

La distracció d'una intel·ligència artificial sobrehumana

quaderndelesidees.press/la-distraccio-duna-intelligencia-artificial-sobrehumana/

May 26, 2019



Imatge de Patxi Ocio i Casamartina.

Bones preguntes fan bona recerca. Aprendre comença amb fer-se preguntes, i una pregunta de recerca, la manera en què la formulem, fixa la nostra atenció, destaca uns aspectes i n'amaga uns altres.

Una de les expectatives sobre la qual es basa el moviment transhumanista és que en un futur no gaire llunyà aconseguirem construir màquines amb una intel·ligència general comparable, o fins i tot superior, a la dels éssers humans, cosa que ens permetrà transcendir la nostra humanitat. És una expectativa que es fonamenta, per una banda, en la suposició que la intel·ligència es basa en el processament d'informació en sistemes físics i, per altra banda, en el fet que la potència computacional dels ordinadors (mesurada en el nombre d'instruccions per segon) només ha fet que créixer de manera exponencial.

Assolir una intel·ligència de tipus general com la humana és de fet l'objectiu últim del programa de recerca en intel·ligència artificial (IA), el qual intenta donar resposta a les preguntes: «És possible construir màquines intel·ligents?» «Podrien aquestes màquines superar els éssers humans en intel·ligència?» Des d'aquesta àrea de recerca, a cavall de la ciència i l'enginyeria, sovint es presenta aquest objectiu com un dels més ambiciosos i transcendents que mai s'ha proposat la humanitat⁽¹⁾. El que no està clar, però, és si aquestes preguntes són realment bones per orientar la nostra recerca científica. Si partim de la hipòtesi 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 (TIC) haurien de jugar per fomentar el bé comú en les societats humanes d'avui dia.

Faig aquesta reflexió perquè aquestes preguntes que es plantegen des de la investigació en IA ens fan veure la intel·ligència de forma dimensional i mesurable, i a més ens fan suposar que la intel·ligència és una característica objectiva, atribuïble a una entitat, ja sigui un organisme vivent o una màquina computacional. Aquesta visió de la intel·ligència queda palesa en el tipus 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 actuïn de forma racional segons el model matemàtic de la teoria de l'elecció racional?» «Podrien aquestes màquines superar els éssers humans en la racionalitat de les decisions i accions que s'ajustin a aquest model?»

Reemplaçar *intel·ligència* per *acció racional* i entendre aquest darrer terme en el marc del model matemàtic en qüestió, de fet, ha estat molt útil per donar rigor a la investigació en IA i per fer-la avançar en termes comparables i mesurables. Això ha permès que les tècniques que s'han anat desenvolupant i refinant en aquesta àrea s'hagin pogut aplicar de forma reeixida en tota mena d'àmbits de la nostra activitat humana, sobretot en aquells en què es donen les suposicions de la teoria racional de l'elecció: robòtica industrial, logística, control d'electrodomèstics, inversió en borsa, recomanació de productes de consum, anàlisi i classificació d'imatges, cerques a Internet, pilotatge automatitzat, consum eficient d'energia, jocs d'estratègia com els escacs i el go, etc. L'abast d'aplicacions de tècniques de la IA és actualment molt gran i confirma l'observació que un dia va fer John McCarthy, un dels fundadors de la disciplina: «Tan bon punt funciona, ja ningú ho anomena IA.»

Aquest enfocament més pragmàtic ha permès caracteritzar de forma precisa què s'entén com a «millora en intel·ligència», concretament una millora dels valors dels paràmetres dimensionals i mesurables que s'ajusten al model matemàtic de racionalitat. Ara bé, en aquells àmbits de la vida en què la teoria de l'elecció racional no és un bon model per guiar les nostres accions, no hauríem de deixar la presa de decisions en mans d'una IA, perquè una IA no és pas intel·ligent, només és racional en el sentit econòmic del terme. 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; 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(2).

L'experiència de la policia de Chicago de guiar la seva presa de decisions per algorismes d'IA és tan sols una mostra del gran perill a què ens enfrontem actualment i que queda reflectit en visions com les que es promouen des del moviment transhumanista: el de prendre el model matemàtic de racionalitat de forma literal i prescriptiva de què s'entén

que és actuar amb intel·ligència, i adaptar les activitats humanes a aquest model. Això soscava de ple la llibertat creativa de la intel·ligència humana i la sotmet totalment a la lògica de l'acumulació de beneficis i de la maximització de la utilitat. Sovint l'anomenada «superintel·ligència» es planteja en termes de màquines que reemplaçaran els éssers humans en la majoria de llocs de treball, i hi ha qui fins i tot s'atreveix a pronosticar que això pot passar abans de trenta anys. Aquest plantejament, però, ja és revelador de la visió que des del moviment transhumanista es té de la intel·ligència: una intel·ligència merament funcional, orientada a l'execució de tasques i a la resolució de problemes; una intel·ligència avaluada en termes de rendiment econòmic en el mercat laboral; i una intel·ligència aplicable a llocs de treball racionalitzats segons el model matemàtic de la teoria de l'elecció racional(3).



