

---

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

---

**JOSEP DOMINGO-FERRER**

Institut d'Estudis Catalans  
Universitat Rovira i Virgili  
*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 a 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 terme *intel·ligència artificial* fou encunyat en una trobada científica tinguda l'estiu de 1956 a Dartmouth College (Estats Units), 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 a manipular símbols per resoldre problemes. En canvi, els connexionistes no representen la informació de manera explícita, sinó que cerquen 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 desmentir que els ordinadors pensen. 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 operar amb 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 es pot *funcionar* com si es fos intel·ligent 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 d'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 de manera autònoma 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 d'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 de 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 que es basen en 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. Els investigadors de Microsoft sostenen que ChatGPT dona «espurnes» d'intel·ligència general artificial (Bubeck *et al.*, 2023). En efecte, ChatGPT pot fer una gran diversitat de tasques de manera competent: generació d'imatges, programació informàtica, ús d'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 entrenar xarxes neuronals profundes. Gràcies a això, som en un moment molt dolç de la disciplina. Els robots conversacionals han fet visible la IA al gran públic, però ja fa temps que la IA es fa servir per a moltes altres 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 l'exemple». A: *Philosophie du langage ou de la logique* [en línia]. Descripció del curs de la Universitat de París-Nanterre. <[https://formations.parisnanterre.fr/plugin/odf-web/odf/\\_content/course-philosophie-du-langage-ou-de-la-logique-fr/Philosophie%20du%20langage%20ou%20de%20la%20logique%20.pdf](https://formations.parisnanterre.fr/plugin/odf-web/odf/_content/course-philosophie-du-langage-ou-de-la-logique-fr/Philosophie%20du%20langage%20ou%20de%20la%20logique%20.pdf)> [Consulta: 1 octubre 2024].
- BUBECK, Sébastien [et al.] (2023). *Sparks of artificial general intelligence: early experiments with GPT-4* [en línia]. <<https://arxiv.org/pdf/2303.12712.pdf>> [Consulta: 22 novembre 2024].
- DESCARTES, René (2000 [1637]). *Discours de la méthode*. París: Flammarion.
- FIRTH, John R. (1957). «A synopsis of linguistic theory, 1930-1955». A: *Studies in linguistic analysis*. Oxford: Blackwell, p. 10-32.
- HOBBS, Thomas (2017 [1651]). *Leviathan*. Londres: Penguin Classics.
- KANT, Immanuel (2015 [1790]). *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». *Artificial Intelligence* [en línia], 212, p. 27-35. <<https://doi.org/10.1016/j.artint.2014.03.007>> [Consulta: 22 novembre 2024].
- MCCULLOCH, Warren S.; PITTS, Walter (1943). «A logical calculus of the ideas immanent in nervous activity». *The Bulletin of Mathematical Biophysics* [en línia], 5, p. 115-133. <<https://doi.org/10.1007/BF02478259>> [Consulta: 22 novembre 2024].
- SEARLE, John (1980). «Minds, brains, and programs». *Behavioral and Brain Sciences* [en línia], 3 (3), p. 417-457. <<https://doi.org/10.1017/S0140525X00005756>> [Consulta: 22 novembre 2024].
- TURING, Alan (1950). «Computing machinery and intelligence». *Mind* [en línia], 59 (236), p. 433-460. <<https://doi.org/10.1093/mind/LIX.236.433>> [Consulta: 22 novembre 2024].

---

# Els reptes de la intel·ligència artificial des d'una perspectiva crítica i sociotècnica

---

**ROSER PUJADAS**

University College London  
r.pujadas@ucl.ac.uk

## 1. De la cosificació de la intel·ligència artificial als reptes de les tecnologies anomenades *intel·ligència artificial*

