

FUNDACIÓ JOAN MARAGALL

INDAGACIÓ
CONTEMPLATIVA
PER A L'ERA DE
LA INTEL·LIGÈNCIA
ARTIFICIAL

MARCO SCHORLEMMER

QUADERNS

121

2019

Marco Schorlemmer (Koblenz, 1970) és doctor en Informàtica per la Universitat Politècnica de Catalunya. Actualment és científic titular de l'Institut d'Investigació en Intel·ligència Artificial del CSIC, on és també el cap del departament de Sistemes Multiagent. Amb anterioritat havia estat investigador a la Universitat d'Edimburg. Recentment ha dirigit el projecte europeu COINVENT sobre la invenció de conceptes i la creativitat computacional, i ha coordinat la publicació dels llibres *Concept Invention: Foundations, Implementation, Social Aspects and Applications* (Springer, 2018) i *Computational Creativity Research: Towards Creative Machines* (Atlantis, 2015). A més, és coautor de més de 100 articles publicats en revistes científiques i actes de congressos especialitzats. A banda de la seva activitat professional, Marco Schorlemmer col·labora voluntàriament amb el Meditatio Centre de Londres de la Comunitat Mundial per la Meditació Cristiana, i també en el grup coordinador de la plataforma País Conscient, amb la voluntat de contribuir al conreu de la dimensió contemplativa de la pràctica tecnocientífica.

Primera edició: novembre de 2019

© 2019, Marco Schorlemmer

Edició: Isidre Ferré

Projecte gràfic: Josep Rom

Fundació Joan Maragall (Cristianisme i Cultura)

València, 244, 2n – 08007 Barcelona

ISBN:

Dipòsit legal: B 25.429-2019

Impress a: I.G. Santa Eulàlia (Santa Eulàlia de Ronçana)

www.igsantaeulalia.com

Amb la col·laboració de:

 **Fundació Bancària "la Caixa"**

ÍNDIX

Ramon Llull i la mecanització del pensament	5
La IA en la nostra vida quotidiana	6
La superintel·ligència i el discurs religiós de la IA	8
Metàfores de la intel·ligència	10
Exageració i por dels progressos de la IA	12
Racionalitat i autonomia	14
El problema de l'alineació de valors i la Declaració de Barcelona	16
El camí de la indagació	19
Les quatre revolucions científiques	21
Tecnociència i llibertat	24
Informació i límits	27
Els contraris i la seva Gestalt	28
La doble via de la indagació	30
Indagació kenòtica per a tecnocientífics	32
Tornant a Ramon Llull	35

RAMON LLULL I LA MECANITZACIÓ DEL PENSAMENT

Permeteu-me començar recordant la figura del filòsof, teòleg i místic Ramon Llull, ja que alguns dels investigadors que l'han estudiat acrediten que és primer intent seriós de mecanitzar el raonament humà.

Llull va néixer a Mallorca, pocs anys després que el rei Jaume I conquerís l'illa als musulmans. Provenent d'una família benestant amb importants vincles amb la cort reial, Llull duia una vida frívola com a trobador quan als trenta anys va tenir una sèrie d'experiències místiques consecutives que van capgirar-li la vida i el van fer deixar família, posició social i totes les pertinences per esdevenir un terciari franciscà i dedicar la vida al servei de Déu amb l'objectiu de convertir els musulmans al cristianisme. Amb aquest propòsit, va aprendre la cultura àrab i islàmica, i en una revelació posterior va imaginar el mètode per dur a terme la seva tasca missionera. Llull va idear un sistema complex basat en la manipulació mecànica de conceptes filosòfico-teològics mitjançant una sèrie de rodes concèntriques (l'anomenada *Ars* o *Art lul·liana*) que permetia construir i analitzar de manera sistemàtica els arguments racionals que sostenien les creences fonamentals del cristianisme. Llull va presentar el seu Art a les principals universitats europees de l'època i també al Papa per obtenir ajuda i finançament per a la seva tasca missionera. Fins i tot va posar l'Art en pràctica en diversos viatges missioners al nord d'Àfrica. La producció literària que va generar durant la seva longeva vida (va viure més de vuitanta anys) és impressionant (unes dues-centes vuitanta obres en català, llatí i àrab) i el seu llegat va influir directament en el pensament de filò-

sofs com Nicolau de Cusa, Giordano Bruno i Gottfried Wilhelm Leibniz, entre molts d'altres.

Avui és àmpliament acceptat que Llull va anticipar algunes idees importants que molts segles després van dur al desenvolupament de la informàtica i la intel·ligència artificial, com la d'una àlgebra formal del pensament amb la qual realitzar anàlisis lògiques i deduccions basades en la realització mecànica de càlculs a través de la manipulació de símbols. Però no fou fins al final del segle XIX i el principi del XX que l'aparell conceptual i els fonaments matemàtics ja estaven prou madurs i clars perquè els ordinadors fossin realment construïts i programats, i perquè en conseqüència sorgís a mitjan segle XX el camp de la intel·ligència artificial (IA) amb l'objectiu explícit de dissenyar i construir màquines que exhibeixin el que anomenaríem *comportament intel·ligent*.¹

LA ·IA· EN LA NOSTRA VIDA QUOTIDIANA

¿On som respecte a aquest intent de mecanització del raonament iniciada per Ramon Llull fa més de set cents anys i que avui dia continuem duent a terme amb ordinadors electrònics amb l'objectiu d'abordar tota mena de problemes i tasques? De fet, la tecnologia basada en la IA ja és present en nombroses i ben diverses activitats, com ara la navegació aeronàutica, la inversió en borsa, el diagnòstic mèdic, el control d'electrodomèstics, la cerca a internet, la logística, la recomanació de productes, el consum energètic eficient, etc., i tots portem IA al mòbil en forma d'assistents personals com *Siri*, *Cortana* o *Google Assistant*. L'abast d'aplicacions amb tècniques de la IA és actualment molt ampli i confirma l'observació que un dia va fer John McCarthy, un dels pioners de la disciplina: «Tan bon punt funciona, ja ningú no ho anomena IA.»²

1. A. FIDORA - C. SIERRA (coord.), *Ramon Llull: From the Ars Magna to Artificial Intelligence*, IIIA-SCIC, 2011.

2. Citació extreta de: B. MEYER, John MCCARTHY, BLOG@CACM, 28 d'octubre de 2011 (<https://cacm.acm.org/blogs/blog-cacm/138907-john-mccarthy/fulltext>; consultat el 23 de juliol de 2019).

Però en la darrera dècada hi ha hagut un interès sense precedents per la IA arran dels èxits que s'han assolit resolent problemes molt complexos que fins ara s'havien resistit a ser abordats mitjançant programes d'ordinador, com el reconeixement de cares i de la parla, la traducció i interpretació de llengües, o el processament i la classificació d'imatges. Aquests èxits s'han produït utilitzant uns algorismes d'aprenentatge automatitzat anomenats *xarxes neuronals*, els fonaments teòrics dels quals es remunten als inicis de la investigació en IA. El que ha canviat darrerament i que ens ha dut als èxits recents és l'elevada velocitat de còmput dels ordinadors de què disposem avui, i l'enorme quantitat de dades disponibles per poder entrenar aquests algorismes perquè aprenguin a resoldre una tasca concreta.

Això ha revifat el debat sobre l'objectiu últim del programa de recerca en IA, el d'assolir una intel·ligència de tipus general comparable a la humana. Sovint aquest objectiu es presenta com un dels més ambiciosos i transcendents que mai s'ha proposat la humanitat, i això ha tornat a posar en primer terme les preguntes de recerca que han orientat la investigació en IA des de l'inici: *¿És possible construir màquines intel·ligents? ¿Podrien aquestes màquines superar els éssers humans en intel·ligència?*³

Aquestes preguntes no tenen una resposta clara i precisa, perquè sovint impliquen una visió del món concreta, una certa mirada metafísica sobre què és un ésser humà i què el diferencia d'una màquina, i sobre allò que entenem per intel·ligència. De fet, no està clar que aquestes siguin realment bones preguntes per orientar la nostra recerca científica i tecnològica. Si partim de la hipòtesi que majoritàriament assumeixen els investigadors en IA (a saber, que la intel·ligència es basa en el processament d'informació), llavors aquestes preguntes esdevenen gairebé retòriques, i de fet ens distreuen del paper que les tecnologies de la informació i la comunicació haurien de tenir a l'hora de fomentar el bé comú en les societats humanes tant d'avui com del futur.

