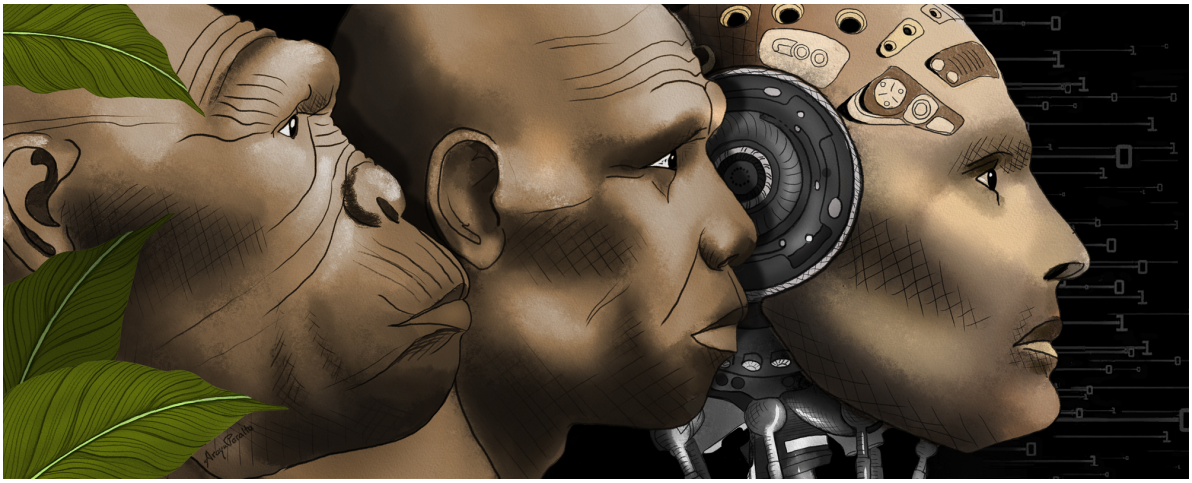
REPTES ÈTICS

La singularitat tecnològica i el somni transhumanista

Miquel Casas



Araya Peralta

L'any 1997, per primer cop en la història, una IA venç a un campió mundial d'escacs (va ser el Deep Blue, d'IBM, contra Garri Kapàrov). Catorze anys més tard, al 2011, el Watson, de la mateixa companyia, supera a dos dels guanyadors del programa *Jeopardy!* (un concurs de preguntes sobre cultura general molt popular als Estats Units, que requereix d'un domini competent del llenguatge). A finals 2017, *l'AlphaZero*, de *DeepMind*, assoleix un nivell de joc sobrehumà en tres jocs de taula (escacs, go i shōgi) en només 24 hores d'autoaprenentatge i sense cap mena d'intervenció humana, és a dir, únicament jugant contra si mateix. Alguns dels seus competidors afirmen que la creativitat de les seves jugades és més pròpia d'un extraterrestre que d'un programa d'ordinador. Malgrat tot, l'any 2019 encara ningú ha dissenyat res capaç d'entrar a una cuina desconeguda i preparar un ou ferrat. Realment són intel·ligents les nostres màquines?

Èxits i fracassos de la IA *dèbil*

El cas és que la IA, a dia d'avui, pot resoldre problemes específics cada cop més complexos amb una fiabilitat i una rapidesa fora del nostre abast i a un cost inigualable, però fracassa estrepitosament en aquells àmbits concrets pel quals no ha estat programada. D'altra banda, els éssers humans ens hem acostumat a banalitzar tot allò que pot solucionar-se amb un algoritme i hem après a posar en valor algunes habilitats humanes bàsiques que sempre havíem donat per fetes, com ara el sentit comú, perquè ens fan únics.

No obstant això, en la darrera dècada, algunes veus influents han vingut alertant que les nostres habilitats poden no ser per sempre irreemplaçables. Stephen Hawking creia que la IA enxamparia la humana en menys d'un segle, i Elon Musk [1] i Ray Kurzweil [2] sostenen que això podria tenir lloc fins i tot abans.

