulació plural no ens deixa reposar enlloc, els *links* dinàmics ens atrauen i mouen sense parar la ment i els sentits perquè nosaltres tampoc no parem de subministrar dades sobre nosaltres mateixos mitjançant clics de ratolí o tocs de pantalla. El resultat és la substitució d'una manera de pensar per una altra: «Calmada, concentrada, sense distraccions, la ment lineal està sent desplaçada per una nova classe de ment que vol i necessita rebre i disseminar informació en esclats curts, descoordinats, sovint solapats, com més ràpidament millor».⁸³

La *recurrència* com a clau d'interpretació de la realitat i el *filtre-bombolla* ens redueixen al que hem sigut fins ara, ens mantenen a prop d'aquells que pensen i actuen com nosaltres, s'anticipen als nostres desitjos suposant que replicarem sempre conductes passades, dificulten que coneguem alternatives i que arribem a pensar diferent. Ens compliquen, en definitiva, anar més enllà de nosaltres mateixos i del cercle dels «nostres» per albirar noves possibilitats: la IA invasiva de la vida personal dificulta, en definitiva, l'exercici de la racionalitat.

⁸¹ N. CARR, *Superficiales. ¿Qué está haciendo internet con nuestras mentes?* Madrid: Taurus, 2011, posició a Kindle: 3384.

⁸² E. SADIN, *La humanidad aumentada*. Buenos Aires: Caja negra Ed., pàg. 41.

⁸³ N. CARR, *Superficiales, op. cit.*, posició a Kindle: 171.

Finalment, l'*ergonomia* o les *interfícies* amables naturalitzen les noves tecnologies, les incorporen al nostre cos, les associen als nostres gestos més senzills, que passen a integrar-se en un procés aliè que no dirigim. Un dels lemes de l'iPad d'Apple és «Aparta la tecnologia del teu camí»: es tracta que no notem la tecnologia perquè aquesta s'ha encarnat en hàbituds que posem en pràctica de manera molt simple i inconscient. El pas següent serà la inserció de dispositius intel·ligents imperceptibles dins del nostre cos per facilitar la monitorització de les nostres dades biològiques i l'orientació permanent d'una existència «saludable», ajustada a uns estàndards que justificaran la autoimposició de pautes de conducta i de consum més sans que algú ens facilitarà a un mòdic preu. Com més suau i amable és la presència de la IA en les nostres vides, més exerceix el seu caràcter totalitari.

d) La IA substitueix l'ésser humà en les seves decisions

Si assignem tasques cada cop més complexes a la IA, és lògic que incloguin resolucions de caràcter típicament moral. A poc a poc, la IA anirà assumint les actuacions humanes que demanen una major inversió de consciència, responsabilitat i llibertat: decretar la condemna d'un acusat o la llibertat condicional d'un presoner, decidir el moviment inversor de milions d'euros a la borsa, la contractació o l'acomiadament d'un treballador, l'ordenació de la feina, l'elecció de parella, l'orientació d'un discurs polític o l'atac amb drons contra un objectiu militar, per citar només alguns exemples. Se suposa també aquí que la màquina no s'altera, no té fòbies ni manies personals o dubtes, i que fins i tot pot ser èticament més ponderada i equitativa que les persones. Es passa per alt que els algorismes incorporen «els prejudicis, les equivocacions i els biaxos humans» dels seus creadors per convertir-se, segons C. O'Neil, en autèntiques «armes de destrucció matemàtica» que ens acaben subjugant.⁸⁴

Sadin diu que «el *superjò computacional* decideix per nosaltres, sota el “postulat tecnològic fonamental” que hi ha una deficiència humana fonamental que serà salvada pels poders de la IA, que augmenten i varien constantment».⁸⁵ S'opera així a poc a poc la dissolució del subjecte modern com a ésser singular i lliure, reflexiu, capaç de prendre decisions, i responsable dels seus actes. Sembla que retrocedim a la «minoria d'edat» preil·lustrada que, segons Kant, hauríem d'haver superat amb l'ús crític de la raó. Silicon Valley «ha destruït –sosté Sadin– el principis fundadors de l'humanisme europeu que afirmen l'autonomia del judici i la lliure elecció i que indueixen un corollari: el principi de responsabilitat i els drets de les societats a decidir el seu destí comú».⁸⁶

⁸⁴ C. O'NEIL, *Armas de destrucción matemática*. Madrid: Capitán Swing, 2017, pàg. 11.