El terme *intel·ligència artificial* (IA) denota una ambició que alguns situen ja a la Grècia clàssica: la de desenvolupar una tecnologia capaç d'imitar o superar les capacitats cognitives humanes. Es tracta d'un enfocament qüestionable, donat que les intel·ligències humanes són múltiples i socials, però que ha esperonat tot un seguit d'innovacions al llarg dels anys. Tanmateix, la consciència col·lectiva sobre les tecnologies anomenades IA sembla que s'ha despertat arran del llançament amb accés obert del bot de conversa (*chatbot*) ChatGPT, d'OpenAI, a finals de 2022, que va anar seguit d'expressions alarmistes sobre els perills de la IA. En concret, experts en IA i líders del sector tecnològic —inclòs el mateix Elon Musk, un dels fundadors d'OpenAI— demanaven una moratòria per tal de preparar-se per afrontar els riscos que generarà la IA. Irònicament, els mateixos responsables del desenvolupament de la IA parlaven de riscos existencials, en un futur, això sí, en què la IA s'assumeix com a inevitable.

Clarament, la IA presenta oportunitats i reptes, però tal com suggereixen els estudis crítics de la IA (Crawford, 2021; Suchman, 2023) per entendre'ls val la pena reflexionar sobre la funció que fan certes concepcions i discursos sobre la IA, qui els produeix i amb quins efectes, i quines qüestions queden relegades. Per començar, el concepte mateix d'IA sembla donar coherència i estabilitat a un fenomen que, de fet, és força difús i distribuït. No és una tecnologia monolítica sinó que requereix diverses tecnologies i pràctiques socials, i es manifesta de manera diversa. Hi ha tecnologies anomenades IA que fa temps que s'usen en molts àmbits i, de manera més o menys visible, hi interactuem quotidianament —per exemple, en la detecció de malalties, la previsió del temps, la identificació de persones, la selecció de personal, en la producció i també en la detecció de notícies falses, o en la personalització i classificació de continguts i ofertes en

portals d'internet. Aquestes tecnologies són diverses, com ho són els seus àmbits d'aplicació, així que qualsevol generalització sobre els efectes, bondats o riscos de la IA amaga més que no pas ajuda a discernir, i tendeix a naturalitzar models dominants d'IA.

Aquesta reïficació tendeix a reduir la IA, els seus potencials i riscos, a qüestions estrictament tecnològiques, i a imbuir la IA d'agència pròpia (Suchman, 2023). A més, tant les visions utòpiques de la IA com les distopies de futur plantejades pels líders del sector tecnològic —predominantment, homes, blancs, adinerats i situats a Silicon Valley— poden distreure'ns dels problemes reals que la IA ja causa. Periodistes, activistes i acadèmics han aportat evidències empíriques, entre d'altres, de la implicació de la IA en l'erosió de la privacitat, el biaix i discriminació racial i de gènere, la perpetuació de desigualtats, la desinformació i la producció de notícies i imatges falses, la manipulació ideològica i de comportament, la concentració de poder o els efectes nocius per al medi ambient (Broussard, 2023; Crawford, 2021; Eubanks, 2018). Aquests problemes no són només estrictament tecnològics, sinó estructurals i ambientals i afecten de manera desproporcionada certs col·lectius, infrarepresentats en el sector tecnològic.

## 2. L'eticoontopistemologia de la IA: quan tecnologia i coneixement configuren realitats i valors

Els estudis de ciència i tecnologia (coneguts com a STS, de l'anglès *science, technology and society*) palesen que la tecnologia per si mateixa no té agència ni té una força intrínseca inevitable. Hi ha moltes versions possibles de les tecnologies i les implementacions que se'n fan. El disseny i l'ús s'han d'emmarcar en contextos socials específics. El desenvolupament tecnològic és el resultat de negociacions i relacions de poder, però també de condicions materials específiques i requereix tasques sovint invisibilitzades. Alhora, la tecnologia és un actor social i polític important, que incorpora i ajuda a reproduir certs valors i influeix de manera important en les pràctiques i estructures socials. Així que és important que com a investigadors i com a societat n'escrudem el desenvolupament i les aplicacions.

