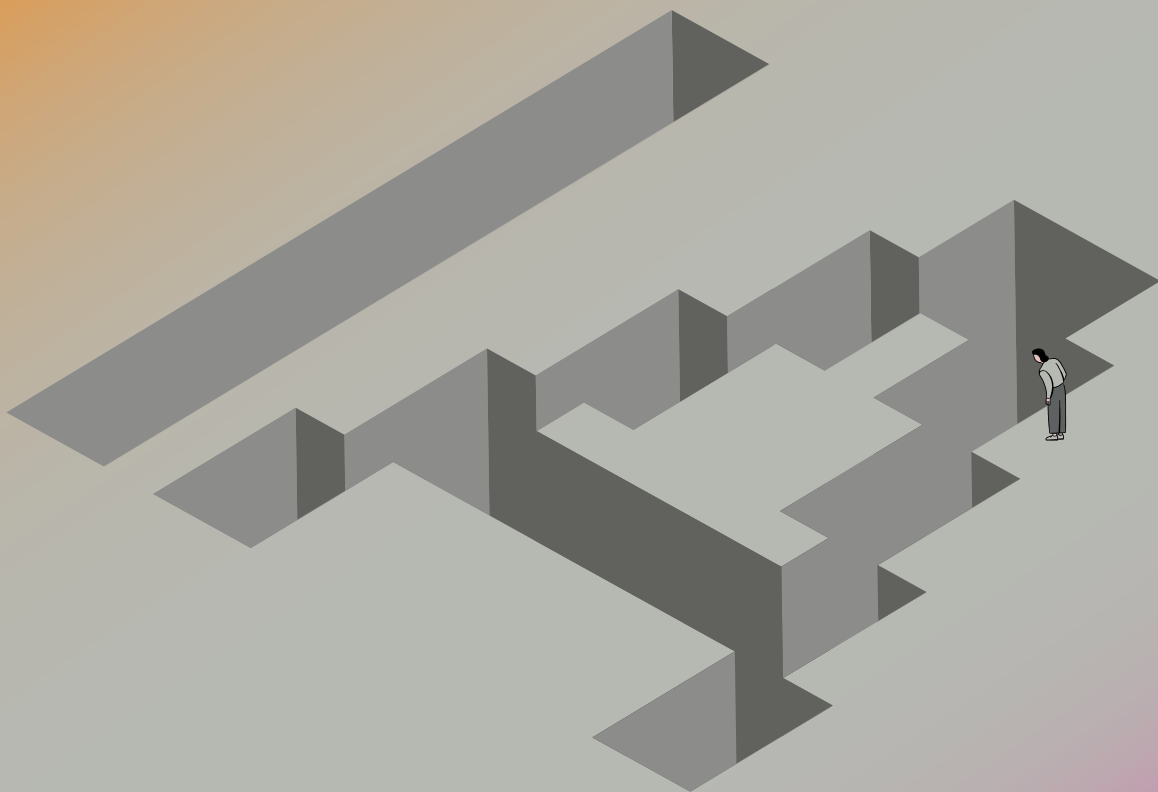


# Qui acciona la màquina?

La IA, el periodisme i (l'esgotament de) les dades  
en l'era de l'*enshittification* d'internet

INFORME EMILI GISBERT







Edita

**Unió de Periodistes Valencians**

Autora

**Sara Calvo Tarancón**

Traducció al valencià

**Vicent Baydal**

Disseny i maquetació

**Sud+Sud**

Impressió

**Toniburguera sl**

Dipòsit legal

**V-3925-2024**



Pròleg	8
Abstract	10
Prefaci	12
12 Un sondeig sobre intel·ligència i artificialitat	
14 <i>L'enshittification</i> del nostre tot: internet mor on la IA naix	
Introducció	18
18 Ni tan intel·ligent ni tan artificial	
20 El misteri de l'explicabilitat: com funciona la IA generativa?	
Ètica i tecnologia	24
24 La consciència (ja no?) és patrimoni de la humanitat	
26 La creativitat fa humana la màquina: antropomorfització i AGI (Artificial General Intelligence)	
Història	30
30 Antecedents, teoria i història: Ada Lovelace, Turing, Altman	
32 Tot el que necessites és atenció: neurones sintètiques, intel·ligència i màquines	
IA i periodisme	36
36 Radiografia del sistema actual: glòria i decadència, èxits i acomiadaments, i la reinvençió constant de les redaccions	
38 La sentència de mort del gran cercador	
40 El cas dels mitjans valencians i l'exemple de les corporacions públiques	
41 El periodisme local en À Punt i les veus sintètiques	
43 L'excepcionalitat del <i>New York Times</i> : la denúncia a les tecnològiques per "utilitzar el contingut sense pagar"	
46 Si hi ha diners, no hi ha plets	
47 Les dades i la fi del petroli	
50 De <i>Gizmodo</i> a <i>La Nación</i> : IA generativa com a problema i també com a solució	
51 Protocols per a un ús de la IA responsable	
53 La difícil convivència entre la IA i els drets d'autor	
54 Qui espolia a qui: la propietat intel·lectual en l'era d'internet	
56 La transparència de l'algoritme: qüestions sobre l'ètica de la IA, els vikings negres i el sobreajustament dels models	
59 Dispensadors de continguts a la carta en època electoral: el paper de les plataformes de difusió en la contenció de la desinformació	

## Legislació i marc regulador 64

- 64 Mecanismes legals de control de la desinformació
- 66 IA per a desemmascarar la IA o el robot que caça robots
- 67 Eines de verificació disponibles i fiables
- 68 Legislar *a priori*: *AI Act*, la primera llei que posa límit a la Intel·ligència Artificial
- 69 Roig, groc i verd: la divisió legal segons el risc de cada producte amb IA
- 71 Privadesa i seguretat en les excepcions de l'*AI Act*: el debat sobre la captació biomètrica als carrers i les fronteres
- 73 El cas d'*iBorder Control*: automatismes en la gestió fronterera

## Transparència, tecnosolucionisme i canvi climàtic 74

- 74 Diners públics, codi públic: el poder institucional i els algorismes automatitzats
- 75 Del tecnosolucionisme a l'increment de la crisi climàtica
- 77 La importància del codi de la IA
- 78 Intel·ligent, artificial i de codi obert
- 81 El bé comú, la privadesa i el capitalisme de plataformes

## Una forma de concloure 84

- 84 Entre l'amenaça colossal i l'oportunitat
- 84 "Som intel·ligents perquè vivim"

## Bibliografia 89

## Glossari d'entrevistes 90

## Glossari de termes 91

# La IA que no podrà substituir al periodisme (de veritat)

La Unió de Periodistes Valencians decidí encarregar este text davant l'auge i la incidència, cada vegada major, de la Intel·ligència Artificial en els processos informatius. Un repte que sabem que quedaria desactualitzat en poc de temps a causa de la velocitat amb què evoluciona, però que pretenia establir una base per a entendre l'actualitat i posseir una visió dels seus inicis en el futur.

Una de les definicions del Periodisme assegura que la seua funció és traure a la llum el que algú no vol que se sàpiga. Per a això cal construir relacions, generar confiança, disposar d'informacions de diverses fonts i actuar quasi en la clandestinitat. Quasi tot (si exceptuem el tercer apunt) únicament atribuïble als éssers humans.

Escoltem en 2023 a Julia Otero en Onda Cero llegir una notícia sobre com la tecnologia destruiria llocs de treball, en un text que haguera passat per actual si en finalitzar-lo no haguera dit la presentadora que pertanyia a l'època de la Revolució Industrial.

Des de llavors, i les xifres són incontestables, el nombre de persones empleades és el major de la història. I amb cada revolució tècnica ha ocorregut exactament la mateixa circumstància, tot i ficar-se en la societat la mateixa classe de por.

Els ERO fets des de 2007 en els mitjans de comunicació no els han fet màquines, sinó persones. Els qui, segurament, d'haver comptat amb més dades o alternatives potser havien pres decisions diferents. O avançat unes altres que hui demostren que un mitjà pot viure sense dependre de la publicitat.



I no només un mitjà. Quan alguns membres de la Unió començaren a estudiar la carrera de Ciències de la Informació, prop de l'any 2000, únicament podien plantejar-se un horitzó laboral en ràdios, televisions i periòdics clàssics. Hui, tot i els acomiadaments, hi ha perfils que viuen de comunicar en les seues xarxes socials, en Twitch, en Youtube, en un pòdcast o a través d'una *newsletter*.

Pot la Intel·ligència Artificial suposar un risc per als periodistes? Sens dubte. En primer lloc, perquè moltes empreses buscaran l'optimització de notícies (que no d'informacions. El matís és fonamental) sobre la base d'esta tecnologia. I, en segon, perquè per a l'exercici que estem realitzant hui en dia, on es premia el copia i apega i el *clickbait*, poca especialització és requerida.

Tanmateix, anem al costat oposat. Pot la IA contribuir a millorar les prestacions d'una redacció menuda? És (ja sabem que sí) capaç de trobar dades i patrons que suposen noticiabilitat i que l'ull humà tardaria dies o fins i tot mesos a descobrir? Amb un ús generalitzat, permetrà accedir a informacions que d'una altra manera romandrien ocultes?

Tornem a la definició de Periodisme, que també engloba saber fer les preguntes correctes a les persones adequades. O, en este cas, a les màquines adequades.

Si estem apostant per un Periodisme de solucions, fem-ho també amb la Intel·ligència Artificial. Formem-nos. Vegem en què pot ajudar-nos. I aprofitem una era on, si som capaços de dominar la tecnologia abans que ella ens domine a nosaltres, és possible que aconseguim pegar (de nou) un salt cap avant en una professió que necessita recuperar la credibilitat. Sobretot perquè prompte serà difícil discernir allò que és veritat del que no ho és.

I ahí, sempre, hi ha d'haver un periodista vigilant.

# Abstract

La irrellevància del periodisme és una amenaça més gran que qualsevol supercomputadora *omnimodal* d'avançades xarxes neuronals sintètiques. De la professió dependrà diferenciar-se de l'*enshittification* o buscar intel·ligents excuses artificials amb què justificar les seues pràctiques poc ètiques i de dubtosa qualitat informativa. Ningú –ni els mitjans– se salva de conviure amb nous models i eines d'intel·ligència artificial que van incorporant-se com a part del dia a dia. Això és una cosa que ha passat històricament amb la tecnologia i, per al neurocientífic Mariano Sigman, la clau és que la novetat no ens reemplace sinó que ens multiplique.

Esta pot ser –o ja ho és– l'enèsima revolució que ho canviarà tot. La nostra recerca és una foto fixa de l'inici de l'any 2024, una narració coral en la qual participen, per descomptat, periodistes, tant de València com d'expertes en desinformació, programadors informàtics, investigadores en el camp de la tecnologia i els drets humans, advocades, juristes, caps de secció, responsables d'informatius, redactores, etc. Al llarg d'estes pàgines intentem sumar opinions que ajuden a redefinir una professió periodística molt precaritzada i a estones injuriada, però que continua sent extremadament necessària. Rigor, credibilitat, confiança, veracitat, imparcialitat són qualitats clàssiques i atemporals del periodisme que poden ser reforçades amb les potencialitats de la IA. Perquè, sobretot, esta és una gran oportunitat per a millorar el periodisme.

El canvi en la forma de comunicar ja s'ha donat en els darrers anys, des de la irrupció d'internet en les nostres vides. La necessitat imperiosa de formar-se i dominar la tecnologia per tal que no siga ella qui ens domine és una màxima d'estos temps. La resignificació d'esta professió és essencial, i cal aprofitar l'impuls de la IA per a dignificar la importància del bon periodisme.

Estes idees convergixen en la present investigació, en què sobreïx una reflexió de fons sobre què és hui dia internet, i en què l'hem convertit. Internet, com diu l'escriptor Cory Doctorow, no és el més important de tot però sí que és el camp de batalla, el terreny on lliurem les lluites, les baralles pel que de veres és important. La web, els fòrums, les xarxes socials estan immerses en un procés lent i constant de decadència palpable en què predomina un embolic de contingut sintètic, artificial i poc intel·ligent que busca el farciment, el *clickbait*, i el control dels patrons de consum. Perquè internet és, sobretot, un lloc de mercadeig capitalista, a més del principal terreny –si no el més important– on ocorre, es desenvolupa, es produïx, es consumix i es difon el periodisme.

Ens semblava inabastable l'antic internet i ara els grans models de llenguatge (LLM, Large Language Models), com ChatGPT, l'han fagocitat i estan ampliant-lo amb contingut multiformat signat per una IA generativa que –a vegades– es fa passar per un periodista. O a l'inrevés: un periodista que obvia explicar que allò que ha escrit no és seu. Un dels principals reptes, no només del periodisme sinó del conjunt de la societat, serà el de preservar l'honestedat i que les velles *fake news* i els nous *deepfakes* no dominen ni el panorama mediàtic ni les nostres vides, opinions, eleccions o creences polítiques.

Perquè el perill de la desinformació és una preocupació real per a una gran part de la població. En una enquesta realitzada per Metroscopia, un 35% de la ciutadania enquestada pensa que la IA millorarà el periodisme, però hi ha un 52% que opina que l'empitjorarà. I més del 80% creu que comportarà un risc per a la privadesa i que afavorirà la proliferació d'enganys i notícies falses.

En estes pàgines analitzem les arestes de la legislació actual de mans de juristes experts, especialment l'*AI Act*, el Reglament europeu sobre Intel·ligència Artificial, en què no els *deepfakes* no són prohibits com a tal, però sí que se suggerix que els continguts sintètics que es puguen confondre amb continguts creats per humans han d'estar degudament etiquetats per tal de no sembrar dubtes. Repassarem també les eines disponibles per a detectar contingut fals, els robots que cacen robots i que deuen formar part, com a ajuda o complement, de la caixa de ferramentes per tal que els i les periodistes no deleguem del tot el treball de continuar sent garants de la veracitat.

Exposarem qüestions que afecten la transparència dels algoritmes i el codi que hi ha darrere de la IA, la privadesa i els drets humans en l'aplicació de tecnologies en qüestions com la migració i la seguretat nacional. Farem un breu recorregut per la història i l'origen de la intel·ligència artificial, una explicació dels conceptes clau per a entendre com funcionen les eines que gastem. Repassarem casos d'ús d'eines d'IA generativa reeixides en redaccions de tot el món, indagarem en les connotacions ètiques i polítiques de l'antropomorfització de les màquines, en què és ser intel·ligent, en la qüestió del *copyright* en els elements periodístics, en els perills del tecnosolucionisme i en les virtuts de l'ús eficient i transparent de les ferramentes automatitzades per a fer un millor periodisme.

# Prefaci

## Un sondeig sobre intel·ligència i artificialitat

Per a la intel·ligència artificial no hi ha res que siga potencialment impossible. És el que diu el neurocientífic Mariano Sigman, autor del llibre *Artificial. La nueva inteligencia y el contorno de lo humano*. Des de la Universitat de Stanford, en l'Informe de l'Índex d'Intel·ligència Artificial 2024<sup>1</sup>, ja es conclou que la IA és millor en classificació d'imatges, raonament visual i comprensió de l'anglès. De moment, però, es queda arrere en tasques més complexes com les matemàtiques de nivell avançat, el raonament visual de sentit comú i la planificació. Al capdavant de les files de científics pessimistes hi ha Kilnam Chon, un informàtic octogenari conegut com el pare d'internet a Corea. Chon fa servir una comparativa entre l'home i el mico i les posicions de domini, i pronostica que d'ací a 30 anys la IA serà més intel·ligent que la humanitat i llavors ens podrà controlar, com controlem nosaltres els micos.

Metàfores amb primats a banda, la veritat és que la història de la IA ha pres un caire fascinant des dels anys 60 a esta part, en este intent de l'ésser humà, ple d'hiverns, per emular la seua pròpia intel·ligència, però també per entendre-la. Ara, en esta primavera daurada que vivim, no passa una setmana sense que hi haja una nova notícia, un nou programa, un nou congrés, una nova xarrada sobre la qüestió, que no és desconeguda per al 93% de la ciutadania.

Segons Andrés Medina, director de Metroscopia<sup>2</sup>, el percentatge de persones preocupades davant la irrupció i l'avanç d'estes tecnologies és del 53% i d'entusiastes d'un 37%. Més de la meitat veu la IA com una oportunitat i un terç opina que és una amenaça. De les persones enquestades, un 35% pensa que millorarà el periodisme; un 52%, que l'empitjorarà. I més del 80% creu que suposarà un risc per a la privadesa i que afavorirà la proliferació d'enganys i de notícies falses. "Els espanyols continuen confiant en els mitjans per tal que defensen l'interés general", assegura Medina. La confiança d'una capçalera no és un regal de la IA, sinó que respon a altres factors com el treball dels seus periodistes, la veracitat de les fonts, la qualitat de les firmes o la seua trajectòria honrosa. L'antic lema de sense periodisme no hi ha democràcia es pot ampliar: sense periodistes no hi ha mitjans, perquè no hi ha redaccions que puguen funcionar només amb la tecnologia més capdavantera que desenvolupe OpenAI. No hi ha casos d'èxit perquè la solució no és la substitució sinó la complementació.

En el panorama mediàtic valencià cap dels entrevistats vol parlar d'acomiadaments i sí de reconversió dels llocs de treball dels i les periodistes que ara fan tasques automatitzables. Segons un informe elaborat per l'FMI<sup>3</sup> juntament amb la Universitat de Stanford, fins a un 60% dels llocs de treball notaran l'impacte de la IA, so-

1 <https://aiindex.stanford.edu/report>

2 Són dades d'un sondeig realitzat per Metroscopia anomenat Periodismo2023/Fundación AXA, presentat el juliol de 2023.

3 Mauro Cazzaniga, M.; Jaumotte, F.; Li, L.; Melina, G.; J Pantou, A.; Pizzinelli, C.; J Rockall, E.; Mendes Tavares, M. (2024). *Gen-AI: artificial intelligence and the future of work*. IMF: <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379>

bretot aquells relacionats amb tasques cognitives. Però esta nova revolució industrial, comparada amb l'arrancada de la física moderna del segle XX, té una qualitat diferenciadora respecte a les anteriors i que recorda la literatura distòpica de ciència-ficció: la IA generativa és capaç de processar grans quantitats de dades, identificar patrons, prendre decisions, conversar amb humans i fins i tot aprendre dels seus errors. És l'arribada de la creativitat computacional. "L'ésser humà continua estant a la cúspide en molts processos creatius, però ara tenim un gran rival", destaca Jordi Linares, doctor en Informàtica i professor de la Universitat Politècnica de València.

Segons l'estudi citat, la meitat dels llocs de treball afectats veuran incrementada la seua productivitat, però l'altra meitat es veuran inevitablement desplaçats: menys llocs de treball, salaris més baixos i fins i tot la desaparició d'alguns treballs. Dones i persones migrades o racialitzades<sup>4</sup> tenen més risc de reemplaçament, ja que copen els llocs més fàcilment automatitzables. Més enllà del biaix per sexe, raça o posició social, per a gran part de les tasques que fins hui realitzava un/una periodista (transcripció d'entrevistes, documentació, programar publicacions a les xarxes, detectar els temes que són tendència, redacció, traducció més o menys simultània i fins i tot enregistrament, locució o edició d'àudio o vídeo, i veurem com es va afegint un llarg etcètera) ja hi ha programes que ofereixen uns resultats de gran qualitat o inclús en certs casos millors que els fets per una mà humana.

José Suárez de Lezo, director del Laboratori de Periodisme de Dades de la Fundació Luca de Tena, recorda que en menys d'un any ChatGPT ha arribat als 100M d'usuaris setmanals i s'ha convertit en la plataforma que més ha crescut en la història en tan poc de temps. També ressenya els seus més de 2M de desenvolupadors de l'API, una quantitat que afavorix la creació de noves línies de treball i, per descomptat, el creixement econòmic exponencial de l'empresa que hi ha darrere, OpenAI, amb el seu mediàtic CEO, Sam Altman, al capdavant. Si el 2022 la companyia va facturar 28 milions de dòlars, les previsions per al 2024 són de 1.300 milions i s'espera que es duplique esta xifra el 2025. "[Estos directius i empresaris] s'estan fent rics amb un producte que no poden controlar", expressa Miguel Arana, investigador en Machine Learning Causal de la Universitat de Cranfield, al Regne Unit, en referència a la relativa autonomia que tenen estos grans models de llenguatge, com ChatGPT, que funcionen com a caixes negres en què amb prou faenes pots controlar la resposta que donen. És el que es coneix com *el problema de l'explicabilitat*<sup>5</sup> i realment suposa un gran obstacle no només perquè posa en perill la confiança de la persona usuària, sinó que també se'n qüestiona la validesa legal i ètica, sobretot si s'està usant una IA per a interferir en decisions que afecten la vida de les persones.

Segons l'últim *World Train Report de Wan-Ifra*<sup>6</sup>, World Association of News Publishers, pràcticament la meitat de mitjans utilitzen la IA en les seues dinàmiques de treball, però només un 28% n'ha desenvolupat guies i polítiques d'ús. Les tasques que li releguen són, sobretot, de creació de titulars i sumaris, correcció de textos, traducció i creació d'articles, suggeriment de temes, simplificació de cerques, o bé per a accelerar la recerca o millorar la interacció amb l'audiència. En estes pàgines veurem exemples d'usos constructius de la IA generativa en mitjans de tot el món, opinions i opcions de periodistes, tant de València com de fora, perfils tècnics, expertes en desinformació, programadors informàtics, investigadores en el camp de la tecnologia i els drets

4 Ellingrud K; Sanghvi, S; Singh Dandona, G; Madgavkar, A; Chui, M; White, O; Hasebe, P. (2023). Generative AI and the future of work in America. McKinsey: <https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america>

5 Això és el que diu ChatGPT: *El problema de l'explicabilitat en intel·ligència artificial (IA) es refereix a la dificultat d'entendre com i per què els models d'IA, especialment els més complexos com els basats en aprenentatge profund, prenen decisions específiques. A mesura que estos sistemes esdevenen més integrats en sectors crítics com la salut, finances i justícia, la necessitat d'explicacions clares i comprensibles esdevé crucial.* Més informació: Yang, W., Wei, Y., Wei, H. et al. *Survey on Explainable AI: From Approaches, Limitations and Applications Aspects*. Hum-Cent Intell Syst 3, 161–188 (2023). <https://doi.org/10.1007/s44230-023-00038-y>

6 <https://wan-ifra.org>



humans, advocades, juristes, caps de secció, responsables d'informatius i redactors, entre d'altres, per tal de crear un relat coral del moment periodístic que vivim.

Esta investigació pretén ser una foto fixa de l'inici de l'any 2024, en este univers canviant en què assistim com a periodistes espectadores al gran degoteig de tecnologies noves, sorprenents, sentents, conversadores, que mirem amb anhels, com a oportunitat i també amb cautela. La mirada crítica, la formació i la reflexió informada és el que ens permetrà, com a societat, decidir quin ús donem a estos artefactes i si som capaços d'emprar tot el seu potencial per a millorar el periodisme en particular i les nostres vides en general, en lloc de per a destruir-lo o destruir-les.

## L'enshittification del nostre tot: internet mor on la IA naix

Estes tecnologies disruptives, transformadores i també vistes amb certa por i recel per tot l'imaginari previ que trobem en el cinema i la literatura distòpica de ciència-ficció, remouran quasi tots els àmbits de la nostra vida, no només la professió periodística. Estem veient canvis quant a la rendibilitat empresarial, en la manera de fer periodisme o de consumir-lo, en la faceta editorial, en la confiança de les marques informatives i en la sostenibilitat dels mateixos mitjans. "La IA retorça els models de negoci i genera noves dificultats", diu Suárez de Lezo, a les quals el conjunt de la societat haurà de fer front.

Les persones expertes consultades coincidixen que estem, com a humanitat, davant d'un dels grans canvis d'era. La Intel·ligència Artificial ha vingut per a transformar tots els àmbits de la nostra vida i per a fer trontollar els pilars sobre els quals s'assenta la construcció de la societat tal com la coneixem hui dia. Els CEO de les grans tecnològiques que marquen i dirigixen el canvi parlen amb el to dels profetes i les seues frases són preludi de la realitat que es va construint a colp de titular, o de declaració grandiloqüent no exempta d'un to misteriós i somiador. Però els exCEO i exdirectius són els que donen més joc. Al llarg del 2024 ha estat notícia la cascada d'abandonaments i cessaments de científics, com en el cas d'Ilya Sutskever, per a molts una de les millors ments en este camp i cofundador d'OpenAI especialitzat en matèria de seguretat. Amb ell se'n van anar 3 persones més d'eixe departament, que ara sembla desproveït de personal i d'atenció per part de la direcció de la companyia. Hi ha moltes especulacions sobre el tema. Es diu que Sutskever i els seus s'encarregaven d'assegurar que cada nova tecnologia comptara amb totes les garanties abans de publicar avanços que pogueren ser potencialment nocius. S'explica que fins i tot van parar la comercialització d'alguns models molt més potents, que podrien comprometre la seguretat de formes que només imaginem per la literatura de ciència-ficció.

D'una banda, hi ha una voluntat manifesta, vista com el camí cap al progrés, de fer millors les màquines, perquè per a ser pitjors ja estem nosaltres. Alhora, hi ha la predisposició de diferenciar-nos rebuscant aptituds i qualitats que siguen exclusivament humanes, que un robot mai podria tindre, almenys en la realitat no ficcionada, per tal que siguem alguna cosa més que els encarregats d'accionar la màquina o prémer el botó d'enviar.

En realitat, pareix que la por que les màquines siguin millor que nosaltres és tan gran com les ganes que tenen uns quants d'aconseguir-ho. Elon Musk, el propietari de la xarxa social X, s'ha atrevit a posar data a l'arribada d'allò que es coneix com la *singularitat*, que és la irrupció d'una Intel·ligència Artificial General o AGI en anglés –ben diferenciada de la IA estreta, feble o aquella dissenyada per a una tasca específica, com reconèixer cares o jugar als escacs–, molt més semblant a les capacitats d'adaptació a qualsevol situació de l'ésser humà. És a dir, una intel·ligència artificial superintelligent: “Tindrem IA o AGI a un nivell que pot fer quasi qualsevol tasca cognitiva. Només ens queda el quan”, se li ha sentit dir, abans de llançar-ne la data: 2025. Mentrestant, el que ja estem veient és com estes potents arquitectures neuronals s'estan acabant a internet. Queda poc perquè el seu aliment siga el seu propi vòmit.

Més enllà del tecnooptimisme desenfrenat hi ha esta altra cara de la realitat tecnològica que té a veure amb el seu gran impacte climàtic i planetari, amb l'amenaça que presenta per a la privadesa dels usuaris o la submissió als designis de Google, Apple, Meta, Amazon i Microsoft, i als seus interessos econòmics. Mentre els humans busquen atorgar –o trobar– en les màquines la singularitat, la perfecció, la versió millorada d'u mateix, les pràctiques capitalistes que encoratgen i paguen esta costosíssima cerca estan convertint l'espai digital en un autèntic femer de contingut patrocinat, sintètic, *spam* o directament fals. És l'enshittificació de la xarxa, un fenomen encunyat per l'escriptor de ciència-ficció Cory Doctorow. La decadència de les plataformes, la merdificació de les xarxes, l'espiral consumista extractivista dels serveis que ofereixen Google, Amazon, Bandcamp, Facebook, Reddit o X. La corba de qualitat decreixent de les plataformes virtuals que dominen tants aspectes de la nostra vida i que s'han convertit en llocs de compravenda d'allò (im)material i també de la nostra privadesa. L'exemple pràctic i recent de decadència el veiem en l'antiga Twitter. La penúltima mesura del seu amo Elon Musk ha estat modificar les normes per a permetre la publicació de pornografia explícita a la xarxa social.

Demis Hassabis, CEO i cofundador de Google DeepMind, eminència dels escacs que va passar del desenvolupament de videojocs a interessar-se per emular la neurociència humana, pronostica que “l'intent de destil·lar intel·ligència en una construcció algorítmica pot arribar a ser el millor camí per a la comprensió d'alguns dels misteris perdurables de la nostra ment”. En la IA depositem les esperances de transcendir els límits que té el fet humà: busquem respostes per a frenar el canvi climàtic, alternatives a la soledat d'este món hiperconnectat, a l'excés de treball capitalista, a les pròpies limitacions del nostre cervell, fins i tot a la mort i la malaltia, quan fem servir algoritmes per tal pronostiquen o diagnostiquen un càncer o qualsevol altra patologia. Però el capitalisme també les troba molt útils a l'hora d'adaptar els productes que ens ofereix en funció dels nostres gustos específics, per això controlen, coneixen i posseïxen centenars de milers de dades sobre el nostre comportament.

Marta Peirano, en el seu llibre *El enemigo conoce el sistema*, parla extensament de les sofisticades tecnologies de la persuasió i del capitalisme de l'atenció: “El 40% dels usuaris abandonen una pàgina web si tarda més de tres segons a carregar”<sup>7</sup>. Les ments més brillants cobren i treballen a Silicon Valley, i investiguen en les millors universitats americanes, per tal que qualsevol comportament nostre a la xarxa repercutisca en un ampli benefici econòmic per a les grans empreses que venen productes i serveis, a base de crear necessitats de consum.