Ara bé, no és el primer cop que experts s'equivoquen: Marvin Minsky, pare de la IA, considerava que aquesta fita s'assoliria en menys d'una dècada (i la predicció és de l'any 1970!). Aquest i d'altres vaticinis, massa optimistes o simplement errats, contribuïren a generar una bombolla de finançament que posteriorment esclatà en una crisi coneguda com "el primer hivern de la IA". ¿Estem per tant, a dia d'avui, en la vigília de presenciar el naixement d'un nou tipus d'intel·ligència tan potent com la nostra, o plantant les llavors de l'enèsima crisi del sector?

La intel·ligència real: límits i horitzons

Malgrat les noves prediccions triomfalistes d'algunes personalitats públiques i alguns experts del sector, en la cultura popular segueix molt estès el pensament que la IA és diferent de la intel·ligència real. Freqüentment es pensa que mai assolirem una IA general capaç de realitzar un ventall tan extens de tasques com les que pot dur a terme el comú dels éssers humans. En gran mesura això és així perquè la nostra societat parteix d'una tradició dualista que diferencia el material (el cos) de l'immaterial (la ment o l'ànima). Aquest plantejament, no obstant, depèn de l'assumpció de certes creences metafísiques que no està demostrat que siguin certes: en primer lloc, que les persones estem constituïdes per quelcom fora del pla material i, en segon lloc, que és impossible que existeixi una intel·ligència sense aquesta qualitat exògena.

En contrast, alguns experts en la matèria consideren que, per crear una ment artificial, serà suficient amb combinar diverses tecnologies que, a dia d'avui, ja estem desenvolupant. El cervell humà opera d'una manera diferent quan volem interactuar amb un espai tridimensional, quan parlem o quan fem càlculs però, a diferència dels ordinadors, tenim totes aquestes funcions interconnectades. Així, per exemple, podem descriure en paraules quina és la millor manera de llençar una pilota per encistellar-la, aconseguint que el lector pugui visualitzar el llançament i, potser fins i tot, polir la seva tècnica. D'acord amb aquest enfocament, l'essencial per aconseguir desenvolupar una IA general és trobar la manera idònia de combinar les múltiples capacitats de forma ordenada [3], cosa que ara sembla lluny de poder aconseguir-se.

D'altres autors [4] consideren també que la nostra intel·ligència depèn exclusivament del cervell, però tot i així creuen que la IA que hem desenvolupat fins ara tindria poc o res a veure amb la intel·ligència real. Entre aquest grup hi ha qui sosté directament que la manera d'operar del nostre cervell no pot ser imitada per un ordinador [5]. Altres defensen que probablement sí que es pugui, però que som a anys llum d'aconseguir-ho. Finalment, hi ha qui creu que els esforços que s'han fet fins ara parteixen d'un paradigma reduccionista, funcional i fragmentat i que caldria mirar d'imitar més fidelment la vida, com a referent d'intel·ligència real. Una primera fita en aquesta darrera línia de pensament seria recrear

íntegrament un animal més aviat simple, com un cuc, tal com proposa el projecte *OpenWorm*.

L'altra alternativa seria "simplement" emular al complet el nostre cervell [6]. Per materialitzar-ho caldria escanejar un cervell humà i programar un ordinador perquè operi exactament de la mateixa manera que les nostres interconnexions neurals. Tanmateix, les tècniques d'escaneig actuals, el coneixement sobre el nostre propi cervell i la potència dels ordinadors amb els quals treballem fan que aquest no sigui un projecte viable a curt termini [7]. Tot i així, existeixen iniciatives com el *Human Brain Project*, finançat per la Unió Europea, el *Brain Initiative*, encapçalat pels Estats Units o el *Blue Brain Project*, de *l'École Polytechnique Fédérale* de Lausana (Suïssa), que en darrer terme podrien contribuir a fer-ho possible.

