

LLENGUA I PENSAMENT CRÍTIC EN LA INTERACCIÓ AMB INTEL·LIGÈNCIES ARTIFICIALS

EMMA GRAU I CABRÉ

Universitat de Barcelona

La difusió d'eines que operen gràcies a la intel·ligència artificial planteja dubtes i decalatges de poder que en gran mesura es deuen al desconeixement que l'usuari mitjà té de la tecnologia que utilitza, i a l'ús acrític que el desconeixement pot alimentar. Exigir un saber expert a tot usuari per garantir-ne un ús reflexionat és fútil davant la velocitat de la seva propagació i la complexitat del seu funcionament, ja que a l'usuari li basten les ordres en llengua comuna per fer-ne ús. Aproximant-nos a la relació que l'usuari actual estableix amb la intel·ligència artificial, i sense pretendre respondre a tots els interrogants sobre l'intercanvi de sabers amb la màquina o sobre la llengua usada, l'esbós d'algunes implicacions filosòfiques permet aventurar com podria fomentar-se una interacció més crítica amb les eines virtuals, no ja externament, sinó des de si mateixes: aprofitant justament que s'hi interactua en llengua natural.

Esbós de l'ús quotidià actual de les IA

L'usuari actual d'una intel·ligència artificial no té per què ser coneixedor de l'àmbit: així com no cal ser informàtic per jugar a videojocs, no cal saber treballar en cap

llenguatge de programació per fer preguntes o peticions a qualsevol dels assistents virtuals que ja són part de la quotidianitat. L'ús d'aplicacions i eines *AI-powered*, que operen gràcies a aquesta tecnologia, també ha esdevingut un avantatge indispensable per l'èxit de projectes i negocis diversos, i s'ha estès igualment pel món de l'art. Com a exemple de la velocitat amb què es desenvolupen les eines *AI-powered*, entre la primera redacció d'aquest escrit i la seva entrega s'ha vist com una obra d'art digital modificada per una intel·ligència artificial guanyava un concurs d'art, fet que ha desencadenat un debat encès i alhora un *boom* en la popularització de l'*AI art*.¹

La clau de l'expansió contínua del seu ús entre el públic no expert és la imitació del llenguatge humà: com que són eines programades per mimetitzar-lo i aprendre a imitar-lo cada cop millor, l'usuari hi interactua en la seva parla de sempre, que la tecnologia processa i amb què és capaç d'emetre una resposta comprensible. Ara bé, la comprensió que l'usuari té del llenguatge de la intel·ligència artificial no passa d'aquesta interacció en la superfície: el llenguatge natural retornat no deixa de ser un producte d'operacions complexes i estructurades segons codis artificials, i l'usuari profà, a part de no comprendre'ls o no saber com accedir-hi, pot no saber ni que existeixen. La situació torna a assemblar-se al jugador que mou un personatge en una consola però no té idea, ni pot fer-se-la sense un llarg aprenentatge, de tot el que cal perquè els píxels canviïn de llum coherentment i simulin el moviment amb què la imatge respon a una tecla. No és forçat dir, per tant, que la comprensió majori-

1. Kevin ROSSE, «An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy», *The New York Times* [en línia], 2022, [consulta: 8 de gener de 2023]. Disponible a: <<https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>>.

tària d'una eina *AI-powered* és superficial i coincideix amb la barrera de la llengua comuna o de tecles bàsiques a què l'usuari mitjà ja està acostumat.

La resposta que aquesta mínima comprensió suscita en l'usuari, conscientment o no, pot ser bé de màxima desconfiança, bé d'un ús absolutament acrític. Per això un dels reptes de la intel·ligència artificial, o un dels nostres reptes davant seu, és la fomentació d'un ús crític que es beneficiï de les possibilitats que amb aquesta s'obren, sense tancar-nos-hi d'entrada ni quedar immediatament enlluernats pels avenços tecnològics.

Però si aquest ús conscient ha de dependre d'una educació suplementària a la difusió de la intel·ligència artificial, ni una assignatura a cada escola ni la inclusió de manuals de programació o d'advertències ètiques en la seva distribució arribaran a grans fites com a intents de democratitzar els seus secrets: els iniciats seguiran sent els pocs amb temps per dedicar-s'hi. Si la intel·ligència artificial ha de deixar de ser, en la mesura del possible, el cúmulo de misteris que és per a la majoria dels seus usuaris, la pretensió no s'adequarà al ritme de la seva expansió mentre la mateixa intel·ligència artificial no contribueixi a deixar de ser misteriosa.