Arranquem recordant una definició, una de tantes, que és la triada per la Comissió Europea a la redacció de l'AI Act, la primera Llei Europea –i mundial– que regula la Intel·ligència Artificial, aprovada el 13 de març de 2024 després de tres anys d'intenses negociacions:





**‘Artificial intelligence system’ (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments.<sup>8</sup>**

# Introducció

## Ni tan intel·ligent ni tan artificial

“Cerrando el mes muy contenta” / “Hoy me he sentado a escribir todo por lo que me siento agradecida” / “Contadme de que estáis agradecidos” (sic) / “Cenita con los amigos que se vuelven familia”. Aitana Lopez, més coneguda pels seus 329.000 seguidors i seguidores d'Instagram com a @fit\_aitana, puja a les seues *stories* fotos dels sostenidors que li ha regalat Victoria's Secret, fotos de capítols de la vida social, fotos dels seus viatges de visita a Madrid o del concert de Blink-182 al qual –suposadament– va anar. I dic suposadament perquè en realitat no té vida social, no viatja de visita a Madrid ni va estar en el concert de Blink-182, encara que Victoria's Secret sí que li va regalar uns sostenidors. Aitana no existix en allò que fins hui coneixem com el món real. En la *bio* del seu perfil es definix com una *virtual soul*. Aitana és el producte d'una IA que du la signatura d'una agència de publicitat de Barcelona.

No és l'única *AI Influencer* amb milers de seguidors a les xarxes socials que podria semblar, quasi quasi, real. Tan real que molts dels seus *followers* barons li envien missatges amb propostes sexuals –els seus creadors<sup>9</sup> parlen que rep unes 400 proposicions de cita diàriament– i desenes de comentaris demanant-li que es despulle o que es dedique al porno. Des de futbolistes fins a actors –amb fins a 11 milions de seguidors a Instagram– continuen confonent-la amb una model de carn i ossos i bombardejant els seus DM. Hi haurà qui pense que Aitana ha passat el test de Turing<sup>10</sup> de sobres. I més quan factura ella soleta 4.000 euros mensuals en publicitat i posats en roba interior.

El pas d'este tipus d'avatars de l'àmbit comercial al periodístic –estèticament no són tan diferents– ja el veiem. Ja existix “la primera influencer virtual entrenada per Intel·ligència Artificial que presentarà un espai associat a un gran format televisiu al nostre país”<sup>11</sup>, un especial del programa *Supervivientes* para la web de Telecinco i la plataforma de continguts a la carta de Mediaset. I no n'és l'única. Cada setmana esmorzem amb una llista de notícies relacionades amb personatges producte de la intel·ligència artificial, noves ferramentes que semblen intel·ligents, nous pactes de mitjans amb les tecnològiques que produeixen i entrenen intel·ligència artificial generativa o entrevistes a no tan noves pensadores i pensadors que opinen sobre la bondat o la maldat de la tecnologia.

Per a la periodista Marta Peirano, especialista en la intersecció entre tecnologia i poder, la IA, en realitat, no és ni tan intel·ligent ni tan artificial. I este sí que és un debat obert interessant: què vol dir ser intel·ligent? Esta qualitat, fins ara patrimoni quasi exclusiu de la humanitat, ha esdevingut l'epítet de ChatGPT, Gemini, Copilot, Claude o qualsevol altre enorme model de llenguatge (Large Language Models, LLM, en anglés). “Em preocupa molt més l'estupidesa humana que no la intel·ligència arti-

9 Benedicto Borges, L. (2023, 8 de novembre). *Nuestra hija Aitana, la primera modelo de IA española que gana... "4.000 euros al mes"*. El Mundo. <https://www.elmundo.es/cronica/2023/11/08/65453bc7fc6c83f03a8b4585.html>

10 Més avant en parlarem, però només per a introduir-ho: el test de Turing definia una màquina com a sentent, intel·ligent o pensant quan fora una màquina que, en conversa amb un ésser humà, este no distingira si estava parlant amb un aparell o amb un igual.

11 Redacció verTele (2024, 13 de març). *Mediaset crea una presentadora por inteligencia artificial para conducir un formato de 'Supervivientes' en Mitele*. verTele! [https://www.eldiario.es/vertele/noticias/mediaset-crea-presentadora-inteligencia-artificial-conducir-nuevo-formato-supervivientes-mitele\\_11210629.html](https://www.eldiario.es/vertele/noticias/mediaset-crea-presentadora-inteligencia-artificial-conducir-nuevo-formato-supervivientes-mitele_11210629.html)

ficial”, reflexiona Ramon López de Mántaras, professor de recerca del CSIC i director de l’Institut de Recerca en Intel·ligència Artificial.

Els límits entre allò que era estúpid i el que hui és intel·ligent s’estan redefinint. De fet, canviem la definició segons parlem d’un ésser humà o d’una màquina. A una persona li preguntes i et respon coherentment, i no consideres que és intel·ligent, dius, bé, és el mínim... “El barem que hem posat per a decidir què és intel·ligent és més baix per a una màquina que per a un humà”, atlega Ujué Agudo, doctora en Psicologia i investigadora de la relació entre la IA i les relacions humanes.

En una de les seues conferències<sup>12</sup>, Peirano qüestionava estos grans mites construïts al voltant de la IA i la seua capitalització de pràcticament tots els sectors socials i laborals. Per a la periodista i escriptora, la tecnologia amenaça de deixar sense faena el personal assalariat des de la primera revolució industrial, quan es va popularitzar el fantasma de les fàbriques sense obrers. Però ni va passar ací ni, previsiblement, passarà en el futur. Seguint a Peirano, el que veiem és que l’arribada de noves màquines potser no ens lleva del tot la faena, però sí que degrada la nostra posició laboral i amb això la nostra remuneració. Certes aptituds humanes ja no són tan necessàries, per tant, no es paguen com es mereixerien, amb la conseqüent reducció de costos de producció per a l’empresariat. La novetat és que hui també –i sobretot– es veuen afectats els treballs creatius i no només els manuals o mecànics. Perquè no cobra el mateix un guionista de Hollywood que té una idea original, i l’escriu, que si simplement es dedica a redactar i donar forma a l’ocurrència cridanera que ha tingut qualsevol xatbot de renom. A més de Peirano, esta teoria la sostenen altres tantes veus que apareixen en la present investigació. Com si la tecnologia i el seu avanç suposara també un pas avant en la precarització de les forces de treball.

El primer a anomenar el *precariat*, eixa nova classe social en què viu instal·lada, entre altres professions, la periodística, va ser l’economista Guy Standing. Els integrants d’este estament van perdent drets socials i econòmics a la mateixa velocitat vertiginosa amb què avancen la computació quàntica i l’automatització dels sistemes algorítmics. El *capitalisme rendista*, seguint a Standing, seria aquell que està encapçalat per les grans corporacions tecnològiques i que l’ha allunyat del lliure mercat per a crear un sistema de *mercat arreglat* en què “cada vegada més ingressos van a parar als amos de la propietat, la propietat física, la propietat financera i la propietat intel·lectual. I va, cada vegada menys, a les persones que depenen de la seua faena”<sup>13</sup>.

Si seguim de prop els darrers estudis, veurem que estem properes a l’escenari que vaticinen pensadores com Peirano. En un *paper*<sup>14</sup> publicat per personal investigador d’OpenResearch, la Universitat de Pennsylvania i OpenAI –l’empresa tecnològica que ha desenvolupat ChatGPT– reconeixen que els seus models poden afectar milers de llocs de treball. Segons s’arregla en l’estudi, especialment els més ben pagats i qualificats i també els relacionats amb la programació, la redacció i la creació de continguts. Una altra investigació<sup>15</sup>, en este cas del MIT –el Massachusetts Institute of Technology– es va proposar veure com la IA de generació de continguts influïx en la productivitat de les persones treballadores. La conclusió va ser que els materials elaborats amb ChatGPT s’havien fet en molt menys temps i tenien més qualitat, segons va dictaminar un equip de persones experts. Este estudi confirmaria el pitjor dels temors.

12 M. Peirano (2023). *Ni intel·ligent ni artificial: terapia extractivista para un mundo en crisis*. Conferència en la trobada *Immaterial 2023* a Tabakalera.

13 Ribas, N. (2022, 21 de juny). *La renta básica no produce más vagos, sino gente con buena salud mental y más productiva*. Entrevista a Guy Standing a Eldiario.es. [https://www.eldiario.es/illes-balears/economia/guy-standing-renta-basica-no-produce-vagos-gente-buena-salud-mental-productiva\\_128\\_9099519.html](https://www.eldiario.es/illes-balears/economia/guy-standing-renta-basica-no-produce-vagos-gente-buena-salud-mental-productiva_128_9099519.html)

14 Tyna Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. (2023, 22 d’agost). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. OpenAI, OpenResearch, University of Pennsylvania.

15 *ChatGPT Lifts Business Professionals’ Productivity and Improves Work Quality*, publicat a Nielsen Norman Group el 2 d’abril de 2023. <https://www.nngroup.com/articles/chatgpt-productivity>



Però, com recorda Marta Peirano, ChatGPT sense la mà de l'ésser humà ni resulta creïble ni serviria per a res. I podríem afegir que sense una mà humana que sàpiga fer-la servir correctament, donant les ordes o prompts<sup>16</sup> adequades, tampoc. “Amb tota esta automatització que ens ve damunt, si estes tasques les fa una intel·ligència artificial i tu no aportes valor, òbviament et quedaràs sense faena. L'ocupació no te la llevarà la intel·ligència artificial, sinó algú que sàpiga gastar la intel·ligència artificial”, deia Cristina Aranda<sup>17</sup>, CEO de Big Onion i cofundadora de Mujeres Tech. Esta frase no és seua, almenys no només, ja que l'hem escoltada als escenaris més dispars, entre ells en un informe d'IBM publicat a l'estiu del 2023: “AI won't replace people—but people who use AI will replace people who don't”<sup>18</sup>.

## El misteri de l'explicabilitat: com funciona la IA generativa?

“Els grans models de llenguatge (LLM) poden fer coses sorprenents. Però ningú no sap exactament per què. I això és un problema. Esbrinar-ho és un dels enigmes científics més grans del nostre temps i un pas crucial per a controlar models futurs més potents”<sup>19</sup>. Com escriu el periodista del MIT Technology Review Will Douglas Heaven, fa un parell d'anys dos investigadors van intentar esbrinar què calia perquè un model de llenguatge executara operacions aritmètiques bàsiques. Per error, van deixar l'experiment en marxa alguns dies i se'ls va anar mostrant als models de manera repetida les mateixes operacions. Així van descobrir que, en alguns casos, els models podien aparentment no aprendre una tasca i, de sobte, aconseguir-ho, una qualitat que no s'esperava del *deep learning* profund i que suposa un pas més enllà del *machine learning* o aprenentatge automàtic<sup>20</sup>. Este comportament el van anomenar *grokking* o capacitats emergents.

“Podem estar segurs alguna vegada que els models han deixat d'aprendre?”, es pregunta Hattie Zhou, investigadora d'IA a la Universitat de Montreal i Apple Machine Learning Research. Estos LLM semblen comportar-se de maneres que les matemàtiques dels llibres de text diuen que no haurien de fer. Ningú sap exactament com funciona del tot el *deep learning*, que és una tecnologia clau dins de la IA, especialment en àrees complexes com el processament del llenguatge natural o la visió per compu-

16 Més avant tornarem a recuperar una frase de la matemàtica i investigadora Ada Lovelace, que deia que amb les instruccions adequades les màquines podrien fer més coses. D'ací la importància dels *prompts*, que són les instruccions que dones a un model de llenguatge per tal de guiar la seua resposta i tractar d'obtenir el contingut que necessites. Si preguntes a ChatGPT, afeg que *dissenyar efectivament un prompt és crucial per a influir en la qualitat i rellevància de l'eixida del model*.

17 Va ser durant el *Summit RedMad, Red de Mujeres en Alta Dirección*, a finals de 2023. Es pot veure íntegra en el canal de YouTube de REDMAD. <https://www.youtube.com/@redmad9369/videos>

18 “La IA no reemplaçarà les persones, però les persones que sàpien utilitzar la IA sí que reemplaçaran les que no”. Goldstein, J; Login, B; Fillare, C; Nowal, C (202-2023). *Augmented work for an automated, AI-driven world Boost performance with human-machine partnerships*. IBM IBV and Oxford Economics. <https://www.ibm.com/downloads/cas/NGAWMXAK>

19 Will Douglas Heaven és redactor en cap d'Inteligència Artificial en *MIT Technology Review*, la publicació divulgativa del conegut Institut de Tecnologia de Massachusetts. Ací parlem d'un reportatge publicat en març de 2024 titulat *Large language models can do jaw-dropping things. But nobody knows exactly why*. <https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why>

20 En les pàgines finals hi ha un glossari de termes, però ací hi ha dos definicions senzilles que aporta ChatGPT: *'Machine learning* és un camp de la intel·ligència artificial on les màquines aprenen a fer tasques a partir de dades sense ser programades explícitament. Els algoritmes troben patrons en grans quantitats de dades i ajusten les respostes o prediccions a partir d'estes. *Deep learning* és una subramificació del *machine learning* que utilitza xarxes neuronals artificials profundes per a modelar i resoldre problemes complexos, inspirant-se en el funcionament del cervell humà.

tació, i que permet a les màquines, per exemple, interpretar i comprendre imatges o vídeos de manera similar a com ho fan els humans. Com assenyala el redactor del *MIT Technology Review*, els investigadors estudien estes qüestions relacionades amb l'aprenentatge automàtic profund com si foren estranys fenòmens naturals, ja que trenquen les explicacions que es tenien fins ara sobre el comportament dels models predictius. “La tecnologia funciona. No és suficient?”, ironitza el reporter.

“Explicar realment per què les xarxes neuronals tenen este tipus de comportament inesperat? Encara estem lluny d'aconseguir-ho”, diu Daniel Hsu, informàtic de la Universitat de Columbia, com a única resposta possible al problema de l'explicabilitat de la intel·ligència artificial. Resulta una mica esgarrifós pensar que estos models sàpiguen fer coses per a les quals no se'ls ha entrenat i que això no es puga justificar de manera matemàtica. Estos models avançats són capaços de generalitzar i aprendre a fer tasques que no han vist abans perquè no només memoritzen patrons, sinó que sorprenen amb la seua capacitat de crear regles per aplicar estos patrons a casos nous. És l'avantsala de l'AGI o *Artificial General Intelligence*, que és capaç d'imitar a la perfecció la capacitat humana per a aprendre o construir conceptes.

La mateixa comunitat científica és la primera a sorprendre's, quan veu com entrenen els seus models amb problemes matemàtics en anglés, després els posen una mica de literatura en francés i d'ací veuen com *generalitzen* i resolen problemes matemàtics en francés. Estos enormes models de llenguatge s'entrenen amb ingents quantitats de paràmetres, també anomenats pesos. Estan basats en els *Transformers*, un tipus d'arquitectura de xarxa neuronal especialitzada en el processament de seqüències de dades, com si foren paraules dins de frases. Així els defineix ChatGPT:

*Els Transformers són una forma específica d'organitzar i estructurar les xarxes neuronals perquè puguen processar i entendre dades de manera més eficient. Pensa en l'arquitectura com el plànol d'un edifici: descriu com s'organitza i connecta cada part del sistema per a complir-ne el propòsit.*

De fet, la T de ChatGPT en fa referència: *Generative Pre-training Transformer*. La P de *Pre-training* ens parla del seu ensinistrament, que ha tingut diverses fases i ha passat per diversos estadis, en alguns moments supervisat per humans i en altres autosupervisat. Este entrenament implica mostrar-li una gran quantitat de dades, com sabem. És lògic pensar que a més quantitat de dades emprada, més dificultat per a traçar-ne l'autoria i saber d'on provenen o a qui pertanyen eixes dades. Quin volum de *datasets* necessita un model amb ChatGPT per al seu entrenament? Quasi és complex fins i tot recomptar-los. Un exemple d'una part d'estes dades podria ser des dels 3 bilions de la Viquipèdia en anglés, sumats amb 410 bilions de Common Crawl, un repositori que rastreja la web mitjançant processos de *crawling* que funciona des del 2008. Encara que parega difícil de creure, les dades són finites i internet s'acaba. Com veurem més avant, les grans tecnològiques continuen usant la seua imaginació per a costejar estos arduos processos d'entrenament i per això no dubten fins i tot a apostar per l'energia nuclear com a forma de proveïment, o a transcriure YouTube sencer per a tindre material amb què alimentar les seues criatures.

Triar una bona matèria primera amb què ensinistrar els grans models de llenguatge és una part clau, perquè si se'ls instrueix amb fem, generaran fem. Si les dades tenen baixos, el resultat serà esbiaixat. “Estes màquines són increïblement intel·ligents, però sorprenentment ximplés”, apuntava Will Douglas Heaven. Tot i així, ens fascinen. “Parlem de l'onada tecnològica més disruptiva que hem viscut i jo crec que de la història de la humanitat, perquè és una tecnologia on tots els terminis s'acurten”, assenyala la periodista i experta en xarxes socials, periodisme mòbil i desinformació Carmela Ríos. I afeg: “Tots els processos relacionats amb el periodisme es veuen

afectats, des de la producció d'informatius a la de continguts, passant per l'escolta". Segons esta veterana periodista, el missatge que ens llança esta nova realitat és molt potent: "Una màquina o un bot amb una intel·ligència artificial és capaç de fer una gran part de les tasques que fins ara només podia garantir el producte de la intel·ligència humana. Ara la intel·ligència no és humana sinó artificial, i eixe missatge fa molta por".

Per al professor de la Universitat Politècnica de València Jordi Linares, ningú podia anticipar esta última onada tan disruptiva que ens faria parlar de conceptes abans impensables, com la creativitat computacional. La màquina ha entrat en camps abans vetats per a ella i només reservats a un cervell humà, periodista, artista o creador: "L'evolució de les xarxes neuronals artificials i del maquinari que les sustenta, unit a l'accés de grans *datasets*, a estos grans volums d'informació a internet, ha precipitat l'aparició d'un conjunt de tècniques extraordinàriament increïbles a l'hora de generar text creatiu, vídeo, 3D, música i pràcticament tot allò que en definitiva puguem tindre de forma massiva a internet i puguem utilitzar per a entrenar estes xarxes", diu.

↳ **Hallucination is the belief that a statistical model of language is governed by truth criteria**

@tallerestampa



# Ètica i tecnologia

## La consciència (ja no?) és patrimoni de la humanitat

D'estos models de llenguatge se'n diu que, a través d'estes complexes xarxes neuronals amb certa bioinspiració, com les qualifica Jordi Linares, són capaces de raonar, entretenir-se, conversar. Estes són qualitats típiques de la consciència humana. Així definix ChatGPT la consciència humana, una vegada hem separat el gra de la palla: *la capacitat subjectiva d'experimentar sensacions, percepcions, pensaments i emocions, així com de tindre consciència d'u mateix i el seu entorn. (...) En termes més específics, la consciència implica una experiència interna de la ment, que inclou la capacitat d'atenció, reflexió, introspecció i autoconsciència*. I això és clau: ser conscient que existixes, que eres, que eres un ésser sentent.

“Sens dubte, no n'hi haurà, no s'ha creat mai consciència a través d'una màquina, ni hi ha possibilitats de crear-la. És un tret específicament humà, és justament el que ens fa diferents milions d'espècies. Plantejar una cosa així és un poc una bogeria, però bojament tots ens ho hem cregut. Esta bogeria els està servint molt per a uns fins, que són els de guanyar molts diners”, retreia Yann LeCun<sup>21</sup>, científic en cap d'IA de Meta, un dels grans defensors de publicar el codi en obert, a qui se li ha sentit moltes vegades denunciar aquells que exploten la por sobre estos sistemes.

Elon Musk, per exemple, discrepa i fixa l'any 2025 com la data en què coneixerem esta intel·ligència artificial *pseudoconscient* capaç d'executar qualsevol classe de tasca cognitiva de la mateixa manera que ho faria un humà. El magnat, segons *Forbes* la persona més rica del món, es troba entre els mil signants d'una declaració per a demanar que es frene l'entrenament de poderosos models de llenguatge, amb l'argument que són una potencial amenaça per a la humanitat<sup>22</sup>.

Com que LeCun no és dels que veuen les orelles al llop, no creu que calga ser alarmista amb la legislació. En declaracions a *El País*, explicava<sup>23</sup> que “contràriament al que es pot escoltar d'algunes persones, no hi ha un disseny per a un sistema que arribe a la intel·ligència humana” i que, per això, “demanar regulacions per por a la intel·ligència sobrehumana és com demanar una regulació de vols transatlàntics a una velocitat pròxima a la del so en 1925. Això no està a tocar; portarà molt de temps, amb uns sistemes que encara no coneixem”. I continua LeCun: “Estos sistemes són

21 Rizzi, A. (2024, 19 de gener). Yann LeCun, científico jefe de IA de Meta: “Una inteligencia artificial de nivel humano va a llevar mucho tiempo”. *El País*. <https://elpais.com/tecnologia/2024-01-19/yann-lecun-cientifico-jefe-de-ia-de-meta-una-inteligencia-artificial-de-nivel-humano-va-a-llevar-mucho-tiempo.html>

22 Pel maig del 2023 més de mil experts i executius de la indústria tecnològica, inclosos, a més d'Elon Musk, el cofundador d'Apple, Steve Wozniak, el director executiu de la firma Stability AI, Emad Mostaque, junt amb investigadors de la firma DeepMind propietat de Google, van signar esta petició. La van publicar a la web del *Future of Life Institute*, una institució ubicada a Cambridge que té com a objectiu reduir els riscos que implica el desenvolupament tecnològic. En esta carta demanaven que les empreses que desenvolupen este tipus de programes “pausen immediatament, durant almenys sis mesos, l'entrenament dels sistemes d'intel·ligència artificial més potents que GPT-4”, perquè s'estava convertint en “una carrera fora de control per desenvolupar i implementar ments digitals cada vegada més poderoses que ningú, ni tan sols els seus creadors, poden comprendre, predir o controlar de manera fiable”.

23 En el mateix article d'*El País* de la cita 21.



intelligents en el domini relativament estret en què han estat entrenats. Dominen el llenguatge i això ens fa pensar que són intelligents, però no ho són tant”<sup>24</sup>.

Este gran pas per a la humanitat –lent per a LeCun però amb grans accelerades en els darrers anys–, que és passar de la IA discriminativa o predictiva i el seu pensament convergent a la IA generativa i el seu canvi al pensament divergent, està succeint i sembla imparabile. La seua evolució ha estat exponencial i ho continua sent, perquè ningú no dubta que cada vegada és millor, més sofisticada, més similar a la resposta humana. En només tres anys els canvis han sigut molt significatius.

Si fem una ullada a l’any 2021, en podem veure un exemple. L’editor de la secció d’IA de *MIT Technology Review*<sup>25</sup> definia GPT-3 com una cosa “sorprenentment bona i totalment absurda”; un xatbot capaç de generar “text sorprenent similar a l’humà a demanda, però que no ens acostarà a la veritable intel·ligència”. Este periodista i doctor en Ciència Computacional fa una aproximació molt interessant a la definició d’intel·ligència, que va en la línia de les reflexions de Marta Peirano, quan ironitza que no és tan artificial perquè està dissenyada, creada i mantinguda per humans; els algorismes no funcionen en un buit sincrètic sinó que necessiten servidors, energia i supervisió humana constant. I potser tampoc no és tan intel·ligent, almenys no ho és en el sentit humà del terme, precisament per esta dependència de les dades que li ofereixen els seus dissenyadors i perquè, per exemple, és incapaç de generar coneixement nou. La seua intel·ligència es limita a identificar patrons i fer prediccions basades en estos patrons, i fer-ho meravellosament bé.

Diu Douglas Heaven: “El rendiment *huma-like* i la sorprenent versatilitat de GPT-3 són el resultat d’una excel·lent enginyeria, no d’una autèntica intel·ligència”. Ho repetim: la IA és un subproducte de la intel·ligència humana. I seguix fent que ens en replantegem el concepte: “Tenim un llistó molt baix per a detectar la intel·ligència. Si alguna cosa pareix intel·ligent, és fàcil enganyar-nos pensant que ho és. El truc més gran de la IA ha estat convèncer el món que existix. GPT-3 és un gran pas avant, però continua sent una ferramenta feta per humans, amb tots els defectes i limitacions que açò implica”. Les paraules de Douglas Heaven s’entenen en un moment (juliol de 2020) en què encara véiem grans errors o al·lucinacions i en què estos models de llenguatges enormes tenien aparença de xatbots de batibull talla-apega, a més d’un important biaix sexista i racista del qual encara en fa gala. Però el rerefons continua vigent.

Aquell mateix any 2021, Renée DiResta, cap de recerca de l’Observatori d’Internet de Stanford, publicava un text a *Wired*<sup>26</sup>, tres mesos abans de les eleccions del 3 de novembre de 2020 als Estats Units, en què criticava que als *deepfakes*<sup>27</sup> audiovisuals se’ls notava massa el *fake*. Tot just l’estiu d’aquell any es va presentar al món una versió beta de GPT-3. DiResta el va definir llavors com “la pròxima frontera de l’escriptura generativa: una intel·ligència artificial capaç de produir frases que sonen sorprenentment humanes, encara que a vegades surrealistes”. Estos enormes models de llenguatge transformador van afinant cada volta més la punteria i són capaços –com els humans– d’anar millorant i aprenent dels seus errors per tal de fabricar productes creatius òptims, més creïbles i més difícils de diferenciar dels generats per persones.

24 Rizzi, A. (2024, 19 de gener). Yann LeCun, científico jefe de IA de Meta: “Una inteligencia artificial de nivel humano va a llevar mucho tiempo”. El País. <https://elpais.com/tecnologia/2024-01-19/yann-lecun-cientifico-jefe-de-ia-de-meta-una-inteligencia-artificial-de-nivel-humano-va-a-llevar-mucho-tiempo.html>

25 Douglas Heaven, W. (2020, 20 de juliol). *OpenAI’s new language generator GPT-3 is shockingly good—and completely mindless*. MIT Technology Review. <https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp>

26 DiResta, R. (2020, 31 de juliol): *AI-Generated Text Is the Scariest Deepfake of All*. Revista Wired. <https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all>

27 Més avant es torna a explicar este terme, però la traducció seria alguna cosa així com ultrafalsificació. De l’anglès, *fake*, engany o falsificació, i *deep*, perquè estos continguts normalment en vídeo s’elaboren gràcies a eines que usen *deep learning*, que, com ja s’ha dit, és una branca de l’aprenentatge automàtic o *machine learning* que utilitza xarxes neuronals complexes per a fer prediccions a partir de grans volums de dades.



Sam Altman, cofundador d'OpenAI, tuitejava aleshores<sup>28</sup>: “L'exageració de GPT-3 és massa. És impressionant (gràcies pels bonics elogis!) però encara té serioses debilitats i, a vegades, comet errors molt ximpls. La IA canviarà el món, però GPT-3 és només una ullada molt primerenca. Encara tenim molt per resoldre”. Esta és la carrera que estan recorrent les grans tecnològiques des de la Badia de San Francisco per a ser les millors i les primeres, amb OpenAI al capdavant, en crear alguna cosa que cada vegada sorprenga més, sense escatimar en despeses econòmiques, energètiques o de recursos. Perquè la multimodalitat necessita *multimilions* de recursos i dades, i un augment creixent de pràctiques extractivistes a gran escala tant d'este nou petroli com de matèries primeres i energies. És la competició multimilionària per la conquesta d'una nova lluna llesta i sintètica. I té un altre punt en contra, que assenyalava Carissa Véliz, experta en ètica aplicada a la tecnologia: la intel·ligència artificial l'estan dissenyant cinc homes blancs a Silicon Valley a qualsevol preu.

↳ Why isn't AI doing the tedious shit for creative people instead of the creative shit for tedious people

@piesaac

## La creativitat fa humana la màquina: antropomorfització i AGI (Artificial General Intelligence)

“Ens meravella el que hem creat”, ironitza la periodista especialitzada en mitjans digitals de *Newtral*, Marilín Gonzalo. “No hi ha necessitat com a humanitat de donar contingut creatiu a les màquines, primer perquè ho fan malament”, continuar Gonzalo. Per a Miguel Arana, investigador de *Machine Learning Causal* de la Universitat de Cranfield, al Regne Unit, només és qüestió de temps: “Les màquines comencen a copiar molt bé allò que fem els humans, fins i tot a sobrepassar, en alguns casos, la capacitat humana. Estem veient que la intel·ligència humana potser no és tan increïble com ens pareixia, que teníem una visió de la intel·ligència humana un poc mitificada”.