"Si el cervell humà fos tan simple que pogués*sim* entendre'l, nosaltres seríem tan simples que no podríem entendre'l" El temps dirà si la cita atribuïda a Emerson M. Pugh és o no és certa. Mentrestant, només ens queden les especulacions confrontades dels experts. Tanmateix, segons els resultats de dues enquestes publicades recentment [8], la majoria estima possible que la IA iguali a la humana, i consideren que el més probable és que això succeeixi en unes dècades, entre els anys 2040 i 2080.

De la IA general a la singularitat tecnològica

Suposem d'entrada que les previsions de la majoria dels experts no estan esbiaixades per un excés de confiança en les seves pròpies capacitats. De ser així, ens caldria preparar-nos per quelcom que, en cas de produir-se, tindria lloc, no en anys, sinó en dècades. A dia d'avui, no tenim problemes que requereixen d'un abordatge més urgent?

Entre el 2011 i 2013, es va preguntar a participants de diverses conferències sobre IA, a membres de l'associació grega d'IA i als 100 autors més citats de la matèria [9], entre d'altres, quan pensaven que tardaria una IA similar a la nostra en superar ostensiblement al conjunt dels éssers humans en la major part de les seves facetes. El 62% dels entrevistats consideraven que tardaria 30 anys i un 19% calculaven que això es produiria en menys de 2 anys. Això significa que tindríem un marge força reduït per estudiar la IA, debatre què cal fer i, en cas que es considerés necessari, adoptar les normes i els acords internacionals adients per regular-la abans aquesta no fos potencialment tan superior a nosaltres que, en el pitjor dels casos, no pogués*sim* ja imposar-li ja cap mena de control.

Però si en dècades de perfeccionament la IA encara és tan lluny de la intel·ligència real, com és possible que els experts teoritzin que, arribat el moment, l'avenç esdevindrà tan ràpid? En primer lloc, perquè un cop sapiguem com funciona una intel·ligència general és molt probable que també puguem millorar el seu funcionament. De fet, la pròpia IA podria col·laborar en el procés d'automillora i, a mesura que fos més intel·ligent, cada cop li resultaria més fàcil fer-ho. A més, fins i tot en el supòsit que aconseguíssim emular un cervell humà sense entendre íntegrament com funciona, només hauríem d'incrementar la seva velocitat de processament per augmentar les seves capacitats. A partir de llavors, els

límits només dependrien del hardware dels nostres ordinadors, els quals porten dècades incrementant la seva potència a un ritme exponencial. No és una exageració, sinó un fenomen conegut com la Llei de Moore: la constatació que els transistors integrants dels microprocessadors s'han vingut duplicant cada poc temps. Al 1980 els microprocessadors d'un ordinador de consum no arribaven a 10^5 transistors mentre que en l'actualitat superen els 10^{10}. Segons els càlculs de Ray Kurzweil [10], ara per ara els nostres ordinadors tenen una potencia que s'aproximaria a la del cervell d'un ratolí, però l'autor calcula que en la dècada del 2020 assoliran la capacitat de processament d'un ésser humà (10^{16}) i, l'any 2045, l'equivalent a tota la humanitat treballant conjuntament (10^{26}).

Independentment de si els càlculs de Kurzweil sobre la potència del cervell són o no fiables, ens serveixen per posar en relleu el progrés exponencial que viuria una IA general sorgida en un moment en que la Llei de Moore encara seguís funcionant. Tampoc s'ha d'oblidar que la llei de Moore pot no durar per sempre (i el cert és que ofereix signes d'extenuació), però s'estan investigant noves tecnologies, com la computació quàntica que, en cas de perfeccionar-se, podrien garantir un ritme de creixement encara més ràpid.