El cas de la IA és particularment punyent, donada la capacitat performativa de la IA, és a dir, de (re)configurar la realitat, és significativa. D'una banda, tot i que la IA es pot fer servir per assessorar humans en tasques complexes, sovint es relaciona amb l'automatització de decisions. Particularment en aquest cas es produeix una delegació d'agència cap a una tecnologia cada cop més difícil d'escrutar. D'altra banda, malgrat que la IA es presenta sovint com una tècnica objectiva, cada versió de la IA incorpora una visió específica de la intel·ligència i de com generar prediccions. I és que, tal com suggereixen els estudis de STS i el feminisme, tot coneixement és situat i parcial. A més, les diverses pràctiques de coneixement formen assemblatges socio-tècnics (Latour, 2005) i són constitutives de certes configuracions de la realitat i sistemes de valors; per tant, tenen conseqüències ètiques i polítiques. Així, cal preguntar-se: quines són les eticoontopistemologies (Barad, 2007) de la IA? És a dir, quines són les epistemologies i

pràctiques de coneixement associades a la IA i quines realitats i valors ajuden a configurar? I amb quines conseqüències ètiques i materials?

### 3. El model d'IA basat en dades massives i en l'extractivisme

Actualment, el terme *IA* sol referir-se a diverses tècniques d'aprenentatge automàtic o *machine learning* (ML). De manera simplificada, el ML es desenvolupa usant un conjunt de dades d'entrenament per produir algoritmes i models matemàtics que representin algun aspecte de la realitat. L'ús del ML permet aplicar aquest model per analitzar de manera automatitzada noves dades i fer prediccions. La idea que una quantitat més gran de dades d'entrenament permet el desenvolupament de models cada cop més acurats, sumada a l'ambició de produir IA d'aplicació general, com ChatGPT, ha acabat provocant l'ús indiscriminat i massiu de dades. Això és problemàtic per diversos motius.

Malgrat les pretensions d'objectivitat i aplicació universal d'alguns models, és una fal·làcia que més quantitat de dades garanteixi una millor representació de la realitat. Qualsevol model d'IA és parcial, té limitacions i contextos específics d'aplicació que poden ser útils. Per exemple, el model de llenguatge extens de ChatGPT prediu estadísticament quina és la combinació de paraules més adient per respondre a una pregunta. Pot ser útil per millorar l'estil d'un text, però a l'hora de respondre una pregunta produeix text versemblant tot i que no necessàriament verídica.

A més, té implicacions ètiques i de justícia social. D'una banda, certes pràctiques d'extracció o compra de dades massives, per exemple del web o les plataformes, violen drets bàsics com la privacitat, o els drets d'autor, i deriven en l'apropiació de coneixement. D'altra banda, les dades recollides no són sempre representatives, es descontextualitzen i per tant perden qualitat o sentit, i alhora contenen els biaixos socials presents en els contextos d'extracció. Aquests biaixos i incorreccions es reproduïxen en els models i prediccions de la IA. Si la IA s'usa en la presa de decisions, pot donar com a resultat decisions injustes i l'amplificació de desigualtats. Hi ha molts exemples d'aplicacions de tecnologies d'IA que reproduïxen biaixos, amb implicacions perjudicials per a col·lectius vulnerables (Eubanks, 2018). Això es veu exacerbat en models de ML que requereixen l'etiquetatge de les dades d'entrenament (Crawford, 2021), com ara en la identificació d'imatges o vídeos, ja que suposen la imposició de categories no neutres i la classificació subjectiva per part d'una mà d'obra sovint precaritzada, i de vegades exposada a material pertorbador, com ara escenes de violència (Gray i Suri, 2019).

Aquest model d'IA no només reproduïx estructures de classe, sinó que també amplifica estructures de discriminació i distribució desigual de capital a escala global. La capacitat d'acumular i processar dades massives per generar aquests models de ML requereix molts recursos naturals, tecnològics, econòmics i capital intel·lectual a l'abast de poques empreses, amb tendències

monopolístiques. Aquestes companyies estan concentrades als Estats Units i a la Xina, on són afavorides per polítiques proteccionistes dirigides a liderar la innovació en IA (Rikap i Lundvall, 2021). Finalment, les implicacions mediambientals són significatives, ja que la gran potència de càlcul necessària per processar dades i generar models consumeix molta energia i cal molta aigua per refrigerar els centres de dades. A més, els components tecnològics contenen minerals que sovint s'extreuen de països del Sud Global.

