

**B-DEBATE**

International Center  
for Scientific Debate  
BARCELONA

**biocat**



Sinopsi

# ARTIFICIAL INTELLIGENCE: DREAMS, RISKS, AND REALITY

March, 7<sup>th</sup> and 8<sup>th</sup>, 2017

COSMOCAIXA BARCELONA. ISAAC NEWTON, 26. BARCELONA

[www.bdebate.org](http://www.bdebate.org)

CO-ORGANIZED BY:



CSIC  
upf.



---

# Intel·ligència artificial: somnis, riscos i realitat

---

Hi ha moltes definicions disponibles per descriure què és la intel·ligència artificial, fet que dona mostres del seu caràcter escorredís. Una definició aproximada seria aquesta: un sistema computacional capaç d'emular funcions pròpies del cervell humà. Plasmada per l'imaginari popular en forma de robots, la intel·ligència artificial va en realitat molt més enllà, i es troba ja present en molts àmbits de la nostra vida. Tot i que encara amb limitacions, la tenim darrere de la selecció de les notícies que llegim, condicionant o almenys influent en els resultats electorals; està darrere de molts dels anuncis que veiem, i ens trobem a les portes de conèixer amb algunes de les seves noves aplicacions, com la dels cotxes autònoms, guiats sense necessitat de conductor.

Però tota oportunitat amaga també un risc. Per debatre unes i altres, diversos dels millors experts internacionals es van reunir a [B·Debate](#), una iniciativa de [Biocat](#) i de l'[Obra Social "la Caixa"](#) per promoure el debat científic. Fruit d'aquesta trobada es va establir la Declaració de Barcelona per a l'adequat desenvolupament i utilització de la intel·ligència artificial a Europa.

## CONCLUSIONS

---

- Resulta complicat predir com evolucionarà la nostra relació amb els robots. Aquesta interacció obre tot un camp nou conegut com roboètica
- La intel·ligència artificial aplicada a la gran quantitat de dades disponibles avui en dia està creant càmeres de ressò i filtres de bombolla a les xarxes socials, amenaçant amb radicalitzar ideologies
- Els algoritmes amb què opera la intel·ligència artificial són en moltes ocasions foscos. Cal aconseguir traduir-los al nostre llenguatge perquè les seves decisions puguin ser ètiques i socialment justes
- Davant de les múltiples oportunitats i perills s'ha establert l'anomenada Declaració de Barcelona per a l'adequat desenvolupament i utilització de la intel·ligència artificial a Europa

# 1. LA INTEL·LIGÈNCIA ARTIFICIAL: DELS ROBOTS A LES DADES

---

Sobre els robots

"La intel·ligència artificial podria escombrar la humanitat quan arribi a ser massa intel·ligent, ja que els humans seran com formigues". Il·lustrat amb la fotografia d'un enorme robot al diari [The Independent](#), aquest titular del físic Stephen Hawking va donar la volta alarmant el món. Els experts, però, no li donen massa importància. Per [Héctor Geffner](#), professor ICREA a la Universitat Pompeu Fabra, els robots són encara maldestres i **"la intel·ligència general de les màquines a un nivell humà solament existeix de moment en les pel·lícules"**. Són capaços de reaccionar, però tenen poca flexibilitat i no tenen vida mental".

[Carme Torras](#), investigadora a l'Institut de Robòtica i Informàtica Industrial del CSIC, comparteix que **"avui dia, els humans són més perillosos que els robots"**, però reivindica el seu desenvolupament actual. El seu grup treballa en robots assistencials que ajuden a vestir-se a persones dependents, i reivindica que ja existeixen robots usant-se en logística o com a recepcionistes, robots que discriminen què ha d'anar a les escombraries o a un rentavaixelles, i fins i tot apunta el fet que hi hagi robots-mascotes, dissenyats per mostrar i provocar emocions. Això últim obre tot un camp ètic que ja s'ha anomenat **"roboètica"**, i que per a Torres està motivada en part per "la humanització dels robots, la percepció que *hi ha un humà a dins*". Per Torres la gran pregunta és: **"Com evolucionarà la naturalesa humana amb l'augment de la interacció entre humans i robots? I, per extensió: pot aquesta evolució predir-se?"**