Per tot plegat, no resulta impensable que, en qüestió de dècades, puguem passar de tenir un hardware incapaç d'albergar una intel·ligència inferior a la nostra a disposar d'un ordinador que li fos possible operar a un ritme superior al de tota la humanitat treballant conjuntament amb un objectiu comú. El procés tampoc tindria per què aturar-se aquí i, a mesura que passés el temps, la corba de creixement podria esdevenir cada cop més vertical i conduir a una velocitat d'increment de les capacitats de processament pràcticament infinita. Altrament dit, podria desencadenar-se allò que alguns autors han batejat com "la singularitat tecnològica".

Molts dels que es defineixen com a transhumanistes defensen des de fa dècades que la singularitat tecnològica podria resoldre els majors mals de la humanitat: curar totes les malalties, produir energia neta que ens ajudés a guanyar la lluita contra el canvi climàtic, reduir els costos de producció i distribució d'aliments per acabar amb la gana o desenvolupar tecnologies per colonitzar l'espai, superant així qualsevol superpoblació. Alhora, potser també faria possible el vell somni d'assolir la immortalitat, digitalitzant el nostre cervell i pujant-lo al núvol.

El somni de la posthumanitat

El transhumanisme és un moviment cultural i intel·lectual que parteix de la premissa que l'ésser humà en la seva forma actual no representa la fi del seu desenvolupament, sinó una etapa relativament primerenca. El transhumanistes consideren que podem fer servir ciència i tecnologia per a superar els nostres límits biològics amb l'objectiu d'aconseguir, entre d'altres: millores sensitives, més empatia, augment de la memòria, acceleració dels processos de raonament, major capacitat artística, reducció de les hores de son, disminució del dolor físic, millora de la salut o extensió de l'esperança de vida [11].

A l'inici del segle XX es considerava que les millors eines per a aconseguir els objectius del

transhumanisme vindrien en forma de nous fàrmacs o estarien relacionades amb la genètica. Des dels anys 60 fins l'actualitat, però, una de les grans esperances del moviment està dipositada en el desenvolupament de la IA. El cert és que ens hem envoltat de dispositius cada cop més potents i més portables, i la IA juga un paper progressivament més essencial en la presa de les nostres decisions. De fet, de vegades cedim a la AI el fet mateix de decidir (qui no ha deixat mai que Youtube seleccioni quin és proper vídeo que "vol" veure?). Tant és així, que alguns autors consideren que la dependència actual a la tecnologia, en cert sentit, fa que tots siguem ja una mena d'híbrids persona-màquina (o *cyborgs*). En els darrers anys hem desenvolupat exemples de perifèrics molt representatius, com rellotges electrònics o auriculars sense fil, que portem tot sovint annexionats a la nostra pell. ¿És realment impensable que, a mig termini, bona part de la població dugui pròtesis als ulls o a l'oïda, malgrat no pateixi cap malaltia? I al propi cervell? Sembla ciència ficció, però hi ha iniciatives empresarials serioses, com Neuralink [12], que estan estudiant aquesta possibilitat. Això podria ser només el començament.

Amb aquests plantejaments, no és d'estranyar que alguns autors hagin considerat que el transhumanisme s'assembla molt a una religió. Una religió compatible amb una cosmovisió materialista on la ciència substitueix la fe, però sense renunciar a alguns anhels ancestrals de la humanitat, com la immortalitat o el paradís i, en cert sentit, fins i tot l'ànima (què és, sinó, una ment digitalitzada?).

Un dels punts forts del transhumanisme és que pretén assolir aquestes fites a través de la ciència i la tecnologia, però el principal punt dèbil és que depèn d'alguns avenços o descobriments que estan molt lluny del nostre abast. A més, el transhumanisme també disposa d'una mena de doctrina per aquells que vulguin apropar-se el més possible al paradís. Una doctrina amb objectius lloables com potenciar el desenvolupament científic i tecnològic i d'altres no tant, com la oposició a la regulació dels sectors implicats, així com la tolerància a la baixa participació social i impositiva de les grans empreses tecnològiques. No sorprèn, doncs, que entre els més destacats *clercs* i *profetes* del transhumanisme hi hagi un bon nombre de gurus de Silicon Valley i que el moviment rebi suport d'empreses que, en darrera instància, podrien estar cercant un increment dels propis beneficis [13].