3. R. LÓPEZ DE MANTARAS - P. MESEGUER, *Inteligencia Artificial*, Madrid, CSIC / Libros de la Catarata, 2017.

Una mostra d'aquesta distracció és la reflexió que científics de renom mundial com Stephen Hawking, Stuart Russell, Max Tegmark i Frank Wilczek van publicar fa uns anys en un article al diari *The Huffington Post*. Per una banda, comentaven que la investigació en IA avançava ràpidament, que les fites recents molt probablement empal·lidirien en comparació amb el que es podria assolir en les properes dècades i que, per tant, es feia difícil predir què podríem aconseguir quan la intel·ligència humana fos ampliada amb les eines que la IA ens podria proporcionar. Els beneficis potencials serien enormes, deien, i l'èxit de crear una IA representaria l'esdeveniment més gran de la història de la humanitat. Per altra banda, però, alertaven que la creació d'una IA podria ser també el darrer esdeveniment de la història humana, tret que aprenguéssim a evitar els riscos: a curt termini, que els exèrcits d'arreu del món desenvolupessin sistemes d'armes letals autònoms, capaces de triar i eliminar els seus propis objectius sense intervenció humana; a mitjà termini, que la IA transformés la nostra economia generant una gran riquesa però també una supressió massiva de llocs de treball, i a llarg termini, que les màquines amb intel·ligència sobrehumana milloressin successivament el seu propi disseny per així eludir els mercats financers, inventar millor que els investigadors humans, manipular els líders mundials i desenvolupar armes que ni tan sols podríem entendre. En conclusió, aquest científics manifestaven que ens estem enfrontant potencialment al millor i al pitjor que li podia passar mai a la humanitat.⁴

He de confessar que, en llegir per primer cop aquest article, em va sorprendre la visió tan aviat utòpica com distòpica que s'hi presentava. En tant que investigador en IA i coneixedor del que s'havia aconseguit fins aleshores i del que continuaven sent problemes molt difícils de resoldre,

4. S. HAWKING - S. RUSSELL - M. TEGMARK - F. WILCZEK, «Transcending Complacency on Superintelligent Machines», *The Huffington Post*, 19 d'abril de 2014.

la meua primera reacció va ser rebutjar les afirmacions i els temors que s'hi manifestaven, per estar completament desvinculats del que es pot aconseguir realment amb la IA en el futur, tant proper com llunyà. Però els autors d'aquest article eren figures molt destacades de la ciència i l'enginyeria, fins i tot amb una àmplia experiència en investigació en IA. Per tant, calia fer una lectura seriosa d'allò que deien per poder avaluar quines són realment les oportunitats i els riscos a què la humanitat s'enfronta amb el progrés recent de la IA.

La idea d'una intel·ligència sobrehumana ha estimulat la imaginació no tan sols de científics i enginyers, sinó també de filòsofs, novel·listes, cineastes i públic en general. Només cal observar les nombroses pel·lícules que tenen la IA com a tema principal: *2001 A Space Odyssey*, *Blade Runner*, *Terminator*, *Matrix*, *Eva*, *Her*, *Ex Machina*, *Transcendence*, etc. En elles podem apreciar com s'explota l'aspecte més controvertit de la IA, i les qüestions que s'hi tracten s'emmarquen en certa manera en una visió de la IA que mostra algunes de les característiques del discurs religiós més mític, en el sentit literal de $\mu\upsilon\theta\omicron\varsigma$ (*mythos*), és a dir de narració. En certa manera són expressions contemporànies de l'antiga història bíblica que narren la creació a imatge del creador, o són manifestacions mundanes del desig d'assolir la immortalitat, identificant el pensament racional amb allò que se suposa que és específicament humà i transferint-lo a una màquina per alliberar així la ment del cos mortal del qual depèn.⁵

Aquest àmbit de debat, tanmateix, és aliè a les preocupacions reals que els investigadors en IA tenim en la nostra tasca científica quotidiana. Per entendre, doncs, la investigació que s'està duent a terme en els centres de recerca i els resultats que realment se'n poden derivar, cal examinar més a fons les metàfores conceptuals subjacents al programa de recerca en IA.

5. D. F. NOBLE, *La religión de la tecnología*, Barcelona, Paidós, 1999. J. BROCKMAN, «The Myth of AI. A Conversation with Jaron Lanier», *edge.org*, 14 de novembre de 2014 (URL: https://www.edge.org/conversation/jaron_lanier-the-myth-of-ai; consultat el 23 de juliol de 2019).

El programa d'investigació en intel·ligència artificial s'originà amb certes metàfores de la ment que constitueixen la base de la filosofia analítica angloamericana. Diem «metàfores de la ment» perquè els científics cognitius han donat en les últimes dècades proves significatives que la majoria dels nostres conceptes, i en particular els conceptes abstractes com *intel·ligència* o *ment*, són en gran part metafòrics i es basen en la nostra experiència sensorial i corporal quan interactuem amb l'entorn.

Si heu estat mai a Grècia, és possible que hagueu vist camions i autobusos retolats amb la paraula *μεταφορά* (*metaforá*). Aquest mot significa 'transport': mitjançant metàfores transportem l'estructura conceptual d'un domini amb què estem familiaritzats a un altre domini més abstracte. En observar detingudament les metàfores subjacents al programa de recerca en IA, podem descobrir la manera en què entenem conceptes com *ment*, *racionalitat* i *intel·ligència* i, per tant, prendre consciència de la manera en què ens veiem a nosaltres mateixos com a éssers intel·ligents.

Una manera molt comuna de parlar de la ment és com si fos un cos: el pensament es conceptualitza com un tipus de funcionament físic que pertany a un cos, com ara moure's, veure, manipular objectes o fins i tot menjar, i les idees són entitats físiques amb una existència independent, com ara ubicacions, coses percebudes, objectes manipulables o menjar. Podem observar aquestes metàfores en les expressions que fem servir habitualment: «Com has arribat a aquesta conclusió?», «Ja veig el que vols dir», «Aquest coneixement està fora del seu abast», «Això que em dius no m'ho empasso pas».

Sovint també utilitzem la metàfora de la ment com una màquina, i més recentment la de la ment com un ordinador. Mitjançant aquesta metàfora conceptualitzem la ment si fos un programa informàtic, els conceptes i les idees com si fossin símbols formals, i el pensament com una manipulació d'aquests símbols.

No hi ha res de dolent en el fet d'emprar metàfores: és pràcticament impossible pensar o parlar de manera seriosa sobre la ment sense concep-

tualitzar-la metafòricament, tal com afirmen el científic cognitiu George Lakoff i el filòsof Mark Johnson. El problema rau, diuen, en les diferents actituds envers la metàfora conceptual que es fa servir. Lakoff i Johnson enumeren tres actituds que es poden adoptar envers la conceptualització de la ment en termes computacionals, tal com es manifesta en la metàfora «La ment és un ordinador»:

a) Es pot identificar la metàfora i estudiar-la al detall, és a dir, estudiar quins tipus de vincles metafòrics es poden establir i com això influeix en la nostra comprensió de la ment.

b) Es pot reconèixer la naturalesa metafòrica de la nostra conceptualització, però agafar seriosament la metàfora com un model científic de la ment. Molts professionals del que s'ha anomenat la *versió feble* de la IA adopten aquesta actitud i fan servir models computacionals com una eina per estudiar la ment, formulant i provant hipòtesis al respecte. La IA feble també s'ha associat amb els intents de crear programes per ajudar i millorar les activitats mentals humanes, en comptes de substituir-les.

c) Es pot prendre la metàfora de manera literal per expressar una veritat científica profunda, és a dir, que els conceptes són efectivament símbols formals, que el pensament és manipulació simbòlica mitjançant el càlcul i que la ment és un programa que funciona amb un *hardware* cerebral. Aquesta és la visió del que històricament s'ha anomenat IA *forta*.

La IA forta, a diferència de la feble, està associada amb l'afirmació que un ordinador adequadament programat podria tenir una ment, i podria pensar com a mínim tan bé com els éssers humans. Hi ha un nombre important d'investigadors en IA que, tot i que en el seu dia a dia investiguen en IA feble, en més o menys grau estan d'acord que l'objectiu final de la investigació de la seva disciplina és assolir una IA forta. Això posa de manifest que, una cop la metàfora «La ment és un ordinador» es pren de manera literal com una definició de l'essència mateixa de la ment, ja no es

considera conscientment com una metàfora, sinó com una veritat mateixa sobre la naturalesa de la ment.⁶

EXAGERACIÓ I POR DELS PROGRESSOS DE LA ·IA·

Sempre que s'han obtingut resultats impressionants i avenços en el disseny i la implementació de sistemes informàtics que exhibeixen algun comportament intel·ligent (èxits de la IA feble), com ara derrotar els campions del món de jocs de taula com els escacs o el go, o vèncer en el concurs televisiu americà *Jeopardy!* (que requereix un ampli coneixement general i la capacitat d'entendre el llenguatge natural), generalment han seguit afirmacions massa optimistes del que es podria fer amb la IA i de quan es podria aconseguir una IA del nivell humà (IA forta). Com a conseqüència d'això, sovint s'expressen visions extremadament distòpiques del que significaria per al futur de la humanitat.