Un dels problemes per estudiar aquesta evolució és la limitació del llenguatge a l'hora de descriure el futur. En paraules de Heidegger: **"És a través de la tecnologia que percebem el mar com a navegable"**. D'aquí que molts dels dilemes es trobin en el terreny de la **ciència-ficció**, un dels papers, segons Torras, és el "d'anticipar possibles escenaris". Escenaris que es troben en llibres com els d'**Asimov, Philip K. Dick o Ray Bradbury**, en pel·lícules com **"Eva"** o en sèries com **"Black Mirror"**. La pròpia Torras és autora del llibre de ciència-ficció **"La mutació sentimental"**, a través del qual està desenvolupant un projecte educatiu.

## Sobre les dades

L'augment en la capacitat de computació dels ordinadors i, sobretot, la instauració d'Internet i l'increment exponencial de les dades disponibles han fet que **la intel·ligència artificial comenci a estar present en molts aspectes de la nostra vida**. Els algoritmes estan darrere de moltes de les notícies que llegim i dels anuncis que ens arriben, però també poden estar darrere de l'anàlisi dels crèdits o les assegurances que demanem, del filtre inicial dels nostres currículums o de l'estudi d'algunes dades de la nostra salut. I no és gens desgavellat pensar que estaran també guiant els nostres propis vehicles, que en un futur passaran segurament a ser completament autònoms.

**Una de les àrees més estudiades és la de les xarxes socials i l'accés a la informació, amb notables implicacions i perills.** Per exemple a l'hora de **condicionar eleccions**, com se suggereix que va succeir en usar la intel·ligència artificial per personalitzar els missatges **a la campanya de Donald Trump**.

Independentment de campanyes puntuals, els algoritmes, segons **Cornelius Puschmann** –investigador al Hans Bredow Institute for Media Research en Hamburg– "són una forma confusa en què a vegades es parla de la intel·ligència artificial", i són els responsables de "triar" –en base al nostre perfil i anteriors preferències– quines notícies tendeixen a aparèixer en els murs de les nostres xarxes socials. Hi ha dues tendències que els algoritmes i les xarxes fan emergir, i cap està exempta de perills. La primera, com va comentar **Camilo Cristancho**, investigador a la Universitat Autònoma de Barcelona, és la formació de "**càmeres d'eco**" en què els individus tendeixen a ajuntar-se amb els seus semblants. L'altra són les "**bombolles de filtre**": bombolles ideològiques com a conseqüència de la individualització de continguts oferts pels algoritmes en base a les preferències passades, i que amenacen de potenciar en excés les idees ja establertes, minimitzant el matís i la pluralitat.

**Hauriem de preocupar-nos per les bombolles de filtre?** Segons Puschmann no excessivament. "Des d'un punt de vista periodístic no és una cosa nova, portem 50 anys estudiant-ho". Per **Walter Quattrociocchi**, cap del Laboratori de Ciència Social Computacional a l'IMT de Lucca, a Itàlia: "encara és objecte de debat si aquestes bombolles són creades per l'algoritme o són una simple tendència de l'ésser humà".

Segons Puschmann, el que sembla més evident és que **el que causa por és la sensació de pèrdua de control**, el fet que alguna cosa no-humana tingui capacitat de decisió. En un experiment amb usuaris de Facebook, **més de la meitat desconeixien que hi havia un algoritme darrere dels seus murs**, i la primera reacció en ser informats va ser de

sorpresa i ràbia. No obstant això, diversos mesos després, i després d'haver-los ensenyat el seu funcionament, la seva satisfacció era semblant a la que tenien abans d'haver-ho "descobert". I el seu ús de Facebook, més gran del que era abans.

