Què és i què pot fer la intel·ligència artificial?

Josep Domingo-Ferrer
Universitat Rovira i Virgili
Institut d'Estudis Catalans
josep.domingo@urv.cat

Introducció

Segons Boden (2018), "la intel·ligència artificial consisteix a fer fer als ordinadors la mena de coses de les quals són capaces les ments". Aquesta definició té dos problemes. En primer lloc, no entra en la natura de la intel·ligència. En segon lloc, sembla que cobreix coses no necessàriament relacionades amb la intel·ligència. La qüestió rau doncs a saber de quina mena d'intel·ligència tracta la intel·ligència artificial (IA). Segons Bonnay (2022), la intel·ligència s'associa a reeixir en fer certs tipus de coses, a fer-les d'una certa manera, i a alguna cosa en nosaltres que ens permet de fer-ho així.

Ja al segle XVII hi havia posicions contraposades sobre si la intel·ligència era exclusiva de l'home o no. Descartes (1637) sostenia que un autòmat no pot fer allò que sap fer un ésser dotat d'intel·ligència. En canvi, Hobbes (1651) deia que raonar no és més que calcular, amb la qual cosa obria la porta a un raonament artificial.

Història de la IA

El nom d'*intel·ligència artificial* fou encunyat en una trobada científica tinguda l'estiu del 1956 a Dartmouth College (EUA), en la qual participaren Shannon, Minsky i altres pares de la informàtica. Ara bé, la IA no partia de zero. Entre els seus fonaments cal esmentar la màquina analítica de Charles Babbage i de Lady Lovelace (1834), el model de neurona artificial de McCulloch i Pitts (1943) i també la màquina i el test de Turing (1950).

Ja des d'un primer moment hi ha hagut dues visions diferents de la IA. Els simbolistes representen la informació per símbols i per llurs relacions. Segons ells, la IA consisteix en manipular símbols per resoldre problemes. En canvi, els connexionistes no representen la informació de manera explícita, sinó que cerquen d'imitar els sistemes biològics naturals a partir de xarxes de neurones artificials.

Passat l'entusiasme inicial sobre les possibilitats de la IA, als anys seixanta va haver-hi pocs avenços pràctics (l'anomenat Primer Hivern de la IA). Durant els anys setanta i vuitanta el simbolisme donà resultats tangibles, com ara els llenguatges LISP i Prolog, i també els sistemes experts (per exemple, el cèlebre MYCIN, que reproduïa els raonaments dels metges per fer diagnòstics). Els anys noranta foren una nova època d'estancament (Segon Hivern de la IA). De l'any 2000 ençà, el connexionisme predomina clarament, gràcies a l'entrenament de xarxes neuronals artificials profundes amb dades massives i amb processadors gràfics.

La IA pensa?

Turing (1950) proposà un test per decidir si una màquina es comportava intel·ligentment. Bàsicament si una persona A no pot distingir entre les respostes escrites que li dona un ordinador B i les que li dona una persona C, es pot dir que l'ordinador B és *funcionalment* intel·ligent.

Searle (1980) presentà l'argument de la cambra xinesa per negar que els ordinadors pensin. La idea és que posem dins d'una cambra tancada una persona que no sap xinès. Per un forat a la paret li introdueixen missatges escrits en xinès. La persona disposa d'un llibre de regles que li permeten d'operar els símbols xinesos d'entrada per produir missatges de sortida que tenen sentit en xinès. Els missatges que produeix els treu per un altre forat de la paret. Tot i que funcionalment pot semblar que la persona de dins de la cambra sap xinès, la realitat és que no en sap, sinó que només manipula símbols. La cambra xinesa és una metàfora d'un programa d'ordinador, que també manipula símbols. La conclusió és que *funcionar* com si es fos intel·ligent es pot fer sense entendre res. Per tant, la definició funcional d'intel·ligència que donà Turing té problemes.

Què saben fer les màquines?

Les màquines, enteses com a ordinadors equipats amb intel·ligència artificial, poden fer diverses tasques.

- Predicció. En aquesta categoria entra la classificació (reconeixement de caràcters, concessió de crèdits, etc.) i també el diagnòstic mèdic (tant el que fan els sistemes experts de tipus MYCIN com el diagnòstic automàtic per la imatge).
- Traducció i transcripció. Aquestes capacitats ja són a l'abast del gran públic de fa anys (Google Translate, transcripció Word i altres).