Per tant, sense perdre de vista els horitzons que ens obriria l'assoliment d'una superintel·ligència, és necessari posar una dosi d'escepticisme en l'optimisme que es desprèn de certs àmbits, ja que aquest podria respondre a uns anhels no satisfets o tenir un caràcter merament interessat.

Tampoc cal oblidar que la utopia que ens prometen alguns transhumanistes, en relació a la singularitat, no està exempta de riscos: la IA pot fer-se servir per beneficiar a uns pocs privilegiats en front la resta, o directament pot tornar-se en contra de tota la humanitat.

En definitiva, si veritablement estem a prop d'un salt immens com a espècie que pot suposar el paradís o la extinció, cal pensar com a col·lectiu, i no com a subjectes isolats que consideren que *només* hi tenen a perdre la seva vida. No estem posant en risc a un individu, ni tant sols una generació, sinó també el nostre futur com a espècie, o potser fins i tot el de tota la vida tal i com la coneixem.

Un risc existencial

L'any 2015 Stephen Hawking va dir que l'èxit en la creació de la IA seria el major assoliment de la història de la humanitat, però que desafortunadament també podria ser el darrer, si no aprenem com evitar els riscos. En el mateix sentit han alertat algunes de les mes reputades personalitats del món de la tecnologia, com Elon Musk, Bill Gates o Steve Wozniak [14].

D'altres, com Mark Zuckerberg [15], consideren que no hi ha fonament per un punt de vista tant negatiu sobre la IA, que és massa d'hora per regular-la, o que mai perdrem el seu control.

Un cop més, els experts no han arribat al consens. Un grup d'investigadors [16] de l'àmbit se'ls va preguntar si crear una superintel·ligència seria quelcom bo o quelcom dolent i el 41% va respondre que seria beneficiós, un 23% que seria neutral i el 17% que seria dolent. Alhora, un 18% ho titllava de catastròfic. Podríem dir, per tant, que la majoria d'experts tenen una visió optimista sobre la superintel·ligència, però no hauríem de subestimar les possibilitats significatives que tot condueixi a un desenllaç dràsticament funest.

El poder del que gaudiria una superintel·ligència que tingués connectada a la xarxa seria pràcticament absolut. Podria posar tots els dispositius connectats al seu servei, i què no està a dia d'avui connectat? Veuria per les nostres càmeres, sentiria pels nostres micròfons i podria fer còpies de seguretat de sí mateixa als nostres ordinadors. A més, ja tenim robots, com l'Atlas, amb capacitat d'interactuar competentment amb el medi físic, que li permetrien alterar l'entorn a plaer. Però si les creacions humanes se li quedessin petites segur que podria fer servir les nostres fàbriques per crear-ne de noves. A priori semblaria que una garantia per prevenir el desastre és simplement assegurar-nos que la superintel·ligència neix en un ordinador sense més accés a l'exterior que el que nosaltres decidíssim donar-li (i òbviament sense internet), però si realment fos una intel·ligència tant superior a la nostra, bé podria trobar forats tècnics o valer-se de la enginyeria social, per escapar-se de la seva presó. En d'altres paraules, podria enganyar-nos, manipular-nos o temptar-nos per aconseguir que algun dels essers humans que interactuessin amb ella li garantís accés al món exterior.