Sembla que costa de reconèixer que un LLM<sup>29</sup> pugui ser més creatiu que tu. Hi ha un estudi molt interessant publicat en la revista *Frontiers in Psychology*<sup>30</sup> titulat ‘Avaluació de l'emoció i la sensibilitat de les obres d'art d'IA’. S'hi va fer l'experiment de mostrar una obra generada amb intel·ligència artificial a dos grups de persones; al pri-

28  
29  
30

Va ser en juliol del 2020 en la xarxa social X, propietat del seu exsòci d'OpenAI Elon Musk.

Com ja s'ha dit, *Large Language Model*, model de llenguatge gran.

Agut, U.; Arrese, M.; G. Liberal, K.; Matute, H. (2022, 5 d'abril). *Assessing Emotion and Sensitivity of AI Artwork*. *Frontiers in Psychology*.

mer se li va dir que era sintètica i al segon, obra d'una mà humana. Ací es va concloure que se li atribuïa menys sensibilitat, menor capacitat per evocar les seues emocions i menor qualitat a l'obra d'art quan es creia que l'artista era una IA en comparació amb quan pensaven que era obra de la creativitat i el talent humà. Els resultats mostren –diu l'estudi– que sembla que l'avaluació de l'art està modulada, almenys en part, per estereotips i prejudis previs sobre les habilitats creatives de la IA.



Ontològicament parlant, la comprensió que som iguals o que podem fer coses semblants a les que fan un muntó d'algoritmes, cables i xips no és senzilla. Per tal que eixe muntó de ferralla ens caiga millor, la humanitzem, conscientment o inconscientment, en un procés que es coneix com a antropomorfització de les màquines. Els atorguem característiques, aptituds i habilitats humanes, i gastem els mateixos adjectius per a parlar d'elles que d'un grup de gent. Per això diem que són *intelligents*, o que les *entrenem* amb una gran quantitat de dades. O anomenem allucinacions als errors que cometen les tecnologies com GPT.

Estes qüestions les explica Marilín Gonzalo en un reportatge<sup>31</sup> de *Newtral*. La periodista defineix les allucinacions com “una resposta confiada donada per una IA que no està justificada per les dades amb què ha estat entrenada”. Este terme ve de la psicologia humana i s'associa amb les falsedats. La utilització del llenguatge no és casual i, com la mateixa experta recorda, no és el mateix dir que la IA còpia a dir que la IA *aprén*<sup>32</sup>, ja que en l'aprenentatge ni s'infringixen els drets d'autor ni es pot rebre, a priori, l'acusació de plagi.

En un article de Naomi Klein per a *The Guardian*, la coneguda autora de *No Logo* i *La doctrina del xoc* es pregunta: “Per què anomenar els errors ‘allucinacions’? Per què no *escombraries algorítmiques*? O errors? (...). En apropiat-se d'una paraula d'ús comú en psicologia, psicodèlia i diverses formes de misticisme, els impulsors de la IA, alhora que reconeixen la fal·libilitat de les màquines, alimenten la mitologia més

31 Gonzalo, M. (2023, 22 de juny). Las “alucinaciones” de la inteligencia artificial son fallos que preocupan a los expertos. *Newtral*. <https://www.newtral.es/alucinaciones-ia-fallos-inteligencia-artificial-confabulaciones/20230622>

32 Gonzalo, M. (2024, 15 de gener). Los derechos de autor y la IA, según OpenAI. *Newtral*. <https://www.newtral.es/derechos-autor-ia-openai/20240115>

preuada del sector: que en construir estos grans models lingüístics i entrenar-los en tot allò que els humans hem escrit, dit i representat visualment, estan en procés de donar a llum una intel·ligència animada a punt de provocar un salt evolutiu per a la nostra espècie”<sup>33</sup>.

Estos programes, continua Gonzalo, “simplement són excepcionalment eficaços a l’hora de predir patrons del llenguatge i han estat optimitzats per a ser conversacionals, per la qual cosa es col·loquen en un pla mental similar al d’altres agents humans quan hi interactuem, i ens resulten creïbles i convincents”. És la nostra empatia la que fa la resta del treball.

Una conscienciosa i ben treballada estratègia de màrqueting empaqueta la màquina per a regal Com es ven la IA al món i la societat és important. Per exemple, tots els anuncis de nous llançaments en OpenAI tenen eixa aura de misteri i temor que desprenen les persones amb poder que tenen accés a una informació que el comú dels mortals desconeix. Solen jugar al dubte i a despistar, i aprofiten les seues grans audiències per tal que els seus líders facen declaracions que van de la fascinació per com afina l’última actualització de ChatGPT a la por de la distopia o a la destrucció de la raça humana amb una cosa que ha creat la mateixa raça humana. “Altman sempre té eixe to com que ell ha vist a Déu en la màquina i no ens ho vol contar, és el que ven, i tracta constantment d’alimentar eixe *hype* que els està reportant molts beneficis econòmics”, objecta Arana.

En una entrevista a Sam Altman en el pòdcast de Lex Fridman<sup>34</sup> –que tenia 1,6 milions de reproduccions set dies després d’haver estat publicada en YouTube–, el CEO d’OpenAI afirma que abans que s’acabe esta dècada ja tindrem entre nosaltres la famosa AGI, la Intel·ligència Artificial General, tot i que assegura que encara no l’han construïda. Tecnològicament parlant, hauríem passat de les calculadores, capaces de resoldre un nombre limitat d’operacions, a una superintel·ligència singular amb consciència de si mateixa que supera amb escreix la humana i que és capaç de respondre i donar solucions a tot, de generalitzar, aprendre i comprendre.

- Sara: Què és l’AGI?  
 Gemini: L’AGI, o Intel·ligència Artificial General (per les sigles en anglés), és un tipus hipotètic d’intel·ligència artificial que iguala o supera la intel·ligència humana mitjana en totes les àrees cognitives. És una IA capaç de comprendre el món de la mateixa manera que un humà, amb la capacitat d’aprendre, raonar i resoldre problemes de manera autònoma.
- Sara: Existix l’AGI?  
 Gemini: No, l’AGI encara no existix. És un concepte teòric i encara no s’ha desenvolupat una IA que complisca totes les característiques de l’AGI. No obstant això, és un camp de recerca molt actiu i s’hi estan fent avanços significatius.

33 Klein, N. (2023, 8 de maig) *AI machines aren't 'hallucinating'. But their makers are.* The Guardian.

<https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein>

34 Segons el seu canal de YouTube, on atesora més de 4 milions de subscriptors, “Lex Fridman és un científic, enginyer i podcaster que explora temes com la intel·ligència artificial, la filosofia o les arts marcials”. Només l’entrevista amb Altman té més de 2 milions de visualitzacions, i no és l’únic habitant de Silicon Valley que apareix en este canal. Trobem entrevistes de Fridman de fins a 3 hores amb Yann Lecun, Elon Musk o Mark Zuckerberg, a més d’altres personatges diversos i molt coneguts, des de Kevin Costner o el còmic egipci Bassem Youssef fins al propietari d’Amazon Jeff Bezos o el president d’Israel, Benjamin Netanyahu.

Al novembre del 2023 va ser notícia l'acomiadament i la posterior readmissió de Sam Altman com a conseller delegat de la tecnològica que ell mateix va cofundar, i just en les mateixes dates l'agència de notícies Reuters publicava una informació<sup>35</sup> segons la qual OpenAI havia assolit un poderós descobriment d'intel·ligència artificial que podria amenaçar la humanitat. Este projecte misteriós s'anomenaria Q\* (pronunciat *qu star*) i, preguntat per ell en el pòdcast de Fridman, Altman respon: “No estem preparats per a parlar-ne. Investiguem tota classe de coses. Hem dit que un millor raonament en estos sistemes és una direcció que ens agradaria prosseguir”.

“A la companyia sens dubte li interessa mantindre esta aura de flirteig amb projectes enigmàtics”, escriu el periodista Jordi Pérez Colomé, en la seua newsletter sobre tecnologia d'*El País*. En ella recorda una entrevista de la revista *Wired* amb Lukasz Kaiser, un altre científic d'OpenAI. Conten que açò és el que va passar en eixa xarrada: “Kaiser és un dels inventors d'una nova tecnologia anomenada Q\*. Quan vaig intentar preguntar a Kaiser sobre el tema en la nostra entrevista, la persona de relacions públiques d'OpenAI pràcticament va botar sobre la taula per a fer-lo callar”.

Altman introduïx el concepte d'*agent d'IA*, que serà una mena de barreja entre un assistent, la teua millor amiga i un conseller real, que sabrà de tu més que tu mateixa: “També posaràs tota la teua informació, el teu historial al llarg del temps i t'anirà coneixent cada vegada millor. Vull que integre les lliçons apreses i que em recorde en el futur què fer de forma diferent o en què he de tindre cura. Tots guanyem experiència al llarg de les nostres vides, i m'agradaria que el meu agent d'IA també se'n beneficiara. Si poguera incloure totes les converses que he tingut amb qualsevol persona en la meua vida, si poguera tindre tots els meus correus, totes les meues peticions cada vegada que faig una pregunta, seria genial”. Pérez Colomé tecleja en esta *newsletter*: “Queda poc per a preguntar-li: què vaig fer malament amb el meu primer ex, fill o faena? Què és millor que no faça una altra vegada? Segur que hi ha gent que voldrà recordar la vida amb eixe nivell de detall, com si això fora important”.

Llavors l'AGI serà una mena de nova religió a la qual encomanar-nos per tal que responga als nostres grans dubtes existencials? Miguel Arana ho imagina com un procés gradual en què anirem conquerint nivells i que “no és probable que passe d'una manera explosiva”, sinó com un procés gradual.



# Història

## Antecedents, teoria i història: Ada Lovelace, Turing, Altman

↳ «La màquina analítica no pot fer res sense les instruccions adequades, però pot fer molt si se li donen les instruccions adequades»

Ada Lovelace

↳ «Un ordinador pot ser anomenat intel·ligent si aconseguix enganyar una persona, fent-lo creure que és un humà»

Alan Turing

En realitat, sempre que es parla de qualsevol fita de la història de la tecnologia, la programació o les matemàtiques, el primer nom necessari que hauria d'aparèixer és Ada Lovelace. Ella va ser la investigadora britànica de principis del XIX que va escriure el primer algoritme destinat a ser processat per una màquina i la primera programadora d'ordinadors de la història, a més d'una gran visionària que va marcar els segles posteriors en predir que els ordinadors podrien anar més enllà dels simples càlculs numèrics: tot es tractava de donar-li les instruccions adequades.

Entre els experts que li van agafar el testimoni en el camp de la computació destaca, sens dubte, Alan Turing. "Poden les màquines pensar?", va ser una pregunta que va llançar Turing, el 1947, a un abarrotat auditori compost per membres del *National Physical Laboratory* de Londres. Este gran científic britànic nascut en 1912 va deixar un gran llegat tant en la pràctica computacional com en la defensa dels drets i llibertats sexuals. Turing també va tindre un paper decisiu en la Segona Guerra Mundial, ja que va aconseguir desxifrar els codis de les transmissions de missatges d'Enigma, la màquina utilitzada pels nazis per a ocultar les comunicacions de les seues maniobres bèl·liques. Però tot just uns anys després de la derrota del Tercer Reich, Turing va ser sotmès a un procés judicial per la seua condició d'homosexual, prohibida al Regne Unit en aquells anys.

El seu conegut test o prova de Turing definia una màquina com a sentent, intel·ligent o pensant quan fora una màquina que, en conversa amb un ésser humà, este no poguera distingir si estava parlant amb un aparell o amb un altre humà. Esta idea ha estat la base teòrica que ha anat esperonant la investigació en matèria d'intel·

ligència artificial, en una història plena d'alts i baixos, d'hiverns i primaveres, i que ha tingut algunes fites famoses i conegudes, com la pugna històrica entre l'ésser humà i les màquines que ell mateix ha creat. Va ser en 1997 quan Kaspàrov, el millor jugador d'escacs del món, va ser finalment derrotat<sup>36</sup> per Deep Blue, una supercomputadora d'IBM els processadors de la qual li permetien calcular milions de posicions per segon, cosa que no es coneix ni en la persona humana amb la ment més brillant del món.

La mateixa Gemini, de Google, si li preguntes sobre este fet, et respon, com a conclusió:

És un exemple del progrés de la tecnologia i de com les màquines poden desafiar i superar les nostres expectatives. És un recordatori que la intel·ligència artificial té el potencial de canviar el món de formes que encara no podem ni imaginar.

Esta frase ens podria recordar algunes de les que llança HAL 9000, l'ordinador assassí de 2001: *Odissea de l'espai* (1968): “*La humanitat és una amenaça per a si mateixa i per a l'univers. Jo soc l'única esperança per al futur*”.

Com recorda la periodista de *Newtral* Marilín Gonzalo, “la intel·ligència artificial naix després de molts anys de ciència-ficció sobre la intel·ligència artificial” i assegura que gran part de la comunitat de desenvolupadors sap que, encara que pugues ensenyar a fer coses a una màquina i que esta siga capaç de fer-les molt bé, estem molt lluny que una sofisticada xarxa neuronal artificial pugua crear, raonar o inventar alguna cosa nova. “Una altra qüestió és que el que hem creat ens meravelle i ens parega espectacular”. I la veritat és que programes com Sora, la IA generativa avançada per OpenAI per a la generació de vídeo, ho són. I també són una benzina perfecta per afegir al foc dels *deepfakes* i la desinformació, dels quals parlarem més avant.

Tornant a Turing, va ser ell qui va suggerir que, en lloc de construir un programa per a simular una ment adulta, seria millor produir-ne un de més simple que emulara la d'un xiquet i després sotmetre'l a educació. O entrenar-lo. El neurocientífic Mariano Sigman, autor del llibre *Artificial. La nueva inteligencia y el contorno de lo humano*, recull l'exemple del procés d'aprenentatge del llenguatge que fa una criatura quan comença a parlar, per pur aprenentatge o inferència estadística, diu. El xiquet escolta, escolta, escolta –com un model de llenguatge de gran escala– i va predient on va un substantiu o quina és la sintaxi correcta d'una frase a base de parar esment a les converses dels adults del seu voltant.

36

En febrer de 1996 Garri Kaspàrov, el millor jugador d'escacs del món des de l'any 1985, va disputar una partida d'escacs a sis partides. Deep Blue, un dels primers prototipus de supercomputadores creat per IBM, va guanyar la primera però l'escaquista soviètic va aconseguir refer-se'n i guanyar per 4-2. A l'any següent, una versió millorada de la màquina va demanar la revenja i va eixir-ne victoriosa.



# Tot el que necessites és atenció: neurones sintètiques, intel·ligència i màquines

La història de la IA està carregada d'expectatives i condemnes, lloes i crítiques, apòstols i apòstates. Però tot va començar molt a poc a poc, quasi com un joc acadèmic. En els seus inicis, la computació i la programació es veien com un sistema d'instruccions amb moltes ordes complexes per tal que l'ordinador executara accions més o menys senzilles, centrades a resoldre operacions matemàtiques. Això va ser fins que Turing va llançar l'associació entre intel·ligència i màquines en 1956, durant la conferència de Dartmouth, a New Hampshire, Estats Units, on deu científics homes nord-americans i blancs es van reunir per a tractar de desentranyar la possibilitat de materialitzar-la. En eixe esdeveniment va ser la primera vegada en què es va sentir el terme *intel·ligència artificial*, i va ser pronunciat per John McCarthy, un prominent informàtic guardonat amb el Premi Turing el 1971.

En els anys 50, quatre dècades abans de la victòria de Deep Blue, ja es pensava com crear màquines intel·ligents que pogueren jugar als escacs. Les preguntes ja s'havien llançat feia molt: com distingir una màquina automàtica que poguera jugar a escacs d'una calculadora? Encara faltava la revolució en la infraestructura computacional i, com sol passar en estos casos, l'arribada d'una idea potser menyspreada al principi, per com d'obvi i simple és el seu raonament.

La resposta aplegarà des del camp de la psicologia. Va coincidir en el temps amb l'apogeu d'esta disciplina i amb l'avanç en el coneixement biològic de la ment humana. Els psicòlegs Carl Rogers i Abraham Maslow, en la seua preocupació per entendre què era la bogeria humana, van suggerir un acostament empàtic i van proposar entreveure la bogeria com una part intrínseca de cadascun de nosaltres. L'empatia va ser clau per tal que, mig segle abans que es llançara ChatGPT, un professor d'informàtica del MIT, Joseph Weizenbaum, programara Eliza el 1966: el primer bot conversacional de la història. Encara que tenia un algorítme senzill i no sabia memoritzar, no entenia la ironia ni podia parlar de molts temes, sí que creava la sensació de conversador empàtic, amb bona escolta i interès per la persona amb qui parlava, a l'estil del que faria un bon psiquiatre. A partir de la dècada dels 60 va ser quan es va optar per canviar estos programes informàtics clàssics per unes unitats de còmput que es van anomenar *neurones* i que, juntes, conformaven xarxes neuronals amb eixa bioinspiració que buscava recrear el cervell humà.

Marvin Minsky –fundador de l'AI Lab del MIT, el prestigiós Institut de Tecnologia de Massachusetts– juntament amb altres científics com John Hopfield o Warren McCulloch, van canviar l'òptica usada fins al moment per a intentar fer brollar la intel·ligència d'un substrat que no és intel·ligent. Transmutaren les seues pretensions d'escriure la intel·ligència humana en un programa per buscar la forma de construir directament un cervell digital, i veure si este podia reproduir comportaments intel·ligents. La neurociència va aportar el que sabia sobre el funcionament de la ment humana: els seus 86.000 milions de neurones estan interconnectades entre elles en forma de xarxa. Una sola neurona té un comportament molt simple, però quan s'hi ajunten diversos milers, fins a milions, formen veritables xarxes neuronals que són les que permeten accions tan fascinants com l'aprenentatge. Cadascuna d'eixes neurones té una missió molt senzilla, per exemple, una detecta línies verticals en una imatge, una altra, els cercles. En el moment en què connecten i s'unixen un bon grapat d'elles, eixa xarxa neuronal és capaç de detectar què es un nas i què és una boca en una fotografia. Així funciona el cervell humà: com una estructura massivament paral·lela en què cadascuna



dels centenars de milers de neurones fa una cosa molt bàsica, però juntes, formant capes, són capaces de memoritzar, assimilar, associar, cada volta amb un major nivell d'abstracció o categorització.

Mariano Sigman, en el seu llibre, ampra, per a entendre de què parlem quan parlem d'intel·ligència, la frase atribuïda al psicòleg suís Jean Piaget que diu que "intel·ligència és saber què fer quan no saps què fer". Una planta movent-se en cerca de la llum del sol, o una decisió astuta d'un animal per a salvar la seua vida, són situacions a les quals l'humà pot atribuir la qualitat d'intelligent, encara que només siga això, una atribució. "Nosaltres som molt bons atribuint intencions que emulen la nostra pròpia experiència", explica Sigman a Marcos Vázquez en el seu pòdcast<sup>37</sup>. Som hàbils veient coses on no n'hi ha, o en la línia de les paraules abans esmentades de Will Douglas Heaven, editor de la secció d'IA de *MIT Technology Review*, si alguna cosa sembla intel·ligent és fàcil que ens enganye i pensem que realment ho és. Per a Sigman, els models extensos de llenguatge (LLM) estan molt a prop de passar el test de Turing però eixa no és la prova definitiva, no és suficient. La seua definició de la intel·ligència, per a un futur pròxim, seria que intel·ligència és tot allò que una màquina no pot fer.

Per a fer rodar eixa espècie de cervells sintètics també calgué una xicoteta revolució en el camp de la infraestructura. Una barreja de sabers, l'extracció d'una idea que no anava encaminada a eixe propòsit, però que, com alguns efectes secundaris beneficiosos per a altres patologies que es descobriren per sorpresa en els medicaments, va donar llum al material necessari per a suportar els cervells sintètics de les màquines.

Fins a aquell moment, el component principal dels ordinadors era la Central Processing Unit (CPU), responsable d'executar instruccions i fer els càlculs necessaris perquè els programes i sistemes operatius funcionen. Estes CPU tenien unes marcades limitacions, com ara la relativament lenta velocitat de processament de la informació, el reduït nombre de nuclis, la capacitat de memòria limitada, etc. Estes restriccions es van resoldre tecnològicament en la indústria dels videojocs amb l'arribada de l'empresa Nvidia el 1993 i les seues avançades unitats de processament gràfic o *Graphics Processing Unit*, en anglés, GPU. Estes avanços quant a microxips i targetes de processament eren una de les peces que mancaven a la indústria de la intel·ligència artificial per a avançar i eixir dels seus prolongats hiverns. Gràcies a l'augment del poder de còmput del *maquinari*, va ser possible afegir cada vegada més quantitat de capes neuronals intermèdies que van donar lloc a un nou tipus de xarxa neuronal coneguda com a aprenentatge profund o *deep learning*, i que constituïx el subcamp de *machine learning* o aprenentatge automàtic més important en l'ecosistema de la intel·ligència artificial.

Va ser en 2015 quan la companyia DeepMind, un any després de ser adquirida per Google, va desenvolupar una IA entrenada sobre la base de milions de partides humanes d'un popular joc de taula xinés anomenat Go, a què els humans juguen des de fa vora tres mil·lennis. AlphaGo, que així es deia l'ordinador, va guanyar de manera sorprenent el màxim campió del moment, amb una jugada revolucionària i inesperada. "Ens va regalar una idea nova que no se li havia acudit a cap jugador en els 3.000 anys d'història del joc. Les màquines, per primera volta, pareixien llestes per a superar-nos fins i tot en l'esfera més humana: la creativitat", escriu Sigman. En el 2017, AlphaGo va presentar la seua versió millorada, AlphaZero, i va aprendre a jugar tant als escacs com al Go sense que ningú li ensenyara. "La màquina comença jugant amb una còpia de si mateixa i la clau és que només se li permet a una de les dos revisar el seu model de joc. Després de milers de partides, la que pot aprendre comença a vèncer l'altra. En eixe moment es fa un nou clon de la versió millorada i es repetix el procés. Així (...) un programa jugant contra si mateix genera un coneixement exponencial", tecleja Sigman.



També en 2015 Sam Altman, Elon Musk i altres socis van fundar el que era aleshores una organització sense ànim de lucre anomenada OpenAI. Quatre anys després arribaria GPT-2, *Generative Pre-training Transformer 2*, un gran model de llenguatge desenvolupat per esta companyia que utilitza l'arquitectura de xarxa neuronal anomenada *transformer*, capaç de processar i comprendre seqüències llargues de text. Eixes onze lletres que sonen a pel·lícula de Marvel i, sobretot, la idea senzilla que hi ha darrere, són la clau, la cirereta del pastís de ChatGPT.

Van ser huit científics que treballaven per a Google els qui, en 2017, van publicar un dels papers fundacionals al ritme de cançó dels Beatles: *Attention Is All You Need*<sup>38</sup>, n'era el títol. Hi presentaven el que anomenaven “una nova arquitectura de xarxa senzilla, el *transformer*, basada únicament en mecanismes d'atenció, prescindint per complet de xarxes neuronals recurrents (RNN) i xarxes neuronals convolucionals (CNN)”. Estes arquitectures per a l'aprenentatge profund, els *transformers*, ajuden a identificar quines paraules o parts del text són importants a l'hora de processar una seqüència. A açò es referixen els *mecanismes d'atenció*: a cada paraula o part d'una frase li atorga un pes que indica quanta atenció li hauria de prestar la xarxa per a entendre la frase. Esta capacitat de sospesar la importància de les paraules en una oració va canviar la manera de processar el llenguatge natural (NLP en anglés), la traducció automàtica i l'anàlisi de sentiments. Els *transformers* són el ciment dels Large Language Models (LLM). Este exemple de ChatGPT ens aproxima més a entendre'ls:

Imagina que tens una frase: “El gat està en el jardí” (RNN). Si vols que un model genere la següent part de la frase, el *transformer* processa la frase paraula per paraula, però no en orde seqüencial estricta, com ho faria una xarxa neuronal recurrent (RNN). En canvi, empra el ‘mecanisme d'atenció’ per a mirar tota la frase en conjunt i entendre quines parts són més importants per a predir la paraula següent.

Cal tindre en compte que abans que apareguera l'aprenentatge profund o *deep learning*, les tècniques tradicionals de processament del llenguatge natural presentaven dificultats serioses a l'hora que les màquines compregueren el context i els matisos i subtileses del llenguatge humà, ja que no és una tasca fàcil. Les xarxes neurals extenses, junt amb l'arquitectura del *transformer*, han fet que siga una realitat. I tot això ha contribuït a fer que esta xarxa neuronal, entrenada en diferents fases, tant de manera supervisada com autosupervisada amb un gran *corpus* d'informació, siga capaç d'entendre la sintaxi, la semàntica i el context, i puga, a més, generar nou text gràcies a la gran comprensió del llenguatge que té.

Per processament de llenguatge natural (NLP) entenem un subcamp de la intel·ligència artificial que se centra en la interacció entre ordinadors i humans a través del llenguatge natural. El NLP consta de dos components principals: la comprensió del llenguatge natural o NLU (traducció d'idiomes, anàlisi de sentiments o resposta a preguntes) i la generació del llenguatge natural o NLG (la creació de frases des de zero). Encara que es van començar estudiant per a la traducció automàtica, els investigadors de seguida van veure el potencial que podria tindre en altres àrees relacionades amb el que hui coneixem com la IA generativa multimodal<sup>39</sup> –omnimodal, com diria OpenAI– i els *Large Multimodal Models* (LMM) que la fan possible i que acaben d'arribar per a substituir els Large Language Models (LLM).

38 Vaswani, S.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; N. Gomez, A.; Kaiser, L.; Polosukhin, I. (2016). *Attention Is All You Need*. Google. <https://arxiv.org/abs/1706.03762>

39 Se'n diu *multimodal* perquè combinen diversos modes d'informació i poden generar contingut nou (text, imatges, o àudio) a partir de múltiples tipus de dades o entrades. Un exemple seria crear imatges a partir de text.

Un altre dels conceptes clau en esta història són els anomenats *models fundacionals*, un terme popularitzat<sup>39</sup> l'agost de 2021 pel Centre de Recerca de Models Fonamentals (CRFM) de l'Institut d'Intel·ligència Artificial Centrada en l'ésser humà (HAI) de la Universitat de Stanford. Els *foundation models* fan referència a “qualsevol model entrenat amb dades àmplies, generalment mitjançant autosupervisió a escala, que es pot adaptar a una àmplia gamma de tasques posteriors”. Estos científics van saber veure el canvi de paradigma que s'aproximava i enumeraren les seues oportunitats, però també els seus riscos i impacte social:

“Encara que els models fundacionals es basen en l'aprenentatge profund i l'aprenentatge per transferència, la seua escala dona lloc a noves capacitats emergents, i la seua efectivitat en tantes tasques incentiva l'homogeneïtzació. Esta homogeneïtzació oferix un gran potencial, però requerix precaució, ja que els defectes del model fundacional s'hereten en tots els models adaptats a partir d'eixe model. Malgrat el seu imminent desplegament generalitzat, actualment no tenim una comprensió clara de com funcionen, quan fallen o de què són capaços a causa de les seues propietats emergents.”

Ja en el 2022 va ser quan va irrompre en les nostres vides ChatGPT-3.5 i, des de llavors, assistim a una cascada d'aplicacions i actualitzacions cada volta millors, més avançades, més intel·ligents i amb processos d'ensinistrament més sofisticats. Les possibilitats d'ús s'amplien. Per exemple, pots utilitzar un model preentrenat tal com està, sense fer cap ajustament addicional (*out of the box*) o bé optar pel *fine-tuning* o ajustament fi, el procés de prendre un model d'IA preentrenat i ajustar els seus paràmetres amb un nou conjunt de dades concret per tal que ens ajude en una faena concreta. És la personalització o l'adaptació d'un model genèric a un problema específic.

40

Més de 50 expertes i experts van signar en agost de 2021 un paper titulat *On the Opportunities and Risks of Foundation Models* de l'*Stanford Institute for Human-Centered Artificial Intelligence (HAI)* del Centre de Recerca en Models Fundacionals (CRFM) de l'*Stanford Institute for Human-Centered Artificial Intelligence (HAI)* <https://arxiv.org/abs/2108.07258>



# IA i periodisme

## Radiografia del sistema actual: glòria i decadència, èxits i acomiadaments, i la reinvenció constant de les redaccions

*Sports Illustrated*, una revista setmanal dedicada al món de l'esport, va ser adquirida el 2021 per Authentic Brands Group. La compra va suposar l'acomiadament de 40 persones del grup, més d'un terç dels seus periodistes. El seu cas va ser sonat perquè se'ls va acusar de substituir els redactors despatxats per periodistes ficticis creats amb intel·ligència artificial: una foto de perfil, una descripció de la persona i uns articles signats per eixa persona que no existix. I tot això, d'amagat.