Quattrociocchi investiga fonamentalment com la informació errònia es propaga per la xarxa, fet que lliga amb la designada paraula de l'any el 2016, la **post-veritat**, un concepte que el seu significat "denota circumstàncies en què els fets objectius influeixen menys en la formació de l'opinió pública que les crides a l'emoció i a la creença personal". I també amb l'anomenat biaix de confirmació, "l'actitud que porta a buscar la informació que confirmi les pròpies creences". Fins i tot el Fòrum Econòmic Mundial ha assenyalat a la massiva informació errònia digital com "un dels principals riscos per a la nostra societat".

Les investigacions de Quattrociocchi confirmen la clara existència de càmeres de ressò, dins de les quals les dinàmiques de reforç són semblants sigui quin sigui el motiu en comú. Potser el més preocupant és que els usuaris més actius i "compromesos" en aquestes càmeres tendeixen a ser **els que defensen teories conspiradores** (a l'estil dels que proclamen l'existència d'OVNIS o la relació entre les vacunes i l'autisme). I encara més preocupant és la següent paradoxa: la publicació dins d'aquestes càmeres d'informació rigorosa que busca desacreditar les seves teories no només no ho aconsegueix, sinó que és contraproductiu. És a dir, els insta a reforçar la seva posició inicial.

## 2. LÍMITS I ÈTICA EN INTEL·LIGÈNCIA ARTIFICIAL

---

Tot i els èxits i les promeses de la intel·ligència artificial, encara compta amb serioses i importants limitacions. Bàsicament, els programes poden funcionar molt bé en aplicacions concretes, però **estan lluny de fer-ho a l'impredictible món real**. Per a Francisco Martín, president de BigML, les tècniques amb prou feines han canviat en els últims anys. Simplement han augmentat les dades sobre les quals poden operar. Segons Martín, "molta gent està reinventant la roda al mateix temps" sense fer veritables avenços. De fet, els processos "requereixen encara molta experiència humana i es fan de forma molt manual".

Una de les limitacions té a veure amb la foscor del procés que segueixen els programes fins a arribar a les seves conclusions. Molts d'ells es basen en el que es coneix com a **xarxes neuronals**. Aquestes xarxes són construccions matemàtiques que s'assemblen

en certa manera com funciona el nostre cervell, on la informació va saltant de nivell i consolidant-se de forma difusa. Són models molt potents, però necessàriament opacs, i han donat lloc a parlar de "**la caixa negra de la intel·ligència artificial**". Alguns experts, [com Pierre Baldi](#), no li donen importància a aquesta foscor. Al cap i a la fi: "Fas servir el teu cervell tot el temps; confies en el teu cervell tot el temps; i no tens ni idea de com funciona el teu cervell". D'altres, com [Marcello Pelillo](#), professor de Ciència Computacional a la Universitat de Venècia, reconeixen que hi ha situacions "**on és important poder explicar les decisions**". Per exemple, si afecta la decisió d'un jutge o d'una institució social. I, en general, per garantir el respecte a l'autonomia i la dignitat individual.

Perquè l'ús de la intel·ligència artificial implica nombrosos dilemes i reptes ètics. Aquests són alguns dels proposats per [Francesca Rossi](#), investigadora a IBM i professora de Ciències de la Computació a la Universitat de Pàdua:

- ¿Substituirà la intel·ligència artificial el treball dels humans?
- Com canviarà la nostra interacció amb altres humans i la societat o l'educació?
- En un futur, què passarà davant la possibilitat de desenvolupar armes autònomes?
- Com serà el desenvolupament ètic de la pròpia intel·ligència artificial?

Per Rossi, **caldrà desenvolupar els sistemes de manera que discriminin els seus resultats de forma ètica**. Això haurà de ser així, òbviament, en el cas dels sistemes autònoms, però també en d'altres. Per exemple, en eines que aconsellin els metges en base a l'anàlisi de dades: "Si aquests suggeriments no segueixen un codi ètic, els propis metges no confiaran en aquest sistema". La clau és com desenvolupar aquest codi, ja que els processos de decisió del programa, com hem vist, són moltes vegades foscos per la seva pròpia naturalesa.