La IA s'ha convertit en una altra manera de fer política tot i que gens democràtica, particularment perquè, en un model econòmic en què preval la llei del més fort, la inèrcia de la versió extractiva —de dades, treball i recursos naturals— de la IA sembla difícil d'aturar. Caldrien polítiques valentes i un activisme social determinat per afrontar els complexos reptes econòmics, epistèmics, ètics, socials, polítics i mediambientals que planteja la IA i repensar quines versions d'IA volem fomentar.

## Referències

- BARAD, Karen M. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Durham, NC: Duke University Press.
- BROUSSARD, Meredith (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. Cambridge, Massachusetts: The MIT Press.
- CRAWFORD, Kate (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press.
- EUBANKS, Virginia (2018). *Automating inequality*. Nova York: St. Martin's Press.
- GRAY, Mary L.; SURI, Siddharth (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Boston: Houghton Mifflin Harcourt.
- LATOUR, Bruno (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford; Nova York: Oxford University Press.
- RIKAP, Cecilia; LUNDVALL, Bengt-Åke (2021). *The digital innovation race: Conceptualizing the emerging new world order*. Cham, Suïssa: Palgrave Macmillan.
- SUCHMAN, Lucy (2023). «The uncontroversial “thingness” of AI». *Big Data & Society*, 10 (2), p. 1-5.



---

# Si és artificial, no és periodisme

---

**JOSEP LLUÍS MICÓ**

Universitat Ramon Llull

[joseplluisms@blanquerna.url.edu](mailto:joseplluisms@blanquerna.url.edu)

La crisi dels mitjans de comunicació ha coincidit en el temps amb l'ascens del màrqueting de continguts, és a dir, de la comercialització de productes i serveis a través de la distribució de materials útils per a uns públics que s'intenten atraure amb aquests mecanismes. Les xarxes socials digitals —Instagram, YouTube, Snapchat, TikTok i un llarg etcètera— s'han erigit en el punt de trobada d'empreses periodístiques i altres companyies que assagen aquestes fórmules. Per aquesta raó, les marques s'estimen més que els professionals de la informació s'encarreguin d'aquestes tasques.

Encara que s'hagin educat com a periodistes, ningú no pot dur a terme aquesta missió promocional amb tanta solvència com ells. Firmes especialitzades en màrqueting de continguts i activitats anàlogues demostren minut a minut una habilitat enorme per resoldre dues tensions del mercat: com fer front a aquest tipus de producció i com ocupar una multitud de periodistes que, si no fos per aquesta solució, estarien condemnats a l'atur.

Semblava que aquesta remodelació laboral era un bon remei per al problema dels redactors que fins fa poc havien treballat en uns mitjans que els havien hagut d'acomiarar o que fins i tot s'havien vist forçats a tancar. En realitat, aquesta transformació amplia la que es va iniciar quan nombrosos reporters van començar a incorporar-se als gabinets de premsa i les agències de comunicació. Aquella transició corporativa, desenvolupada per vocació o simplement per necessitat, no va estar exempta de polèmica. I l'actual, amb la intel·ligència artificial i la resta de sistemes de les anomenades *quarta i cinquena revolucions industrials*, tampoc.

La digitalització ha desencadenat una sofisticació que encara introdueix més interrogants ètics (Shilton, 2018: 107-171). Els mitjans estan incorporant la intel·ligència artificial, l'aprenentatge

profund (*deep learning*) i l'aprenentatge automàtic (*machine learning*) a les rutines dels seus departaments: màrqueting, publicitat, redacció, infografia, fotografia... Gràcies a l'automatització de la recollida, l'anàlisi i el tractament de la informació comercial i periodística, el sector de la comunicació, que encara arrossega les greus conseqüències dels últims sotrats econòmics —del col·lapse de la bombolla immobiliària de 2008 a la pandèmia del coronavirus—, pretén assegurar-se el futur (Guzman, 2019: 83-96).