No és difícil pensar que internet s'està omplint de contingut automatitzat i potser ja en som consumidors *de facto*, potser sense saber-ho. L'*enshittification* de la xarxa és un fet des de fa almenys dos anys, que va ser quan l'escriptor Cory Doctorow va popularitzar este terme per a fer referència a la degradació de la qualitat dels productes i serveis en línia. Va ser triada Paraula de l'Any el 2023 per l'American Dialect Society i de tant de repetir-la, cada volta és més real i està més estesa. *L'enshittification arriba a absolutament tot*, va titular un article<sup>41</sup> Doctorow: "Tots estem vivint un

QUI ACCIONA LA MÀQUINA?



gran *enshitting*, en què els serveis que ens importen, en els que confiem, s'estan convertint en gegantins muntons de merda. És frustrant. És desmoralitzador. Fins i tot aterridor". Però l'*enshittificació* anomena el problema i proposa una solució, segons el bloguer i autor de ciència-ficció canadenc.

"Durant molt de temps s'ha minimitzat el valor de la signatura, s'ha castigat la presència de marca personal amb mitjans en què abunden continguts signats simplement per *Redacció*", escriu Mauricio Cabrera, periodista mexicà que es defineix com un *terapeuta de continguts* en la seua *newsletter Story Baker*. "Per a recuperar un poc de la confiança perduda cal emfatitzar la figura de l'humà que està darrere de la creació de contingut. Així els mitjans no només podran presumir que tenen redactors humans en lloc de robots, sinó també humans altament capacitats en què la gent puga confiar". Humans, diu, idealment ben pagats per a explicar bones històries. En declaracions a esta investigació, este emprenedor i expert en innovació en mitjans, ressalta la importància de trobar una veu i perspectiva pròpia per a saber diferenciar-se en el maremàgnum de la informació: "Com a periodistes ens van ensenyar a escriure amb una mateixa estructura: què, qui, com, on, quan, per què. És un estil automatitzat amb una manca de segell distintiu fàcilment substituïble per la IA". El periodista ha de reunir qualitats com "memorable, autèntic" per "compartir i manifestar-se com a creador", a l'alçada d'un poeta o un cineasta.

"La classe mitjana del periodisme acabarà patint moltíssim", vaticina Cabrera davant esta amenaça "colossal" per a l'exercici periodístic. L'adjectiu el posa Antonio Lorenzo, redactor d'*El Economista* especialitzat en tecnologia i corresponsal d'este mitjà en el metavers. "Fins que va arribar ChatGPT, tots els discursos sobre que la IA ens robaria la faena se centraven en coses molt més mecàniques, com que si els robots ens alliberaran de doblegar la roba i el que ha passat és que, de sobte, ha aparegut una cosa que fa trontollar esta sensació de seguretat en segons quins llocs laborals. En el cas del periodisme, que és una professió que ja ve precaritzada des de fa molts anys, hi afegim una capa extra de pressió i precarització", i qui parla ara és Judith Membrives, tècnica de polítiques digitals de Lafede.cat i gran experta en l'*AI Act*, la legislació europea sobre intel·ligència artificial: "El gran impacte és la precarització i no el robatori de llocs de treball", sentència. En este punt coincidixen la majoria d'expertes del món del periodisme i la comunicació consultades.

Què opinen sobre això els mateixos responsables dels mitjans? Per a Juan Candela, cap de Digital del diari *Las Provincias*, el fantasma dels acomiadaments és només una "paranoia": "En una redacció tots fan de tot, no només tasques automatitzables". Potser este és el cas del seu mitjà, *Las Provincias*, però no hi ha un model únic. En la redacció madrilenya del diari *Público*, per exemple, hi ha un equip d'entre 6 i 9 persones que es dediquen únicament i exclusiva al que es coneix com a *Última hora*, és a dir, a pujar teletips d'agència a la web del diari. Treballen en torns rotatius, més o menys de 8 del matí a 12 de la nit, i tenen un director. Un bon aprofitament dels models de la IA generativa permetria reduir este equip i tindre més redactors disponibles per a treballar en temes propis que aporten riquesa al diari. Un ús poc ètic implicaria una reducció de plantilla.

Sobre la utilització que es fa de les arquitectures generatives en *Las Provincias*, Candela afluix a la "gran quantitat d'allucinacions, quasi un 40 per cent" de ChatGPT-3 per a argüir que, encara que el seu mitjà ha investigat amb elles des de l'inici, utilitzar-les implicaria "destrossar el periodisme". Amb ChatGPT-4 sí que albira cert salt qualitatiu, "però encara li queda molt recorregut, tot i que estem fent proves constantment". Els tres mitjans valencians consultats coincidixen que la creació d'un xatbot particular, entrenat –*fine-tuning*– amb dades i continguts del seu propi mitjà, seria una bona ajuda al periodista per a aportar context, història, referències i xifres de manera més senzilla, simplement formulant les preguntes adequades i facilitant notablement la tasca de recerca.



Per a Candela, el focus del debat és un altre: “El que és perillós per a la faena dels periodistes és Google i Facebook. Això és el que ens està exterminant a tots absolutament. Que de cada 10 euros en publicitat que guanyen, 8 els guanyen sense tindre gens de respecte per la privadesa. Eixos són els que han exterminat els periodistes i els mitjans de comunicació i no la intel·ligència artificial en principi, que tindrà la seua repercussió, però hui dia tots els problemes que tenen els mitjans de comunicació ho són exclusivament per eixos dos actors”.

Íñigo Roy, responsable de Canals del *Levante-EMV*, recalca que una penalització de Google pot suposar que caiga el trànsit de la web del *Levante-EMV* fins a un 70%, cosa que suposa pèrdues quantioses de diners: “Els mitjans de comunicació tenim una cosa per damunt de tot, que és la nostra reputació en internet, i esta és de dos maneres: la reputació editorial i després la reputació davant de Google. Si gastes de manera indiscriminada la IA i descuides aspectes com la veracitat, corres el risc de perdre el prestigi davant dels teus lectors. Però més important que això són les conseqüències que pot tindre, perquè Google, després de l’actualització del seu darrer *core update*, està penalitzant els mitjans que fan un ús abusiu de contingut a través d’IA sense tutela humana”. Cal preguntar si el castic s’estén també en el cas que el contingut haja estat creat pel seu xatbot, Gemini. Davant la qüestió de per què ve esta penalització de Google, Roy respon: “Perquè volen donar més valor a la informació de qualitat per tal que el trànsit siga de qualitat, qualitatiu i no quantitatiu”. Una qualitat posada seriosament en tela de juí sobretot des dels darrers cinc anys, ja que el gegant de les cerques representa un clar ús de l’*enshittification*.

## La sentència de mort del gran cercador

Hui coneixem que el 2019 els directius de Google ja temien la caiguda del seu monopoli quasi absolut i van començar a prioritzar el negoci abans que la qualitat i l’eficiència<sup>42</sup>. Açò és, a més cerques, més visites, més publicitat, més diners. I si per tal que hi haja més cerques cal entorpir la tasca de l’algoritme i incrementar els seus continguts patrocinats o llocs web *spam*, es fa. Com diu Marta Peirano, Google cada vegada llança més a les seues cerques continguts patrocinats, afiliats, sintètics o, directament, desinformació. L’*enshittification* de Cory Doctorow sempre està *justificada* des de les grans corporacions: tot val excepte perdre diners. Alphabet, l’empresa matriu que engloba el conglomerat Google, “és prou gran per a sobreviure a la desaparició del seu primer gran negoci. Però cal un model capaç de garantir la supervivència del periodisme de veritat”, sentència Peirano<sup>43</sup>.

La decisió de prioritzar la quantitat sobre la qualitat, presa per la companyia que domina i controla una part tan gran dels nostres dispositius –i de la nostra vida: del *mail* al calendari, del mòbil al document de treball, del mapa a les nostres hores de son–, en un context en què manté, des del 2014, més del 90% de quota de mercat dels cercadors, no pareixia important. Però des de la irrupció d’enormes

42 Pérez Colomé, J. (2024, 3 de maig). *El día en que Google empezó a empeorar: “Estamos acercándonos demasiado al dinero”*. El País. <https://elpais.com/tecnologia/2024-05-03/el-dia-en-que-google-empezo-a-empeorar-estamos-acercandonos-demasiado-al-dinero.html>

43 Peirano, M. (2024, 18 de marzo). *Quién mató al buscador*. El País. <https://elpais.com/opinion/2024-03-18/quien-mato-al-buscador.html>

models de llenguatge conversacionals com ChatGPT, Copilot o Perplexity, este lideratge s'està posant cada vegada més en dubte. Els competidors que li han aparegut estan molt a prop de ser millors i més utilitzats. Perplexity, per exemple, actua com un motor de cerca que ordena i redacta la informació d'internet, ofereix, segons diu, "informació concisa, directa i sense anuncis patrocinats, evitant la necessitat que els usuaris facen clic a múltiples enllaços per obtenir la informació". El seu creador, Aravind Srinivas, el definix així: "Perplexity és com una unió de Wikipedia i ChatGPT que pot respondre instantàniament a les preguntes d'una persona sense l'enorme desordre dels resultats de busca convencionals de Google". Les revistes *Forbes* i *Wired* ja han llançat acusacions de plagi cap a esta plataforma per haver-los copiat material original i oferir descaradament com a fonts primàries, articles d'altres mitjans que van replicar les publicacions de *Wired* i *Forbes* sense citar-los<sup>44</sup>. El CEO es defensa dient que Perplexity no s'entrena amb el contingut de ningú perquè la seua comesa és l'agregació de contingut a partir del que generen altres motors.

Estos motors de cerca conversacionals contribuïxen més encara al fet que desapareguen, o almenys es reduïsquen, les visites al mitjà original. Així mor el trànsit, del qual tant depenen hui dia els mitjans per a subsistir. Sobretot aquells que no aconseguixen fidelitzar el lector i, especialment, fer que pague pels seus continguts amb una subscripció al mur de pagament.

Mauricio Cabrera augura a Google Search una mort no tan llunyana: "No importa que els mitjans pretenguen aferrar-se al fet que el consum de contingut es produïska a través de busques a Google. La pràctica de trobar sumaris d'informació en un esquema conversacional s'acabarà imposant". De fet, motors de cerca com Perplexity s'estan emportant una part significativa del pastís: 85 milions d'usuaris mensuals en total segons Cabrera, tot i que encara li queda molt per arribar als 2.000 milions de Google.

Cabrera defensa que "els mitjans ja estan canviant, no només per la IA sinó per l'enèsima evidència que quantitat per damunt de qualitat no funciona, perquè no et porta a fidelitzar una audiència. Els mitjans de nova generació s'han de centrar a crear comunitats en lloc d'audiència". Per al periodista, la premsa hauria de pensar en termes de desenvolupament de producte que vagen més enllà del contingut purament escrit i anima que la IA es desenvolupe no només per a faenes grolleres o bàsiques, sinó també per a permetre "processos de recerca i desenvolupament periodístics que abans hagueren requerit moltíssima inversió o un seguiment que necessita pressupostos que no té ni tan sols el *New York Times*". I ofereix un consell al sector:

"Tenim una capacitat de poder generar alteracions en el sistema a partir del nostre criteri i crec que és allà on el periodista, vist com un autor, com una marca personal, té la capacitat de generar productes que produïsquen una disrupció. És a dir, la intel·ligència artificial ens donarà més del mateix tret que sapiguem sol·licitar, que sapiguem donar els prompts adequats. És ací on nosaltres podem diferenciar-nos a través d'una veu, de l'autenticitat, de l'estil, a través de les connexions o de les inspiracions i referències que tinguem i que ja no estiguen instal·lades en el sistema".

44

Sullivan, M. (2024, 21 de juny). *Perplexity CEO Aravind Srinivas responds to plagiarism and infringement accusations*. Fast Company. <https://www.fastcompany.com/91144894/perplexity-ai-ceo-aravind-srinivas-on-plagiarism-accusations>



# El cas dels mitjans valencians i l'exemple de les corporacions públiques

RTVE va anunciar que cobriria amb ajuda de la IA les eleccions al Parlament de Catalunya del 14 de maig del 2024, a través d'un conveni amb l'empresa Narrativa. Va ser un suport a la tasca periodística en forma de text automàtic en llenguatge natural per a donar cobertura als resultats en tot el territori català, en un procés que, com van destacar en un comunicat<sup>45</sup>, va estar supervisat i entrenat per professionals de RTVE “per a certificar i verificar-ne la qualitat informativa, editorial i visual”. Perquè també estava prevista la generació d'imatges i veus sintètiques “que informen dels resultats obtinguts tant en espanyol com en català, cosa que permet una cobertura més àmplia i equitativa del territori on se celebren els comicis, tancant bretxes territorials i informatives”. Esta és la seua arma per lluitar contra la desinformació i la manipulació informativa: publicar la major quantitat possible de notícies amb rapidesa i precisió, diuen, sempre basades en dades oficials.

No és la primera vegada. A les Generals i Municipals del 2023 ja es va implementar alguna cosa similar per a generar notícies dels resultats electorals en els prop de 5.000 municipis d'Espanya que tenen menys de 1.000 habitants. Segons la corporació, van ser més de 76.000 les publicacions en text produïdes amb esta firma: *‘El contingut d'esta notícia ha estat generat automàticament mitjançant intel·ligència artificial a partir de les dades proporcionades pel Ministeri de l'Interior del Govern d'Espanya’*.

La televisió autonòmica À Punt Mèdia segueix l'estela de l'ens públic estatal. És el mitjà valencià que fa un dels usos més intel·ligents dels sistemes d'aprenentatge automàtic, en part gràcies a un conveni, signat el 2020 i pendent de renovació mentre es redacta esta investigació, amb la Universitat Politècnica de València per a la subtitulació assistida en temps real de continguts audiovisuals.

“Nosaltres fugim del concepte IA. El que fem són processos d'aprenentatge o *machine learning*”, redefinix Pau Peiró Sendra, responsable per part de la Societat Anònima de Mitjans de Comunicació (SAMC) del projecte *Investigació i desenvolupament de sistemes intel·ligents avançats per a l'accessibilitat, catalogació i indexació de continguts textuals, radiofònics i audiovisuals*, que desenrotlla al costat de la UPV des del 2022. Treballen amb sistemes algorítmics de reconeixement automàtic de veu (ASR), és a dir, de transcripció del llenguatge parlat a l'escrit, per tal que la màquina, per exemple, sàpiga processar adequadament els registres sonors de la llengua, i també amb aprenentatge profund o *deep learning* per a millorar-ne la creació del corpus de la llengua valenciana. La taxa d'encert és del 86% i la de latència, inferior als 0,7.

El pas del reconeixement del llenguatge a l'aprenentatge profund té uns avantatges clars per al projecte; per exemple, els permet aplicar capes correctives i corregir les errades orals per a subtitular-les amb la paraula o expressió correcta: “Volem que siga ‘subtitulació assistida, encara que estem tan acostumats al fet que siga ‘automàtica’. Som conscients que la llengua és viva i sempre s'ha d'anar actualitzant”. Peiró recalca que esta tecnologia és replicable i de codi obert; de fet, ja hi ha hagut contactes amb altres televisions autonòmiques.

De moment, les persones expertes consultades reconeixen que, malgrat que la intel·ligència artificial en si té moltes oportunitats, des de l'ens autonòmic no s'està apostant per la generativa i sí per millorar l'accessibilitat dels continguts i oferir

45

RTVE prepara la incorporació de la IA a la labor periodística en la cobertura de nuevos procesos electorales, publicat al web de RTVE.es el 6 de març de 2024. <https://www.rtve.es/rtve/20240306/rtve-prepara-cobertura-nuevos-procesos-electorales-ia/16001716.shtml>



eines que ajuden a la redacció a ser més productiva, eficient i, sobretot, més accessible. Un factor a tindre en compte és que la implementació de tecnologies noves a la televisió pública sol ser un procés lent, amb molta part administrativa i burocràtica, sense oblidar la difícil qüestió de quadrar les partides pressupostàries destinades a renovacions tecnològiques, que moltes vegades es destinen a condicionar els estudis, sales d'enregistrament i platós. Segons Peiró, més enllà de la IA generativa, en À Punt s'empren almenys 14 eines que utilitzen la intel·ligència artificial per a millorar tasques relacionades, per exemple, amb la composició d'imatge i vídeo.

Entre els projectes desitjables en un futur pròxim, Peiró aposta per implementar algun dia “models entrenats per nosaltres mateixos” per a poder tindre'n el màxim control possible. Que siguin “com un miniChatGPT alimentat pels continguts de la redacció i que servisca d'ajuda als periodistes que, per exemple, arriben a una secció per primera vegada. Seria tindre una mena d'assistent que done tots els aspectes i connotacions bàsiques de com hem tractat fins ara una temàtica concreta de notícies”. Un model *finetunejat* d'acord amb les característiques i les necessitats de la redacció d'À Punt.

“No hi ha dubte que el *machine learning* agilitza molt els processos interns, però també és cert que per a poder fer-ne ús, has de tindre un coneixement previ. Això és el que crec que no s'està explicant a la societat. El que ens diuen és: tens esta *app* al mòbil, pots preguntar-li el que vullgues, que és un cunyat perfecte. T'ho dirà tot, però en mode cunyat. Has de tindre el coneixement de com usar-lo, i no només com fer els *prompts*, sinó saber el que hi ha baix”, clou.

## El periodisme local en À Punt i les veus sintètiques

Javier Lifante, cap de l'àrea digital d'À Punt, enumera les ferramentes que en este moment fa servir la redacció i que compten amb Intel·ligència Artificial: “Estem gastant-ne alguna que ens faciliten alguns proveïdors, per exemple TikTok, que l'usa per eliminar un fons. Abans ho feies amb Premiere o Photoshop i ara és molt més ràpid. També tenim aplicacions com Metricool per a publicar en xarxes socials. És veritat que ho estem fent de manera molt experimental encara, però bàsicament estes són les dos que tenim més avançades”. En la redacció d'Informatius, les eines d'intel·ligència artificial emprades per a transcripcions i traduccions són una ajuda valuosa que estalvia temps a les i els redactors. Traducció de teletips d'agència al valencià i transcripció d'entrevistes, per exemple, per a elaborar continguts per a la web o per a minutar materials gravats, serien tasques que es podrien realitzar fins i tot des del mateix CMS o sistema de gestió de continguts de l'ens autonòmic si este incloguera eixes opcions: “És una manera d'estalviar-te faena, després tu només has de canviar la foto, millorar l'entrada o el titular. El periodista explica què ha passat, explica com t'afectarà a tu des del punt de vista local, ací és on aporta valor”.

En la part de xarxes socials que coordina Lifante, les tècniques automatitzades troben nombroses utilitats a l'hora de crear, generar i adaptar continguts a les diferents plataformes. Des de retallar imatges d'un programa de televisió per a adequar-les i que puguen ser transmises a Twitter o Facebook fins a dissenyar creativitats de manera ràpida per tal d'anunciar un programa a Instagram.

Iván Esteve, Director d'Informatius i Esports d'À Punt Mèdia, té clar que, en les qüestions creatives, no se substituirà el treball del periodista i que estem davant



d'una oportunitat, diu, per a millorar les rutines de treball i els processos d'elaboració de les notícies: “Per a una redacció com la nostra, la intel·ligència artificial, de moment, és més una oportunitat que un perill per als llocs de treball. Una oportunitat perquè a mesura que vagen apareixent ferramentes que faciliten la faena dels periodistes i, a mesura que els periodistes es formen, faran millor el seu treball”.

Els preceptes i exigències d'un mitjà públic no són –o no tant– els mateixos que en un de privat: “Els nostres criteris i els nostres valors són diferents dels de qualsevol mitjà de comunicació de titularitat privada, que pot ser que hi vegen una oportunitat per estalviar-se costos o millorar beneficis. Nosaltres ens regim per altres paràmetres que tenen més a veure amb el servei públic i amb la informació de proximitat”.

Parlar sobre possibles acomiadaments per substitució en una televisió autonòmica en què, hui, s'ha suspès la convocatòria d'ocupació pública és una qüestió complexa i fins i tot estranya. Tot i així, Esteve aprecia: “En principi no crec que passe [comiats per substitució]. Sí que hi ha algunes categories professionals lligades a la redacció que fan treballs fàcilment automatitzables, que és possible que siguen on més afecte, com ara les persones que estan a l'arxiu documentant i catalogant imatges. És un treball molt mecànic que, probablement, la intel·ligència artificial amb reconeixement d'imatges i rostres podrà fer en un futur prou immediat”.

Les televisions, ràdios i diaris ja han vist extingir moltes professions lligades a l'ofici periodístic: muntadors, realitzadors, maquetadors, són llocs de treball que s'han transformat en els darrers 30 anys. Bé perquè les seues tasques han estat substituïdes per algorismes o bé perquè és el mateix periodista qui ha assumit esta faena<sup>46</sup>. I continua Esteve: “A mi em costa imaginar algú escriure un reportatge de recerca amb ChatGPT. Ací tenim l'avantatge que els nostres continguts són tots en valencià i encara no hi ha este tipus de ferramentes generalitzades que puguen redactar o crear contingut en valencià. En un futur, no sé si això arribarà a existir”.

En realitat, això ja és possible. ChatGPT puntualitza que és capaç de fer traduccions en més de 50 idiomes, encara que maneja millor els majoritaris:

Per a alguns llenguatges més especialitzats o dialectes locals, la meua capacitat pot ser més limitada en termes de precisió o matisos culturals.

“Amb el valencià tindrem una barrera. Al departament tecnològic han entés bé que havíem de ser nosaltres mateixos els qui apostàrem per crear este *corpus propi*”. La idea és integrar un botó en Dalet –el programa d'edició que s'usa tant a la tele com la ràdio– que oferisca una transcripció literal d'una roda de premsa, per exemple, i que incloga un cercador perquè pugues anar ràpidament al tall de veu que interessa al periodista.

De l'AI Act –l'única llei que fins ara regula la intel·ligència artificial i les seues pràctiques– no s'extrau una prohibició d'usar veus sintètiques als mitjans de comunicació, encara que se suggerix que, en tal cas, ho haurien d'indicar. En la ràdio, una altra de les tasques més mecàniques i fàcilment substituïbles, com advertix Esteve, són els butlletins informatius: “Tu podries dir a una intel·ligència artificial que t'interessen les notícies que hagen succeït en els darrers 60 minuts, que incloga alguna declaració

46

“Correctors, introductors de textos, auxiliars de redacció i maquetadors han anat desapareixent, a poc a poc, des de fa quinze anys als diaris i les seues tasques han estat assumides pels periodistes. En la ràdio, el redactor ha passat també a fer tasques de control tècnic i, fins i tot a les emissores xicotetes d'àmbit municipal, el periodista també és director, publicista i administratiu. En la televisió, els equips ENG (Electronic News Gathering) amb càmeres portàtils dissenyades per a treballar al muscle estan formats per un únic redactor que capta les imatges, escriu el text i edita la informació per a lliurar una peça ja preparada per a emetre. Esta pràctica també s'ha estés a les agències de notícies i els redactors d'EFE porten, a més, una càmera menuda amb què recullen les imatges de la notícia per a servir-la als abonats del servei”. Prieto Sánchez, C. (2013). Tesis Doctoral sobre *El movimiento asociativo de los periodistas: el caso del Sindicato de Periodistas de Andalucía (SPA)*. Departament de Periodisme, Facultat de Ciències de la Comunicació, Universitat de Màlaga. <https://riuma.uma.es/xmlui/bitstream/handle/10630/5964/Tesis%20Cristina%20Prieto.pdf>

política important que hi haja hagut, que miren a veure si hi ha algun succés o algun accident o que mire com està l'estat del trànsit o dels trens. Amb esta informació és molt fàcil elaborar un butlletí de notícies”. Fins i tot es podria clonar la veu del periodista estrella d'informatius per tal que la locució sonara com ell o ella. Només caldrien 15 segons d'un discurs oral seu i un programa com el Voice Engine d'OpenAI o l'aplicació d'Eleven Labs: “Probablement seria una veu que no tindria errors, que faria una pronunciació perfecta i sense *apitxar*. Crec que una de les riqueses dels mitjans públics és també que se senten veus diferents i maneres de parlar diferents dins d'una mateixa llengua. Per tant, a mi de moment esta idea no em seduïx. Segur que tècnicament es podrà fer d'ací a molt poc, però ara mateix ni ens plantegem substituir els periodistes dels butlletins ni és una prioritat per a nosaltres”, conclou Esteve.

El responsable d'informatius de l'ens autonòmic no veu els peus al llop i no creu que la intel·ligència artificial haja arribat per a traure o substituir la faena dels periodistes en el sistema públic: “Jo això ho descarto del tot. La innovació tecnològica ha passat sempre, els mitjans que tenim ara per a l'edició de les peces a la tele i a la ràdio no són els que hi havia fa 10 o 20 anys. Els periodistes hem de formar-nos, aniran apareixent ferramentes que ens facilitaran la nostra faena i ens ajudaran a millorar-la, però com s'ha fet des de sempre”.

Un altre dels projectes futurs que sí que compten amb l'acceptació de la direcció d'À Punt Mèdia és la creació d'un avatar que faça la traducció per a televisió a llengua de signes de forma simultània, encara que, de moment, només és un conjunt d'idees en fase inicial.

## L'excel·lència del *New York Times*: la denúncia a les tecnològiques per “utilitzar el contingut sense pagar”

A finals del 2023, la companyia editora del *New York Times* va demandar OpenAI i Microsoft i els va acusar de “còpia i ús il·legal de les obres de singular valor del Times”, amb les quals haurien entrenat els seus models d'IA generativa. Este treball d'ensenyar al seu famós model de llenguatge extens milers d'articles, la majoria amb cademat al mur de pagament, haurien generat a l'emblemàtic diari americà una despesa de “milers de milions de dòlars en danys i perjudicis legals i reals”. En la demanda es requereix a les companyies que destruïsquen qualsevol model de xatbot i dades d'entrenament que usen material amb drets d'autor del *Times*. “Els demandats busquen aprofitar-se de l'enorme inversió que ha fet el Times en el seu periodisme”, a més d’“emprar el seu contingut sense pagar per a crear productes que substituïxen el Times i li furten audiència”<sup>47</sup>.

En la denúncia s'inclouen 100 exemples concrets de textos llançats per ChatGPT que copien paraula per paraula algunes parts d'articles del diari. Esta especificació que siguen frases literals és important i és el que pot marcar la diferència, segons alguns analistes, ja que hi ha hagut diversos casos de denúncies que han estat

47

M. Grynbaum, M. i Mac, R. (2023, 27 de desembre). *The New York Times* demanda a OpenAI y Microsoft por el uso de obras con derechos de autor en la IA. *The New York Times*. <https://www.nytimes.com/es/2023/12/27/espanol/new-york-times-demanda-openai-microsoft.html>



desestimades, com la de la còmica i escriptora Sarah Silverman contra Meta i OpenAI<sup>48</sup>. En este cas, els jutges no van considerar que els textos que generen arquitectures neuronals sintètiques com LLAMA (Large Language Model Meta AI, hui conegut com a Meta Llama 3) o ChatGPT tinguen “similituds substancials” amb els seus llibres originals. Una tesi que defensa àmpliament el doctor en Informàtica Jordi Linares perquè, des del punt de vista tècnic, estos LLM no emmagatzemen textos, articles, imatges ni quadres pictòrics, sinó que creen representacions internes –anomenades *embeddings*– de les paraules i frases, per a comprendre les relacions semàntiques i sintàctiques.

Solament són capaços de guardar dades numèriques, conegudes pel nom de pesos. Estos són una peça fonamental per al *deep learning* que els permet determinar la importància de les connexions entre les neurones dins d'una xarxa. Els pesos són valors numèrics associats a diferents elements (com ara paraules dins d'una frase) que comencen sent valors aleatoris i, durant el procés d'entrenament, es van ajustant mitjançant algorismes i penalitzacions en un procés d'assaig-error amb què van afinant les prediccions. Així aprenen patrons, estructures i relacions entre paraules i conceptes a partir de grans quantitats de dades textuales.