Imatge de Patxi Ocio i Casamartina.

Hi ha un consens creixent que la IA jugarà un paper molt important en el futur de l'economia i del funcionament de les societats humanes, i els diversos governs ja han elaborat plans estratègics específics preveient aquest impacte. Però la gran majoria d'esforços que actualment s'estan fent per garantir un desenvolupament de la IA que sigui beneficiós per a la humanitat s'emmarquen principalment en una visió funcional de la intel·ligència i en una concepció dels valors humans ajustada als paràmetres del model matemàtic de racionalitat. S'estudia com «alinejar» les funcions d'utilitat emprades pels sistemes algorísmics amb una caracterització formal dels objectius i els valors humans(4). I d'aquesta manera ens distraiem intentant resoldre els problemes que ens pot portar una IA sense qüestionar el paradigma de fons que amenaça les nostres societats.

Només cal veure quins són actualment els actors principals de la recerca en IA. No són pas les universitats o els centres de recerca d'utilitat pública i sense ànim de lucre, sinó 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(5).

En els darrers anys hi ha hagut un interès sense precedents per la IA per part d'aquestes grans empreses tecnològiques, arran dels èxits que hi ha hagut en resoldre problemes molt complexos que fins ara s'havien resistit a ser abordats mitjançant programes d'ordinador, com ara el reconeixement de cares i de la parla, la traducció i interpretació de llengües, o el processament i classificació d'imatges, entre altres. Aquests èxits s'han assolit 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 d'avui dia i l'enorme quantitat de dades disponibles per poder entrenar aquests algorismes perquè aprenguin a resoldre una tasca concreta.

Veient aquests progressos, és molt probable que l'impacte més immediat de la IA en la nostra condició humana no sigui pas en forma d'una IA que superi la intel·ligència d'un ésser humà, sinó a través del que ens oferiran sofisticades aplicacions informàtiques, especialitzades en dominis molt concrets, executant-se en tota mena de dispositius de tot tipus i en pràcticament totes les àrees de l'activitat humana. La visió de l'anomenada *Internet de les coses*, amb què es pretén posar sensors a tot i interconnectar-ho tot —ja siguin electrodomèstics, llibres, la roba que portem, els vehicles amb què ens desplaçem, etc.— farà possible que, amb les enormes quantitats de dades que generarem en la nostra vida quotidiana, puguem afinar encara més aquestes capacitats de processament d'informació amb què haurem dotat totes les nostres coses.

S'ha emprat el terme *transició digital* per anomenar aquest camí que la humanitat està fent vers la hiperconnectivitat tecnològica, aquest procés social que sorgeix del desplegament i l'ús generalitzat de les TIC i que ja està transformant significativament la nostra quotidianitat(6). Evidentment, podem albirar moltíssimes oportunitats i beneficis en aquesta transició: en l'àmbit de la salut, en el transport, en l'eficiència energètica, en l'ajuda mútua entre membres d'una comunitat, etc. Ara bé, també hi ha riscos en les tecnologies que desenvolupem i que ens impulsen a fer aquesta transició. Si prenem consciència de qui són aquells a qui regalem cada dia totes aquestes dades que generem en el nostre dia a dia, i de qui té la capacitat de càlcul per processar-les, podrem copsar el risc que pot comportar el desenvolupament d'una IA basada en la racionalitat econòmica, la qual pot acabar sent de profit només per a una minoria de la població humana i continuar enfortint les distribucions actuals de poder, amenaçant d'aquesta manera les nostres democràcies.

Llavors, ¿com hem d'orientar la recerca en IA? Quin ha de ser el rol de les TIC en les nostres societats humanes perquè siguin un instrument per viure la nostra humanitat amb plenitud, fins i tot abans de plantejar-nos una hipotètica transhumanitat? Per respondre aquestes preguntes evidentment cal fer una recerca en profunditat sobre el sistema de valors que ha de dirigir la nostra indagació científica i el nostre

desenvolupament tecnològic. Així, doncs, amb les següents reflexions el que pretenc és contribuir a l'eixamplament de la nostra mirada sobre l'activitat tecnocientífica per tal de poder abordar aquesta recerca axiològica tan necessària.

La indagació científica està fonamentada en la informació, en la nostra capacitat d'*in-formar*, de donar forma a la realitat. Tanmateix, aquesta capacitat d'*in-formació* també ens permet adonar-nos que aquesta realitat plena de formes està al mateix temps buida de formes. És una realitat lliure. I no es tracta que una realitat sigui «més real» que l'altra —la lliure més que la *in-formada*, o viceversa. No són dues realitats, però tampoc és una de sola. Perquè no hi ha un fragment de la realitat que siguem capaços de conèixer i un altre fragment que no. En tota la realitat hi ha al mateix temps un saber i un no-saber. Nosaltres 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.

La ciència cognitiva contemporània ens mostra que les estructures de l'experiència viscuda són essencialment cognitives, i a més 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 el nostre entorn. És més, sovint en el nostre raonament hi participen també de forma 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 al seu torn determinen l'estructura cerebral(7).