Molts dels responsables d'aquestes organitzacions han pensat que si els gegants tecnològics —Google, Meta, Apple, Amazon, Microsoft, IBM...— han apostat per aquesta tendència amb convicció i fortes inversions, ells també ho haurien d'intentar (Moran i Shaikh, 2022: 1756-1774). En el marc de la indústria 5.0, la intel·ligència artificial generativa, el funcionament de la qual consisteix a recopilar informació sobre diversos assumptes que després servirà per donar lloc a noves idees, està tan de moda que fins i tot la indústria audiovisual hi confia per assegurar-se el millor futur possible.

Aquest és un dels pronòstics que es poden llegir en un estudi recent elaborat pels experts de la companyia Dentsu, els quals sostenen que aquesta tecnologia serà fonamental a l'hora de revitalitzar les fórmules creatives sobre tots els suports: la televisió, l'ordinador, la tauleta tàctil, el telèfon intel·ligent, la consola i altres dispositius de la internet de les coses.

Els analistes afirmen que la intel·ligència artificial generativa transformarà la manera en què els usuaris accedeixen als continguts, tant a les plataformes com als motors de cerca (Pavlik, 2023: 84-93). A més, els publicitaris, els guionistes, els productors de continguts i els influenciadors disposaran d'un ajut inèdit en la història de la comunicació, capaç de saber quin efecte tindrà el seu material amb un marge d'error mínim. De fet, aquestes figures ja són conscients que la intel·ligència artificial generativa els permetrà fabricar imatges, sons i altres recursos que, sense la seva intervenció, mai no existirien. Alhora, els consultors estan convençuts que la competència entre les corporacions, els mitjans, els canals i els operadors d'estríming s'intensificarà en les pròximes temporades, entre altres factors, per la irrupció d'aquest puntal de la cinquena revolució industrial.

L'explicació del fenomen és ben senzilla: si els competidors fan servir màquines idèntiques, els resultats tendiran a assemblar-se. Potser massa... Per això, es multiplicaran les maniobres per individualitzar les propostes comercials, de Hollywood a Barcelona, de Bollywood a Estocolm. En aquest context, els anunciants estan buscant formats inèdits amb l'objectiu d'aconseguir captar l'atenció d'un públic sotmès a un volum d'estímuls creixent (O'Neil, 2017).

Sigui com sigui, les veus —humanes— més autoritzades alerten que l'audiència es cansarà aviat d'habitar en un ecosistema en què allò que és promocional serà gairebé omnipresent. Llavors, què podria preservar les marques de la manca d'interès dels espectadors, els oients, els lectors i els navegants? Doncs els especialistes ho tenen clar: l'autenticitat i el compromís amb els valors socials i ecològics, elements que, per cert, deixaran de ser aliens a la intel·ligència artificial generativa algun dia.

No en va, aquesta tecnologia acapara l'interès de les empreses i les institucions, monopolitza els titulars i fins i tot suscita recels entre el gran públic. Els acadèmics més crítics la conceben com la principal amenaça per continuar entenent l'economia, la societat i la cultura com han estat compreses fins ara. En qualsevol cas, la valoració general d'aquesta innovació és altament positiva (Stray, 2019: 1076-1097). És així fins i tot si es corre el risc que l'entusiasme per l'acollida d'aquesta tecnologia reveli una actitud acrítica i esnob.

Ara per ara, no abunden els estudis rigorosos que permetin passar de la intuïció i la cridòria de les tertúlies o els col·loquis a l'anàlisi freda i fonamentada de la ciència. Una de les poques excepcions en aquest aspecte es refereix a la utilització de la intel·ligència artificial per part dels productors audiovisuals. I la conclusió dels responsables de l'enquesta de Yahoo Entertainment i YouGov de la qual s'extreuen aquestes idees és tan contundent com desconcertant.