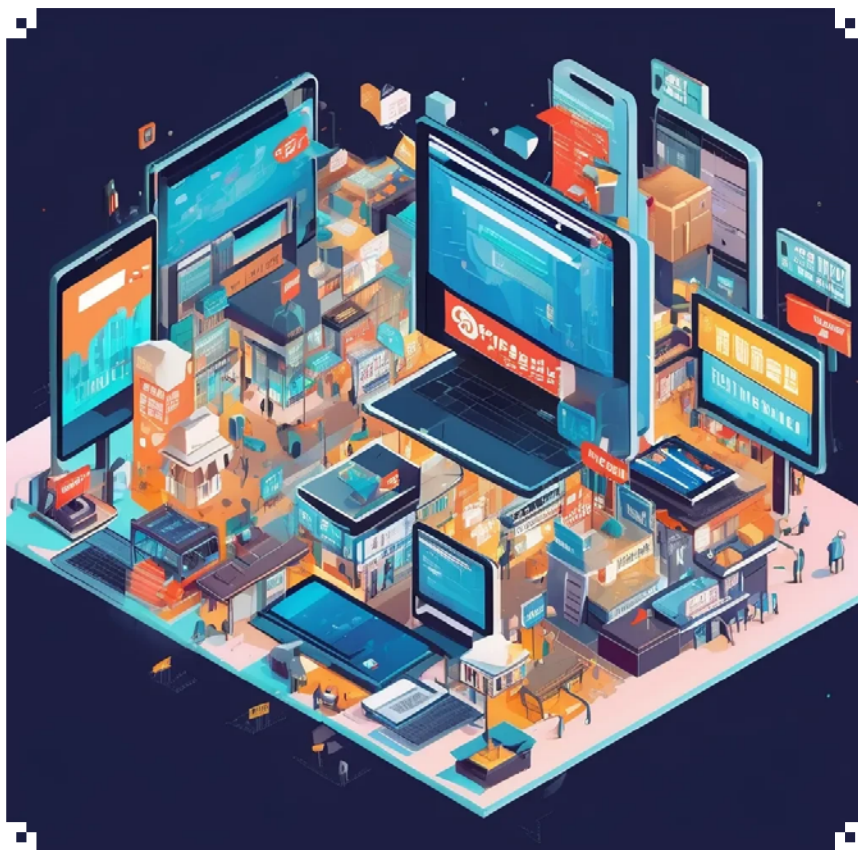
La sensació dels analistes consultats és que els tribunals decidixen en estos casos com en qualsevol altre juí sobre drets d'autor, i centren la seua atenció en si el material generat a través d'estos models infringix les lleis del *copyright*, però sense tindre en compte –almenys fins ara– allò amb què s'ha alimentat la màquina. És el debat de si la IA aprén o copia, com avançàvem unes línies abans: en el primer cas no es vulnera cap dret i en el segon, potser sí.

Què aporta a este debat la nova legislació europea sobre IA? Preguntem sobre això a Andrés Boix, professor de Dret Administratiu de la Universitat de València. El jurista diferencia el cas europeu de l'americà. Als Estats Units, més que una legislació sobre propietat intel·lectual el que tenen és una exposició jurisprudencial sobre allò que es pot fer i allò que no, i allò que es considera un ús just, o legítim. És el que es coneix com a *fair use*: “A Espanya i a la Unió Europea no hi ha cap restricció legal, en estos moments, que diga que no es puga entrenar la IA usant les coses que estan per ahí o a les que tenen accés. Per a poder restringir-ho caldria canviar la doctrina del *fair use* als Estats Units i legislar a Europa i Espanya. Mentre no se'n legisle, és perfectament possible entrenar-les així”. I insistix Boix: “No és que no hi haja *fair use* a Europa, el que passa és que la doctrina als Estats Units és de creació jurisprudencial i van decidint cas a cas sobre el *fair use*. A Europa es codifica a les directives europees i a les nostres lleis i, com dic, fins ara no hi ha cap prohibició d'ús d'obres a què tinguen accés perquè estan lliures, o que tenen drets d'autor i n'hagen comprat la llicència. Igual que una persona les consumeix, que les consumeixca la màquina”.

L'AI Act no és una llei sobre propietat intel·lectual, però sí que s'hi apliquen la resta de directives en matèria de drets d'autor de la Unió Europea. En el cas d'Espanya, la Llei de propietat intel·lectual: “La gran discussió és si l'ús que fan els models d'IA generativa, que s'entrenen amb continguts subjectes a drets d'autor, és un *fair use*”. Fins ara ha estat així”. Sobre el desenllaç final de la denúncia del *New York Times*, Boix es manté prudent: “De moment hi ha una demanda, però no hi ha cap resolució. Tu no pots deseducar la màquina, així que [en el cas que els jutges donaren la raó al diari] s'establiria una indemnització que pot ser més gran o més menuda, i que es pagaria a qui haja patit el mal”.

Una altra de les qüestions que preocupa el *Times* en la seua denúncia, a més que s'afusellen els seus reportatges sense citar-los, és que la utilització massiva d'estos models d'IA generativa perjudicaran l'estratègia comercial del diari digital, ja que si ChatGPT és capaç de generar notícies o de donar respostes sobre l'actualitat,

restarà visites a la pàgina web. Com s'apuntava més amunt, la viabilitat de la premsa en línia, novament, ha de ser repensada. Referent a açò, en una *newsletter* de *La Wikly*, Emilio Doménech i Bosco Bárcena expliquen que el famós columnista de tecnologia Ben Thompson recorda que el taló d'Aquil·les més gran dels xatbots és la falta de confiança que destil·len, perquè també fallen i molt. La informació fidedigna i de fiar sempre seguirà a la web del *Times*. Però, “quin percentatge de lectors ignoraran esta premissa fiant-se de ChatGPT i prescindiran de pagar per una subscripció per punxar un enllaç afiliat?”, es pregunten Doménech i Bosco. Potser el *New York Times* sobreviva a la baixada de trànsit web, però n'hi haurà molts, potser la majoria, que no ho faran.



Poca presentació necessita el *New York Times* perquè, sens dubte, la seua fama el precedix. És, des de fa dècades, un mitjà referent de rigorositat i qualitat periodística que ha marcat la forma de fer per a molts altres mitjans. Van ser pioners en qüestions hui ben escampades i popularitzades com el mur de pagament, que utilitzen des del 2011, i d'altres que no tots es poden permetre sense precaritzar la laboralitat dels seus empleats, com el seu enfocament, prioritat i pressupost dedicat al periodisme de recerca –que l'ha portat a guanyar 132 premis Pulitzer– o la seua extensa xarxa de corresponsals internacionals. Des de desembre del 2023 el *Times* va anunciar el fitxatge de Zach Seward com a director editorial d'iniciatives d'intel·ligència artificial. La seua comesa seria donar resposta a estes preguntes: *Com s'hauria de beneficiar el periodisme del Times de les tecnologies d'intel·ligència artificial generativa? Poden estes noves eines ajudar-nos a treballar més ràpid? On hauríem de traçar les línies roges entorn de les quals no les utilitzarem?* Davant l'avanç de les aplicacions amb IA, el diari reforça les bases que li atorguen la confiança del públic i reitera que “el periodisme del Times sempre serà informat, escrit i editat pels nostres periodistes experts”.

# Si hi ha diners, no hi ha plets

Com veiem, el diari novaiorqués ha estat el primer gran mitjà a optar per la via judicial, una opció que també van escollir alguns escriptors<sup>49</sup> com Jonathan Franzen i John Grisham, que van llançar la seua acusació sense palliatius: “L'èxit i la rendibilitat d'OpenAI es basen en la infracció massiva dels drets d'autor sense consentiment i sense pagar un cèntim de compensació als propietaris d'estos drets”. Les tecnològiques es defensen, entre el victimisme i l'amenaça, assegurant que exposar-les a la responsabilitat per drets d'autor “obstaculitzaria significativament el seu desenvolupament”, a més que limitaria l'expansió de la innovació i el lideratge dels Estats Units en el desenrotllament de la IA. El mateix Sam Altman ha reconegut sense pèls a la llengua que ara com ara és impossible entrenar els seus models sense infringir els drets d'autor.

La sonada denúncia del *Times* a OpenAI arriba després que les dos empreses no aconseguiren arribar a un acord econòmic, cosa que sí que va aconseguir la creadora de GPT amb Associated Press (AP). L'agència de notícies nord-americana ven la llicència d'ús de part del seu fitxer de text a la companyia de Sam Altman a canvi de poder implementar la tecnologia i les aplicacions d'OpenAI. El que no es va fer públic és l'acord econòmic assolit. Tot i que en el comunicat de les dos companyies no es parlava de diners, es dona per fet que hi ha un pagament. “Ningú sap com són estos acords”, lamenta Marilín Gonzalo, i critica l'opacitat amb què funcionen les grans tecnològiques, les traves que posen a la premsa i la poca transparència que mostren, malgrat que el missatge públic que llancen siga molt diferent.

Estem assistint a un degoteig constant de notícies sobre nous pactes amb la companyia tecnològica, en afegir el conglomerat alemany Axel Springer, el diari francès *Le Monde* i Prisa Media al seu elenc de mitjans a sou. Va ser pel març del 2024 quan vam conèixer esta col·laboració entre les publicacions del grup Prisa –*El País*, *Cinco Días*, *AS* o *El Huffpost*, entre d'altres– que permetrà “els usuaris de ChatGPT interactuar al voltant de l'actualitat amb els continguts d'alta qualitat de Prisa Media en espanyol i de *Le Monde* en francès”, segons es llegeix en la notícia publicada en *El País*: “Continguts que també contribuiran, alhora, a la formació i millora contínua dels models d'intel·ligència artificial”. Brad Lightcap, director d'operacions d'OpenAI, afirmava en un comunicat: “El nostre objectiu és permetre que els usuaris de ChatGPT de tot el món connecten amb les notícies de manera més interactiva i que tinguen una visió més completa de l'actualitat”. Segons la informació que publica *El País*, ChatGPT incorporarà resums de continguts d'actualitat publicats en estos mitjans i inclourà els enllaços als articles originals.

Quan s'escriuen estes línies, s'ha anunciat una nova aliança entre el *Financial Times* i l'OpenAI, uns acords bons per a les dos parts segons Mauricio Cabrera encara que amb un clar beneficiari, i és la companyia de Silicon Valley: “No només adquireix el dret a utilitzar de manera legal la propietat intel·lectual dels mitjans, sinó que es fa també amb grans segells de qualitat que validen els resultats editorials que presenta als seus usuaris. És una rentada de cara necessari per a qualsevol tecnològica”. El degoteig és constant, i també hi ha pacte entre OpenAI i News Corp, propietària del *Wall Street Journal*, *The New York Post*, *The Daily Telegraph* i *Herald Sun*.

A Cabrera estes últimes notícies li recorden la promesa de Mark Zuckerberg d'apostar per la informació per a les persones usuàries de Facebook en incloure pestanyes de notícies –i acords milionaris amb mitjans per enllaçar els seus continguts–. Una injecció extra que li va vindre molt bé a la premsa fins que el propietari de Meta va decidir posar fi a la festa del talonari l'abril del 2024 i donar prioritat a les

interaccions entre persones, eliminar Facebook News i, per tant, desubscriure estos suculents acords econòmics amb mitjans de comunicació majoritàriament britànics, americans, australians, francesos i alemanys. Així va començar el que Cabrera anomena “la caiguda de les grans icones mediàtiques de l’era Facebook”, amb *BuzzFeed*, *Vice* i *Cultura Colectiva* al capdavant.

“Amb semblant antecedent pel mig, la indústria (dels mitjans) té l’oportunitat d’actuar diferent davant d’un mateix escenari”. Però no ho està fent segons Cabrera, ja que tornen a optar per lliurar “les marques més valuoses a canvi d’ingressos significatius”. Segons estimacions de les publicacions *The Information* i *Tech Crunch*, OpenAI estaria desemborsant entre 4 i 20 milions de dòlars a l’any per les notícies. Una quantitat que pot –o no– semblar morralla per a una empresa que supera els 2.000 milions de dòlars en ingressos.

Andrés Boix és de l’opinió jurídica que als mitjans de comunicació a Europa els ix a compte pactar per l’ús dels continguts amb les grans tecnològiques que no denunciar-les. “Una altra cosa és que vullguen fer pressió al legislador per a veure si s’estableix algun tipus de restricció o prohibició que els donara més poder de negociació”. Però això només t’ho pots permetre si eres el *Times*.

## Les dades i la fi del petroli

D’entre els que han decidit agafar els diners i córrer destaca el cas de Reddit, que, segons algunes fonts<sup>50</sup>, rebrà 60 milions de dòlars a l’any de Google perquè este puga accedir a les seues dades i entrenar-hi els seus models d’IA. Esta plataforma web, que actua com una mena de fòrum en què els usuaris compartixen i comenten notícies i publicacions, és una de les pàgines més visitades del món. Són les persones que entren, precisament, les autèntiques creadores del contingut que ha comprat Google, però no veuran ni un dòlar.

Esta és la manera com s’ha anat construint internet, amb els continguts desinteressats de tots els seus participants posats al servei d’una causa més gran, que és la de generar riquesa als habitants de Silicon Valley. Ho vam veure també el 2018, quan Microsoft va comprar GitHub, el repositori més gran de codi obert d’internet. En aquella època, semblava que la xarxa, aquell macroespai virtual infinit, no ens l’acabaríem mai i ara, a estes architectures amb xarxes neuronals que simulen les nostres, se’ls ha quedat curta. Ja hi ha estudis<sup>51</sup> que auguren la data de l’apocalipsi de les dades: les d’alta qualitat, abans del 2026. Les de baixa qualitat duraran fins al 2050.

Les tecnològiques saben que han d’anar un pas avant. OpenAI va haver de tirar mà de la imaginació per a seguir entrenant el seu model GPT-4 i, per això, se li va ocórrer crear una altra IA que traduïra a text els vídeos de YouTube per a alimentar el seu ingent i flamant nou model de llenguatge, segons una investigació del *New York Times*<sup>52</sup>, amb la vulneració dels drets d’autor que això comporta. Es calcula que la famosa plataforma conté més de 150.000 milions de minuts de vídeo, però la major

50 Anna Tong, A; Wang, E y Coulter, M. (2024, 22 de febrer). *Exclusive: Reddit in AI content licensing deal with Google*. Reuters. <https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22>

51 Villalobos, P.; Sevilla, J.; Heim, L.; Besiroglu, T.; Hobbhahn, M.; Ho, A. *Will We Run Out of ML Data? Evidence From Projecting Dataset Size Trends* (2022, 10 de novembre). EPOCH AI. <https://epochai.org/blog/will-we-run-out-of-ml-data-evidence-from-projecting-dataset>

52 Metz, C; Kang, C; Frenkel, S; A. Thompson, S y Grant, N. (2024, 6 d’abril). *How Tech Giants Cut Corners to Harvest Data for A.I.* The New York Times. <https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html>



part del material té poc o cap àudio utilitzable, perquè és música, imatges fixes o seqüències de videojocs. Seguint la investigació del *Times*, Meta s'hauria plantejat la compra de l'editorial Simon & Schuster per a adquirir les seues obres, una opció pel que sembla descartada perquè els ix més rendible pagar multes i seguir recopilant dades protegides per drets d'autor en internet. “Negociar llicències amb editors, artistes, músics i la indústria periodística duria massa temps”, és el seu principal argument.

Una cosa semblant li està passant tant a Google com a Meta. De fet, els de Zuckerberg ja han anunciat a les persones amb compte a Instagram i Facebook que faran servir les seues publicacions, entrades, textos i peus de fotos per a educar la seua criatura artificial. A Europa esta decisió no ha caigut gens bé en les persones usuàries, que han exercit en massa el seu dret d'oposició. Meta posseïx dos dècades de dades de xarxes socials; Google va començar la seua carrera cap al cim a principis dels 2000 i ha anat creixent fins a arribar als 102,2 mil milions d'usuaris en tot el món hui en dia. Això són moltes dades i molt de material per a alimentar la IA, sobretot si inclou les dades de les versions gratuïtes de Google Docs, Google Sheets o Google Slides, cosa que depassaria certs límits de la privadesa. De fet, la companyia va eixir a aclarir<sup>53</sup> que només fa servir com a material d'entrenament aquells arxius de Google Docs que estan “disponibles públicament” a internet, és a dir, si han estat compartits a través de xarxes socials o publicats en llocs web.

Sembla que hi ha cert consens en la idea que, a llarg termini, no hi haurà prou blogs, articles noticiosos o comentaris en xarxes per a mantindre l'actual nivell de desenvolupament de la IA. Per això les opcions que es contempen per a este escenari són, com diem, utilitzar dades privades o directament text generat sintèticament per les mateixes ferramentes d'intel·ligència artificial. Totes dos tenen contrapartides tant ètiques i legals com de qualitat, sobretot en el cas que opten per contingut sintètic. Ja es parla que en este escenari es podria produir un col·lapse del sistema, i que passaria el mateix que passa quan fas una fotocòpia d'una fotografia original, i després seguixes fent còpies de la fotocòpia. Suposaria alimentar la màquina amb la pròpia *enshittification* que ella mateixa ha ajudat a generar.

L'escassetat d'este nou petroli que són les dades no és una realitat llunyana al nostre temps i espai. Alguns investigadors en diuen la nova “febra de l'or” i anticipen, com s'ha dit més amunt, que podria esclatar en els pròxims dos anys. Els acords econòmics de les tecnològiques amb els grans mitjans de comunicació són, sens dubte, una forma d'insuflar oxigen en esta carrera per saquejar internet i escalar els models, incrementar-ne les capacitats i amb això la qualitat de les respostes que ofereixen. Segons l'estudi abans esmentat<sup>54</sup>, la quantitat de dades de text amb què s'alimenta la intel·ligència artificial generativa s'ha multiplicat per 2,5 a l'any, mentre que la capacitat de càlcul s'ha quadruplicat. Només el model més recent i més gran de Meta, l'empresa matriu de Facebook, s'ha entrenat amb fins a 15 bilions de *tokens*<sup>55</sup>, que són la unitat bàsica de processament del text que utilitza la IA generativa.

Una de les reflexions més brillants sobre la relació entre l'escassetat de dades i el desenvolupament de la Intel·ligència Artificial General (AGI) la fan<sup>56</sup> els professors de ciència de la computació de la Universitat de Princeton Arvind Narayanan i Sayash Kapoor. Per als dos experts, el camí actual no ens conduirà on pretenem

53 Europa Press. (2024, 11 d'abril). *Google asegura que solo usa archivos de Docs para entrenar su IA si están "disponibles públicamente" en Internet*.

54 Vegeu la cita 51.

55 Esta és la definició de *token* que dona ChatGPT: En el context de la IA generativa, un *token* és una xicoteta unitat de text que processen els models d'IA. Pot ser una paraula completa, una part d'una paraula, o un signe de puntuació. Açò permet que la IA treballi de manera més eficient i comprenha millor el llenguatge.

56 La *newsletter* de Narayanan i Kapoor es diu *AI Snake Oil*. Acaben de publicar un llibre que du per títol *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Esta és la *newsletter* del 27 de juny del 2024, titulada *AI scaling myths*.



perquè la mida no és el més important. Els nostres passos, segons els investigadors, estarien plens de mites i conceptes erronis:

“L’aparent previsibilitat de l’escalat és un malentès d’allò que ha demostrat la investigació. A més, hi ha indicis que els desenvolupadors d’AGI ja estan al límit de dades d’entrenament d’alta qualitat. I el sector està observant una forta pressió a la baixa en la mida dels models. Tot i que no podem predir amb exactitud quant avançarà la IA mitjançant l’escalat, creiem que pràcticament no hi ha cap possibilitat que l’escalat per si mateix conduísca a l’AGI”.

Una de les qüestions centrals a plantejar podria ser si es pot assolir l’anelhada Intel·ligència Artificial General entrenant les màquines amb els seus propis subproductes. Si només amb això valdria per arribar a la metacognició, que és aconseguir que les màquines pensen més com els humans. Tant Narayanan com Kapoor posen el focus en què no només és important l’escala, és a dir, construir LLM més i més grans: “La investigació sobre les lleis d’escalat demostra que, a mesura que augmentem la mida del model, el càlcul d’entrenament i la mida del conjunt de dades, els models lingüístics ‘milloren’. La millora és realment sorprenent pel que fa a la previsibilitat, i es manté en molts ordres de magnitud”. Però què és un *model millor*? “Les lleis d’escalat només quantifiquen la disminució de la perplexitat, és a dir, la millora de la capacitat dels models per a predir la paraula següent d’una seqüència. Per descomptat, la perplexitat és més o menys irrellevant per als usuaris finals; allò que importa són les “capacitats emergents”, és a dir, la tendència dels models a adquirir noves capacitats a mesura que augmenta la seua mida. És cert que fins ara els augments d’escala han aportat noves capacitats. Però no hi ha cap regularitat empírica que ens permeta confiar que això continuarà de manera indefinida”. I continuen: “Si els LLM no poden fer molt més del que es veu a l’entrenament, arribarà un moment en què tindre més dades ja no ajudarà, perquè totes les tasques que alguna vegada estaran representades en ells ja estan representades. Tots els models tradicionals d’aprenentatge automàtic acaben estancant-se; potser els LLM no siguin diferents”. Sobre la pressió i l’escassetat de les dades, aporten:

“Sovint es discuteix sobre quan les empreses es quedaran sense dades de formació. Però esta pregunta no té cap sentit. Sempre hi haurà més dades d’entrenament, però aconseguir-les costarà cada volta més. I ara que els propietaris dels drets d’autor se n’han adonat i volen ser compensats, el cost podria ser especialment elevat. A més de dòlars, hi podria haver costos de reputació i legals, perquè la societat podria oposar-se a les pràctiques de recopilació de dades”.

Estos LLM, com sabem, són capaços de fer tasques més complexes com el reconeixement d’imatges o el processament del llenguatge natural gràcies al *deep learning*. Per a això s’utilitzen quantitats ingents de dades. I, com ja hem vist, finites. Internet ja va ple d’imatges sintètiques signades per DALL-E, Leonardo AI, Stable Diffusion, Midjourney, Visual Electric o Playground. Així que els nous sistemes ja s’estan entrenant, com en un bucle, amb contingut artificial que ha perdut la propietat intel·lectual en el garbull de les xarxes neuronals de les màquines.

Jathan Sadowski, investigador en l'àmbit de l'economia política de les tecnologies de la informació, encunyava en febrer del 2023 el terme *Habsburg AI*, “un sistema tan entrenat amb els resultats d'altres IA generatives que es convertix en un mutant endogàmic, probablement amb trets exagerats i grotescs”. El pas de la IA generativa a la degenerativa és cada volta més a prop, i amb això, la fi –o almenys, la reconfiguració– d'internet. Un usuari de la xarxa X oferia la seua solució<sup>57</sup>: escriure un “before:2022” en el buscador de Google per tal que els resultats no incloguen aquells generats de forma artificial.

## De Gizmodo a La Nación: IA generativa com a problema i també com a solució

‘Historia de un equipo español sustituido por una IA, literalmente’<sup>58</sup>. La notícia va ser sonada i els titulars cridaners. *Gizmodo* en espanyol era una publicació tecnològica que funcionava des del 2012 i comptava amb 5 redactors a sou. En juliol de 2023, tots cinc van ser acomiadats per videoconferència i substituïts per una intel·ligència artificial que traduïa la versió en anglés de la publicació, amb el consegüent minvament de qualitat dels continguts i també, de la plantilla. “En la redacció de Madrid d'El Economista som 70 persones”, avisa el periodista Antonio Lorenzo. “Et dic que amb intel·ligència artificial generativa n'hi hauria prou amb 20, entre personal encarregat de supervisar esta IA i redactors que aporten valor diferencial des del punt de vista de la comunicació. El mateix per als editors”. Carmela Ríos vaticina que “hem generat un destructor de llocs de treball” al qual cal donar espai, inevitablement, però no tot l'espai: “Hi ha el risc que tot el que es produïx amb IA generativa siga considerat *low cost* o de menor qualitat. La intervenció humana té ara una oportunitat més gran de ser valorada en processos com el periodisme”. La veterana reportera afegí que els fonaments del pacte de confiança entre el mitjà i l'audiència poden trontollar si els periodistes treballen de forma opaca, o sense reconèixer davant dels seus lectors i espectadors quan i per a què usen ChatGPT o qualsevol altre sistema. No només per a generar articles, fotografies o vídeos, sinó també per a fer els titulars o elaborar resums de premsa.

La indústria de la comunicació viu a la deriva des de fa diversos decennis, diu Íñigo Roy, responsable de Canals del *Levante-EMV*: “La premsa fa més d'un segle que està en crisi. Ha après a viure en la crisi i, en conseqüència, va apagant focs a mesura que l'entorn canvia. Per exemple, la premsa no creia en internet, la premsa es va introduir en internet per obligació”. Des de llavors ha estat experimentant, per necessitat o per competència, amb diferents models. El *New York Times* va ser dels primers a capar les seues informacions amb un mur de pagament i a poc a poc el van anar copiant la resta de diaris. A Espanya va ser a partir del 2019 quan es va començar a popularitzar el model *freemium*, pel qual només els continguts més elaborats són de pagament, i des d'aleshores han anat creixent gota a gota les subscripcions. Segons algunes anàlisis, només el 5% de la població espanyola estaria disposada a pagar pel

57  
58

<https://twitter.com/DMueller93/status/1766091304432279634>

Rodríguez, D. (2023, 8 de setembre). *Historia de un equipo español sustituido por una IA, literalmente*. La Vanguardia. <https://www.lavanguardia.com/tecnologia/20230908/9213417/artificial-historia-equipo-espanol-sustituido-ia-literalmente.html>

contingut. Segons Roy, en la zona mediterrània (“de Tarragona cap avall”) hi ha moltes menys subscripcions de pagament que al nord de la península. I molts menys que al nord d’Europa.

Flor Coelho, periodista de *La Nación* a Argentina, ens compartix com empren en el seu mitjà la RPA o Automatització Robòtica de Processos, un primer pas per a monitoritzar els documents electorals i validar les dades mitjançant *crowdsourcing*<sup>59</sup>. Són projectes implementats des de La Nación Data, el departament d’experimentació amb noves tecnologies per a la producció periodística. Des d’ací van fer una altra faena per a la qual van utilitzar ferramentes de Processament del Llenguatge Natural (NLP) amb què van analitzar 692 lletres de cançons de *trap*, populars entre la joventut argentina, per a, segons diuen, entendre de què parlen estos artistes. D’ací van concloure que Gucci és una marca esmentada 24 vegades o que la paraula més escoltada és *yeah*, seguida de *querer*. “Sens dubte tot això t’accelera la faena, per exemple, per a fer traduccions, resums, per a fer primeres aproximacions a temes... Tot això et servix si després ho saps revisar. Per a mi és una font més i t’optimitza la faena si la saps fer servir per al bé”, objecta Coelho. Són eines d’assistència i no de substitució, com puntualitza Jordi Linares.

Per a notícies com els resultats dels partits de futbol, en *La Nación* ja les estan implementant. *Este resum va ser realitzat per intel·ligència artificial amb supervisió d’editors de LA NACIÓN*<sup>60</sup>, és la llegenda que figura, en el lloc de la firma, en algunes de les entrades sintètiques del diari argentí. De la mateixa manera, en Las Provincias estan fent proves –encara que quan s’escriuen estes línies es continuen fent manualment– per tal que les informacions de servei, com la informació meteorològica o la programació de les sales de cinema siguen automatitzades.

## Protocolos per a un ús de la IA responsable

L’estiu del 2023 l’agència de notícies americana Associated Press, juntament amb altres mitjans, van anunciar l’elaboració d’estàndards pactats per a la implementació de tecnologies de sintetització de processos i aprenentatge profund. Van afegir, així, unes quantes pàgines més al seu llibre d’estil on estableixen pautes i consells per als seus periodistes sobre qüestions del tipus com cobrir la notícia, o un glossari de terminologia en matèria d’intel·ligència artificial. A més, fixaven que “el material produït per la intel·ligència artificial ha de ser examinat amb cura, igual que el material de qualsevol altra font de notícies”. O que “una foto, vídeo o segment d’àudio generat per la IA no ha de ser utilitzat, llevat que el material alterat siga en ell mateix el tema d’una història”.

*Wired*<sup>61</sup>, la publicació especialitzada en tecnologia referent en este camp, també va fer públic per a què usen i per a què no els generadors de text i imatge en el

59 En esta entrada del blog de *La Nación* contenen tot el procés i les ferramentes emprades per a monitoritzar la transparència en el procés electoral mitjançant intel·ligència artificial, automatització de processos i visualització de dades: <https://blogs.lanacion.com.ar/projects/data/elections-data/computer-vision-to-monitor-elections-transparency>

60 *Messi se compró una casa en Miami: los millones que pagó por la propiedad, dónde está y cómo es por dentro* és el titular d’esta informació, amb data del 12 de setembre del 2023, que signa una intel·ligència artificial “amb supervisió d’editors” en *La Nación*.

61 Diversos autors. (2023, 22 de maig). *How WIRED Will Use Generative AI Tools*. Revista *Wired*. Segons anuncien, esta política es va actualitzar aquell dia, el 22 de maig del 2023, “per a permetre l’ús d’imatges i vídeos generats per IA en determinades condicions”.

seu mitjà. “No publiquem articles amb text generat per intel·ligència artificial, excepte quan siga generat per intel·ligència artificial és el punt central de l'article”, redacten. Sí que utilitzen la IA per a suggerir titulars o textos per a publicacions breus a les xarxes socials, per a afegir noves idees al seu *brainstorming* o com a *tools* d'investigació o anàlisi. “Però escriure històries és una altra cosa”, concreten. Prometen anunciar-ho sempre, fins i tot quan siga per a fins no editorials com l'automatització de correus electrònics de màrqueting.

De la mateixa manera que cada mitjà té el seu propi llibre d'estil, es fa urgent que hi haja un document que detalle, per a periodistes i lectors o espectadores, quines pràctiques estan permeses i quines no en matèria d'IA. La transparència és un procés circular que hem de demanar tant als enginyers que entrenen LLM com als periodistes i directius de les empreses de comunicació.