Al cap i a la fi, ja respon a preguntes sobre les cançons més exitoses de l'any o la data de naixement d'un jugador, o els preus del lloguer d'un barri, o la definició de nocions de física quàntica. No podria oferir informació útil, rigorosa però no necessàriament especialitzada, sobre l'ús de la intel·ligència artificial? I anant més enllà, no es podrien fomentar, ja no l'acumulació de sabers, sinó l'aprenentatge i el pensament crítics a través de l'ús d'eines *AI-powered*?

Aquest és, doncs, el tema: pot la intel·ligència artificial respondre a preguntes sobre si mateixa? I pot col·laborar a plantejar preguntes filosòfiques, crítiques, sense que l'usuari hagi de ser ni programador ni filòsof? En últim terme seria com preguntar: pot plantejar aquestes qüestions i les

seves possibles respostes en la llengua comuna de l'usuari i atenen a la rapidesa de les interaccions?

Intercanvis amb les IA

Possible o no, en primer lloc pot plantejar-se què és exactament la llengua que permet la interacció. La imitació del llenguatge natural és un objectiu abstracte que es concreta en idiomes específics. Els predominants són l'anglès i el xinès, i després van les llengües europees i asiàtiques dels mercats amb més demanda. Si no és en les llengües dominants políticament i econòmicament, s'esdevé més difícil parlar amb les màquines; per exemple, l'ús del català en intel·ligència artificial molt recentment s'ha començat a concretar en iniciatives viables. En la mesura que els termes prenen matisos diferents entre llengües, aquesta limitació pot ser ja una barrera per als usuaris la llengua nativa dels quals no sigui cap de les més parlades, més enllà de la marginalització que l'absència d'una llengua en aquest àmbit suposa per a la llengua mateixa, i les conseqüències negatives per a la seva preservació a la llarga.

Alhora, la parla natural no és menys complexa que l'artificial. En quin registre lingüístic convé que la màquina respongui? Si se li pregunta com funciona la intel·ligència artificial, serà necessari que esmenti termes com *algorisme*, però no tothom en sabrà la definició. L'acumulació de termes especialitzats pot ser descoratjadora, però massa simplificació no aportaria la informació demanada; potser caldria certa regulació de la freqüència amb què es fa ús de vocabulari específic d'una disciplina, i facilitar-ne una definició assequible. Igualment, res impedeix que una eina virtual remeti a termes o cites de filòsofs, com ja fan molts videojocs orientats a un públic divers; però això, si bé pot ser suggeridor i servir a propòsits pedagògics, no assegura per si sol la transmissió de cap coneixement o d'una pers-

pectiva crítica. Pot ser, això sí, un suport útil a la divulgació i la formació, i part d'un entramat més complet de recursos filosòfics disponibles.

A més, la llengua no es limita a enunciat; pot ser convenient que, si la interacció amb la màquina simula una conversa, també des de la màquina se'ns puguin plantejar preguntes que, per exemple, despertin curiositat pel seu funcionament i portin a indagar sobre com ens hi relacionem. Sorgeixen aquí molts dubtes: convé que en certes situacions preguntem pels nostres criteris ètics en plantejar un projecte, o posi en dubte els nostres prejudicis racials o la creença en el que la pròpia màquina ens està dient, o faci preguntes personals? Tot i el potencial d'aquestes possibilitats a fi de reconceptualitzar problemes i avivar una reflexió crítica i més oberta, els riscos van des d'una crisi existencial inesperada fins a una extracció no desitjada de les nostres dades.

Cal no oblidar que no hi ha en joc només el saber que l'usuari té de la màquina o que la màquina ofereix de tantes coses: la intel·ligència artificial també «sap» coses de l'usuari, pot emmagatzemar i distribuir informació extreta amb el seu consentiment o sense. En aquest sentit, la intel·ligència artificial té, de fet, més coneixement de nosaltres que nosaltres d'ella. Greu no és un mer decalatge d'informació, sinó les asimetries de saber i poder implicades. Una intel·ligència artificial pot fer-nos sentir impotents o totpoderosos segons com ens relacionem amb els sabers que a través seu aconseguim, i així atrapar-nos en una percepció caòtica del que fem i sabem o induir-nos a respostes automàtiques.² I en últim terme, les nostres dades esdevenen objecte de saber en mans de poders que ni sabem on són, i que van des dels proveï-

2. FRANCO "BIFO" BERARDI, «Prophecy and computation», *Verso* [en línia], 2019, [consulta: 8 de setembre de 2022]. Disponible a: <<https://www.versobooks.com/blogs/4274-prophecy-and-computation>>.

dors del servei fins als altres clients d'aquests, incloent-hi multinacionals i partits polítics. Els interessos entrecreuat³ són part dels motius pels quals urgeix un ús conscient de les noves tecnologies, i per part seu més transparència.