Aquí s'evidencia que al públic de la principal potència mundial, els Estats Units, no li ve de gust veure pel·lícules o programes de televisió generats mitjançant la intel·ligència artificial, en especial si als films hi apareixen personatges que no són de carn i ossos. En síntesi, l'audiència i els professionals estan d'acord en allò essencial. A l'hora de preparar un reportatge, entrevistar els principals actors de l'actualitat, construir una crònica sobre el terreny, dissenyar una anàlisi especialitzada o improvisar en directe a partir d'un succés inesperat, no hi ha discussió possible: allà han de ser-hi els informadors. I allà mateix seguiran, perquè si és artificial, no és periodisme.

## Referències

- GUZMAN, Andrea L. (2019). «Beyond extraordinary: Theorizing artificial intelligence and the self in daily life». A: PAPACHARISSI, Zizi (ed.). *A networked self: Human augmentics, artificial intelligence, sentience*. Nova York: Routledge, p. 83-96.
- MORAN, Rachel E.; SHAIKH, Sonia J. (2022). «Robots in the news and newsrooms: Unpacking meta-journalistic discourse on the use of artificial intelligence in journalism». *Digital Journalism* [en línia], 10 (10), p. 1756-1774. <<https://doi.org/10.1080/21670811.2022.2085129>> [Consulta: 28 maig 2022].
- O'NEIL, Cathy (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Nova York: Broadway Books.
- PAVLIK, John V. (2023). «Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education». *Journalism & Mass Communication Educator* [en línia], 78 (1), p. 84-93. <<https://doi.org/10.1177/10776958221149577>> [Consulta: 28 maig 2022].
- SHILTON, Katie (2018). «Values and ethics in human-computer interaction». *Foundations and Trends in Human-Computer Interaction* [en línia], 12, p. 107-171. <<https://doi.org/10.1561/11000000073>> [Consulta: 28 maig 2022].
- STRAY, Jonathan (2019). «Making artificial intelligence work for investigative journalism». *Digital Journalism* [en línia], 7 (8), p. 1076-1097. <<https://doi.org/10.1080/21670811.2019.1630289>> [Consulta: 28 maig 2022].



---

# Dret i intel·ligència artificial generativa. Dilemes sobre autoria, originalitat i atribució de drets de propietat intel·lectual

---

**ANTONI RUBÍ PUIG**

Universitat Pompeu Fabra  
antoni.rubi-puig@upf.edu

Els avenços en intel·ligència artificial (IA) generativa i la popularització als darrers anys de models i aplicacions com ChatGPT, Stable Diffusion o Midjourney han esperonat la recerca acadèmica sobre les seves conseqüències per als drets de propietat intel·lectual, han servit per a endegar iniciatives legislatives en diverses jurisdiccions i també han donat lloc a diversos litigis, la majoria dels quals pendents de resolució.

Els sistemes d'IA generativa permeten, a partir d'una entrada d'informació feta per un usuari —habitualment, un text o una indicació (*prompt*, en anglès)—, generar nous continguts, com ara una imatge, un escrit, un vídeo o el codi font d'un programa d'ordinador. Aquests resultats o generacions són molt semblants als tipus de continguts que són reconeguts amb drets d'autor: presenten les característiques usuales de les obres literàries i artístiques, que són l'objecte primari de protecció del dret de la propietat intel·lectual, com ara un poema, un dibuix, una fotografia o un programari. Per això, els sistemes d'IA generativa susciten diverses qüestions amb relació a aquesta branca del dret. En aquesta contribució, m'ocupo de la qüestió de si els resultats de la IA generativa poder ser protegits mitjançant drets d'autor, però hi ha més temes d'interès pràctic i teòric, com són les possibilitats d'entrenar-ne els models subjacents a partir d'obres protegides o els riscos d'infracció si les generacions són massa similars a obres preexistents (Lee *et al.*, 2024).