Ara vivim novament un d'aquests períodes en què s'estan obtenint resultats impressionants en tecnologia de la IA (de la IA feble), acompanyats d'especulacions sobre quan es podrà assolir la IA de nivell humà (la IA forta) i, un cop superada, sobre els riscos d'arribar a aquest punt en el desenvolupament de la IA. Tanmateix, no hi ha precedents de l'actual interès per la IA i les aplicacions que pot tenir, i la diferència amb períodes anteriors, amb exagerades visions distòpiques, és que la tecnologia actual de la IA feble està aconseguint per a determinades tasques cognitivament complexes un rendiment comparable a l'humà, i fins i tot millora aquest rendiment en àrees que eren tradicionalment considerades molt difícil d'abordar computacionalment.

En conseqüència, ara la indústria està fent una aposta important per posar al mercat productes basats en tècniques de la IA, i les grans empreses tecnològiques (Google, Facebook, Amazon, Microsoft, Apple, etc.) han creat departaments i unitats específiques que fan recerca, desenvolupa-

6. G. LAKOFF - M. JOHNSON, *Metáforas de la vida cotidiana*, Madrid, Cátedra, 1986. IDEM, *Philosophy in the Flesh*, Basic Books, 1999.

ment i innovació en IA. Actualment hi ha un ampli consens sobre la creença que la IA serà de gran importància per al futur de l'economia, i que tindrà un fort impacte en el funcionament de les societats humanes. Una pregunta immediata que moltes persones es plantegen actualment és com la tecnologia basada en IA afectarà el mercat laboral: ¿els robots ens prendran la feina?

Quan es parla de *superintel·ligència*, sovint es fa referència a la possibilitat que les màquines reemplacin els éssers humans en la majoria dels llocs de treball, i hi ha qui fins i tot s'atreveix a pronosticar que això pot passar tot just en els propers vint o trenta anys. Basat en un estudi realitzat per investigadors de la universitat d'Oxford juntament amb la consultora Deloitte, la BBC ha creat un lloc web que permet calcular el risc d'automatització que tenen determinades professions. Així, per exemple, als taxistes i xofers els assigna una probabilitat d'automatització del 57%, i als bibliotecaris, del 97%. En canvi, per a les infermeres o llevadores el risc d'automatització se situa en el 2%. Plantejar prediccions d'aquest tipus, però, revela la visió que hom té de la intel·ligència: una intel·ligència merament funcional, orientada a executar tasques i resoldre problemes; una intel·ligència avaluada en termes de rendiment econòmic en el mercat laboral; una intel·ligència aplicable a llocs de treball racionalitzats segons la teoria de l'elecció racional. Així, doncs, per saber l'impacte que la IA tindrà en l'automatització de les nostres feines i en el funcionament de les nostres societats humanes, serà útil examinar les hipòtesis bàsiques de la IA en el disseny de sistemes computacionals que presenten un comportament intel·ligent.⁷

7. N. BOSTROM, *Superinteligencia. Caminos, peligros, estrategias*, Madrid, TEELL, 2016. BBC NEWS – TECHNOLOGY, «Will a robot take your job?», 11 de setembre de 2015 (URL: <http://www.bbc.com/news/technology-34066941>; consultat el 23 de juliol de 2019). C. B. FREY - M. A. OSBORNE, «The future of employment: How susceptible are jobs to computerisation?», *Technology Forecasting and Social Change* 114 (2017), p. 254-280.

El paradigma principal que s'adopta en la majoria d'investigacions que es fan en IA és el d'un *agent autònom i racional*. És el paradigma descrit en el llibre de text sobre la IA més utilitzat actualment en les universitats d'arreu del món. Deixeu-me descriure aquest paradigma en termes molt generals. En l'àmbit de la IA, es pren com a objectiu dissenyar una entitat o sistema computacional que percebi l'entorn i les altres entitats situades en ell (és a dir, rep *inputs* en els seus sensors), i, basant-se en aquesta percepció actual, en la memòria que guarda de percepcions anteriors i en algun tipus de model intern de l'entorn i altres coneixements previs, pren la decisió de manera autònoma d'actuar d'una determinada manera (és a dir, produeix un *output* en els seus actuadors). D'aquí el nom d'*agent*. ¿Quin és el principi a seguir a l'hora de triar una acció determinada? Per això, en IA se sol adoptar una comprensió idealitzada del que significa que un agent actuï de manera intel·ligent, reduint la *intel·ligència* a la *racionalitat*; i així, per *actuar racionalment* s'entén *triar l'acció que produeix el millor resultat*, segons una mesura de rendiment prèviament definida (o quan hi ha incertesa, tal com succeeix en la majoria dels casos, triar l'acció que produeix el millor resultat esperat).⁸

Aquest enfocament ha estat molt reeixit per al progrés en la investigació de la IA, ja que evita nocions més complexes com *pensar*, *ment*, *consciència*, etc., i parteix d'una noció d'intel·ligència dotada de dimensió i per tant mesurable científicament. A més, suposa que la intel·ligència és una característica objectiva, atribuïble a una entitat, sigui un organisme vivent o una màquina computacional. Aquesta visió de la intel·ligència queda palesa en la mena de recerca que realment s'ha anat desenvolupant en els centres d'investigació i els departaments universitaris, i que intenta donar resposta a unes preguntes molt més pragmàtiques i menys transcendents que les que hem anunciat abans: *¿És possible construir màquines que*

8. S. RUSSELL - P. NORVIG, *Inteligencia Artificial: Una enfoque moderno*, Madrid, Anaya, 2004.

actiuïn de manera racional segons el model matemàtic de la teoria de l'elecció racional? ¿Aquestes màquines podrien superar els éssers humans en la racionalitat de les decisions i les accions que s'ajustin a aquest model?

Així un agent racional i autònom que juga als escacs percep el tauler d'escacs, les figures i els moviments del seu oponent, i tria el moviment que amb major probabilitat el duu a guanyar la partida. L'agent que sap triar els millors moviments i que, per tant, guanya el major nombre de partides se'l considera «més intel·ligent» a l'hora de jugar als escacs. Si a més escull millors moviments que els humans, llavors direm que és «més intel·ligent» que els humans per a aquesta tasca en particular. Un agent racional i autònom que té com a objectiu classificar les imatges de teixits tumorals per diagnosticar el càncer de pell i que classifica les imatges amb un índex d'error inferior al dels humans es considera «més intel·ligent» que els humans per a aquesta tasca particular de classificació d'imatges. Un agent racional i autònom que té com a objectiu neutralitzar una posició enemiga escollint l'objectiu per disparar un projectil explosiu en particular i ho fa amb menys danys col·laterals pel que fa al nombre de baixes civils es considera «més intel·ligent» que els humans per a aquesta tasca particular de neutralització enemiga.

Òbviament, la racionalitat perfecta rarament s'aconsegueix (per exemple, explorar *tots* els moviments d'escacs possibles, els propis i els de l'oponent, és computacionalment inviable), de manera que en la IA se suposa una *racionalitat limitada*, i un problema d'investigació important és el de dissenyar bons algorismes per fer front a aquesta situació de racionalitat limitada. Aquesta comprensió de la racionalitat ha estat molt útil per abordar moltes de les tasques per a les quals s'han desenvolupat tècniques d'IA fins ara i que he esmentat al principi d'aquest assaig: la navegació aeronàutica, la inversió en borsa, el diagnòstic mèdic, el control d'electrodomèstics, la logística, els recomanadors per a consumidors, els jocs d'ordinador, etc.

Tanmateix, aquesta noció de racionalitat i la noció d'autonomia associada a ella (és a dir, que escollim lliurement les accions que volem fer a partir de la nostra decisió racional) es basen fonamentalment en les as-

sumpcions que han caracteritzat la modernitat i que formen el nucli del sistema de valors de les societats industrials. Algunes d'aquestes assumpcions són:

- Les societats les formen persones i organitzacions que miren per si mateixes i que intenten aconseguir els seus objectius interactuant amb altres agents i amb l'entorn.
- Les persones i les organitzacions tenen ben definides les seves preferències, les quals es poden quantificar i comparar.
- Les persones i les organitzacions són autònomes i escullen l'acció que cal fer de manera racional, és a dir, maximitzant la utilitat i el benefici esperat.
- La racionalitat és incorporada, és a dir, no depèn de la nostra naturalesa corporal.
- El coneixement és informació creada i emmagatzemada en les ments i en altres artefactes humans com històries, llibres i ordinadors interconnectats en xarxa.
- El coneixement és informació processable i constitueix la base per als nostres pensaments i accions.
- El coneixement és una mercaderia: es produeix, es dona i es rep.
- La qualitat de les nostres accions es pot mesurar i comparar, fet que dona lloc a la noció d'excel·lència, a saber, quan la qualitat d'una persona o organització està per sobre de la d'una altra.

Aquestes assumpcions es basen en una sèrie de metàfores conceptuals que constitueixen el model socioeconòmic que determina avui moltes de les activitats que fem en les nostres societats: la teoria de l'elecció racional.