⁸⁵ E. SADIN, *La siliconización del mundo*, pàg. 37.

⁸⁶ *Ibid.*, pàg. 40.

I no tan sols les resolucions més conscients, també les petites decisions més intranscendents se'ns prenen. La IA aplicada a internet té «l'objectiu de crear i perfeccionar una teoria sobre qui ets, el que faràs i el que desitjaràs a continuació».⁸⁷ L'objectiu declarat dels fabricants d'IA és avançar-se a les nostres necessitats, estalviar-nos les preguntes i les demandes: «Google ja considera la cognició humana com quelcom gastat i inexacte, un procés biològic que és millor que faci un ordinador. “Crec que d'aquí a uns anys la majoria de les cerques seran contestades sense ni tan sols haver de preguntar”, afirma Ray Kurzweil».⁸⁸ No caldrà esperar tant: el mòbil ens recorda els aniversaris i ens planifica el dia, el frigorífic que es va buidant encomana per internet els productes que ens falten, el biberó intel·ligent indica als pares la millor inclinació de la mà, la polsera d'activitat ens programa les millors rutines esportives i l'horari d'anar a dormir, la impressora encomana els cartutxos de tòner, els llums de la casa s'encenen i s'apaguen sols quan cal, Google Instant prediu el que vols comprar i et selecciona les millors ofertes, Netflix sap quines sèries vols veure. Es consagra en definitiva un «home no-pràctic», que queda expropiat de la seva praxi completa; aquesta ja no li pertany, l'executa en molts casos però ja no és seva sinó impulsada, orientada, ordenada i quasi decidida per sistemes automàtics intel·ligents. Es despulla progressivament les persones de la seva autonomia. L'ésser humà, a base de fer-se a si mateix cada cop més prescindible, corre el risc de perdre's a si mateix. No cal que temem una invasió de robots maliciosos: ja ens haurem rendit i batut en retirada molt abans davant una IA hegemònica i invasiva.

Conclusió

Hem passat revista a les principals maneres de concebre la IA: l'equiparació, la superació, la distinció o la imitació, i hem reflexionat sobre la seva tendència imparable a substituir l'home i les seves capacitats més característiques.

Cal reconèixer les aportacions positives que la IA ja està fent en molts terrenys; tots ens n'hem aprofitat personalment i els seus beneficis són evidents en l'economia, les comunicacions, els serveis públics o la salut. No hem de renunciar-hi, per descomptat. Però cal veure quin preu haurem de pagar en termes antropològics si la IA acaba ocupant *tot* el conjunt de la nostra vida personal i social i si es desenvolupa sense un control social democràtic i lúcid. Serà tan important evitar que els algorismes d'IA gestionin la nostra vida personal al servei d'interessos espuris (polítics, comercials, etc.), com demarcar i regular els àmbits de col·laboració entre la IA i les persones d'una manera socialment profitosa. La IA pot ser una «eina» al servei de la societat, però no pot ser una

⁸⁷ E. PARISER, *El filtro burbuja*. Madrid: Taurus, 2017, posició a Kindle: 158.

⁸⁸ N. CARR, *op. cit.* (AT), posició a Kindle: 3036.

«arma» descontrolada al servei dels poders econòmics i polítics. Ha de contribuir a l'eficiència econòmica, però sobretot i molt principalment a la justícia.