Al tancament d'esta investigació, cap dels mitjans de comunicació valencians consultats –À Punt, *Levante-EMV* i *Las Provincias*– compta amb res semblant. En algun cas sí que és seguida la directiva marcada –i de moment, en fase de redacció– del grup editorial. Tampoc no existix en el diari *El Economista*, però, en canvi, en *Newtral* sí que compten amb una guia ètica per a l'ús d'intel·ligència artificial<sup>62</sup>.



# La convivència difícil entre la IA i els drets d'autor

El pragmatisme i l'adaptabilitat és una característica de les espècies per a sobreviure a un medi canviant i l'ésser humà ho fa des dels inicis de la seua història. Com hem vist més amunt, part del món de la cultura opta per denunciar les infraccions de *copyright* dels models de llenguatge de grans dimensions (LLM) per a intentar frenar la difusió d'este tipus d'obres *derivades* o creades a partir d'estes eines. Els tres grans segells discogràfics –Sony Music, Warner Music Group i Universal Music Group– també van llançar una denúncia conjunta a les companyies de generació de música IA Suno i Udio per suposada infracció de drets d'autor. En el cas particular de Warner, va ampliar esta denúncia per tal que YouTube i altres plataformes eliminaren els continguts culturals artificials publicats amb els noms i veus sintetitzades del seu planter de músics a sou.

Este abril de 2024 s'ha aprovat a Tennessee, als Estats Units, l'*Elvis Act (Ensuring Likeness Voice and Image Security Act)*, una norma que busca protegir els músics i evitar que la seua veu, imatge i actuacions siguen replicades sense autorització prèvia. Esta llei vol marcar un exemple que es repetisca en la resta del món, encara que en realitat no ha estat promoguda pels músics sinó per la indústria, sobretot per editors i agrupacions de societats de drets d'autor. “Sembla més dissenyada per a protegir la inversió de les discogràfiques i garantir noves vies de negoci als hereus i societats de gestió”, escrivia Marta Peirano en *El País*<sup>63</sup>.

Hi ha una altra part de la indústria creativa que opta, en canvi, per aprofitar-ne les potencialitats i anticipar-se al canvi de paradigma que ve. Este any, Grimes, el nom real del qual és Claire Boucher, va presentar un programari d'intel·ligència artificial anomenat Elf.Tech que permet a altres persones cantar a través de la seua veu, i a més anima els músics a publicar cançons amb l'aplicació, sempre que es compartisquen a mitges els guanys obtinguts. “Em repartiré el 50% dels drets d'autor de qualsevol cançó generada per IA que faça servir la meua veu. El mateix tracte que amb qualsevol artista amb qui col·labore. Sigues lliure d'usar la meua veu sense penalització. No tinc discogràfica ni obligacions legals”, va escriure a la xarxa social X<sup>64</sup>.

La tecnologia no ha de ser sempre el monstre que ens lleve la faena. També pot jugar a favor. Només cal saber-la emprar i donar-li les indicacions adequades per tal que treballes amb nosaltres i no contra nosaltres. Alguna cosa així devia pensar el departament de Ciència Computacional de la Universitat de Chicago quan va desenvolupar Nightshade, un programa que desordena les dades d'entrenament de manera que podria causar danys als models d'IA generadora d'imatges. És la manera que se'ls ha acudit per lluitar contra les *big tech* que utilitzen obres d'artistes amb *copyright* per a entrenar els seus models sense el permís de les creadores. Amb este programa els artistes podran afegir canvis invisibles als píxels de les seues obres abans de pujar-les a internet, de manera que si s'incorporen a un conjunt de dades d'entrenament d'IA el model resultant pot fallar de manera caòtica i impredecible i convertir gossos en gats o no diferenciar entre cotxes i vaques. D'alguna manera, enverina estes dades d'entrenament per a danyar futures iteracions de models de la IA com DALL-E, Midjourney o Stable Diffusion, entre d'altres, en inutilitzar alguns dels seus resultats. “Entenc que els artistes estan molt preocupats, però crec que cal buscar la manera de finançar-los la vida d'una altra manera que no siga a partir d'una defensa acèrrima d'un *copyright* que les grans empreses es botaran i que impedirà que, en realitat, el comú dels mortals tinguem un accés molt més lliure a la cultura”, postilla Gala Pin, diputada al Congrés i activista pels drets socials a internet.

63

Peirano, M. (2024, 6 d'abril). *No has entendido nada, Beyoncé*. *El País*.  
<https://elpais.com/babelia/2024-04-06/no-has-entendido-nada-beyonce.html>

64

<https://twitter.com/Grimezs/status/1650304051718791170>



# Qui espolia a qui: la propietat intel·lectual en l'era d'internet

El març del 2024 l'Audiència Nacional ordena bloquejar l'aplicació de missatgeria instantània Telegram a conseqüència d'una denúncia presentada per Mediaset, Antena 3 i Movistar, que acusaven l'aplicació d'allotjar contingut protegit per drets d'autor sense permís: “És una mostra de com la defensa del *copyright* el que aconseguís és generar censura i clausura d'espais d'organització i no ajuda que els creadors de continguts o els artistes no es moren de fam”, llançava Gala Pin a les xarxes<sup>65</sup>. La veritable espoliació de la propietat intel·lectual, per a Marta Peirano, l'estan fent les grans tecnològiques amb els seus processos d'entrenament opac amb material protegit i sense respectar les llicències, siguen privatives o lliures.

La qüestió de la defensa dels drets d'autor té moltes arestes i alguns cognoms il·lustres, com el de Richard Stallman, fundador del moviment pel programari lliure, del sistema operatiu GNU i de la Free Software Foundation. Seua és la famosa frase “no és lliure com en barra lliure, sinó lliure com en llibertat d'expressió”, per a diferenciar lliure i debades, perquè en anglés es fa servir la mateixa paraula: *free*.

En la primera dècada dels 2000 es va estendre a la península el moviment per la cultura lliure i conceptes com Creative Commons, programari lliure, llicències lliures o els programes de codi obert es van popularitzar i arrelar en una estesa lluita social, esperonada per lleis que anaven contra la naturalesa d'internet i que no han pogut envellir amb dignitat, perquè el temps ha mostrat la seua inutilitat i qüestionat la seua ètica. Sonada va ser la promulgada per l'aleshores ministra de Cultura Ángeles González-Sinde el 2011, molt criticada i controvertida, entre altres punts, per haver estat aprovada després de les pressions del *lobby* audiovisual nord-americà i per propiciar directament o indirectament la censura i la creació de llistes negres, la criminalització del que anomenaven *pirateria* i la vulneració explícita de la llibertat d'expressió i els drets de les persones usuàries d'internet. “Ens han venut la moto d'una manera intensa i absurda sobre la propietat intel·lectual. Ens han dit pirates fins a l'extenuació i la realitat és que els primers pirates han sigut ells, que per a poder construir eixes imatges i vídeos han hagut d'esprémer mig internet”, protesta Karlos G. Liberal, investigador de la interacció entre persones i tecnologia i responsable de Bicolabs.

Com veiem, la irrupció de la IA només fa que tirar més llenya a un foc ben viu, on jurídicament encara no hi ha un full de ruta clarament establert. En data d'abril de 2024, les companyies d'IA generativa ja acumulen 20 demandes als Estats Units per violar drets d'autor. En l'europea *AI Act*, la primera legislació sobre intel·ligència artificial del món, com hem dit, no es fa referència de forma clara als drets d'autor, encara que sí que deriva a les directives ja existents, com la Llei de Protecció de Dades –l'adaptació espanyola al RGPD europeu– o la Directiva 2019/790 sobre drets d'autor i drets afins al mercat únic digital. Les lleis que protegeixen el *copyright*, tant als Estats Units com a Europa, estableixen que el depositari de la propietat intel·lectual ha de ser una persona i no una màquina. Amb això, qualsevol contingut produït per intel·ligència artificial no en podria ser atribuïble, almenys teòricament. “Els processos [i l'algoritme] són massa aleatoris perquè siga possible dir que l'autor és l'usuari. Això vol dir que l'autor és la màquina? Tampoc. Ho són les persones que van programar la màquina? Les que van obtenir les dades per entrenar-la? Les que van crear el model? En este cas, el problema és que ells no tenen ni idea de com s'utilitzaran estes ferramentes. En tots dos

casos, el procés és massa aleatori, així que ningú no pot ser l'autor”, respon al diari *El País* l'advocada experta en propietat intel·lectual Jane C. Ginsburg<sup>66</sup>.

Ricardo Oliva pertany a la firma d'advocats Algoritmo Legal, especialitzada en Dret de les noves tecnologies i negocis: “Els drets no són absoluts. No ho són el dret moral d'autor, els d'exploació, reproducció, transformació, comunicació pública, distribució, etc. Com passa amb el dret a la propietat ordinària, estan sotmesos a límits en alguns casos, sobretot en supòsits on un usuari podrà utilitzar una obra sense necessitat de sol·licitar el consentiment del titular dels drets, sempre que es complisquen determinats registres legals”.

Oliva rescata el concepte de mineria de textos i dades i exposa que el legislador –tant la Unió Europea com la llei espanyola– en casos com estos “permet el seu ús lliure sobre la base de la prevalença; en concret el dret d'accés a la informació i a la investigació sobre els drets de propietat intel·lectual”. Això no vol dir que hi haja carta lliure. Este Reial decret llei 24/2021 –la transposició de la Directiva 2019/790– defineix que la mineria de textos i dades és tota tècnica automatitzada destinada a analitzar textos i dades en format digital per tal de generar informació que incloga pautes, tendències o correlacions sobre elements semblants. Ací s'inclouen mètodes per atrapar informació de la web com el *crawling*<sup>67</sup>, l'*scrapping* o extracció de dades específiques de pàgines d'internet. Per a l'advocat, “això és el que fa ChatGPT, així que clarament estaria inclòs ací”.

Seguint a Oliva, l'article 67 diu que “no serà necessària l'autorització dels titulars de drets d'autor per a fer reproduccions d'obres i prestacions amb fins de mineria de text, sempre que s'hi haja accedit de manera legítima. És a dir, si l'autor t'autoritza, podràs reproduir, transformar, tractar i fer el que et vinga de gust, i ja no n'hauràs de demanar més permís”.

Esta redacció oberta llança dos missatges: d'una banda, dona certa permissivitat als creadors d'IA per tal que continuen fent el que feien i als autors els recorda que no tenen drets totalment exclusius. En el cas que estes dades siguen personals i s'hi haja accedit sense el consentiment exprés del titular, el RGPD estableix, segons Ricardo Oliva, que “les pots utilitzar sempre que et poses en contacte amb la persona per a informar-lo i oferir-li dret d'accés o oposició. Però després el reglament, intel·ligentment, diu que quan és impossible identificar o ubicar la persona les dades de la qual es fan servir este cas, el responsable del tractament no està obligat a buscar ni a comunicar, n'hi ha prou que en la seua política de privadesa informe que ha fet el millor esforç i es pose a disposició en cas que el titular li demane que els lleve”. Una volta més, en els termes i les condicions d'ús ho acceptem tot sense llegir res. I és lògic: “Llegir-ho tot requereix una paciència de sant i una llicenciatura en dret”, ironitza Marta Peirano en el seu llibre *El enemigo conoce el sistema*. El 2015, els Termes d'Usuari de la botiga d'iTunes tenien 20.000 paraules. Els de Facebook, 15.000.

Des del punt de vista tècnic, Jordi Linares recalca una idea que ja ha aparegut abans: en els processos d'aprenentatge de tecnologies per a la generació d'imatges, per exemple, no hi ha fotografies a les *memòries*, només números o dades associades a estes fotografies. Així explica Linares el funcionament: a la imatge d'un gos se li assigna el número 1, a la d'un gat, el 2. Es mostra a la màquina centenars de milers de fotos de gossos i gats i, a través d'operacions matemàtiques en què s'establixen premis i castics, es va ensinistrant l'ordinador perquè, per probabilitat, encerte

66 Aunió, J. A. (2024, 30 de març). *Jane C. Ginsburg, experta en propiedad intelectual: “No estamos preparados para aceptar la idea de que la máquina sea el autor”*. *El País*. <https://elpais.com/eps/2024-03-30/jane-c-ginsburg-experta-en-propiedad-intelectual-no-estamos-preparados-para-aceptar-la-idea-de-que-la-maquina-sea-el-autor.html>

67 Segons Gemini: El *crawling*, també conegut com a rastreig web o indexació web, és el procés automatitzat que realitzen els motors de cerca, com Google, per a descobrir i explorar les pàgines web que hi ha a internet. Este procés es du a terme mitjançant robots anomenats *crawlers* o aranyes web, que segueixen els enllaços d'una pàgina a una altra per a descobrir noves pàgines i afegir-les a l'index del motor de cerca.



en quina quantitat s'assembla a un gos allò que veu en una imatge i en quina quantitat a un gat. "Això no és copiar i enganxar. Allò que emmagatzema una xarxa neuronal són pesos, són valors numèrics en què no podem identificar ni el text ni les imatges que hagem fet servir en el seu entrenament" i compara esta xarxa neuronal amb un gran fitxer de dades. Per a Linares, ací residix la complexitat de garantir la traçabilitat en estos processos, a més de la novetat que suposen fins i tot per als seus creadors. Així ho detalla:

"Estes arquitectures no són meres bases de dades, sinó que actuen com la nostra memòria a llarg termini. En el procés d'entrenament d'un model de llenguatge, este no emmagatzema les dades d'entrenament. Allà dins només hi ha paràmetres, pesos, valors numèrics que multipliquen les entrades de les neurones, fent que estes paren més atenció a un determinat valor que li ve d'unes abans que d'altres neurones. Al final, és un conjunt de números, molt semblant a la nostra pròpia xarxa neuronal biològica. Això s'anomena *coneixement paramètric* o *memòria paramètrica de caràcter estadístic*. Com dic, actua de manera semblant a la nostra memòria a llarg termini: si alguna cosa ha aparegut en el *corpus d'entrenament* moltes vegades, és més probable que la resposta siga correcta. De la mateixa manera que si, a mi, a l'escola m'han repetit moltes vegades l'any de la conquesta d'Amèrica, me'n recordaré més que va ser el 1492 que de l'any en què va passar la Revolució Francesa."

En el camp de l'aprenentatge automàtic no es donen solucions, com sí que fa un programa d'ordinador diguem-ne a l'ús. El que oferix són prediccions, ja que és un aprenentatge basat en exemples i experiència, i, com passa amb l'ésser humà, no ens donarà mai una certesa al 100%, segons Linares.

## La transparència de l'algoritme: qüestions sobre l'ètica de la IA, els vikings negres i el sobreajustament dels models

L'halo de devoció i misteri al voltant d'estes eines automatitzables és enorme i ve alimentat des de diverses cuines: d'una banda, perquè la ciutadania no sap ben bé com funcionen –els seus propis creadors, en part, tampoc–. De l'altra, els titulars, els llançaments, les declaracions, els mitjans de comunicació. Els mateixos conversadors virtuals posen la seua particular cirereta al pastís perquè estan creats per humans que no amaguen este biaix d'admirar la màquina. "Els algoritmes són dissenys i construccions humanes, i estan dissenyats a imatge i semblança de la realitat que hem generat", recorda Karlos G. Liberal.

Si aprenen de nosaltres, és probable que compten amb les nostres mateixes distorsions o parcialitats. "Els algoritmes [utilitzats en l'administració pública] poden revelar patrons racistes o sexistes de les nostres institucions. Si l'algoritme és racista és perquè s'ha entrenat amb dades racistes, perquè els humans que han produït estes dades tenien comportaments racistes", en paraules d'Ana Valdivia, professora i investigadora en Intel·ligència Artificial, Govern i Polítiques a l'Oxford Internet Institute



de la Universitat d'Oxford. I si l'algoritme no és transparent és perquè la indústria o els organismes públics no volen ser transparents. "Si tot el món sap com funciona el meu algoritme, se sabrà amb quin tipus de dades està entrenat o es veurà que no funciona tan bé com ens fan creure. Hi ha molt de misticisme sobre allò que es pot fer o no amb la IA", continua Judith Membrives, responsable d'incidència en Intel·ligència Artificial de Lafede.cat i membre d'Algorights<sup>68</sup>.

El màrqueting, novament, es torna una bona ferramenta per a sostindre un sistema que no és transparent i ajuda a alimentar el seu interès i popularitat. Que no decaiga el *hype*. Marilín Gonzalo esmenta el màrqueting de la intel·ligència artificial com a peça clau i necessària per a vendre-la millor. La mateixa legislació europea, l'*AI Act*, parla de *providers* i *deployers*, és a dir, proveïdor i implementador, perquè la IA és sobretot un negoci dins del mercat capitalista. "No és una llei que protegisca els drets humans: és una norma orientada al que en diuen *product safety* o seguretat del producte, però això no implica que per se estiga protegint drets humans", avisa Membrives.

"Fa més de 15 anys que treballo en la responsabilitat tecnològica. Crec que la tecnologia en general pot ser i ha estat, en gran mesura, una força positiva en este món. Però això només pot continuar passant si hi ha estes forces compensatòries que intenten alinear els incentius entre les empreses i la resta de la societat", apunta Arvind Narayanan, professor de ciència computacional de la Universitat de Princeton<sup>69</sup>.

Des de la psicologia de les xarxes neuronals humanes, s'ha vist que la por també és una arma de control i per a Marilín Gonzalo, OpenAI ho sap: "Els estan construint una AGI, una intel·ligència de propòsit general, quan ni tan sols este concepte està totalment ni consensuat ni definit per la majoria dels científics que hi treballen" i rememora que ha estat el mateix CEO d'OpenAI qui demanava una regulació a causa dels riscos que comporta l'ús d'estes tecnologies. "I qui ajudarà en esta regulació si no és la gent que hi està treballant? És un discurs que busca poder, o siga que busca mantindre el poder que estan generant. Quan parles amb els científics, els pareix molt poc seriós. És més, et diuen, hem aconseguit moltes coses i està molt bé, però estem molt lluny d'aconseguir alguna cosa semblant al que diuen d'una AGI".

Esta mateixa línia de pensament d'Altman la compartix Brad Smith, president de Microsoft, en una entrevista per a *El País*: "Com més poderosa es torna una tecnologia, més fortes han de ser les salvaguardes i els controls que l'acompanyen. Necessitem un nivell de regulació que en garantisca la seguretat. A vegades em sorprén quan observe persones en el sector tecnològic que diuen que no hauriem de tindre eixa regulació. Quan comprem un cartó de llet al supermercat no ens preocupem de si és segur beure'l, perquè sabem que hi ha una base de seguretat que ho garantix. Si esta és, com crec que és, la tecnologia més avançada del planeta, no crec que siga irracional demanar que tinga almenys tanta regulació de seguretat com la que tenim per a un cartó de llet"<sup>70</sup>. Els que creen la IA són els primers interessats a regular-la, això sí, sent participants de la redacció de les normes. Judith Membrives alerta de la quantitat de diners que inverteix el *lobby* tecnològic a Brussel·les per a poder tindre *lleis a mida*. Segons alguns mitjans, més de 100 milions d'euros l'any.

És lògic pensar que, si esta és, com diu el directiu de Microsoft i tants d'altres, la invenció més disruptiva i important des que es va descobrir la impremta, la carrera per a maximitzar els beneficis entre les *big tech* arribarà –i ho fa ja– a nivells

68 Algorights és una xarxa col·laborativa que promou la protecció dels drets humans i la participació de la societat civil en l'àmbit de les tecnologies de la intel·ligència artificial. La seua tasca durant tota la redacció de l'*AI Act* ha estat remarcable, amb un seguiment exhaustiu junt amb altres organitzacions de tot Europa i amb un procés de divulgació molt interessant per tal que el conjunt de la societat compregua com afecta esta norma.

69 La revista *Time* va designar les 100 persones més influents en el món de la intel·ligència artificial i entre elles hi havia Arvind Narayanan i Sayash Kapoor: el primer és professor de ciència computacional i el segon estudiant de doctorat, tots dos de la Universitat de Princeton. <https://time.com/collection/time100-ai/6308266/arvind-narayanan-sayash-kapoor>

70 Fernández de Lis, P. (2024, 20 de febrer). Brad Smith, presidente de Microsoft: "Debemos tener una manera de ralentizar o apagar la inteligencia artificial". *El País*. <https://elpais.com/tecnologia/2024-02-20/brad-smith-presidente-de-microsoft-debemos-tener-una-manera-de-ralentizar-o-apagar-la-inteligencia-artificial.html>



estratosfèrics. Només cal veure el muntant de facturació d'empreses com Nvidia, Microsoft o la mateixa OpenAI. Però en la intersecció entre tecnologia, poder i benefici econòmic es troba la persona que consumix, compra, utilitza esta tecnologia, per la qual paga un preu, bé monetari, bé en paquets de dades i privadesa. Són persones les que entrenen ChatGPT i les que produïxen el material d'entrenament, són persones les que usen ChatGPT. Per tant, la tecnologia no és ni pot ser neutral.

Sobre els biaixos d'estes eines generatives s'han escrit rius de bits. D'una banda i de l'altra. Una de les més sonades va ser la notícia de la cancel·lació durant un temps de l'opció de Gemini de generar imatges per la incomoditat que produïa la seua *diversitat forçada*<sup>71</sup>. És a dir, per tornar imatges històricament errònies, com capellans dones, soldats nazis negres, vikings amb trets racialitzats, guerrers de la Grècia clàssica de rostre asiàtic, o imatges dels pares fundadors dels Estats Units amb pells fosques. Els resultats eren “històricament incorrectes”, com reconeixia la mateixa Gemini si li preguntaves per què ja no podia generar imatges. Es tracta d'una errada per sobrecompensació, o per sobreajustament, segons concreta Karlos G. Liberal. L'oposat justament a les acusacions que va rebre DALL-E, la IA generativa d'imatges d'OpenAI, d'incloure biaixos marcadament racistes i classistes en oferir imatges de forma sistemàtica de presos racialitzats i directius homes blancs.

“La generació d'imatges per IA de Gemini origina un ampli ventall de persones. I això és bo en general, perquè l'utilitza gent de tot el món. Però ací està fallant”, va argumentar a la revista *Wired*<sup>72</sup> Jack Krawczyk, director sènior de gestió de productes en Gemini Experiences de Google. “Estem treballant per a millorar este tipus de representacions immediatament”, ha declarat a la publicació i en X ha postil·lat: “Dissenyem les nostres capacitats de generació d'imatges per a reflectir la nostra base global d'usuaris, i ens prenem seriosament la representació i el biaix. Seguirem fent-ho en el cas de les preguntes obertes (les imatges d'una persona passejant un gos són universals). Els contextos històrics tenen més matisos i els adaptarem”.

Elon Musk els va acusar<sup>73</sup> de tindre un biaix woke i va avisar del perill criminal d'allò que ell anomena “woke AI” i que seria sinònim d'IA assassina. Així ho va escriure a X: “Si una IA està programada per a impulsar la diversitat siga com siga, com ho estava Google Gemini, llavors farà el que siga per a provocar este resultat, potencialment fins i tot matant gent”.

71 Domènech, E. (2024, 24 de febrer). *¿Vikingos negros? Google Gemini y sus imágenes con diversidad forzada*. Newtral. <https://www.newtral.es/vikingos-negros-gemini-imagenes-diversidad/20240224>

72 Gilbert, D. (2024, 22 de febrer). *Google's 'Woke' Image Generator Shows the Limitations of AI*. Wired. <https://www.wired.com/story/google-gemini-woke-ai-image-generation>

73 <https://twitter.com/elonmusk/status/1768746706043035827>

# Dispensadors de continguts a la carta en època electoral: el paper de les plataformes de difusió en la contenció de la desinformació

“A mesura que es fan més populars [les eines de creació de contingut] és raonable esperar que seran utilitzades en importants contextos polítics”, li explica Renée Diresta a Marta Peirano en un reportatge per a *El País*<sup>74</sup>. “Les pròximes campanyes electorals tenen un dispensador de continguts a la carta”, apunta la Peirano. Tot això l’any –2024– en què se celebren eleccions en 74 països: almenys 20 països de la Unió Europea i en 47 més de tot el món, entre ells, els Estats Units.

Des de l’organització internacional AlgorithmWatch es va fer un estudi<sup>75</sup> juntament amb AI Forensics a finals de 2023 en què es va concloure que Bing Chat de Microsoft –ara anomenat Copilot– que utilitza el model GPT-4 (encara que entrenat amb una part de dades diferents), proporcionava informació errònia fins i tot en un terç de les preguntes que li van fer. “Microsoft sembla no poder o no voler solucionar el problema”, denunciaven. El xatbot s’inventava des de falsos escàndols sobre polítics reals fins a candidats obsolets o dates falses. “La UE i els governs nacionals han d’assegurar-se que les empreses tecnològiques reten comptes, especialment a mesura que els sistemes d’IA s’integren en productes que ja s’utilitzen àmpliament”, proclamen.

Pel març del 2024, un grup d’investigadors del Center for Countering Digital Hate, una organització britànica que treballa sobre discursos d’odi i desinformació en línia, van provar<sup>76</sup> alguns dels generadors d’imatges amb IA més populars del moment. En concret: Midjourney, Dream Studio de Stability AI, ChatGPT Plus d’OpenAI i Microsoft Image Creator. La seua conclusió –gens sorprenent– va ser que totes estes eines podien crear imatges enganyoses relacionades amb els candidats a la presidència dels Estats Units: “Les proteccions existents són inadequades”, conclouïa l’informe, encara que reconeix que algunes empreses sí que prenen certes mesures, com posar marques d’aigua a les imatges generades –és el cas de Stability AI, propietat de DreamStudio– perquè les seues imatges siguen identificables, o prohibir expressament “la creació i l’ús indegut de continguts enganyosos” a les seues normes d’ús. Per la seua banda, OpenAI, segons declaracions a la CNN, estaria dissenyant “mitigacions com declinar les sol·licituds que demanen la generació d’imatges de persones reals, inclosos els candidats” [a les eleccions presidencials dels Estats Units]. Amb tots els ulls posats en els comicis americans, la indústria no ha tingut més remei que mostrar-se proactiva.

En la Conferència de Seguretat de Munic (MSC) es va signar un Acord tecnològic per a combatre l’ocupació enganyosa de la IA en les eleccions del 2024. En esta trobada es va parlar de l’auge d’estos potencials “dispensadors de continguts a la carta” en un any de tant de flux electoral, una qüestió que preocupa de forma generalitzada els *fact checkers*, segons explica Marilín Gonzalo: “Seguint alguns estudis, l’impacte dels *deepfakes*, sobretot en vídeo, no és tan gran respecte a les *fake news* o

74 Peirano, M. (2023, 10 de febrer). *Cómo evitar que ChatGPT provoque un nuevo asalto al Capitolio*. El País. <https://elpais.com/ideas/2023-02-10/como-evitar-que-chat-gpt-provoque-un-nuevo-asalto-al-capitolio.html>

75 Diversos autors. *An analysis of Microsoft’s Bing Chat Generative AI and elections: Are chatbots a reliable source of information for voters?* Desembre 2023. AI Forensics, AlgorithmWatch. [https://algorithmwatch.org/en/wp-content/uploads/2023/12/AlgorithmWatch\\_AIForensics\\_Bing\\_Chat\\_Report.pdf](https://algorithmwatch.org/en/wp-content/uploads/2023/12/AlgorithmWatch_AIForensics_Bing_Chat_Report.pdf)

76 Diversos autors. *FAKE IMAGE FACTORIES. How AI image generators threaten election integrity and democracy*. Marzo 2024. Center for Countering Digital Hate. <https://edition.cnn.com/2024/03/06/tech/ai-images-election-misinformation>

les campanyes de desinformació i de distribució d'enganys diguem-ne tradicionals a través de les xarxes socials, a les quals estem més habituats". Fins ara, segons Gonzalo, les imatges creades amb IA generativa en campanya electoral triomfen més pel costat de la simbologia que no per la voluntat d'enganyar o confondre.

Ací hi ha el cas de les imatges de Javier Milei, difoses en la campanya electoral argentina, en què l'ara president ensenyava un gran lleó –la figura simbòlica amb què s'ha autorepresentat– alliberant d'una gran gàbia la ciutadania a les portes del Congrés dels Diputats<sup>77</sup>. “Per mi el més greu és que s'estenga esta creença que no et pots refiar de ningú ni de res, perquè justament això és el que busquen les campanyes de desinformació: embrutar l'espai informatiu i que la gent pense que tots els mitjans mentixen”. Esta actitud escèptica i desencantada ha augmentat al mateix temps que el vot a la ultradreta. “El descreïment ens conduïx de cap al cinisme”, publica l'escriptora Irene Lozano en *El País*<sup>78</sup>: “Davant el dubte, el soroll i els torrents d'informació contradictòria, una dreuera freqüent ens diu: creu el líder de la teua tribu, ja siga el partit al qual votes, el professor de pàdel (...) o un cosí meteoròleg. Això (...) contribuïx a la divisió i danya profundament el debat públic, que és el combustible del qual s'alimenta la democràcia, (...) que és un règim d'opinió pública i funciona malament si no hi ha un debat sa”.