EL PROBLEMA DE L'ALINEACIÓ DE VALORS I LA DECLARACIÓ DE BARCELONA

¿Què ens diuen aquestes assumpcions i metàfores conceptuals que fonamenten la noció d'un agent autònom i racional sobre l'impacte potencial

de la IA en les societats humanes, fins al punt de convertir-se realment en una amenaça per a la humanitat? Stuart Russell, professor de Ciències de la Computació a Berkeley i coautor de l'article del *The Huffington Post* que he mencionat anteriorment, descriu el problema de la següent manera: (a) la funció d'utilitat respecte a la qual un agent racional autònom pren decisions d'alta qualitat pot no estar perfectament alineada amb els valors de la humanitat, que són molt difícils de definir, i (b) tot sistema intel·ligent prou capacitat tractarà d'assegurar la seva pròpia existència continuada i d'adquirir els recursos físics i computacionals necessaris, no pas per a si mateix, sinó per reeixir en la tasca assignada (és a dir per aconseguir la màxima utilitat).

L'objectiu de la investigació en IA ha estat millorar la qualitat dels algorismes de presa de decisió (segons el paradigma de l'agent racional i autònom), independentment de la funció d'utilitat que hom escull. Això podria dur-nos, diu Russell, a produir l'*efecte Rei Mides*: obtenim exactament allò que demanem i no pas el que en realitat volem. I aquí, diu, rau el perill. Per tant, Russell identifica correctament que els valors han de formar part explícita dels resultats que cerquem en la nostra investigació: la tecnologia basada en IA hauria de donar lloc a sistemes d'IA que estiguessin alineats amb els valors humans. Ho anomena el *problema de l'alineació de valors*.⁹

En un mateix sentit, diversos experts mundials van pactar a Barcelona una declaració per al bon desenvolupament i ús de la IA, que va ser el resultat d'un debat sobre els somnis i els riscos de la IA. Aquesta declaració proposa un codi de conducta per als professionals de la IA, tant investigadors, com desenvolupadors o implantadors:¹⁰ ser *prudents* (establir prerequisits molt estrictes abans d'aplicar una tècnica basada en IA), garantir la *fiabilitat* (dissenyar metodologies adequades de verificació i validació que

9. S. RUSSELL, «Inteligencia artificial de beneficios probados», dins *El próximo paso: la vida exponencial*, Madrid, BBVA OpenMind, 2017.

10. L. STEELS - R. LÓPEZ DE MANTARAS, «The Barcelona declaration for the proper development and usage of artificial intelligence in Europe», *AI Communications* 31(6) (2018), p. 485-494.

certifiquin una aplicació d'IA abans que s'utilitzi de manera generalitzada), garantir que es *reten comptes* (proporcionar explicacions de les decisions perquè els humans puguin qüestionar-les amb arguments raonats), demanar *responsabilitat* (que sigui clar obligatòriament quan una interacció es realitza amb un ésser humà o amb un sistema d'IA), implementar una *autonomia limitada* (circumscriure adequadament els sistemes d'intel·ligència autònoms de manera que estiguin alineats amb els valors humans) i confiar en els *éssers humans* (continuar ensenyant, desenvolupant i exercint coneixements humans, perquè els sistemes d'IA depenen de manera crítica de la intel·ligència humana).

L'objectiu d'una «intel·ligència artificial de beneficis provats» i la Declaració de Barcelona, que proposa un codi de conducta, són passos valuosos cap a un desenvolupament i un desplegament més conscients dels sistemes d'IA. No obstant això, aquestes propostes —i les accions realitzades fins ara que les tenen en compte— continuen derivant de les mateixes hipòtesis i es troben emmarcades dins de les mateixes metàfores conceptuals en què s'ha dut a terme el programa de recerca en IA. Perquè abordar el problema de l'alineació de valors proposat per Russell mantenint-nos en la conceptualització de la intel·ligència com a acció racional, i de l'acció racional com a maximització de la utilitat, porta a veure els valors humans segons la metàfora «Benestar és riquesa», que és part constitutiva del model d'elecció racional. Però encara que puguem oferir conceptualitzacions dels nostres sistemes de valors humans per raonar amb ells —i això és essencialment el que fa l'ètica—, aquestes conceptualitzacions no seran capaces d'evitar les metàfores i les inferències que se'n deriven. Els valors és allò que sentim en el nostre cor, no pas el que hi ha al nostre cap. En conseqüència, cal abordar també els valors vivencialment i fenomenològica, en primera persona, de com *sentim* els valors, i no tan sols conceptualment i objectiva, en tercera persona, de com *pensem* i *raonem* sobre els valors.

Si volem prendre mesures efectives per abordar el problema de l'alineació de valors per tal de plantejar els possibles riscos dels sistemes d'IA que puguin ser nocius per a les societats humanes, primer de tot hem de saber

mirar a través de les assumpcions actuals de la IA i de les metàfores conceptuals subjacents, per avaluar si aquestes assumpcions i metàfores segueixen sent aptes per a les societats actuals i les formes de viure-hi. Si continuem operant en el mateix sistema conceptual i en el mateix sistema de valors, podem acabar empitjorant les coses, encara que de manera involuntària i de bona fe. I és aquí on pren rellevància la contemplació.

EL CAMÍ DE LA INDAGACIÓ

S'ha argumentat sovint que el progrés en ciència i tecnologia entra en conflicte amb l'espiritualitat, que la ciència i la tecnologia han desplaçat l'espiritualitat del centre de les nostres vides i que ara confiem tot allò que creiem i fem a la ciència i a la tecnologia en comptes de confiar-ho als déus. Això certament és així amb les expressions tradicionals de l'espiritualitat que podem trobar en els sistemes de creences que hem heretat de les religions convencionals i institucionalitzades. N'hi ha que pensen que aquest desplaçament és positiu, que és un progrés per a la humanitat, en què es passa de creences irracionals cap a una major il·lustració; i n'hi ha que pensen que és negatiu, una gran pèrdua per a la humanitat, ja que elimina el valor, la profunditat, el significat i la importància de la nostra existència, tal com els proporcionen els marcs religiosos. Crec que aquest conflicte d'interpretacions, positiva *vs.* negativa, de l'avanç de la ciència i la tecnologia és enganyosa, perquè en realitat la investigació científica genuïna és fonamentalment espiritual. Perquè investigar és obrir-se al desconegut, deixar que la curiositat i la meravella guïïn la investigació i deixar-nos transformar per ella: cal saber que no sabem i que ens cal indagar.

De fet, això és del que tracta la fe. Com va afirmar el filòsof Richard Kearney, «perquè la fe significa saber que no sabem absolutament res sobre els absoluts». Així, al contrari del que hom pot pensar, la fe no tracta sobre creences, ni sobre la veritat, la certesa o el dogma; la fe tracta, en darrera instància, de la indagació, ja que a l'univers no existeixen ni la certesa, ni la fixació, ni l'aïllament. Per tant, el nostre camí humà de la vida, el

nostre viatge espiritual, és el de la indagació: és el camí de la recerca; és a dir, mantenir l'èmfasi en la recerca i no en allò que s'està cercant: l'objectiu és el camí. Aquesta és la raó per la qual la investigació científica genuïna és essencialment contemplativa en el seu nucli:

- és vocacional, impregnada d'una actitud de sorpresa i meravella;
- es basa en una comunicació oberta pel fet de compartir i confiar, que crea comunitat i la construeix, perquè la investigació científica es fa en comunió;
- requereix silenci, atenció, paciència, cura i respecte, i hom no s'hi pot precipitar;
- és gratuïta i amorosa;
- és transformadora, «purifica» la nostra experiència i ens allibera de les concepcions mítiques (*μῦθος*) de la realitat,
- i és profundament humil.

El desplaçament per part de la investigació científica de les formes convencionals d'espiritualitat és en realitat una forma de creixement espiritual: l'ateisme científic ha estat per a moltes persones, inclòs per a mi mateix, un pas important per anar més enllà d'una comprensió convencional de l'espiritualitat. Kearney fa servir el neologisme *aposta anateïsta* per anomenar aquest compromís amb el viatge espiritual que ens allunya del teisme (quan estem lligats a alguna concepció de Déu) i ens porta a través de l'ateisme (quan ens desfem del tot de Déu) cap a l'anateïisme (quan tornem a Déu «després de Déu»). El prefix *ana-* del mot *anateïisme* significa 'de nou', però és «de nou un teisme» en termes de Déu, no pas en termes propis, ja que realment ens n'hem després del tot, d'ell. Per tant, el programa de recerca en IA, com qualsevol altra investigació científica, s'ha de situar en aquest context més ampli de l'aposta anateïsta, que és en darrera instància una aposta per la indagació, per obrir-se al desconegut i al foraster, i per deixar-se humiliar i transformar.¹¹

11. R. KEARNEY, *Anatheism. Returning to God after God*, Columbia University Press, 2010.

Fa aproximadament uns cent anys, Sigmund Freud va identificar tres revolucions científiques importants que han canviat substancialment la comprensió de nosaltres mateixos: Nicolau Copèrnic va revolucionar la nostra comprensió del cosmos (va demostrar que no estem al centre de l'univers), Charles Darwin va revolucionar la nostra comprensió de la vida (va demostrar que no som el cim del regne biològic) i el mateix Sigmund Freud va afirmar que ell va revolucionar la nostra comprensió de la consciència: va demostrar que no som els amos de la nostra ment, de la nostra consciència.