El fet de saber apreciar que tota la ciència i el coneixement que hem anat construint són essencialment de naturalesa metafòrica basada en el nostre sentir corporal és el que ens permet saber que la realitat és lliure del nostre sentir i del 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 el cultiu d'aquesta capacitat de transcendir la dualitat de les formes si no vol quedar estancada en una determinada manera d'entendre i percebre la realitat i així contribuir als canvis de paradigma de què parlava Thomas Kuhn. Podríem anomenar aquest cultiu *indagació contemplativa* o *indagació alliberadora*. És la indagació que es fa a través de les pràctiques contemplatives com aquelles que trobem en totes les grans tradicions de saviesa de la humanitat, pràctiques com la meditació, que desfan les formes, que ens deslliuren de quedar-hi atrapats i d'intentar fonamentar la nostra felicitat i plenitud de vida sobre unes formes o unes altres. La indagació científica sempre serà informativa, impregnada per les formes, però no serà indagació si no és cultivada per la indagació alliberadora.

Què ens diu això sobre la recerca en IA? Doncs que no hauríem de quedar atrapats en els nostres models metafòrics de què entenem com a intel·ligència. Com he dit abans, no resoldrem els riscos d'aquesta tecnologia si els debats ètics queden circumscrits en

un nivell conceptual i intel·lectual dins del marc dels models basats en la teoria de l'elecció racional, com em sembla a mi que passa sovint ara, quan s'intenta incloure una conceptualització dels valors i de què significa ser beneficiós per als ésser humans dins dels models matemàtics de racionalitat econòmica.

La metàfora mecanicista de la nostra ment i la hipòtesi que la intel·ligència es basa 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 en el desenvolupament futur d'aquesta tecnologia, potser ha arribat el moment de cercar noves metàfores que s'ajustin millor al sentir i al viure de les societats actuals. La IA —i la ciència de la computació en general—, de fet, es fonamenten en una filosofia i una manera de veure el món que són bàsicament anteriors al 1930, basades principalment en explicacions mecanicistes i una separació rigorosa entre ment i matèria, entre cognició i acció. En canvi, corrents del pensament com els que fonamenten les nocions de *cognició corporitzada* i *enacció*, i que estan fortament influenciades pels moviments filosòfics de la fenomenologia i el pragmatisme, poden ajudar-nos a reorientar la recerca en IA perquè sapiguem posar l'èmfasi en la millora de la interacció de la nostra intel·ligència humana amb els sistemes computacionals de processament d'informació(8).

En aquest sentit potser també hauríem d'alliberar-nos 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. Hi ha corrents de la psicologia i la ciència cognitiva contemporànies que posen seriosament en qüestió la visió objectivista de la intel·ligència i plantegen el cervell no pas com un mecanisme creador de la intel·ligència, sinó més aviat com 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 dels organismes amb el seu entorn i entre ells. La nostra corporeïtat participa a donar i crear sentit a la nostra experiència, i evidentment el cervell hi juga un rol important, però no és el creador del sentit(9).

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 «corporitzades» 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 el seu entorn. Tanmateix, una «IA corporitzada» en un sistema robòtic no deixa de ser un processador d'informació, 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 basat en 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.

És evident que la tecnologia que desenvolupem sota el nom d'IA necessàriament transforma la nostra realitat informativa. Crea noves realitats en tant que modifica la interacció entre nosaltres i el nostre entorn, i per tant també modifica aquesta intel·ligència ecològica que se'ns manifesta en la intersubjectivitat de les nostres interaccions. És en aquestes noves realitats, dinàmiques i en canvi continu, amb les noves formes de viure i de sentir que hi sorgeixen, on hem de saber viure amb plenitud. Per tant, expressar l'objectiu últim de la IA com el de dotar màquines d'intel·ligència, d'una intel·ligència general com la humana, per a mi manca de sentit. En conseqüència, tampoc veig sentit en la visió que el moviment transhumanista té sobre la direcció que hem de prendre com a humanitat. No vull afirmar amb això que no es pugui assolir allò que planteja el transhumanisme; simplement no crec que aquest moviment es basi en bones preguntes de recerca capaces d'orientar la nostra indagació. En canvi, sí que té sentit per a mi, i molt, indagar sobre com orientar la nostra recerca en les TIC en la direcció que fomenti la llibertat creativa dels ésser humans, de tots els altres éssers, del nostre entorn, de tota la realitat en si mateixa, una direcció que ens ajudi a assolir la nostra plenitud humana com a éssers que vivim immersos en la intersubjectivitat i dualitat de la realitat informativa que creem, i també en la llibertat creativa de la realitat *ab-soluta* i no dual.

Marco Schorlemmer

Marco Schorlemmer

Científic Titular a l'Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC). Doctor en Informàtica per la Universitat Politècnica de Catalunya, ha realitzat una investigació a la IIIA-CSIC i a l'UPC a nivell nacional. A nivell internacional ha realitzat estades d'investigació a EUA i al Regne Unit. La seva investigació es centra en [...]

[Llegir més](#)