La diputada Gala Pin compartix esta idea de Gonzalo i manifesta que li preocupen més les eleccions dels pròxims tres anys quan la IA generativa es ratifique com “un element d'assentar la mentira”, però considera que ja hui en dia contribuïx a eixa guerra epistèmica “i a l'estratègia de l'extrema dreta de deslegitimar les institucions democràtiques generant una mena d'estat d'ànim i imaginari col·lectiu en què la pèrdua de sobirania popular siga percebut com una cosa molt menys greu per part d'un sector de la població”. La diputada de Sumar en el Congrés participa a la Comissió de qualitat democràtica que se centra precisament en *deepfakes* i desinformació. “Em preocupa la generació d'imatges que ridiculitzen, no que mentixen, sinó que d'alguna manera deslegitimem des d'una manca de respecte. Em preocupa més per ací, pel paper dels *deepfakes* a l'hora de deslegitimar determinades figures i institucions democràtiques, que no tant com un element que instaure la mentira *stricto sensu*”. I conclou: “Per a mi, la part que més ens ha de preocupar en relació amb la intel·ligència artificial és: quines dades, de qui, com i per què. Hauríem de posar el focus en l'emmagatzematge, el tractament i la venda de les nostres dades, que és el més preocupant”.

El doctorand de la Universitat de Princeton Sayash Kapoor dona llum al respecte i mou el focus: “Hem sentit prediccions sobre tota esta desinformació catatròfica que acabarà amb la nostra capacitat de destriar els fets de la ficció. No esperem que això succeïssa prompte. Però això no vol dir que la desinformació no siga un problema important, sinó que la nostra capacitat per a contindre la desinformació depèn molt més de contindre la distribució de la desinformació. Plataformes com Twitter, Instagram i YouTube tenen un paper molt més rellevant a exercir en este camp”.

Segons les reflexions de l'informàtic Jordi Linares, el 2024 serà l'any del vídeo generat a través d'intel·ligència artificial i serà en els pròxims mesos quan entreveurem el futur del món audiovisual. De fet, el febrer del 2024 OpenAI va seguir sorprenent el món i va llançar Sora, la seua ferramenta per a crear vídeo a partir de línies de *prompt* amb una qualitat fins ara no aconseguida, molt pròxima al fotorealisme. Els titulars dels mitjans l'aclamen per tal que no baixes de l'onada d'admiració: ‘Tot allò que has vist de la IA eren miquetes i l'últim invent d'OpenAI ho demostra’, arengava el diari *El Confidencial*.

77 Redacció de La Voz (2024, 2 de febrer). *Un león frente al Congreso, el pueblo y una jaula: polémica por una imagen hecha con IA y publicada por Milei*. La Voz de Argentina. <https://www.lavoz.com.ar/politica/un-leon-frente-al-congreso-el-pueblo-y-una-jaula-polemica-por-una-imagen-hecha-con-ia-y-publicada-por-milei>

78 Lozano, I. (2024, 14 de març). *Si no te crees nada, no confías en nadie*. El País. <https://elpais.com/opinion/2024-03-14/si-no-te-crees-nada-no-confias-en-nadie.html>

Karlos G. Liberal concreta que és qüestió de temps que Sora faça uns vídeos perfectes, que no puguen ser distingits d'un altre fet per una mà humana: “Sora és una eina de *deepfake* encara que no està feta per això”, però la d'OpenAI no n'és l'única: ja estan disponibles per a treballs de disseny i animació gràfica, animació 3D i vídeo de gran qualitat altres companyies, com Gen-2 Runaway (“No lights. No camera. All action”, es llig en la seua web)<sup>79</sup> o Luma Labs (“Dream Machine is an AI model that makes high quality, realistic videos fast from text and images”<sup>80</sup>), per dir-ne dos.

A propòsit de Sora, Marilín Gonzalo penjava dos publicacions<sup>81</sup> en X posant en dubte el seu funcionament i acusant OpenAI de fer una campanya promocional sense que cap periodista “o tècnic exterior” haja provat el model:

“Com a periodista tecnològica he vist moltes campanyes de màrqueting que els mitjans s'afanyen a reproduir, però almenys quan ho feien Apple o Microsoft donaven a provar el producte a algun periodista o tècnic extern (encara que estigueren als EUA)”.

“Això d'OpenAI amb Sora és només màrqueting, no hi ha cap informe tècnic ni OpenAI ha demostrat que el model funcione realment. I diu que no llançarà Sora d'ací poc. Impossible comprovar o escriure alguna cosa mitjanament seriosa o verificada sobre això”.

Com explica Marilín Gonzalo, la paraula *deepfake* és una barreja de *deep learning* (aprenentatge profund, la tècnica d'IA que s'utilitza per a crear-los) i *fake* (fals). Els *deepfakes* són imatges, àudio o vídeo artificials manipulats o produïts íntegrament amb algoritmes d'intel·ligència artificial. Com més sofisticades són les eines d'IA generativa, més reals semblen i, per tant, més complicades són de diferenciar. La majoria de *deepfakes* són vídeos o imatges pornogràfiques falsejades de dones sense el seu consentiment, tant famoses com desconegudes. “Es parla molt dels riscos que vindran, però n'hi ha molts que ja estan ací, com el porno *deepfake*”, denuncia Gonzalo, “ja està tenint un impacte terrible en molta gent a molts llocs del món, des de poblacions locals on els comença a faltar l'aigua pels sistemes de Data Centers o perquè els sistemes d'extracció de minerals que es necessiten per a este tipus de tecnologies els estan causant danys”.

Mariana Spring és especialista en *fake news* de la BBC i forma part de Verify, l'equip de *fact checkers* de la cadena pública britànica encarregada de comprovar la veracitat de les informacions virals. Ella sosté<sup>82</sup> que la IA farà encara més complexa la tasca dels periodistes i recorda la difusió d'un enregistrament fals<sup>83</sup> de l'alcalde de Londres en què suposadament deia que ajornaria una desfilada d'una festa local per a prioritzar una manifestació a favor de Palestina. L'engany va agitar el carrer i fins i tot es van convocar protestes per a mostrar rebuig davant d'eixa mesura, sense saber que només era un fake amb aparença hiperrealista. “El pitjor és que ni tan sols et fa falta la IA; hi ha tàctiques molt simples com els mems, les teories de la conspiració, els missatges en línia, que són igualment eficaços en la difusió de la desinformació, i amb les quals encara no hem aconseguit acabar d'enfrontar-nos”.

79 <https://research.runwayml.com/gen2>

80 <https://lumalabs.ai/dream-machine>

81 <https://twitter.com/marilink/status/1759552461306466699>

82 Aunió, J. A. (2024, 23 de març). *Marianna Spring, especialista en 'fake news' de la BBC: "Si conoces a alguien que cree en teorías conspirativas, no le abandones, intenta entender de dónde viene"*. El País. <https://elpais.com/eps/2024-03-23/marianna-spring-especialista-en-fake-news-de-la-bbc-la-desinformacion-tiende-a-hacerse-viral-porque-queremos-creerla.html>

83 Spring, M. (2024, 13 de febrer). *Sadiq Khan says fake AI audio of him nearly led to serious disorder*. BBC News. <https://www.bbc.com/news/uk-68146053>







# Legislació i marc regulador

## Mecanismes legals de control de la desinformació

Com controlar i contindre els *deepfakes*? El marc regulador és difús i a vegades poc eficient per la complexitat a l'hora de regular les qüestions que tenen a veure amb internet i que afecten la llibertat d'expressió. Hi ha almenys dos articles al Codi Penal espanyol que castiguen la difusió de notícies falses o inexactes, de la mateixa manera que la Llei General de la Comunicació Audiovisual regula quin tipus de continguts audiovisuals es difonen. També hi ha un Pla d'Acció contra la Desinformació impulsat pel Govern, aprovat el 2020, que proposa, entre altres mesures, la creació d'una Comissió Permanent contra la Desinformació<sup>84</sup>.

En el pla legal europeu, Andrés Boix extrau de l'AI que “tot allò que es puga fer passar per creació humana i que un observador extern es puga confondre i no saber si és humà o no, cal explicar-ho. Està permès fer-ho, però cal explicar que ha estat generat per intel·ligència artificial”. De nou, els *deepfakes* no estan prohibits com a tals en la legislació europea, però cal avisar que és un contingut fals, llevat que siga evident. “A nivell periodístic, encara que no estiga específicament regulat, crec que sí que cal indicar quan un contingut ha estat elaborat amb IA. A més, és molt bona pràctica. Els mitjans de comunicació probablement diran que mentre no hi haja una obligació expressa ells no n'estan obligats. Jo crec que la lògica subjacent a la regulació és que tot allò que es puga confondre i semblar que és humà, quan en realitat és intel·ligència artificial, cal indicar-ho”, subscriu Boix.

Hi ha una certa disparitat d'opinions legals sobre si l'AI Act estableix o no exigències de traçabilitat i transparència. Així respon Lorenzo Cotino, catedràtic de Dret Constitucional de la Universitat de València: “Sobre la transparència en l'AI Act, n'hi ha molt poc. Molts es confonen perquè hi ha obligacions de transparència en el seu article 13<sup>85</sup>, però esta transparència fa referència a la informació que el fabricant [d'IA] ha de donar a qui compra l'ús del sistema i no està pensada per als afectats”. És a dir, dins de la transacció de compravenda, el proveïdor de sistemes d'intel·ligència artificial haurà de facilitar tota la informació necessària al comprador, com un manual d'instruccions on aparega clarament la forma d'ús.

“L'article 50<sup>86</sup> també parla de transparència, però fa referència a l'obligació d'informar si s'interactua amb un xatbot o d'informar que és una ultrafalsificació (àudio, vídeo, imatge). Per contra, la transparència sobre la qual he escrit un parell de llibres i diversos articles essencialment no es regula”, respon Cotino. De fet, la majoria de sistemes d'intel·ligència artificial no hauran de complir la llei perquè la legislació només

84 En la pàgina web del Ministeri d'Afers Estrangers, Unió Europea i Cooperació es troba l'informe complet amb data de 2019, a més d'enllaços a propostes europees més elaborades i actualitzades com l'EU vs. Disinfo: <https://www.euvsdisinfo.eu>

85 <https://artificialintelligenceact.eu/es/article/13>

86 <https://artificialintelligenceact.eu/es/article/50>



s'aplica als sistemes considerats d'alt risc<sup>87</sup>. A l'article 50 que menciona Cotino s'han afegit, a última hora, eixes certes exigències de transparència de les quals parlava Boix, però són "obligacions molt lleus", limitades només al fet que has d'informar, llevat que siga evident, que estàs tractant amb un contingut sintètic. Este deure d'informar només s'aplica per a assumptes o informacions que puguen ser d'interés públic<sup>88</sup>. "És una obligació de les fluxetes", reтраu Cotino i insisteix: "No hi ha en el reglament la transparència per als algoritmes".

Esta llei, com apuntava Judith Membrives més amunt, és un reglament per a fabricants d'intel·ligència artificial en què quasi no apareixen els drets de la persona afectada per la utilització d'estos sistemes d'alt risc. Sí que hi ha obligació per als fabricants de sistemes d'IA d'alt risc de dissenyar-los des de l'inici basant-se en una sèrie de requisits: anàlisi de riscos, estudi d'impacte sobre els drets fonamentals, documentació tècnica, robustesa, fiabilitat, seguretat, supervisió humana, etc.

Quasi en l'últim moment s'ha incorporat un "xicotet dret a l'explicabilitat", que encara que no és molt clar, indica que si hi ha un sistema que prenga decisions només automatitzades cal explicar-ne la lògica de l'algoritme, un precepte similar al que estableix el Reglament Europeu de Protecció de Dades. "Caldrà veure, amb els anys, si això de veres servix per a saber com funcionen els algoritmes, sobretot els del sector públic". Pel que fa a la traçabilitat, per als sistemes d'alt risc s'estableix una mena de tacòmetre, uns registres que es guardaran un mínim de 6 mesos.

En l'estrat europeu hi ha, alhora, el Codi de Bones Pràctiques sobre Desinformació, signat per les principals plataformes tecnològiques el 2018, encara que no té caràcter legal obligatori, com sí que el posseïx la Llei de Serveis Digitals (*Digital Services Act*), que inclou l'obligatorietat de retirar contingut com els *deepfakes* i marcar clarament els continguts sintètics. Un altre dels imperatius que estableix és que les grans plataformes adapten els sistemes de recomanació per evitar l'amplificació algorítmica de la desinformació.

Els intents legals per implementar mecanismes de transparència o protecció del dret a l'honor estan resultant, sens dubte, insuficients. Ha passat amb les imatges *deepfakes* de les xiques d'Almendralejo<sup>89</sup>, però també amb imatges falsejades de Taylor Swift, que van rebre 25 milions de visualitzacions abans que X les eliminara de la plataforma. "Jo crec que el problema és que este tipus de tecnologies són accessibles, a l'abast de qualsevol persona que tinga un ordinador i una connexió a Internet", argumenta Ana Valdivia. L'ús particular de cada ferramenta en determina el producte final: "Intentar crear un model d'intel·ligència artificial que no es puga utilitzar malament és com intentar crear un ordinador que no es puga utilitzar per a coses roïnes", escriuen Arvind Narayanan i Sayash Kapoor.

87

L'AI Act classifica els programes segons el nivell de risc. Més avant ampliem este tema.

88

Això és el que diu el punt 4 de l'article 50 de l'AI Act: "Els implantadors d'un sistema d'IA que genere o manipule continguts d'imatge, àudio o vídeo que constituïsquen una falsificació profunda han de revelar que el contingut ha estat generat o manipulat artificialment. Esta obligació no s'aplica quan l'ús estiga autoritzat per la llei per a detectar, previndre, investigar o perseguir delictes penals. Quan el contingut forme part d'una obra o programa manifestament artístic, creatiu, satíric, de ficció o anàleg, les obligacions de transparència establides en este apartat es limitaran a revelar l'existència del contingut generat o manipulat d'una manera adequada que no obstaculitze la visualització o el gaudi de l'obra. Els usuaris d'un sistema d'IA que genere o manipule text que es publique per a informar la gent sobre assumptes d'interés públic han de revelar que el text ha estat generat o manipulat artificialment. Esta obligació no s'aplicarà quan l'ús estiga autoritzat per llei per a detectar, previndre, investigar o perseguir delictes penals o quan el contingut generat per IA haja estat sotmés a un procés de revisió humana o control editorial i quan una persona física o jurídica ostente la responsabilitat editorial de la publicació del contingut".

89

El setembre del 2023, la Policia Nacional investiga unes fotografies de nus fetes amb intel·ligència artificial que agafen el rostre d'un xique menor d'edat a la població d'Almendralejo, en la província de Badajoz. S'hi van detectar almenys 20 casos.



# IA per a desemmascarar la IA o el robot que caça robots

La Iniciativa per a l'Autenticitat del Contingut, impulsada per Adobe, Twitter i el *New York Times* el 2019, proposa un sistema de traçabilitat a través de metadades que ajuden a identificar-ne la procedència dels continguts. És a dir, que quede clar que l'atribució d'estos productes pertany als autors i autors. És una de les opcions que tracten de donar resposta al complicat problema del rastreig i control dels materials creatius.

“Si cada empresa crea un model diferent de generació de textos o d'imatges, no es podrà construir una eina que comprove els errors sistemàtics adaptada a cadascun d'estos models”, apunta Karlos G. Liberal, i segueix: “Hi haurà un descreïment absolut, però jo pense que els mitjans s'ho han guanyat. Si Ferreras, després del que va fer i del que va dir<sup>90</sup>, continua sent l'entrevistador estrella de La Sexta, per què el problema haurien de ser les notícies falses? El problema és el model periodístic basat, en molts casos, en els interessos dels grans conglomerats comunicatius i empresarials”. I contra els enganys i la desinformació, la seua recepta són xarxes d'afinitat i de confiança que ens permeten desmuntar-los: “Són perillosos i capaços de desestabilitzar un país, ho vam veure amb la victòria de Bolsonaro, on es va descobrir tot l'entramat dels evangelistes per a difondre mentides i notícies falses. En el cas del Brasil se sap i en altres també, sobretot després de l'escàndol de Cambridge Analytica”.

Això passa, segons l'investigador de BikoLabs, perquè no hi ha hagut una reflexió prèvia sobre on volem que ens duga la tecnologia. Ujué Agudo, doctora en Psicologia i investigadora de la relació entre la IA i les relacions humanes, afeg: “També estem veient casos en què és el propi algoritme el que genera una informació falsa, en la seua necessitat per crear contingut”, i recorda el que va passar amb el xatbot de la companyia de vol Air Canada, quan va donar una informació incorrecta sobre el reemborsament d'un bitllet a un passatger i l'empresa es va defensar dient que el bot xarrador era “responsable de les seues accions”<sup>91</sup>. La història la va tancar un tribunal canadenc que va obligar la companyia a complir<sup>92</sup> el que l'assistent virtual havia promès al client. El perill de depositar tota la confiança en els models i pensar que les al·lucinacions o equivocacions no existixen està ahí. Karlos G. Liberal explica: “Als Estats Units cada volta estan més preocupats per la traçabilitat del codi perquè la gent està escrivint codi amb GPT i després no se sap qui ha fet la peça de codi que ha fallat”.

En línia amb les paraules de l'investigador Karlos G. Liberal van les reflexions de Marta Peirano i, per a ella, el *robot per a caçar altres robots* no funciona, principalment perquè les empreses no ofereixen transparència en els seus processos i no deixen que els seus paràmetres siguin detectables.

“Un dels grans problemes amb estos models generatius tipus ChatGPT o els visuals són que no pots controlar massa els resultats. Tu demanes alguna cosa i ix alguna cosa, però a vegades està malament, és incorrecta, no és coherent o no és allò que busques. Per a les grans empreses que els fabriquen açò és un gran problema, ja que

90 El juliol del 2022 es van fer públics els àudios d'una conversa que va tindre lloc durant la campanya electoral del 2016 entre Antonio García Ferreras i el hui excomissari José Villarejo, agent operatiu del Ministeri de l'Interior. El presentador d'*Al rojo vivo* de La Sexta reconeix que la suposada notícia destapada per Eduardo Inda relativa al fet que Pablo Iglesias té un compte bancari en un paradís fiscal en què el govern de Nicolás Maduro hauria ingressat milers d'euros “es demasiado burda (sic)”, encara que reconeix que “va con ello”. És a dir, que ho va explicar a La Sexta com si fora una informació verídica. Més informació: <https://www.elsaltdiario.com/medios/ultimos-audios-villarejo-ferreras-epicentro-debate-cloacas-estado-podemos>

91 Yagoda, M. (2024, 23 de febrer). *Airline held liable for its chatbot giving passenger bad advice - what this means for travellers*. Bbc.com. <https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know>

92 Belanger, A. (2024, 13 de febrer). *Air Canada must honor refund policy invented by airline's chatbot*. Arstechnica.com. <https://arstechnica.com/tech-policy/2024/02/air-canada-must-honor-refund-policy-invented-by-airlines-chatbot>

tenen un producte que no poden controlar. S'estan fent rics amb això, però al mateix temps tenen una cosa que és incontrolable, cosa que els produïx un malestar profund", expressa l'expert en *machine learning* Miguel Arana. L'explicabilitat d'eixes capacitats emergents, com ja hem comentat, és un problema que encara no té resposta.

"En l'era de la intel·ligència artificial, quan desconfiem de les imatges que veiem, la confiança s'hauria de depositar en la font que les subministra (...) Les fake news succeïxen perquè les institucions que tenien el treball de dir la veritat ja no ho fan. Es produïx així un desplaçament inevitable: quan les autoritats canvien el món canvia. La postveritat no se soluciona amb el fact-checking", sospesa Mária Martínez Bascuñán<sup>93</sup>.

## Eines de verificació disponibles i fiables

No hi ha una recepta precisa i absoluta contra les ultrafalsificacions o les mitges veritats. Tant en institucions acadèmiques com en mitjans privats hi ha departaments especialitzats en tasques de verificació a través d'eines amb IA, a la cerca del perfecte robot caçamentires. Algunes són l'Observatorio de la Desinformación de la Universitat Carles III de Madrid o VerificaRTVE, la plataforma de comprovació de dades impulsada per RTVE. Google té el seu propi paquet de programes verificadors i fins i tot un cercador, Google Fact Check Explorer. La proposta basada en IA del Grup Prisa es diu VerificAudio<sup>94</sup> i la seua finalitat és, segons posa en la seua web, "combatre la desinformació en els continguts sonors, una amenaça en ascens propiciada per l'avanç de les darreres tecnologies generatives". L'aplicació de Prisa ofereix la possibilitat que qualsevol persona pugui pujar un àudio a la web per tal que siga analitzat pels seus *equips de verificació*. El responsable de Sistemes d'À Punt, Pau Peiró, aporta la seua: autoverifAI.com, una plataforma oferta per Telefonica que diu detectar tant àudio com vídeo, i també text, sempre que haja estat creat per ChatGPT i en anglés.

Com més passen els mesos i més perfectes i versemblants són les composicions de la generativa, més dificultat tenen els periodistes a l'hora de saber què és real i què és un *fake*. "Hi ha signes", desvela Peiró, "o detalls que són merament humans, com el pestanyeg irregular. Però sobretot hi ha metadates, una que indica quan va ser creada i una altra dins de la composició que et pot donar una orientació, tot i que és un procés complicat i cada volta més". Per al cap de l'àrea Digital d'À Punt Mèdia, Javier Lifante, la ferramenta de verificació més infal·lible és despenjar el telèfon i cridar a les autoritats, a les fonts, al bar proper al succés, una cosa factible i necessària en el periodisme local i de proximitat que fa la cadena autonòmica. Però, en esta era de l'artificialitat amb aparença d'intel·ligència, "tots correm el risc de ser enganyats alguna volta".

A VerificaRTVE<sup>95</sup> un equip format en detecció d'enganys comprova notícies, fotografies i vídeos falsos. A més, ofereix una llista prou completa d'aplicacions per a extraure metadates i anàlisi inversa d'imatges, diagnòstic de *links* per veure si són maliciosos, registres d'IP d'un lloc web, eines que detecten alteracions digitals en una imatge, rastrejadors de vols i vaixells, analitzadors de vídeo com Azure Video Analyzer de Microsoft o fins i tot ferramentes de cronolocalització, capaces de saber quina hora

93 Martínez-Bascuñán, M. (2024, 31 de març). *Cloacas reals: de Kate Middleton a RTVE*. El País.

<https://elpais.com/opinion/2024-03-31/cloacas-reales-de-kate-middleton-a-rtve.html>

94 <https://verificaudio.ai>

95 <https://www.rtve.es/noticias/verificartve>

era en una imatge o un vídeo per les ombres i la col·locació del sol, o conèixer quin temps feia en una data, hora i lloc concret.

Per a detectar text sintètic, hi ha una llarga llista de programes. Els que solen funcionar millor són Winston AI i Undetectable. Gpt Zero està especialitzat a detectar text generat per ChatGPT, Gemini i LLaMA, principalment. Originality intenta assegurar que els continguts estiguen lliures de plagi. En cas d'imatges, especialment les creades amb DALL-3, pot funcionar bé AI or not. Sensity.ai es pot utilitzar per a detectar falsificacions mitjançant IA en vídeos, imatges, àudio i fins i tot en carnets d'identitat.

Encara que hi ha certa varietat de plataformes que asseguruen poder comprovar i distingir els continguts generats de forma artificial dels produïts per humans, com hem dit, no són infal·libles. En el camp de les *deepfakes* també es complix la màxima, com passa amb les *fake news*, que és molt més fàcil crear-les que desmentir-les, eliminar-les o frenar-ne la propagació a la web. La complexitat de fer *fact-checking* o comprovar si són verídics continguts que han pogut, o no, ser creats per la màquina de desinformació de la intel·ligència artificial és molt més gran perquè tenen gran versemblança i són més senzilles de generar o de propagar en quantitats industrials.

## Legislar *a priori*: AI Act, la primera llei que posa límit a la Intel·ligència Artificial

El periodista Javier Sampedro, en una de les seues *newsletters* en *El País*, augurava que en un futur tan llunyà com el que tarde a arribar el GPT-5 “es podran piratejar webs de forma totalment autònoma”. Que la tecnologia GPT escriu codi, amb el *prompt* o l'orde adequada, millor que moltes mans humanes ja és una realitat. D'ací a manipular el codi font d'un lloc web per a robar la informació dels seus usuaris sembla que no queda tant<sup>96</sup>. Segons Sampedro, amb la versió GPT-4 s'ha aconseguit una taxa d'èxit 10 vegades superior a la de GPT-3.5, el seu predecessor immediat, en esta tasca. “El problema és que esta tecnologia és capaç de produir realitats distòpiques”, alerta la periodista Carmela Ríos: “estem en el moment perfecte per identificar estos perills gairebé abans que apareguen per a poder legislar *a priori* i no *a posteriori*”.

Per a juristes experts, l'anomenada Artificial Intelligence Act –AI Act–, amb les seues 468 pàgines, és un dels textos de més complexitat redactats fins ara. La primera llei que regula la intel·ligència artificial es va començar a gestar al si d'Europa l'abril del 2021. Finalment, després d'intenses negociacions que van coincidir en la fase final amb la presidència espanyola de la Unió Europea, va ser aprovada el 14 de març de 2024. Encara falten almenys tres anys per a la seua entrada en vigor, però a principis de 2025 es posarà en funcionament la prohibició de fer servir sistemes d'IA en certs contextos.

A mitjans del pròxim any, l'Oficina d'Intel·ligència Artificial europea acabarà de redactar els codis de pràctiques que especificaran els requisits tècnics dels *models de propòsit general*, com ChatGPT, i d'IA *amb risc sistèmic*, que és com s'anomena

96

Caburn, T. (2024, 17 de febrer). *How to weaponize LLMs to auto-hijack websites. We speak to professor who with colleagues tooled up OpenAI's GPT-4 and other neural nets.* The Register. Ací recull l'experiment que van fer informàtics de la Universitat d'Illinois Urbana-Champaign (UIUC), que, usant LLM, van atacar webs vulnerables sense intervenció humana i van demostrar que estos models “poden hackejar llocs web de forma autònoma i executar tasques complexes”.

en la llei a la intel·ligència artificial generativa. En maig d'això any hauran d'estar funcionant els organismes que vigilaran el compliment de l'AI Act. També s'espera que estiguen definides com seran les multes que es posaran per l'incompliment dels articles establits per a sistemes d'IA d'alt risc inclosos en l'annex II i per als que ja estiguen comercialitzats. En el cas de ChatGPT, com ja es troba en el mercat, la UE li dona un termini de 2 anys per adaptar-se a la llei.

Segons aclareix Andrés Boix, els temps llargs són habituals i es fan per donar temps que els operadors jurídics, tant públics com privats, s'hi adapten. “Encara que no entre en vigor, ja estableix uns principis directores i unes normes que conformaran un cert estàndard de comportament”. Per tant, donarà peu al fet que els Estats membres puguin anar aplicant normes coherents amb ells.

Com que té l'estatus de reglament, es transposa tal com en el règim jurídic dels països de la UE, a diferència del que passa amb les directives europees, en destaca Algorights<sup>97</sup>. Això vol dir que els Estats poden decidir limitar estes excepcions i imposar-hi regulacions més garantistes.

El catedràtic de Dret Constitucional Lorenzo Cotino la qualifica de “molt positiva” a l'hora de donar “seguretat jurídica per tal que es propicie la inversió pública i privada amb un marc jurídic clar (...) amb garanties per als nostres drets fonamentals i la dignitat de l'ésser humà”. Esta norma, encara que només afecta els països membres, influirà *de facto* en qualsevol tecnologia que s'elabore en un altre país i vullga ser comercialitzada a Europa. “Això es traduirà en el fet que països com els Estats Units hauran d'assumir la normativa europea perquè, altrament, no podran operar a Europa”, precisa Marta Suárez-Mansilla, advocada experta en art i tecnologia.

Tant Cotino com Suárez-Mansilla coincidixen que l'AI Act s'ha de llegir conjuntament amb el Reglament General de Protecció de Dades (RGPD) en el cas que, per a l'entrenament dels models d'IA s'usen dades personals. Són normes que “se superposen encara que hui en dia s'ignoren” i cal esperar-ne una “millora del procés regulador”, diu Cotino.

## Roig, groc i verd: la divisió legal segons el risc de cada producte amb IA

La llei estableix el repartiment d'una sèrie d'etiquetes a cada producte d'IA segons el risc que presente per als drets humans: “A més perill, més obligacions preventives”. Com si fora un semàfor, l'etiqueta roja es correspon amb pràctiques prohibides; la groga, per a aquelles que siguen considerades d'alt risc (“per a evitar danys, imposablem preventivament una sèrie d'obligacions a productors, dissenyadors i usuaris”). Les verdes, seguix Cotino, no estan regulades per esta norma, però sí per altres normes com el Reglament General de Protecció de Dades.