Aquestes tres revolucions han desmantellat progressivament la comprensió antropocèntrica que teníem de la realitat i han desplaçat la humanitat del centre a la perifèria de l'univers. I tot humiliant l'ésser humà, aquestes revolucions han ampliat alhora la llibertat creativa de la humanitat, ja que només a través d'aquesta llibertat creativa, l'ésser humà va ser capaç d'abordar i de buscar respostes a la pregunta: *qui sóc?* El filòsof de la informació d'Oxford Luciano Floridi ha defensat recentment una quarta revolució. Afirmar que Alan Turing va revolucionar la nostra comprensió de la intel·ligència, de la racionalitat: va demostrar que no tenim l'exclusivitat del pensament racional. Va afirmar que els ordinadors també poden raonar.¹²

Probablement no tenim la perspectiva històrica per jutjar si realment estem vivint una nova revolució científica de la mateixa importància i el mateix abast que les tres anteriors que hem citat. Però hi ha una part de veritat en el fet que la investigació en IA està transformant realment la nostra comprensió de la cognició i, de retruc, de tota la realitat. El nostre progrés en IA ens ajuda a definir millor quines funcions cognitives de la nostra ment es poden caracteritzar en termes de processament d'informació i quines no, i per tant quines es podrien mecanitzar i quines estan fora

12. F. WEINERT, *Copernicus, Darwin, & Freud: Revolutions in the History and Philosophy of Science*, Wiley-Blackwell, 2009. L. FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford University Press, 2016.

de l'abast de les màquines. En conseqüència, la investigació en IA ens hauria de portar a revisar la noció dominant de *racionalitat* (una racionalitat conceptualitzada metafòricament com la maximització de la utilitat), i fins i tot ens hauria de portar a revisar la concepció que la racionalitat és allò que ens distingeix com a éssers humans, ara que tenim prou evidències que les màquines són millor essent racionals (segons la metàfora conceptual dominant avui). La mateixa recerca en IA s'hauria de replantejar les assumpcions i les metàfores sobre les quals ha desenvolupat el seu programa de recerca. No hauríem de quedar atrapats en els nostres models metafòrics de què entenem per intel·ligència. Com he dit abans, no resoldrem els riscos d'aquesta tecnologia si els debats ètics queden circumscrits a un àmbit conceptual i intel·lectual en el marc dels models basats en la teoria de l'elecció racional, com em sembla que passa sovint avui; és a dir, quan s'intenta incloure una concepció dels valors i del que significa ser beneficiós per als éssers humans dins dels models matemàtics de racionalitat econòmica.

La metàfora mecanicista de la nostra ment i la hipòtesi que la intel·ligència es fonamenta en el processament d'informació segurament han estat útils per iniciar i orientar el programa de recerca en IA dins del context socioeconòmic que vam heretar de la industrialització. No endebades les màquines són entitats que coneixem bé en tant que són creacions nostres. Però davant dels reptes amb què ens trobem amb el desenvolupament futur d'aquesta tecnologia, potser ha arribat el moment de cercar unes noves metàfores que s'ajustin millor a la manera de sentir i de viure de les societats actuals. La IA i la ciència de la computació en general es fonamenten de fet en una filosofia i en una manera de veure el món que són bàsicament anteriors a l'any 1930, basades principalment en explicacions mecanicistes i una separació rigorosa entre ment i matèria, entre cognició i acció. En canvi, corrents del pensament com aquelles que fonamenten les nocions de *cognició corpòria* i *enacció*, i que estan fortament influïdes pels moviments filosòfics de la fenomenologia i el pragmatisme, poden ajudar-nos a reorientar la recerca en IA perquè així sapiguem posar l'èmfasi en la millora de la interacció de la nostra

intel·ligència humana amb els sistemes computacionals de processament d'informació.¹³

Les revolucions científiques, com ara les quatre esmentades, haurien de ser avaluades per tant per la seva contribució a la llibertat creativa de tots els éssers. I aquí rau la situació crítica a què s'enfronta la humanitat, i que és realment una qüestió de supervivència. El perill actual a què ens enfrontem és el d'oblidar-nos de la naturalesa metafòrica de les nostres concepcions de *ment*, *intel·ligència* i *racionalitat*, i prendre-les literalment, deixant de veure-les com a metàfores. Dotaríem així els nostres models científics d'una realitat ontològica. Quan això succeeix, els nostres models deixen de ser només descriptius i predictius i amenacen de convertir-se en prescriptius: si la intel·ligència és computació, i raonar és calcular, només manca un petit pas per intentar prendre qualsevol procés en què ens cal raonar (per exemple, el diagnòstic mèdic, les sentències judicials o la selecció de personal) i estandarditzar aquests processos en termes de procediments algorísmics el resultat dels quals es pot mesurar i comparar adequadament per aconseguir un millor rendiment, i així esforçar-se per assolir l'excel·lència.

A mesura que «racionalitzem» inconscientment les nostres activitats humanes en termes de computació algorítmica, per la nostra incapacitat de distingir les nostres metàfores conceptuals, els humans anirem ajustant de manera imperceptible les nostres decisions i accions a aquests processos «racionalitzats», sense qüestionar-nos-en l'aptitud. I com que les màquines són millors (i més ràpides) a l'hora de ser «racionals» (segons la conceptualització metafòrica dominant), els humans es podran substituir fàcilment per sistemes informàtics, i fins i tot es consideraria poc ètic deixar de fer-ho. Si un sistema de diagnòstic basat en IA realitza diagnòstics mèdics amb un índex d'error molt menor que un metge, ¿no hauríem de deixar a les seves mans la tasca de diagnosticar? O si un vehicle autònom basat en IA transporta mercaderies i persones amb una taxa

13. P. DOURISH, *Where the Action Is. The Foundations of Embodied Interaction*, MIT Press, 2004.

d'accidents molt menor que un camioner o un taxista, ¿no hauríem de deixar a les seves mans la tasca de conduir?

TECNOCIÈNCIA I LLIBERTAT

Aquesta situació crea la paradoxa següent. Durant els darrers tres segles, la tecnociència s'ha convertit en l'instrument i el símbol predominant de l'alliberament humà, i ha proporcionat cada vegada més llibertat creativa a les persones i les societats a l'hora de decidir com volien dur els seus assumptes a la vida: l'economia, el govern, la llei, la sanitat, l'educació, etc. Però «racionalitzant» els nostres processos a causa de l'avanç de la ciència i la tecnologia, estem eliminant la llibertat creativa originalment concedida a les decisions preses pels comerciants, els responsables polítics, els jutges, els metges, els mestres... Els fem encabir en procediments algorítmics.

Les contribucions de la ciència i de la tecnologia no haurien d'avaluar-se tan sols en termes d'utilitat, de creixement econòmic o de maximització d'una mesura de rendiment, sinó en relació amb la forma d'ajudar i ampliar —i no pas obstaculitzar o suplantar— la nostra llibertat creativa. I no només de les persones, sinó de tota la creació. I això també s'aplica a com abordem la investigació en IA.

Si ens mantenim en el marc actual heretat de les societats industrials, centrat en la producció i la utilitat material, governat per una economia de consum, amb el benestar conceptualitzat com a riquesa i mesurat en termes de valor monetari, i si ens quedem en una conceptualització de la IA en termes d'agents autònoms i racionals que maximitzen la utilitat esperada, aleshores la tecnologia basada en IA pot continuar beneficiant tan sols una minoria de la població humana, pot seguir reforçant la distribució actual de poder, i pot continuar alimentant l'economia de consum, que no condueix a l'alliberament sinó més aviat a l'esclavitud. El programa de recerca en IA i la quarta revolució científica de Floridi que he esmentat anteriorment poden convertir-se en la millor aliada d'un model social impulsat per la teoria de l'elecció racional i per les metàfores computacionals de la racionalitat a què la llibertat es restringeix cada cop més.