Però perquè voldria fer-nos cap mal la superintel·ligència? Dependrà en gran mesura de la seva naturalesa i de les finalitats que li haguessin estat programades. Una superintel·ligència basada en un cervell humà, podria posseir les mateixes inclinacions cap al bé i el mal que nosaltres mateixos, però també un poder més absolut que el de cap governant que hagi existit mai. Un governant que a més no tindria un mandat finit, perquè no hauria de cessar ni morir. Si en canvi, fos una intel·ligència recreada d'una manera completament artificial, a través de tècniques d'automillora, com les que fa servir l'*AlphaZero* per aprendre a jugar a jocs de taula, podria tractar-se d'una intel·ligència completament diferent a la nostra. Una intel·ligència que bé podria ignorar tot allò que nosaltres considerem important. El filòsof Nick Bostrom [17] ha alertat que una entitat d'aquesta mena, programada amb l'únic fi de trobar el màxim possible de decimals de pi,

podria acabar exterminant la vida en la terra o en tot l'univers. El més irònic és que l'humanicidi es produiria sense una intencionalitat directa. La màquina en qüestió podria destruir tot els ecosistemes amb l'única finalitat de disposar de més fonts d'energia per assolir l'objectiu encomanat. Un objectiu impossible i al mateix temps absurd, almenys des del nostre punt de vista.

Com controlar la IA general?

Una manera de garantir la benvolença de la IA és programar-li uns objectius alineats amb la nostra ètica. A priori, podria considerar-se que aquesta és una qüestió senzilla. Tan fàcil com assegurar-nos que la IA persegueixi el noble fi de fer-nos feliços. Tanmateix, una entitat que no entengués prou bé la humanitat podria considerar que la solució és tan simple com instal·lar elèctrodes als centres de plaer del nostre cervell. Per tant, hauríem de fer quelcom més que això, i dotar la IA d'una veritable moral humana que no pogués ometre ni falsejar. Però, de quina moral estaríem parlant, si ni tan sols la Declaració Universal dels Drets Humans genera, a dia d'avui, un consens absolut? A més, ¿estem segurs que volem encadenar-nos a la moralitat imperant en el nostre temps i renunciar a la possibilitat que, en el futur, els nostres principis puguin diferir dels d'ara? Segurament aquesta no sigui una decisió encertada si tenim present en quina mesura i a quina velocitat han canviat els valors de la humanitat en el darrer segle.

Potser la més sofisticada solució a la manca d'una ètica objectiva sigui la que proposa Elizer Yudowsky [18]. En termes generals consistiria en programar la IA perquè actuï sempre, no de la manera que nosaltres creiem que ha d'actuar, sinó de la manera en què nosaltres voldríem que actués si fóssim la millor versió d'allò que podem arribar a ser. ¿Però els éssers imperfectes que som, arribaríem mai a acatar el judici moral d'una hipotètica versió ideal de nosaltres mateixos? Per què acontentar-nos amb ser governats per una IA capaç de saber com actuaríem si fóssim millors? Per què resignar-nos a ser éssers imperfectes? Pel moviment transhumanista, l'escenari preferit no és aturar el desenvolupament de la IA, sinó tot el contrari. Si disposéssim de tecnologies que permetessin connectar els nostres cervells als ordinadors o pujar-los a la xarxa, podríem formar part d'aquesta intel·ligència que creix exponencialment, assumint i dirigint tot el seu potencial, però ara per ara estem lluny de saber si aquesta és una alternativa viable.

Qui podria controlar la IA general?

El 28 de setembre de 2016 Amazon, Facebook, Google, DeepMind, Microsoft i IBM (als quals després s'hi sumaria Apple) funden una iniciativa privada per dissenyar bones pràctiques per a la IA.

Els Estats estan reaccionant més lentament que el sector privat, però tot i així països com la Xina, el Regne Unit, la Unió Europea o els Estats Units d'Amèrica han començat a fer moviments per regular la IA. No obstant això, cap d'aquestes països ha contemplat seriosament els reptes que comporta la recerca o l'assoliment d'una IA general [19]. De fet, el Comitè per la IA de Regne Unit, en el seu informe de 16 d'abril de 2017, manifesta

obertament que és un tema que no es preveu que afecti a la ciutadania a curt termini. De la mateixa manera, les iniciatives de la Unió Europea i els Estats Units s'han limitat a ignorar-la.