Impressiona veure com la matemàtica Cathy O'Neil descriu el desplegament als EUA de la IA associada al Big Data, aplicada a la publicitat individualitzada i depredadora, que exprimeix el consumidor fins al darrer dòlar; a la predicció de delictes, que acaba penalitzant els més pobres i marginats; a la planificació dels horaris dels treballadors, que els priva d'una vida personal i familiar ordenada; a la prevenció de riscos financers, que estigmatitza les classes baixes i justifica que paguin més interessos que ningú o es quedin sense la cobertura de les assegurances; a les campanyes electorals, basades en la microsegmentació de l'electorat per enviar-li missatges i *fake news* personalitzats. Tot plegat en nom de l'eficiència i la racionalitat, quan al darrere hi ha la voracitat capitalista i l'ambició dels poderosos. Els algorismes d'IA esdevenen així «armes de destrucció matemàtica» caracteritzades pel seu abast universal, la seva opacitat i el seu poder de fer mal, aliments de totes les dinàmiques de la desigualtat.⁸⁹

L'*entusiasme* ingenu amb la IA revela una profunda desconfiança en relació amb les persones, el rebuig de la finitud, la corporalitat i la fal·libilitat de l'ésser humà; posa de manifest, en paraules de H. Arendt, «el desig d'escapar de la condició humana», tan característic del nostre temps. Quan l'educació i la política ja no semblen recursos a qui confiar l'orientació i la millora de la nostra existència individual i social, apareix la IA com a pilar d'una utopia tecnològica i salvadora, la sospitosa promesa de la ideologia «tecnoliberal» hegemònica entre les elits polítiques i econòmiques.

Tal i com es va desplegant dins la societat tecnoliberal, la IA promet el guiatge científic i complet de les nostres vides quan el que es persegueix molt sovint és el control polític, la colonització comercial de tots els espais de la nostra existència. Sadin diu que s'aspira a convertir la IA en la «indústria de la vida», la darrera etapa d'un capitalisme salvatge que fa negoci amb les nostres dades personals.

Les sensacions simultànies i falsament compatibles de tenir-ho tot a les mans i d'estar en les millors mans pot anar lligada al major dels esclavatges i a la pèrdua gradual de la nostra humanitat i la nostra llibertat. Per això el debat sobre la IA té una forta dimensió política i caldrà fer-lo en moltes instàncies socials i de manera molt interdisciplinària, lúcida, participativa i democràtica. Per *agafar el timó* de la IA ens haurem d'encomanar novament a la política i l'educació com a recursos per a la preservació de la vida humana en la seva integritat i dignitat. La dignitat humana i la justícia, no pas el negoci, són les finalitats principals del sistema polític i econòmic.

Davant la IA, la filosofia ha d'estar sempre de guàrdia, evitant condemnes taxatives i irracionals, però exercint el seu característic poder d'anàlisi i revisió crítica de les actuacions i dels discursos humans. Ens pot ajudar a entendre en què consisteix el supo-

⁸⁹ C. O'NEIL, *op. cit.*

sat caràcter «intel·ligent» de la IA a mesura que vagi evolucionant i a esbrinar tot allò que es posi en joc quan interfereixi i moduli la nostra praxi. La filosofia pot contribuir al discerniment ètic de tot allò relacionat amb la IA, dels guanys que ens proporciona i els riscos que suposa. És obligatori continuar aquestes reflexions en diàleg amb totes les disciplines vinculades a la recerca i el desenvolupament de la IA. Per descomptat que no es tracta de fer desaparèixer la IA, sinó de «co-regir-la» i de supeditar-la a les més pregoneres necessitats humanes.⁹⁰

Joan Albert VICENS
Facultat de Filosofia (URL)
joanalbert.vicens@gmail.com

Article rebut: 30 de setembre de 2020. Article acceptat: 1 de febrer de 2021

⁹⁰ L'esborrany d'aquest article fou objecte de lectura i debat amb els estudiants de l'assignatura «Veritat i Raó», del Màster de recerca de Filosofia i Humanitats de la Facultat de Filosofia de la URL, el curs 2019-20. Els seus comentaris i objeccions, que vull agrair aquí, em van servir per esmenar-lo i aclarir-lo.