Només cal veure quins són actualment els actors principals de la recerca en IA. No ho són pas les universitats o els centres de recerca d'utilitat pública i sense ànim de lucre, sinó que ho són les grans empreses tecnològiques, que seran les més beneficiades de la creixent racionalització i quantificació de totes les nostres activitats humanes. Les activitats d'aquestes empreses han anat creant un mercat que es fonamenta en la predicció cada cop més acurada dels nostres comportaments a través de l'extracció i l'anàlisi de grans quantitats de dades. Els beneficis econòmics que genera aquest mercat podrien ser encara més grans si s'aconseguís incidir en el nostre comportament futur per modificar-lo, reduint així la impredictibilitat inherent de la llibertat creativa de l'ésser humà.¹⁴

Un exemple és la tendència cap a l'*algorocràcia*, el govern dels algorismes, és a dir la creixent dependència dels processos de presa de decisions d'algoritmes. Les inversions en borsa i els mercats financers en general es regeixen gairebé exclusivament per algorismes, i no cal recordar com el sistema financer actual del nostre món ha restringit les llibertats individuals i socials i ens ha abocat a una crisi financera global. Un altre exemple més local seria la vigilància governada per algorismes. A Chicago, per exemple, la policia va fer servir algorismes basats en IA per generar uns *heat maps* ('mapes de calor') per decidir on identificar persones que molt probablement podrien estar involucrades en un tiroteig; un estudi posterior, però, va mostrar que l'ús d'aquests mapes no reduïa pas el nombre de crims comesos, però en canvi sí que generava una excessiva activitat policial en comunitats marginals i augmentava la probabilitat que la policia arrestés persones de determinats col·lectius. Això també fa minvar la nostra llibertat.¹⁵

14. S. ZUBOFF, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, Profile Books, 2019.

15. J. DANAHER, «The Threat of Algorocracy: Reality, Resistance and Accommodation», *Philosophy & Technology* 29(3) (2016), p. 245-268. J. SAUNDERS - P. HUNT - J. S. HOLLYWOOD, «Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot», *Journal of Experimental Criminology* 12(3) (2016), p. 347-371.

Nodrir la llibertat és un signe de creixement espiritual, ja que la llibertat mateixa és una dimensió espiritual. De fet, podem veure la llibertat com el símbol secular de l'Esperit. Aleshores, ¿com podem fer que la llibertat brilli a través de la investigació en IA, així com de qualsevol altre esforç de recerca en tecnociència, perquè aquesta investigació sigui un veritable testimoni del camí d'indagació de la humanitat cap a la plenitud de vida? Potser primer de tot cal que ens alliberem de la idea que la intel·ligència és quelcom atribuïble a un organisme o entitat, des d'una perspectiva purament objectiva, de tercera persona, i començar a aproximar-nos al fenomen de la intel·ligència situant-la en la interrelació *entre* organismes, des d'una perspectiva intersubjectiva, de primera i segona persona; i fins i tot explorar la seva *intra-acció*, prenent l'acció intel·ligent com a principi constitutiu dels organismes. Hi ha corrents de la psicologia i la ciència cognitiva contemporànies que posen seriosament en qüestió la visió objectivista de la intel·ligència i no plantegen pas que el cervell sigui un mecanisme creador de la intel·ligència, sinó més aviat un òrgan de mediació d'una intel·ligència ecològica, una intel·ligència de la realitat mateixa en la seva totalitat, que s'estén més enllà d'un sol organisme i que se'ns manifesta en la interacció i *intra-acció* dels organismes amb l'entorn i entre ells. La nostra corporeïtat participa a l'hora de donar i crear sentit a la nostra experiència, i evidentment el cervell hi té un paper important, però no és el creador del sentit.¹⁶

Aquesta visió més ecològica de la intel·ligència fins i tot va més enllà de la investigació que s'està duent a terme en propostes «corpòries» de la IA, com les que es fan des de la robòtica, i que es basen en la visió que per assolir una IA cal tenir en compte el context de la interacció que sistemes robòtics tenen amb altres entitats i amb l'entorn. Tanmateix, una IA corpòria en un sistema robòtic no deixa de ser un processador d'informa-

16. K. BARAD, *Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning*, Duke University Press, 2007. T. FUCHS, *Ecology of the Brain. The Phenomenology and Biology of the Embodied Mind*, Oxford University Press, 2018. M. JOHNSON, *The Meaning of the Body. Aesthetics of Human Understanding*, The University of Chicago Press, 2007.

ció, de la informació que rep dels seus sensors i que, tant per tant, podria provenir d'una realitat virtual. El sistema d'IA continua fent un processament purament objectiu d'acord amb un model de la realitat, molt allunyat de l'enacció directa i subjectiva amb la realitat que tenen els éssers vius a través dels seus sentits.¹⁷

INFORMACIÓ I LÍMITS

La investigació científica és el tipus d'indagació que ens duu a una manera de conèixer que es fonamenta en la informació: és una forma de coneixement utilitària. Per això, la investigació científica és també el motor de la nostra tecnologia. *Tecnologia* és un terme força recent per allò que abans del segle XX es coneixia com les *arts útils*. La informació mateixa sorgeix de la nostra capacitat per establir distincions en la nostra experiència. Fixeu-vos que vivim en un món ple de contraris. Organitzem l'espai físic en què ens movem: amunt i avall, dins i fora, gran i petit, aquí i allà, esquerra i dreta; proporcionen contingut a les nostres creences: veritat i fals, ser i no ser, aparença i realitat; i també estructuren els nostres valors: bo i dolent, plaer i dolor, llibertat i servitud.

¿D'on provenen tots aquests contraris? Els científics cognitius contemporanis afirmen que la nostra experiència corpòria amb el món crea en el nostre inconscient cognitiu certes estructures prelingüístiques anomenades *esquemes d'imatge* amb què edifiquem els nostres sistemes conceptuals. Per exemple, quan érem nadons i interactuàvem amb l'entorn posant unes coses dins d'altres, amagant-les en caixes i agafant-les de nou per treure-les, etc., el nostre inconscient cognitiu va construir una estructura mental del que significa ser un recipient que conté coses (l'esquema d'imatge anomenat RECIPIENT, que determina la relació que hi ha entre un interior i un exterior, i la frontera que els delimita). En fases subsegüents del nostre desenvolupament cognitiu, utilitzem les estructures

17. R. BROOKS, *Cambrian Intelligence. The Early History of the New AI*, MIT Press, 1999.

cognitives bàsiques adquirides a partir de la nostra experiència sensorial i motora (com ara l'esquema d'imatge RECIPIENT) i les fem servir per estructurar conceptes més abstractes, com ara la noció de *categoria*.¹⁸

Des de la nostra experiència d'observar que les coses que comparteixen propietats solen estar a la mateixa regió limitada (arbres en un bosc, raïm en vinyes, ramats d'animals, estols d'ocells, etc.), estructurarem el concepte abstracte de *categoria* utilitzant l'esquema RECIPIENT, de manera que «posem dins» de la categoria *raïm* allò que identifiquem com instàncies particulars de raïm, mentre que les coses que no identifiquem com a raïm les «col·loquem fora» de la categoria *raïm*. I així classifiquem les coses dins o fora de les categories que anem creant: *cavall*, *arbre*, *ocell*, etc. I a mesura que desenvolupem les nostres habilitats cognitives, comencem a crear categories cada cop més abstractes, però sempre estructurades en darrera instància pel nostre esquema d'imatge RECIPIENT que vam adquirir quan érem infants i que estableix un límit entre interior i exterior. D'aquesta manera sorgeix el món de contraris: gran/petit, veritable/fals, bo/dolent, plaer/dolor, etc.

Aquesta lògica binària de l'esquema RECIPIENT i la noció de classificació en categories es troba en el nucli mateix de la noció d'*informació* i en conseqüència del coneixement que obtenim a través de la investigació científica. Informar és, literalment, 'posar en una forma', on la paraula del llatí *forma* significa 'contorn, límit'. La paraula *ciència*, al seu torn, prové del llatí *scindere*, que significa 'tallar, dividir'.

ELS CONTRARIS I LA SEVA GESTALT

No obstant això, els esquemes d'imatges, com ara RECIPIENT, formen una estructura denominada Gestalt: no existeix un interior si no es conceptualitza respecte a un exterior i una frontera que delimiti l'interior de l'exterior; el mateix passa amb l'exterior, que no existeix de manera aïllada sense

18. G. LAKOFF - M. JOHNSON, *Philosophy in the Flesh*, Basic Books, 1999. (Adoptem la convenció d'escriure els esquemes d'imatge en versaletes per distingir-los dels conceptes del mateix nom.)

un interior i una frontera que els delimiti; i sempre que establim alguna frontera, creem també un interior i un exterior. Aquesta estructura de Gestalt es transporta (μεταφορά) als contraris que hem conceptualitzat en termes de recipient, de manera que no hi ha plaer sense dolor ni dolor sense plaer, així com no hi ha bondat sense maldat ni maldat sense bondat. Ara bé, com hem dit abans, un esquema d'imatge com a RECIPIENT és una estructura prelingüística del nostre inconscient cognitiu. I quan a més hi afegim la parla, aleshores succeeix quelcom meravellós que ens atorga el gran poder que tenim.