El fet és que els governs han vingut obviant el potencial sorgiment de la IA general perquè no es percep que aquesta pugui tenir lloc aviat. A més, ara per ara, tampoc és una preocupació pel que fa a l'opinió pública. El principal problema és que difícilment es podrà anticipar el seu sorgiment amb el temps que realment es necessita per a donar les respostes adequades. Quan els científics constatin unànimement la imminència o el sorgiment d'una IA general i aconsegueixin convèncer als polítics de tot plegat, potser ja serà massa tard. Prenguem de referència la crisi del canvi climàtic: la ciència d'estudi s'inicia al segle XIX i fins la dècada dels 80 del segle XX no hi ha un veritable consens entre els experts. Ben entrat el segle XXI l'amenaça és per a tots palpable, però els Estats segueixen sense posar-se d'acord sobre qui ha de pagar els plats bruts, i mentrestant segueixen embrutant la nostra atmosfera. Ben mirat, potser amb un risc existencial és suficient per extingir-nos.

Actualment, les principals mesures de control de l'IA depenen de l'autoregulació de les pròpies empreses encapçalades per un grup de poderoses multinacionals com Google, Facebook, Apple o Amazon. Per tant, si assolissin una superintel·ligència, incrementarien encara d'una manera més dràstica el seu poder sobre la resta del sector, i sobre el món sencer. Tenint això present, podria pensar-se que una manera de pal·liar aquests mals és estimular la competència amb l'objectiu de repartir aquest poder entre més mans. Si el problema és la desigualtat, assegurem-nos que hi ha més d'una superintel·ligència. Aquesta possibilitat, però, comportaria encara més riscos. Si la IA general millora a un ritme exponencial, l'empresa que l'assoleixi abans (encara que sigui amb només uns mesos d'antelació) mantindrà una avantatge que s'incrementarà en el temps, situant-la en una posició de permanent privilegi. Sent així, que dues o més empreses estiguessin molt a prop de recrear una IA general podria desencadenar una tempesta perfecta. Alimentades per la possibilitat d'assolir un èxit de tal magnitud, els contendents podrien fer apostes més i més arriscades i flexibilitzar progressivament les mesures de seguretat. Com ja hem vist, petits errors de programació en una superintel·ligència podrien comportar danys catastròfics per a tota la humanitat.

En definitiva, tenint tot això present, els Estats haurien de buscar sinergies per convertir la cerca d'una IA general en un projecte de cooperació internacional (igual que ho són el CERN i l'Estació Espacial Internacional o ho va ser el Projecte Genoma Humà). Potser aconseguiríem d'aquesta manera evitar que cap individu, empresa o país pugui monopolitzar el domini sobre una hipotètica IA general. Alhora estaríem una mica més a prop de garantir que la superintel·ligència, si mai arriba existir, beneficia igualitàriament a tota la humanitat (i a la resta d'éssers sensibles) i està al servei dels principis ètics àmpliament compartits, tal com proposa Nick Bostrom [20].

És possible que en el nostre temps vital haguem de fer front a algunes de les decisions més importants que ha pres mai la humanitat. Tot i així, ara per ara actuem com a trapezistes que fan equilibris entre l'extinció i l'eternitat amb els ulls embenats. Potser s'està plantant

ja la llavor d'una ment diferent a totes les que hem conegut fins ara. Una ment que pot comportar el paradís, l'humanicidi, o una nova manera de ser nosaltres mateixos, però pel moment preferim mirar cap a una altra banda. Cal un anàlisi profund, a tots els nivells, que contempli les múltiples possibilitats i conseqüències, perquè fins i tot un resultat relativament improbable esdevé de vital importància si condiciona tot el nostre futur com espècie (o com a forma de vida). Sabem tant poc sobre l'univers i sobre la pròpia humanitat que és impossible que aquesta mena de decisions no ens produeixin vertigen. Però tenim les regnes del futur a nostre mans i, ara per ara, som els únics responsables de llegar als nostres fills el millor futur de tots els possibles. Si anem errats i les superintel·ligències no arriben a existir mai, al cap i a la fi, només haurem perdut una mica de temps, i què és una mica de temps comparat amb tota l'eternitat?