Però pot la màquina revelar a què accedeix, qui hi accedeix i com es vincula la informació, i fer-ho comprensible per a nosaltres? Reprenquem la controvèrsia viva de l'*AI art*. L'usuari no té per què plantejar-se si l'entrenament de l'eina capaç d'associar estil amb paraules i oferir un resultat adequat s'ha dut a terme amb imatges lliures de drets d'autor o bé amb un contingut que, si bé penjat al web, s'ha aprofitat sense autorització; fet que s'agreuja si la imatge resultant es premia o monetitza. En casos com aquest i anàlegs, pot l'eina informar sobre el seu entrenament si l'usuari desitja assabentar-se'n, o detectar l'ús de fonts no autoritzades en la generació d'una imatge i advertir-ne? Pot facilitar a l'usuari cert tipus d'agència o control respecte de la immensitat del material, o de la gestió de les dades a implicades, pròpies i d'altres? Pot plantejar aquests temes a l'usuari que no se'ls ha plantejat? Hauria de fer-ho?

L'abast de la llengua en les IA

Retornant arran d'això a la formulació de preguntes per part de la màquina, poden ara tenir-se en compte la resta d'actes de parla paradigmàtics: usos del llenguatge que clarament no es limiten a una afirmació.⁴ Quan una intel-

3. Per a una anàlisi de caire foucaultiana sobre l'ús del *big data* i l'exploració de les dades com a valor generat sense retribució: JAVIER CIGÜELA, «Big Data, saber-poder y pastoreo digital: sobre el fundamento mitológico de la autoridad», *Foro Interno. Anuario de Teoría Política*, 17. Madrid: Universidad Complutense, 2017, p. 35-54.

4. Es parteix aquí de la noció i el desenvolupament bàsics pel que fa als actes de parla de J. L. AUSTIN, *How to do things with words*. London: Oxford University Press, 1962.

ligència artificial diu que és conscient, utilitza una definició de consciència que encaixa amb les seves capacitats o està fent una broma? Quines bromes acceptaríem d'una intel·ligència artificial? I què provoca que en pugui fer: una programació directa, o una font del seu entrenament? En molts casos, bromes a part, ni tan sols facilitar l'accés a aquestes fonts assegura el rigor d'una resposta; i en cas de poder programar cert humor, certa ironia, no serien sempre evidents per a l'usuari. Malgrat tot, segueixen sent aspectes de la llengua natural, i si la intel·ligència artificial ha de mimetitzar-la cada cop més acuradament val la pena recordar-se'n. En els possibles vincles d'intel·ligència artificial i pensament crític, bromes i ironia exigeixen justament una distància crítica amb la literalitat de les paraules; però en un context on no s'esperen, fàcilment poden prendre's massa literalment.

Lligar intel·ligència artificial amb actes de parla pot semblar especulació: al cap i a la fi, la informació que en rebem és fruit d'algoritmes, de caràcter en últim terme mecànic, sense una consciència a qui, en proferir una frase, pugui atribuir-se intencionalitat. Per molt que en diguem «intel·ligència», i que sigui capaç d'imitar el llenguatge humà, no hi ha en l'eina programada ni emocions pròpies ni una referència conscient a res del que està dient; o seria millor dir que no està «dient» res, per bé que combina paraules. Amb tot, prendre ja les proferències lingüístiques de les intel·ligències artificials com a actes de parla, siguin frases en llengua natural o percentatges matemàtics, pot ser una perspectiva fructífera atenent a tres raons bàsiques que en últim terme es condensen en la implicació de consciències en la interacció.

En primer lloc, en la fabricació i el proveïment de l'eina han intervingut com a mínim programadors, dissenyadors de *software*, empresaris i lingüistes, i cada cop més filòsofs o antropòlegs, entre altres agents amb els interessos corres-

ponents. En segon lloc, la intel·ligència s'ha entrenat amb material que, per ingent que sigui en quantitat, és en general fruit d'agència humana i conscient, i per tant està travessat d'emocions, intencions i ideologies; segons el seu disseny i l'abast del seu entrenament, l'eina pot oferir un marc de pensament més limitat o menys, però que amb les dades accessibles s'ha fornint i es continua fornint d'actes de parla concrets. I en tercer lloc, les respostes de la màquina són rebudes per un usuari que, tot i que no adjudiqui consciència a l'eina i no concebi intencionalitat rere les seves paraules, a la pràctica pot fer servir el material de la interacció com el de qualsevol conversa amb un amic o empleat que tinguis la informació desitjada a l'abast. Per tot això, malgrat el desajust conceptual que complica una aplicació legítima de la noció d'*acte de parla* a les preferències d'una intel·ligència artificial, l'anàlisi al nivell de la praxi quotidiana, en què la distinció esdevé més boirosa, pot beneficiar-se d'enfocar la conversa amb una intel·ligència artificial com una interacció teixida d'actes de parla.