Amb la parla definim i identifiquem les categories i els contraris que creem, i això ens permet evocar amb la simple pronunciació d'un mot com *dolor* tot l'univers d'experiències que hem classificat com a doloroses. Però aquesta capacitat també ens allunya de cada vivència particular i singular de l'experiència de dolor, l'experiència prèvia a la seva classificació dins de la categoria *dolor* i fora de la categoria *plaer*. Però aquesta capacitat nostra de la parla pot crear la il·lusió que existeix una separació, ja que ens capacita a raonar amb el mot que defineix i identifica una categoria determinada, per exemple *plaer*, tot construint arguments que només fan servir la paraula *plaer*, sense esmentar mai explícitament la paraula *dolor*. I això, al seu torn, ens fa pensar que hi ha la possibilitat de desfer-nos d'una de les meitats.¹⁹

Aquesta capacitat de manipulació que ens ofereix la parla ha estat tan útil per a la prosperitat de la nostra espècie, i ens ha atorgat tant de poder, que prenem el món tal com sorgeix d'aquesta *in-formació* i la seva subsegüent manipulabilitat com a món *real*. «El que és real és informatiu i el que és informatiu és real», afirma Floridi.²⁰

Però això només és la meitat de la història. El poder que hem guanyat mitjançant la categorització i la parla ve acompanyat d'una pèrdua d'immediatesa: el coneixement que obtenim amb aquesta *in-formació* porta en si mateixa la pèrdua de l'experiència immediata *no in-formada*. Parado-

19. K. WILBER, *La consciència sin fronteras*, Barcelona, Kairós, 1985.

20. L. FLORIDI, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, 2016.

xalment, però, en virtut de la mateixa capacitat que tenim de categoritzar i anomenar —cosa que crea el món informatiu—, també tenim la consciència que els contorns, amb els seus interiors i exteriors i els contraris que creen, no existeixen intrínsecament en el món. En realitat només són relatius a la cognició de l'observador. La nostra capacitat de parla és, doncs, una espasa de doble tall: d'una banda, divideix el món en contraris i ens porta a apreciar el món tal com se'ns mostra, ple de dualitats, però, de l'altra, ens fa conscients que allò que apreciem i coneixem sempre serà relatiu a l'experiència corpòria que modela el nostre inconscient cognitiu, i per tant permet adonar-nos que aquesta realitat plena de formes és alhora buida de formes. És en el fons una realitat lliure. I no es tracta que considerem una realitat «més real» que l'altra —la lliure més real que la *in-formada*, o viceversa—. No són dues realitats, però tampoc és una de sola. Amb la nostra cognició contínuament creem la realitat (la realitat *in-formativa*) en què vivim. No hi ha res a fer: és la realitat humana relativa a la nostra experiència com a organismes vivents amb la nostra corporeïtat específica. Però no hi ha un fragment de la realitat que siguem capaços de conèixer i un altre fragment que no. No es tracta de tornar a traçar una frontera que estableixi de nou uns contraris. En tota la realitat hi ha alhora un saber i un no-saber.

LA DOBLE VIA DE LA INDAGACIÓ

Per això, la recerca científica només és mitja indagació. La investigació científica es basa intrínsecament en el fet d'establir límits i categoritzacions. Crea coneixements dualistes. Es basa en la informació. I genera la mena de coneixement que després hom pot utilitzar a través de la tecnologia. La investigació científica és també el motor que ens duu al progrés, perquè el progrés requereix una categorització, en tant que fa que ens allunyem del que qualifiquem com a *dolent* per apropar-nos a allò que qualifiquem com a *bo*. Així doncs, «progressem» a la recerca d'una de les meitats del parell de contraris (salut, plaer, bellesa, riquesa...), intentant erradicar la meitat contrària (malaltia, dolor, lletjor, pobresa...), impulsats per les pa-

relles de contraris que constitueixen el nostre sistema de valors. Però aquesta naturalesa intrínsecament dualista de la recerca científica és la que la fa inadequada a l'hora de proporcionar-nos l'orientació i la motivació necessàries en la nostra tasca d'investigació.

Els contraris que constitueixen el sistema de valors que impulsa la nostra investigació científica i les metàfores conceptuals sobre les quals es basen les nostres conceptualitzacions només es poden indagar sabent mirar a través de la il·lusió dels contraris, transcendint-los. Això és, en darrera instància, el que significa alliberar-se. La llibertat és l'aprehensió directa, no pas mitjançada, de la dimensió absoluta de la realitat, la realitat que no és informativa, la realitat indivisa, no cognoscible, no dual, més enllà dels contraris. Ara bé, la investigació científica continuarà produint «progressos» en medicina, agricultura i tecnologia, que seran «progressos» en el sentit que estaran motivats i impulsats pels contraris que constitueixen el sistema de valors que fan que els apreciem com a «progressos» d'una meitat cap a l'altra.

La vida humana necessita contraris per trobar orientació. Ho portem al cos. Com hem dit, la ciència cognitiva contemporània ens mostra que les estructures de l'experiència viscuda són essencialment no conscients i a més estan totalment condicionades per la nostra corporeïtat. Les nostres conceptualitzacions del temps, de la causalitat, de la nostra ment, del jo, totes elles són metafòriques, fonamentades en la nostra experiència corporal amb altres organismes vivents i amb l'entorn. És més: sovint en el nostre raonament hi participen també de manera determinant les sensacions sentides en el nostre cos, sensacions que tenen un sentit molt concret encara que no el conceptualitzem simbòlicament. I aquestes estructures cognitives determinen al seu torn l'estructura cerebral.²¹

Cal continuar fomentant la recerca científica per progressar en la orientació que li donem a la nostra vida. Però el fet de saber apreciar que tota la ciència i tot el coneixement que hem anat construint és essencial-

21. E. GENDLIN, *Experiencing and the Creation of Meaning. A Philosophical and Psychological Approach to the Subjective*, Northwestern University Press, 1997.

ment de naturalesa metafòrica, basats en el nostre sentir corporal, és el que ens permet saber que la realitat és lliure respecte al nostre sentir i al nostre conceptualitzar, i que és això el que ens permet fer indagació científica. La indagació és l'expressió d'aquesta llibertat. Podem estar oberts a aquesta realitat lliure que ens porta a canviar unes formes per unes altres, i que en definitiva ens *trans-forma*. És per això que la indagació científica necessita del cultiu d'aquesta capacitat de transcendir la dualitat de les formes per així contribuir als canvis de paradigma de què parlava Thomas Kuhn, si hom no vol quedar-se estancat en una determinada manera d'entendre i percebre la realitat.

La indagació científica necessita anar de la mà de la mena d'indagació que transcendeix la il·lusió dels límits i les fronteres i dels contraris que creen, i ens allibera del nostre aferrament a aquest contraris, i no fonamentar la nostra felicitat i plenitud de vida en aquests límits i aquestes fronteres. Podríem anomenar aquest tipus d'indagació una indagació *alliberadora*, *no dual* o *advaita*, o, per utilitzar un terme de la tradició cristiana, una indagació *kenòtica*, fonamentada en la *κένωσις* (*kénosis*), que significa literalment 'vacuïtat'. És a dir, la indagació kenòtica és la indagació que ens buida i ens allibera de les formes. La indagació científica sempre serà informativa, impregnada per les formes, però no serà indagació si no va cultivada per la indagació kenòtica, alliberadora.

INDAGACIÓ KENÒTICA PER A TECNOCIENTÍFICS

¿Com seria aquesta indagació kenòtica, aquesta indagació alliberadora? Doncs bé, no ens cal inventar aquest tipus d'indagació del no res, sinó que podem seguir els passos d'aquells que ens han precedit en aquest camí de la indagació, de la mateixa manera que ho fem amb la nostra recerca científica. Qualsevol forma d'oració contemplativa és indagació kenòtica. La meditació és probablement la més òbvia. El silenciament dels nostres pensaments, imatges i desitjos a través de la fidel repetició d'un *mantra*, o posant l'atenció en la respiració, fa que la dimensió contemplativa i no dual de la realitat brilli a través de totes les nostres activi-

tats. La meua experiència personal amb la pràctica diària de meditació és el que em va ajudar a posar-me de nou en contacte amb el nucli contemplatiu de la recerca científica: amb el sentit de meravella i esbalaïment que el condueix; amb la necessitat de silenci, atenció, cura i respecte; amb l'amor pel subjecte investigat; amb el plaer de compartir idees amb els companys; amb la humilitat que nodreix la investigació científica.

Per tant, els científics i els enginyers ens beneficiaríem enormement en la nostra tasca diària de recerca científica i desenvolupament tecnològic si sabéssim integrar aquesta tasca amb una pràctica diària d'indagació kenòtica com la meditació. És irònic que aquest tipus de pràctica, que alguns científics ortodoxos consideren *descervellada*, és realment allò que ens fa més lliures i creatius, i que és per tant la característica essencial de la nostra humanitat. Però no tan sols ens en beneficiaríem científics i enginyers. La humanitat es veuria enriquida amb els resultats d'una tecnociència feta amb contemplació.