En dret europeu, les obres originals són protegibles per drets d'autor. D'acord amb una jurisprudència consolidada del Tribunal de Justícia de la Unió Europea (UE), per a ser original, una obra hauria de constituir una creació intel·lectual pròpia d'un autor o, en altres termes, hauria de reflectir-ne la personalitat a partir de decisions lliures i creatives (Rubí Puig, 2024). No és gens clar que una generació reflecteixi la personalitat de l'usuari de la IA o la d'un altre subjecte, com

ara un enginyer desenvolupador del model o el compilador de la base de dades que es va fer servir per a entrenar-lo. A més, hom també pot defensar que les generacions no són originals, atès que estarien determinades per consideracions tècniques o les mateixes regles de funcionament del model. En altres termes, no hi hauria expressió personal pròpiament dita, sinó l'obtenció d'un resultat a partir d'un model d'optimització estadística o probabilística.

En aquests moments, no hi ha cap sentència o resolució en un estat membre de la UE que s'hagi pronunciat sobre aquesta qüestió jurídica. Trobem, en canvi, resolucions en altres països influents, en especial als Estats Units d'Amèrica (EUA). Segurament, el cas més rellevant és l'assumpte *Kashtanova*, decidit el 21 de febrer de 2023 per l'Oficina del Registre de Propietat Intel·lectual: la senyora Kristina Kashtanova havia utilitzat Midjourney per a crear diferents imatges, que després va arranjar per fer un còmic titulat *Zarya of the dawn*. Va sol·licitar la inscripció del còmic com a obra protegida al Registre i quan l'Oficina descobrí que les imatges havien estat generades automàticament va denegar-ne la protecció. Més ben dit, limità la salvaguarda només a aquells aspectes contribuïts per la sol·licitant —el text del còmic i l'arranjament de les diferents imatges per a narrar una història—, però no les imatges individuals de cada vinyeta. La resolució recorre a una sèrie d'arguments per a negar la protecció dels continguts generats per IA que il·lustren molt bé els dilemes entorn de llur originalitat, autoria i atribució.

En primer lloc, assenyala que els sistemes jurídics de drets d'autor recullen un model antropocèntric en què només són protegibles creacions fetes per humans (Blaszczyk, 2023). Esmenta, en aquest sentit, *Naruto vs. Slater*, un cas en què un tribunal de San Francisco resolgué que aquest model antropocèntric impedia l'atribució de drets a obres generades per animals no humans. En el cas en qüestió es tractava d'una fotografia feta per un macaco crestat d'una illa d'Indonèsia. Una ONG defensora dels drets dels animals, People for the Ethical Treatment of Animals, va iniciar un plet en nom i representació del mico contra David Slater, el fotògraf que havia propiciat la foto i l'havia comercialitzat després. Sense entrar a valorar si els drets corresponien al demandat, el tribunal rebutjà, però, que el macaco pogués ser-ne considerat autor.

En segon lloc, la resolució apunta que els resultats de la generació no són suficientment previsibles per part de l'usuari de la IA generativa. En altres termes, quan un usuari escriu un *prompt* no pot anticipar l'expressió final que produirà el sistema. En aquest sentit, la resolució també esmenta una altra sentència, la dictada en el cas *Sarony*, un assumpte sobre un retrat d'Oscar Wilde resolt pel Tribunal Suprem dels EUA fa cent quaranta anys. En aquell cas, el tribunal entengué que una obra fotogràfica podia ser protegida per drets d'autor. El procediment tècnic emprat no era rellevant, segons el Tribunal, sinó que l'important era la concepció de l'obra i que aquesta concepció es veiés reflectida en el resultat final obtingut. El tribunal elaborà allí la doctrina denominada de la *master mind*: si un subjecte pot concebre i anticipar tots els aspectes expressius d'una obra, n'ha de ser considerat autor amb independència de l'execució final o del procés tècnic seguit. Si s'aplica aquesta doctrina al cas actual, la resolució nega que l'usuari d'una IA generativa sigui efectivament una *master mind*, ja que no pot preveure els aspectes expressius finals a partir de les seves instruccions.