Això no és senzill. D'una banda per un tema de flexibilitat. El que a nosaltres ens pot semblar una mica de **sentit comú** no té per què ser-ho per a una màquina. "Un robot de cuina no hauria de cuinar el nostre gat si no hi ha res a la nevera, encara que consideri que aquest seria un dinar acceptable per a nosaltres", exemplifica Rossi.

Una solució podria ser programar segons els codis professionals. "Però aquests tenen molts buits que nosaltres omplim amb sentit comú. Una cosa que és difícil d'introduir en una màquina".

Es pot fer, doncs? Segons Rossi, sí. "Hi ha diverses maneres que ara mateix s'estan estudiant". El procés, a més, amaga avantatges. "Fer-ho pot ajudar-nos a prendre consciència i a comportar-nos nosaltres mateixos d'una manera més ètica".



### 3. LA DECLARACIÓ DE BARCELONA

---

Impulsat pels líders del B-Debate [Luc Steels](#) –professor ICREA a l'Institut de Biologia Evolutiva de Barcelona– i Ramon López de Mántaras –director de l'Institut d'Investigació en Intel·ligència Artificial de Barcelona–, durant la trobada es va proposar i es va consensuar la "Declaració de Barcelona per a l'adequat desenvolupament i utilització de la intel·ligència artificial a Europa", **un manifest obert a noves firmes i comentaris** que pretén servir de guia per al progrés de la intel·ligència artificial. El manifest reconeix la gran importància d'aquest camp en el futur de l'economia i les societats, però també la preocupació pel possible ús inadequat, prematur o maliciós de les noves tecnologies. En base a això, proposa un codi de conducta articulat al voltant de sis punts:

- **Prudència:** malgrat el desenvolupament de la intel·ligència artificial i les seves possibilitats, molts problemes fonamentals encara no estan resolts i requeriran avenços radicals. Fins i tot és possible que alguns problemes mai es resolguin.
- **Fiabilitat:** els nous mètodes han de ser verificats abans d'estar disponibles. La nova agència per a la robòtica i la intel·ligència artificial creada pel Parlament Europeu podria assumir aquesta responsabilitat.
- **Transparència:** quan un sistema d'intel·ligència artificial pren una decisió, les persones afectades han de poder rebre una explicació de per què s'ha pres aquesta decisió. No obstant això, moltes de les tècniques que "expliquen" els processos interns del sistema encara no estan ben desenvolupades.
- **Responsabilitat:** hi ha una preocupació creixent pels sistemes de missatgeria automàtica o xat-bots que operen a internet i són capaços de manipular a l'opinió pública. Cal deixar clar als usuaris si els missatges procedeixen d'un sistema d'intel·ligència artificial.
- **Autonomia restringida:** els sistemes d'intel·ligència artificial, quan estan inclosos en sistemes físics com els cotxes autònoms, tenen el potencial d'actuar a través de les seves decisions en el món real. Cal tenir regles clares que limitin la seva autonomia, així com aclarir qui és el responsable en cas de fallada.
- **El paper humà:** l'entusiasme per la intel·ligència artificial dóna lloc en ocasions a la impressió que la intel·ligència humana ja no serà necessària. Això és un error important. Tots els sistemes depenen de la intel·ligència humana, de manera que l'experiència de les persones ha de continuar sent ensenyada, desenvolupada i exercida.

Finalment, donada la importància creixent de la intel·ligència artificial en l'economia, els signants del manifest criden a les agències de finançament europees i les empreses a invertir en el seu desenvolupament de forma adequada i estructural, creant les condicions per estimular l'emprenedoria. A més, **subratllen la importància de desenvolupar plataformes de recursos oberts**, una infraestructura comuna compartida que afavoreixi el desenvolupament de les properes innovacions.