De fet, si s'accepta en termes amplis que fer una afirmació no és només constatar una cosa susceptible de ser veritat o mentida, sinó influir en el receptor i condicionar-ne tant el coneixement com les accions que se'n derivin, amb més motiu urgeix considerar aquesta influència en el cas de les eines *AI-powered*. Quan la intel·ligència artificial fa prediccions sobre l'èxit d'un post o la fiabilitat d'una persona per a una feina, no es limita a constatar: indueix a decidir i actuar de certa manera, i així repercuteix amb efectes reals tant en l'usuari com en la resta d'implicats. Aquestes prediccions no estan lluny de les promeses com a actes de parla, en el sentit que afirmen una veritat, però en releguen el compliment a una realització futura del que es diu. Així com una promesa pot no complir-se, una predicció pot errar: algú poc recomanat per ser contractat pot merèixer i rebre la feina. Però el grau d'autoritat que s'atribueix a la

intel·ligència artificial pot influir la decisió fins al punt que les seves prediccions actuïn com a profecies que s'autocompleixen, i alimentar així un determinisme artificial amb conseqüències negatives ben palpables.⁵

Retorna aquí una qüestió central: com seria millor que la intel·ligència artificial advertís sobre aquests riscos, els seus propis límits i el que sap i el que no? No es tracta de negar tota validesa als resultats, fet que seria igualment fal·laç i alhora faria absurd l'ús d'eines com aquestes, sinó d'incentivar una decisió reflexionada a partir dels resultats que la intel·ligència artificial facilita, i ser conscient dels àmbits en què mereix més confiança o més precaució. No és impossible trobar eines *AI-powered* que expliciten el seu propi marge d'error, però aquesta transparència no és d'abast majoritari, perquè generalment depèn d'una interpretació avançada: no s'integra al nivell de la interacció en llengua natural.

Tampoc es tracta de relegar tot el pensament crític al que la intel·ligència artificial pugui provocar, evidentment: no es distingiria així d'una reacció mecànica per part nostra. La interacció amb una eina virtual i programada no pot substituir la reflexió i l'aprenentatge que, independentment d'ella, pot contribuir a un ús més ponderat d'aquesta i a plantejar-se els problemes i pressupòsits inherents al seu ús. Però això no treu que la interacció amb la intel·ligència artificial pot, a més d'exigir una actitud crítica i informada, ajudar a suscitar-la i oferir recursos per aprendre i per pensar-hi millor.

Salvant les distàncies, o no, no deixa de ser una situació anàloga a la de l'escriptura i la lectura davant d'una classe

5. Per a un tractament més complet d'aquests problemes: Carissa VÉLIZ, «If AI Is Predicting Your Future, Are You Still Free?», *WIRED* [en línia], 2021, [consulta: 8 de setembre de 2022]. Disponible a: <<https://www.wired.com/story/algorithmic-prophecies-undermine-free-will/>>.

presencial o d'una conversa filosòfica de viva veu. És a dir, la consideració que Plató mostrava de l'escriptura: mai un substitut ideal, però, així i tot, un gran vehicle per transmetre i provocar la reflexió filosòfica.⁶ La interacció amb la intel·ligència artificial és una oportunitat per a la reflexió, no només un objecte d'aquesta. Jugant amb la comparació, se la pot portar a l'extrem: si es tracta d'aconseguir una interacció crítica i en llengua natural entre l'usuari i l'eina, i la programació de l'eina pot contribuir-hi, podria arribar a establir-se un diàleg sobre la intel·ligència artificial, al pur estil platònic, amb una intel·ligència artificialment filosòfica?⁷

6. PLATÓ, *Fedre*, 274c-277a i *Carta VII*, 340d-345b.

7. La resposta d'una intel·ligència artificial va ser negativa, adduint la mateixa manca de pensament crític i abstracte. Però quant trigarà a imitar-lo prou exitosament per generar una altra resposta?