- Jocs d'estratègia. El 1997 l'ordinador Deep Blue va guanyar als escacs el campió mundial Kaspàrov. El 2011 l'ordinador Watson va guanyar els campions del joc Jeopardy. El 2017 l'ordinador AlphaZero va aprendre tot sol fins a esdevenir campió d'escacs, de *shogi* i de *go*. Ara bé, aquests ordinadors tenen limitacions en llur raonament. Per exemple, Watson diria que un cocodril no pot saltar obstacles perquè no ha vist a Internet que ho faci, mentre que una persona raonaria que no pot perquè té les potes curtes i és feixuc (Levesque, 2014).
- Creació de continguts. Aquí podem distingir la creació de continguts artístics feta amb models generatius (*generative artificial networks,* GAN) i la generació de discurs feta pels grans models de llenguatge (ChatGPT i altres). Hi ha un debat filosòfic sobre si les màquines poden ser creatives o no. Kant (1790) deia que "les belles arts només són possibles com a producte del geni [...] No són un talent que es pugui aprendre amb regles." Kraus (1986), en canvi, concep la creativitat com a derivació i recomposició, amb la qual cosa seria automatitzable.

Els grans models de llenguatge

BERT (Google AI, novembre del 2018) és un model de llenguatge preentrenat que dona representacions numèriques per a frases i paraules, que es poden refinar per a tasques específiques. Les representacions numèriques (*embeddings*) són vectors i exploten el fet que les paraules que tenen significats semblants ocorren en contextos semblants. D'aquesta manera, representen el significat com a coocurrència: com més coocurrència de dues paraules o frases, menys distància entre els vectors que les representen. Aquest procediment té el seu fonament en el treball dels lingüistes. Ja fa més de seixanta anys, Firth (1957) va afirmar que "coneixeràs una paraula per qui li fa companyia".

ChatGPT o Chat Generative Pre-trained Transformer (Open AI, 2022) és un robot conversacional basat en un gran model de llenguatge. Ha estat entrenat amb tot el contingut d'internet, en particular la Viquipèdia. Pot explicar conceptes amb paraules senzilles, donar idees per fer regals d'aniversari o fer programes d'ordinador. En una conversa, recorda el que l'usuari li ha dit anteriorment i adopta les correccions que aquest li fa. Ara bé, de vegades pot donar informació incorrecta, perjudicial o esbiaixada. Investigadors de Microsoft sostenen que ChatGPT dona "espurnes" d'intel·ligència general artificial (Bubeck *et al*. 2023). En efecte, mostren que ChatGPT pot fer una gran diversitat de tasques de manera competent: generació

d'imatges, programació informàtica, habilitats matemàtiques, interacció amb el món, interacció amb humans, etc.

Conclusions

La IA té una història d'alts i baixos. L'actual abundància de dades i els processadors gràfics han permès d'entrenar xarxes neurones profundes. Gràcies a això, som en un moment molt dolç de la disciplina. Els robots conversacionals ha fet visible la IA al gran públic, però ja fa temps que es fa servir per a moltes coses (traduir, seleccionar continguts, etc.) Malgrat tot, no és clar que siguem a prop de la intel·ligència general artificial, que s'assemblaria a la intel·ligència dels humans.

Referències

BODEN, Margaret A. (2018). *Artificial Intelligence. A Very Short Introduction*. Oxford: Oxford University Press.

BONNAY, Denis (2022). "Que peuvent les machines ? Petite philosophie de l'IA par exemple". Curs de filosofia del llenguatge o de la lògica. Université de Paris-Nanterre.

BUBECK, Sébastien et al. (2023). *Sparks of artificial general intelligence: early experiments with GPT-4*. <https://arxiv.org/pdf/2303.12712.pdf>

DESCARTES, René ([1637] 2000) *Discours de la méthode*. París: Flammarion.

FIRTH, John R. (1957). "A synopsis of linguistic theory, 1930-1955", a *Studies in Linguistic Analysis*, p. 10-32.

HOBBES, Thomas. ([1651] 2017). *Leviathan.* Londres: Penguin Classics.

KANT, Immanuel ([1790] 2015). *Kritik der Urteilskraft*. Colònia: Anaconda Verlag.

KRAUS, Rosalind E. (1986). *The Originality of the Avant-Garde and Other Modernist Myths*. Cambridge, Massachusetts: The MIT Press.

LEVESQUE, Hector J. (2014). "On our best behavior", a *Artificial Intelligence*, *212*, p. 27-35. <https://doi.org/10.1016/j.artint.2014.03.007>

McCULLOCH, Warren S.; PITTS, Walter (1943). "A logical calculus of the ideas immanent in nervous activity", a *The Bulletin of Mathematical Biophysics*, *5*, p. 115-133. <https://doi.org/10.1007/BF02478259>

SEARLE, John (1980). "Minds, brains, and programs", a *Behavioral and Brain Sciences*, *3*(3), p. 417-457. <https://doi.org/10.1017/S0140525X00005756>

TURING, Alan (1950). “Computing machinery and intelligence”, a *Mind*, *59*(236), p. 433-460. <https://doi.org/10.1093/mind/LIX.236.433>