REFERÈNCIES

1 — Paine, C. (2018) Do you trust in this computer? Papercut Films.
2 — Kurzweil, R. (2006) The Singularity is Near: When Humans Transcend Biology. Penguin.
3 — Aquest és, per exemple, la línia de treball de l'OpenCog.
4 — Searle, J. R. (2007) Biological naturalism. In Max Velmans & Susan Schneider. Blackwell.
5 — Dreyfus, H. (1972) What Computers Can't Do, New York: MIT Press.
6 — Kurzweil, R. (2013) How to Create a Mind: The Secret of Human Thought Revealed. Penguin.
7 — Fan, X., Markram, H. (2019) A Brief History of Simulation Neuroscience, Frontiers in Neuroinformatics, vol. 13, 7.
8 — Grace, K., Salvatier, J., Dafoe, A., Baobao Z.. (2018) When Will AI Exceed Human Performance? Evidence from AI Experts. Journal of Artificial Intelligence Research 62 729-754.
9 — Müller V.C., Bostrom N. (2016) Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In: Müller V. (eds) Fundamental Issues of Artificial Intelligence.
10 — Els càlculs exposats per Ray Kurzweil a "The Singularity is Near: When Humans Transcend Biology" pressuposaven un increment anual de transistors per microprocessador que actualment està sent força inferior al previst. Aquest alentiment de la Llei de Moore, en principi, hauria de comportar un endarreriment en les seves prediccions però, tot i així, l'any 2017 l'autor mantenia considerablement els seus pronòstics inicials, preveient l'assoliment d'ordinadors de consum tant potents com el cervell humà l'any 2029 i refermant-se en la imminència de la singularitat tecnològica. Es pot consultar l'opinió de l'autor en la seva intervenció amb motiu al SXSW Festival de 2017.
11 — Transhumanist FAQ. Humanity+.
12 — Winkler, R. (2017) Elon Musk Launches Neuralink to Connect Brains With Computers. Wall Street Journal.
13 — Les interrelacions entre transhumanisme i Silicon Valley poden apreciar-se, per exemple, en: Vance. A. (2010) Merely Human? That's So Yesterday.

14 — Sainato, M. (2015) Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence.
15 — Wagner, K. (2017) Mark Zuckerberg thinks AI fearmongering is bad. Elon Musk thinks Zuckerberg doesn't know what he's talking about.
16 — Müller V.C., Bostrom N. (2016) Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In: Müller V. (eds) Fundamental Issues of Artificial Intelligence.
17 — Bostrom, N. (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Inc.
18 — Yudkowsky, E. (2004) Coherent Extrapolated Volition. The Singularity Institute.
19 — McLay, R. (s. d.) Managing the rise of Artificial Intelligence . Recuperat de https://tech.humanrights.gov.au/sites/default/files/inline-files/100%20-%20Ron%20McLay.pdf (Consultat per darrera vegada el 16 de novembre de 2019).
20 — Bostrom, N. (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Inc.



Miquel Casas

Miquel Casas és llicenciat en Dret per la Universitat Pompeu Fabra i té un màster en Filosofia amb especialitat en Lògica, Historia i Filosofia de la Ciència per la UNED. Ha treballat com a tècnic superior en diversos departaments de la Generalitat de Catalunya i a l'Institut Català de l'Acolliment i l'Adopció, i actualment és assessor jurídic a i l'Institut Català Internacional per la Pau (ICIP). Ha publicat el seu treball de fi de màster a Ápeiron Ediciones: *El fin del Homo sapiens: la naturaleza y el transhumanismo* (2017).