A més de la meditació, aquí mostro alguns exemples addicionals de pràctiques d'indagació kenòtica que els científics i els enginyers podríem —i hauríem de— conjugar amb la nostra recerca científica:

a) *Llegir i escriure contemplativament.* Les pressions de la vida acadèmica per maximitzar el nombre de publicacions i de citacions és tan elevada, i la quantitat de documents que es generen i que cal llegir i revisar és tan nombrosa, que els científics perdem el contacte amb el valor que s'obté d'una lectura lenta i pausada d'un text, que permet alternar moments de reflexió i de silenci —gairebé com una *lectio divina*—. I també el valor d'escriure lentament. Escriure un diari de recerca pot ser una bona pràctica contemplativa: ser breu, però escriure diàriament, i cultivar la *taciturnitas*, l'actitud d'expressar tan sols allò que és realment necessari comunicar, i fer-ho amb actitud amorosa. Trobo molt encertat el proverbi gaèlic que diu: «Abair ach beagen is abair gu math e» ('Digues poc i digues-ho bé').

b) *Escoltar amb profunditat i atenció plena.* He vist, quan he

participat en conferències i congressos científics, que cada cop s'escolta menys. Mentre un ponent fa la seva xerrada i projecta les diapositives de la seva presentació, una gran part de l'audiència està ocupada treballant amb els propis ordinadors portàtils, ja sigui responent correus electrònics o acabant de polir la presentació que hauran de fer uns minuts després. Els científics ens oblidem d'escoltar-nos amb atenció, amb respecte. Als nostres centres de recerca i laboratoris hauríem de practicar una escolta atenta: fer regularment l'exercici d'explicar-nos mútuament la nostra recerca i les nostres preguntes obertes, i practicar el fet de saber reproduir de la manera més fidedigna possible allò que ens han comunicat els companys, fins que se sentin veritablement escoltats. I també hem de tornar a aprendre a saber apreciar el «paisatge» més ampli. Sovint els científics estem massa preocupats en l'elaboració del petit detall i no ens deixem captivar per la visió general i àmplia de les coses. Cal aprendre a situar-nos davant del nostre tema de recerca i saber dissoldre la frontera entre subjecte i objecte per tal que l'experiència immediata enriqueixi la investigació científica, i complementar d'aquesta manera la divisió de subjecte i objecte necessària per dur a terme les tasques més analítiques.

c) *Acollir al foraster*. Cultivar l'hospitalitat intel·lectual és també una pràctica contemplativa important en la investigació científica. Sant Benet, en la seva *Regla per als monjos*, diu que «tots els forasters que es presenten han de ser acollits com el Crist» (cap. LIII, 1): el científic hauria de saber acollir les noves idees que inicialment podrien semblar estranyes. Benet afirma que cal que «es mostri la màxima sol·licitud en l'acolliment dels pobres i dels pelegrins, [...] que el respecte que infonen els rics es fa honorar d'ell mateix» (cap. LIII, 15). No hi ha res d'extraordinari en el fet d'acceptar i elogiar la feina d'un investigador famós o d'un equip que treballa en una universitat prestigiosa o un laboratori de recerca molt conegut; però és la recerca del científic menys conegut la que necessita que li dediquem més cura i atenció.

Per tant, el camí de la indagació que fa la humanitat, perquè sigui complet, és un doble camí d'indagació, científica i kenòtica, de saber i de no-saber, en si mateix un parell de contraris que també cal transcendir: ni dos, ni un.

TORNANT A RAMON LLULL

Això em porta de nou a Ramon Llull, el místic «pioner» de la IA. Sovint es crea una separació artificial entre l'èmfasi que Llull posava, per una banda, en la racionalitat i en el desenvolupament del seu Art per mecanitzar el raonament i, per altra banda, en l'experiència mística i la vocació que el portaren a aplicar el seu Art en el diàleg i l'argumentació amb els musulmans amb l'objectiu de convertir-los al cristianisme. Així doncs, freqüentment hom s'ha centrat sovint en l'estudi del sistema lògic de Llull independentment de la seva experiència mística, perquè se la veia innecessària per avaluar-ne les contribucions a la mecanització del pensament.²²

Però el propi Llull no separava la racionalitat (el domini de la investigació científica i del coneixement dualista) de l'experiència directa del diví (el domini de la investigació kenòtica i del no-saber no dual). Perquè, per tal de construir tot l'edifici del raonament racional dirigit a la tasca de conversió, començava amb allò que els cristians i els musulmans compartien: l'experiència directa del diví, que les dues tradicions expressaven a través d'atributs associats a aquesta experiència directa (bondat, grandesa, eternitat, poder, saviesa, voluntat, virtut, veritat i glòria). Tot l'Art lul·lià hauria estat tan sols una manipulació buida de símbols abstractes si Llull no l'hagués fonamentat en aquesta aprehensió directa de l'Absolut. Així, doncs, el raonament i la racionalitat només tenien la capacitat de convèncer i convertir si en definitiva estaven vinculats a l'experiència de la reali-

22. E. BONET, *Comments on the Logic and Rhetoric of Ramon Llull*, dins A. FIDORRA - C. SIERRA (coord.), *Ramon Llull: From the Ars Magna to Artificial Intelligence*, IIIA-CSIC, 2011.

tat. La seva indagació també fou, per tant, una doble indagació que anomenava *sciència* i *amància*, i Llull va idear mètodes per a ambdues: l'*ars inventiva* o art de trobar, i l'*ars amativa* o art d'estimar. Mentre la primera guia l'intel·lecte vers la veritat, la segona l'orienta, guiant la voluntat cap al bé.²³

Llull explica com aquestes indagacions es troben íntimament emparellades i ens adverteix que no podem considerar pas que aquests dos mètodes pertanyin a àmbits diferents de la nostra tasca indagadora. Com a místic fou capaç de mirar a través dels contorns i límits que imposen a la realitat i així captar-ne la dimensió no dual. Deixeu-me concloure, doncs, aquest assaig amb una citació del més conegut dels seus escrits místics:

Foll, digues, què és amor. Respòs que amor és aquella cosa qui los francs met en servitud, e a los serfs dóna llibertat. E és qüestió a qual és pus prop amor: o a llibertat, o a servitud (*Llibre d'amic e amat*, 294).

23. A. BONNER, *L'Art i la lògica de Ramon Llull: manual d'ús*, Barcelona, Universitat de Barcelona, 2012.

QUADERNS
DE LA FUNDACIÓ JOAN MARAGALL

96. D. JOU: *Trenta poemes sobre ciència i fe*
97. A. BORRELL: *El viatge interior. De la recerca de si mateix a l'encontre amb l'Absolut*
98. J. MOLAS: *Maragall i Azorín*
99. I. MORETA: *Joan Maragall, pensador religiós. Antologia comentada de textos*
100. A. MATABOSCH: *Testimoniatsges. Identitat cristiana i diàleg interreligiós*
101. R. TORRENTS: *Entorn de Maragall i Gaudí*
102. M. T. ARECES - S. BUENO - R. JORBA: *Llibertat religiosa i societat actual*
103. J. MARTÍN VELASCO: *Fidelitat al Vaticà II en el segle XXI*
104. C. LLUCH: *Joan Sales i la novel·la catòlica*
105. D. JOU: *Esperances davant d'incerteses. Tecnologia, humanitats i cristianisme en les crisis d'avui*
106. J. CASASSAS: *El canvi cultural del segle XX*
107. I. BOADA: *L'home modern: un balanç*
108. A. MATABOSCH: *Del diàleg ecumènic al diàleg interreligiós. Característiques i possibilitats*
109. D. ABADÍAS - I. FERNÁNDEZ TERRICABRAS: *Església i poder. Els casos d'Anglaterra i França*

110. J. REQUESENS: *La presència de santa Teresa en la poesia catalana. Tres hipòtesis i una constatació*
111. F. TOUS: *Els provebis de Ramon Llull o el repte de comunicar la veritat*
112. F. TORRALBA: *Humanisme, transhumanisme i posthumanisme. Assaig de discerniment*
113. G. AMENGUAL: *Identitat, memòria, alteritat. Per una fenomenologia de l'existència cristiana*
114. N. COMADIRA - J. FERRER - D. JOU - P. LLUÍS FONT - P. PASTOR - A. PUIG I TÀRRECH: *Bíblia i cultura. Sis mirades*
115. R. PUIGDOLLERS I NOBLOM: *Les cartes de sant Pau. Introducció a l'epistolari paulí*
116. J. CABÓ: *Maurice Blanchot, testimoni de l'impossible. Una introducció*
117. A. BALCELLS: *Secularització i confessionalitat. Un recorregut per l'anticlericalisme a la Catalunya contemporània*
118. F. TORRALBA - J. L. MARTÍNEZ: *L'intel·lectual cristià avui*
119. L. TORCAL: *Conversió ecològica integral. Una resposta cristiana al problema del canvi climàtic*
120. C. DI SANTE: *Per una espiritualitat reconciliada. Un cristianisme en temps de contradiccions*