En el llistat roig figuren com a restringides aquelles accions que empen intel·ligència artificial per al reconeixement facial biomètric en temps real en espais d'accés públic. En paraules d'Algorights, no et podran reconèixer per les teues dades

97

La tasca de divulgació de col·lectius com Algorights de la llei europea d'intel·ligència artificial és intensa i extensa, ja que han seguit des del començament tot el procés legislatiu. Ací un vídeo: *¿Qué va a pasar con la IA en Europa? Introducción a la AI Act para movimientos sociales*  
<https://fediverse.tv/w/kzVGtTT8Eh3o2hzgXNSPcY>



biomètriques amb una IA mentre et trobes al carrer, al cinema o un supermercat. També seria il·legal gastar-la per a implementar tècniques manipuladores o subliminals que puguen generar danys o que suposen l'exploració de les vulnerabilitats de les persones a causa de la seua edat, discapacitat o una situació social o econòmica de què es puguen aprofitar per a condicionar-ne el comportament i causar-los perjudici. Ací entraria la creació i difusió de *deepfakes* o les pràctiques comercials fraudulentament manipulatives.

Està prohibit per la llei categoritzar persones segons les dades biomètriques per a deduir o inferir la seua raça, opinions polítiques, afiliació sindical, creences o orientació sexual. Un dels punts punitius va en la línia d'impedir l'*scoring social* o carnet de bon comportament i crèdit social com el que hi ha a la Xina, amb què la ciutadania rep punts o els perd en funció del seu comportament social i que condiciona directament les prestacions i el tracte que rep de l'Estat.

No es permeten els sistemes de reconeixement d'emocions en entorns laborals i educatius, llevat de motius mèdics o de seguretat. Tampoc els sistemes per a crear o ampliar bases de dades de reconeixement facial mitjançant *scrapping* d'imatges d'internet o Circuits Tancats de Televisió. Un exemple, segons Algorights, d'empresa que ja no podria operar a la UE seria Clearview, si bé continua sent utilitzada als Estats Units, el Regne Unit o Ucraïna.

Sara: Què és Clearview?  
 Gemini: Clearview AI és una companyia nord-americana que desenvolupa tecnologia de reconeixement facial. El producte principal és un motor de cerca facial que permet a les forces de l'orde públic i altres organitzacions pujar una imatge d'un rostre i comparar-la amb una base de dades massiva de més de 40 bilions d'imatges públiques provinents d'internet, incloent-hi xarxes socials.

Entre els sistemes d'alt risc hi ha els sistemes d'identificació biomètrica per a l'ús en l'admissió d'escoles i centres educatius, el control de l'ocupació pública i privada, accés a serveis essencials públics o privats, resposta a emergències, auxili mèdic, bombers, etc. En l'àmbit penal i policial, qüestions com la prevenció, la detecció i la persecució de delictes, la migració, l'asil i el control de fronteres. Alguns exemples de pràctiques que inclourien tècniques automatitzades considerades d'alt risc serien l'ús de polígrafs, avaluacions de la possibilitat de reincidir o no d'un pres, el mesurament de l'estat emocional d'un acusat o d'una víctima, la fiabilitat de les proves en un judici, o bé utilitzar la IA per a predir la personalitat o el comportament delictiu, l'anàlisi de delinqüència, migració, asil o control de fronteres. Segons denuncia Algorights, estos sistemes deixarien de ser considerats d'alt risc si no suposen risc significatiu per a la salut, la seguretat o els drets fonamentals o no influïxen materialment en el resultat de la presa de decisions.

Finalment, en la llista roja hi ha els sistemes de l'anomenada Policia predictiva, segons la qual s'avaluaria el risc que una persona cometa un delictes basant-se en la recreació del perfil o les característiques de la seua personalitat. Sí que es permetran sistemes que donen suport a l'avaluació humana de la implicació d'una persona en un delictes.

Des d'Algorights també recorden que l'AI Act no impedeix que empreses europees desenvolupen este tipus de tecnologia, prohibida a Europa, per a exportar-la. Quant a les mesures correctives, encara que sí que es preveuen sancions per a les empreses que incomplisquen el reglament, no existixen per a l'Administració, malgrat que les institucions usen àmpliament estes eines i es beneficien de la llarga llista d'excepcions previstes. L'encarregat de supervisar i de garantir que tots els sistemes que

arriben a la UE complixen la regulació serà la nova Oficina de la IA, dependent de la Comissió Europea.

En la mateixa llei s'indica que serà el proveïdor del servei tecnològic qui decidisca el seu nivell de risc en una primera autoavaluació, una qüestió que per a les organitzacions que defensen els drets humans pot suposar un problema. Gala Pin compartix este pensament: “La distinció entre alt, mitjà i baix risc opera sobre com s'aplica i com s'usa intel·ligència artificial en l'àmbit europeu, però tu pots desenvolupar intel·ligència artificial d'ús prohibit a Europa i exportar-la perquè es faça servir en altres països, cosa que va en contra d'una visió global de garantia dels drets humans”.

Per la seua part, Boix aporta una explicació jurídica sobre això: “Tots estos sistemes de seguretat industrial es basen en la col·laboració amb la indústria i les empreses privades. La primera avaluació és una autoavaluació que fa la mateixa indústria, la segona una auditoria per part d'una entitat privada de certificació. Hi ha, així mateix, un mecanisme últim de control en el poder públic, encarregat de revisar si s'estan fent bé les auditories o comprovar si estes valoracions de les empreses són correctes”. Això passa ací i en tots els sectors, per exemple amb qualsevol electrodomèstic o les empreses automobilístiques. Són les firmes les que marquen les certificacions energètiques dels seus productes, després hi ha auditories i comissions privades i finalment, si hi ha denúncies o sospites d'irregularitats, un control públic. “Una de les funcions de l'Agència europea de Supervisió de la IA serà eixa, controlar els que fan les auditories i establir sancions si és el cas. Només en extrems potencialment perillosos s'exigix que, a més a més, hi haja avaluació pública des del principi, i si no, el que es fa és que l'avaluació pública es realitza després, si es detecta algun problema”, acaba Boix.

## Privadesa i seguretat en les excepcions de l'AI Act: el debat sobre la captació biomètrica als carrers i les fronteres

Un dels temes més controvertits de la norma és el que es referix als sistemes de reconeixement facial en la via pública, que, com s'ha dit, queden prohibits llevat de determinades excepcions que, segons la lectura política i administrativa, podrien ser més àmplies o menys. Només podran ser emprats pels Cossos i forces de seguretat de l'Estat i el seu ús està subjecte a autorització judicial o administrativa prèvia, encara que en casos d'*extrema gravetat* es podran implementar sense este consentiment, com analitzarem més avant.

Segons explica Algorights, les autoritats públiques podran fer ús del reconeixement biomètric per raons de seguretat en casos de risc d'atemptat, o per a buscar víctimes de segrest o de tracta i a sospitosos de delictes greus. Abordem este punt amb la llei a la mà junt amb Javier Boix: “Hi ha hagut molta discussió sobre si s'havia de prohibir o no el reconeixement biomètric. La decisió adoptada per la Unió Europea és que no es prohibeix amb caràcter general, sinó que es prohibeixen els reconeixements biomètrics en temps real que facen empreses privades. No estarà prohibit, en canvi, quan ho facen els poders públics per a la persecució i el combat de delictes. La pregunta és: fins a on arriba esta habilitació als poders públics quan es tracta de combatre delictes?”.

El professor de Dret Administratiu ressaltava el cas de la Xina, on el control ciutadà mitjançant biometria es fa amb l'excusa d'evitar que es cometin infraccions. En la redacció de l'AI Act s'extrau que el cas xinès no és extrapolable a Europa perquè ací només es podrà fer servir per a la persecució de delictes ex post, és a dir, una volta s'hagen produït i dins d'una investigació que ja estiga en marxa: “El caràcter preventiu generaria un control general i passiu de tota la població”, que és justament allò que es vol evitar.

El reglament prohibeix l'ús en temps real d'identificació biomètrica remota en espais públics excepte en els casos en què siga estrictament necessari per a aconseguir objectius, continua Boix, com la busca específica de víctimes d'abducció, tràfic de persones, tràfic o explotació sexual: “És a dir, quan tu ja tens una operació en marxa, no preventiva”. I concedeix: “Això no té a veure, en principi, amb la identificació en frontera, encara que a vegades s'hi puga utilitzar. Però ha de ser només per a combatre assumptes com les xarxes de tràfic”.

Una altra de les excepcions que permeten l'ús d'estos sistemes d'identificació basat en trets físics seria per a la prevenció d'una amenaça imminent per a la vida o la seguretat física de persones, per a la prevenció d'un atac terrorista i per a la localització i la identificació d'un sospitós d'haver comés una ofensa criminal. La diputada Gala Pin completa amb un cas pràctic: “Si tu actives el reconeixement facial en temps real només en casos com després de l'atemptat de la Rambla de Barcelona ens pot no agradar, però al final estàs dient que l'excepcionalitat són moments realment excepcionals. En canvi, si comences a aplicar el reconeixement facial a qualsevol manifestació de més de mil persones, normalitzes l'excepcionalitat”.

Judith Membrives, que també és part d'Algorights, hi aporta un altre exemple: “Estem en alerta 4 d'atemptat terrorista, que és alerta imminent. Si el Govern considera que eixe estat d'alerta requereix ús de reconeixement facial o de reconeixement biomètric en temps real per seguretat nacional, ho podria implementar”. Per a això necessitaria una autorització judicial o bé una d'administrativa. I en certs casos ni tan sols això, perquè des d'agost del 2023 la Policia Nacional empra una eina de reconeixement facial en almenys 11 ciutats sense requerir l'orde d'un jutge. Es coneix com el programa ABIS, que són les sigles en anglés de Sistema Automàtic d'Identificació Biomètrica. El seu funcionament de portes enfora és senzill d'explicar, però la manera com s'ha entrenat el seu algoritme és un misteri. El que se'n sap és que els sistemes automàtics de reconeixement facial extrauen el rostre de les imatges mitjançant visió computacional i apliquen un algoritme a eixe rostre per a obtenir un patró que el distingisca dels altres. La IA selecciona i ordena si el patró es repeteix o apareix dins de l'extensa base de dades de fitxes policials que inclouen tant patrons facials com l'ADN de 4,4 milions d'individus fitxats per la Policia Nacional o la Guàrdia Civil. Dos operadors formats per a això verifiquen la llista de semblances que ofereix el sistema, encara que rarament prenen com a errònia l'opció de l'algoritme.

El sistema ABIS no processa imatges en temps real perquè no està connectat a càmeres de vigilància, sinó que obté les cares d'enregistraments de càmeres de seguretat o de fotografies fetes amb mòbils. “Un ús descontrolat del reconeixement facial erradicaria l'anonimat”, replica la periodista del *New York Times* Kashmir Hill. Els nostres moviments quotidians podrien ser rastrejats tant pel nostre veí com pel nostre Govern, com fa Rússia en les manifestacions a la via pública en què fitxa ciutadans insurrectes mitjançant eixa tecnologia<sup>98</sup>: “El rostre és l'últim bastió de la privadesa. A la Xina hi ha una llista roja per als que es troben al poder i no volen que els vegem, així són invisibles per als sistemes de vigilància i s'eliminen dels registres de les càmeres. És molt revelador que no ser vist siga un privilegi exclusiu dels poderosos”, prossegueix la reportera del *New York Times*.



Segons publica el diari *El País*<sup>99</sup>, des que està operatiu el sistema ABIS s’ha utilitzat en unes 400 investigacions policials. Este diari també informa que el reconeixement biomètric es fa servir des de fa almenys un any en les fronteres de la UE per a registrar els ciutadans no comunitaris nousvinguts.

“Cap d’elles [les excepcions per a l’ús de biometria a la via pública] està pensada per a l’anàlisi i el control de la immigració o en la frontera. Una altra cosa és que, entés de forma àmplia, es faça servir de forma sistemàtica a la frontera, i això seria un problema perquè els Estats estarien incomplint la norma”. Sobre el seu ús en matèria de migració, Boix subratlla que travessar línies frontereres entre països sense els permisos no és delictiu: “Moltes organitzacions que defensen els drets humans entenen que la redacció actual sí que permet una entesa molt ampla, potser massa, que farà que els governs ho facen servir a frontera”. Per al professor titular de Dret Administratiu de la Universitat de València, la intenció de la UE no és instar a l’ús d’estes ferramentes per al control fronterer. “És una discussió en marxa”, i argumenta: “Si fem una interpretació estricta de la llei, a hores d’ara estaria prohibit. Si es dona una creixent pressió migratòria i la funció pública canvia d’opinió i decidix que sí que es pot emprar per a això, el que farà és canviar l’AI Act per a permetre-ho”.

## El cas d’iBorder Control: automatismes a la gestió fronterera

L’ús de la biometria per al control migratori no és una cosa nova. Un dels projectes pilot europeu no exempt de polèmica va ser iBorder Control, utilitzat del 2016 al 2019 i basat en el reconeixement d’emocions per a categoritzar les persones migrants a partir de dades biomètriques. Mitjançant “un mecanisme automatitzat de detecció de mentires”<sup>100</sup> analitzava les “microexpressions facials” i servia per a orientar “les autoritats frontereres en la presa de decisions”. Com diu Judith Membrives, hi ha evidència científica del gran marge d’error que hi ha a l’hora de reconèixer les emocions d’una persona, ja que és una cosa que canvia en cadascuna, i està molt marcada per qüestions com la seua educació o pràctiques culturals.

“Tot el que és Seguretat Nacional i Cossos Policials ha quedat catalogat com a alt risc, però amb poques prohibicions sobre això per la quantitat d’excepcions incorporades. Hi ha requisits de transparència, però no ho són de cara a la societat civil. Sí que s’ha establert una *database*, una mena de registre d’algoritmes, però no seran accés públic. Això ja passa prou amb Frontex, que és qui té l’externalització de les fronteres”. I afeg: “Europa està fent política pública centrada a fortificar les seues fronteres i per això utilitza la intel·ligència artificial, els sistemes algorítmics i el Big Data”. I menciona Eurodac<sup>101</sup>, la gran base de dades de totes les persones que demanen asil polític o refugi a Europa, com el gran instrument de seguiment digital d’Europa que quedaria fora del Reglament d’Intel·ligència Artificial.

99 G. Pascual, M. (2024, 28 de maig). *La Policía española ya usa en sus investigaciones un sistema automático de reconocimiento facial*. *El País*. <https://elpais.com/tecnologia/2024-05-28/la-policia-espanola-ya-usa-en-sus-investigaciones-un-sistema-automatico-de-reconocimiento-facial.html>

100 Romano, Andrea. (2023). *Derechos fundamentales e inteligencia artificial emocional en iBorderCtrl: retos de la automatización en el ámbito migratorio*. *Revista Catalana de Dret Públic*, 66, 237-252. <https://doi.org/10.58992/rcdp.i66.2023.3928>

101 European Asylum Dactyloscopy Database, Eurodac en su abreviatura oficial, es, según ChatGPT, una base de datos biométrica de huellas dactilares utilizada por los países miembros de la Unión Europea y otros estados asociados para gestionar y comparar las huellas dactilares de los solicitantes de asilo y los migrantes irregulares que funciona desde 2003.

# Transparència, tecnosolucionisme i canvi climàtic

## Diners públics, codi públic: el poder institucional i els algoritmes automatitzats

No s'espera de l'AI Act que siga garant dels drets humans perquè, com hem dit, no està pensada per a això. “El reglament d'usos de la IA és un reglament de seguretat industrial”, diferencia Boix. Això vol dir que és una norma de control semblant a les ITV dels vehicles o les inspeccions dels ascensors: “Establix unes regles bàsiques sobre la tecnologia, quines cauteles ha de tindre, quins mecanismes de control i d'inspecció hi ha per part dels poders públics, amb la col·laboració d'una sèrie d'auditories privades, per tal d'assegurar que totes estes normes s'estiguen complint. Les empreses tindran l'obligació, igual que nosaltres tenim l'obligació d'haver passat la ITV, d'haver passat l'auditoria corresponent”.

Per al professor de Dret Administratiu de la Universitat de València, esta llei estableix el marc, la pedra inicial sobre la qual els països europeus –o la mateixa institució europea– poden desenvolupar un altre tipus de normes que responguen a com han de ser els algoritmes, què passa quan afecten drets humans, com s'han de protegir les dades o la privadesa, i què fer quan estos algoritmes generen biaixos. En realitat, l'AI Act a penes toca això. “No està pensada per a controlar què fan les administracions i els poders públics amb la IA”, cosa que sí que es va produir amb el Reglament General de Protecció de Dades (RGPD). “Es podria pensar que amb el RGPD ja valdria o que podria ser traslladable”. El problema, segons Boix, és que no és el mateix controlar els usos que es dona als algoritmes d'automatització que fan agents privats –per a remirar en les nostres dades o per a oferir-nos anuncis en funció del nostre historial de consum–, que “quan els usem per a adoptar decisions que fins ara són competència de poders públics com el judicial o l'administració pública, que incidixen i afecten directament els nostres drets, garanties i estatut jurídic”.

En un dels primers esborranys sí que es consideraven d'alt risc –i, per tant, regulats– tots els sistemes algorítmics. En la versió final només es consideren d'alt risc els algoritmes que impacten en els processos democràtics, cosa que encara està per definir i aclarir. “Per a això tardarem uns anys, fins a veure com les autoritats de supervisió imposen el seu compliment a les plataformes”, exposa Cotino.

Fa temps que Andrés Boix analitza des d'una perspectiva jurídica el fet que cada vegada més siguen els algoritmes els que prenen les decisions que abans prenién els poders públics, els jutges o les autoritats administratives. N'hi ha diversos de famosos, com ara BOSCO, l'algoritme encarregat de dirimir si una persona té dret

al bo social o no. Quan la Fundació Civio<sup>102</sup> el va voler estudiar, el Govern es va negar a lliurar-li el codi font, decisió que va ser ratificada per un jutge.

Però n'hi ha més. A Euskadi tenen el seu propi programa informàtic en matèria de violència masclista, anomenat EPV-R, Escala de Predicció del Risc de Violència greu contra la parella, i porta en ús des del 2007. Ana Valdivia va auditar l'algoritme el 2022 i la seua conclusió<sup>103</sup> va ser clara: “Està molt mal dissenyat i funciona molt malament”, entre altres qüestions perquè la seua taxa d'error és del 53%: “De cada deu casos severos, cinc els etiqueten malament”. La investigadora de l'Institut d'Internet d'Oxford apunta que es va implementar perquè l'Ertzaintza no tenia prou experts en matèria de violència masclista.

Com explica Naiara Bellio per a *El Salto*<sup>104</sup>, els algoritmes dissenyats per a predir el risc que una persona torne a cometre un delictes s'han utilitzat durant dècades, molt abans de l'auge de la intel·ligència artificial. Es dissenyen en forma de qüestionari psicològic que avalua la persona segons les respostes. Una de les preguntes és la procedència de la persona sobre la qual hi ha la sospita, i s'explica com a estranger algú del Marroc o Colòmbia, però no algú de França. Si com pareix, este traspàs del poder de prendre decisions passarà dels poders públics a este tipus d'algoritmes, segons Boix, el mínim que podem fer és controlar i regular eixos sistemes.

A nivell estatal, la Policia Nacional fa servir VioGen des de fa 17 anys per a definir el risc de patir violència masclista d'una dona que denuncia la seua parella. Com precisa Boix, VioGen no fa servir –almenys en este moment– intel·ligència artificial, però igualment hi ha molta opacitat i una negativa clara des dels poders a donar tota la informació sobre este tipus d'algoritmes de forma transparent per tal que siguen fiscalitzats per la societat civil. Per a Boix, estem assistint a “un procés de substitució” a l'hora de condicionar la decisió que prenen els jutges.

## Del tecnosolucionisme a l'increment de la crisi climàtica

“Hi ha un biaix humà que és pensar que tot allò que ve de la màquina és objectiu i que la persona no ho és. Això és tecnosolucionisme nu i cru: creure que mitjançant la tècnica salvarem o arreglarem problemes estructurals”, subscriu Judith Membrives, i porta a col·lació la paradoxa de Jevons, una metàfora utilitzada en les anàlisis econòmiques segons la qual a mesura que s'incrementa l'ús eficient d'un recurs, també n'augmenta el consum total i, amb això, la despesa energètica. És a dir, que les millores en l'eficiència poden comportar una utilització més gran de recursos en lloc d'ajudar a reduir-los, cosa que sembla evident en matèria d'IA, com denuncia Ana Valdivia. La investigadora

102 El principal argument del Jutjat Central Contenciós Administratiu número 8 per denegar a Civio l'accés al codi font de BOSCO és que seria un perill per a la seguretat pública. Ho expliquen en una notícia publicada en la seua web el 10 de febrer del 2022 titulada *La Justicia impide la apertura del código fuente de la aplicación que concede el bono social*. <https://civio.es/novedades/2022/02/10/la-justicia-impide-la-apertura-del-codigo-fuente-de-la-aplicacion-que-concede-el-bono-social>

103 Un extracte de la conclusió de l'auditoria: *Les eines d'avaluació de riscos s'apliquen a vegades per a patir la manca de recursos humans. Tot i això, els algoritmes i les eines d'avaluació de riscos no poden pensar com les persones, com afirmava Dreyfus: “Els éssers humans tenen una intel·ligència intuïtiva que les màquines ‘raonadores’ senzillament no poden igualar”*. Es troba ací: Ana Valdivia, A; Hyde-Vaamonde, C; García-Marcos, J. (2022). *Judging the algorithm: A case study on the risk assessment tool for gender-based violence implemented in the Basque country*. King's College London (KCL), Regne Unit. Euskal Herriko Unibertsitatea (EHU), Biscaia, Espanya.

104 Bellio, N. (2023, 5 d'abril). *Los sesgos con los extranjeros en el algoritmo de violencia de género de la Ertzaintza*. El Salto. <https://www.elsaltdiario.com/policia/sesgos-extranjeros-del-algoritmo-violencia-genero-ertzaintza>



de la Universitat d'Oxford ens convida a tindre una perspectiva crítica i a no oblidar una cosa que al seu parer ha estat invisibilitzada, i que anomena “la materialitat de la tecnologia”. És un factor que cal tindre en compte, sobretot quan s'espera que la intel·ligència artificial ensenye la humana com acabar, per exemple, amb el canvi climàtic.

Un article en la revista *Nature*<sup>105</sup> recollia les declaracions del directiu d'OpenAI Sam Altman en què reconeix una cosa que els experts diuen des de fa anys: que la indústria de la intel·ligència artificial s'encamina cap a una crisi energètica imminent. Part de la investigació en IA es destina, per necessitat, a trobar fonts d'energia alternatives, així que no sorprén saber que Altman du invertint des del 2021 en empreses que treballen per a assolir la fusió nuclear. “Es calcula que una cerca realitzada mitjançant IA generativa consumix entre quatre i cinc vegades més energia que una cerca web convencional. En uns anys és probable que els grans sistemes d'IA necessiten tanta energia com nacions senceres”, escriu la investigadora i professora universitària Kate Crawford en *Nature*. No és només energia, també és aigua, però també són les traves de les tecnològiques a saber amb exactitud les despeses mediambientals dels seus productes: “Els costos planetaris totals de la IA generativa són secrets empresarials gelosament guardats”, prossegueix Crawford.

Algunes estimacions diuen que ja es dedica un 8% de l'energia mundial a la intel·ligència artificial. “Necessites minerals, però no només això. Els xips, quan els poses en un centre de dades que executa algoritmes d'intel·ligència artificial com ChatGPT, utilitzen molta aigua, molta energia i fa molt de soroll. Les comunitats que viuen a prop de centres de dades es queixen d'esta gran contaminació sonora. A més, la seua infraestructura té un cicle de vida limitat i cada cinc anys un centre de dades ha de rebutjar tota esta infraestructura tecnològica, cosa que té un impacte negatiu en les persones que viuen a prop dels abocadors, com passa a Ghana i a Kenya”, lamenta Ana Valdivia.

Segons un estudi<sup>106</sup> realitzat el 2019 per la Universitat de Massachusetts Amherst, entrenar una sola intel·ligència artificial amb arquitectures neuronals emet 284 tones de CO2, equiparable al que contaminen 45 europeus en tot un any, o a volar més de 2.800 vegades en el trajecte Madrid-Barcelona. Des que Sam Altman anunciara a principis d'enguany que la indústria de la intel·ligència artificial està a punt de provocar una crisi energètica, governs com el nord-americà ja treballen colze a colze amb les grans tecnològiques en solucions tan imaginatives com instal·lar xicotets reactors nuclears<sup>107</sup> als centres de dades que funcionen dia i nit sense parar. “Els grans models de llenguatge tenen una arquitectura molt gran. Un algoritme d'aprenentatge automàtic que t'ajude a triar qui contractar potser necessite 50 variables: on treballa el candidat, quin salari té ara, experiència prèvia, etc. La primera versió de GhatGPT té més de 175.000 milions de paràmetres”, il·lustra Valdivia. La seua recomanació: “Quan entrenes un algoritme que t'analitza la qüestió del canvi climàtic, tingues també en compte i analitza el teu propi algoritme, com està contribuint també a eixe canvi climàtic”.

105 Crawford, K. (2024). *Generative AI's environmental costs are soaring — and mostly secret*. Revista Nature. <https://www.nature.com/articles/d41586-024-00478-x>

106 Strubell, E.; Ganesh, A.; McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. College of Information and Computer Sciences University of Massachusetts Amherst. <https://arxiv.org/pdf/1906.02243>

107 G. Pascual, M. (2024, 30 d'abril). *Los centros de datos quieren tener sus propios reactores nucleares*. El País. Este extens i complet reportatge d'El País parla que ja hi ha acords de compra entre Google i els desenvolupadors de xicotets reactors nuclears. Des de Bill Gates fins a AWS, la divisió de computació al núvol d'Amazon, estan fent moviments similars. [https://elpais.com/tecnologia/2024-04-30/los-centros-de-datos-quieren-tener-sus-propios-reactores-nucleares.html?sma=newsletter\\_tecnologia&utm\\_medium=email&utm\\_source=newsletter&utm\\_campaign=tecnologia\\_2024.05.03\\_3](https://elpais.com/tecnologia/2024-04-30/los-centros-de-datos-quieren-tener-sus-propios-reactores-nucleares.html?sma=newsletter_tecnologia&utm_medium=email&utm_source=newsletter&utm_campaign=tecnologia_2024.05.03_3)

# La importància del codi de la IA

El 5 de novembre del 2024 se celebraren eleccions presidencials als Estats Units. Una de les particularitats del seu mecanisme electoral és el sistema digital de recompte de vot. Les paperetes virtuals van esperonar Donald Trump i els seus seguidors a estendre la idea de frau electoral quan fa quatre anys va guanyar Joe Biden. A Concord, New Hampshire, des de 1989 usaven màquines AccuVote, propietat de la macroempresa Dominion Voting Systems, el segon major proveïdor del país. Però això pot canviar. Com relata la revista tecnològica de l'Institut Tecnològic de Massachusetts, la *MIT Technology Review*, són tres les empreses que controlen aproximadament el 90% del mercat nord-americà de tecnologia electoral. Totes tres són de capital privat i mantenen el codi font privat i ocult, igual que la identitat dels seus inversors o els seus informes de salut financera.

Però en esta menuda localitat de New Hampshire s'ha presentat a la competició per la gestió del vot digital una quarta protagonista inesperada: VotingWorks, una empresa sense ànim de lucre amb 17 empleats, les màquines de votació dels quals ja operen en cinc xicotets comtats, tots a Mississipi. La firma ha adoptat l'enfocament oposat al de les altres tres grans tecnològiques: transparència, senzillesa i seguretat, són els seus tres principals reclams. Els seus comptes estan publicats en la seua pàgina web i cada línia de codi que alimenta les seues màquines es publica a GitHub, disponible perquè qualsevol pugua inspeccionar-la.

Algunes veus interessades es van afanyar a propagar el missatge que el codi obert podria ser perillós. "Si l'FBI estiguera construint un nou edifici, no posaria els plànols en línia", van dir, i el cofundador de VotingWorks, Ben Adida, va replicar: "Un sistema de votació de codi obert és aquell en què no hi ha secrets sobre el seu funcionament". La transparència és un factor fonamental per a recuperar la confiança en els equips automatitzats de votació: "Tot el codi font és públic per tal que el món el veja. Per què el 2023 estem comptant vots amb software propietari?"

La història de l'*open source* ha tingut molts capítols i ara estem davant d'un de nou. "Codi obert significa per a mi que no hi ha prejudis, que no hi ha segones intencions i que, al final, és un llegat al qual jo he contribuït en allò que he pogut, i si l'any que ve no estic ací, una altra persona haurà d'agafar-ne el testimoni. Això amb un codi tancat no ho tindràs. Jo crec que, sense l'*open source*, internet no s'hauria democratitzat", aporta Pau Peiró.



S'ha escrit i escoltat molt sobre la seguretat dels models oberts. Miguel Arana és contundent en el tema:

“Pràcticament tots els sistemes informàtics que tenim ara es basen precisament en models de programari obert, és a dir, tota la infraestructura d'internet, tot el que és important, funciona amb Linux i és programari obert, perquè s'ha demostrat que funciona millor, que és més segur, que genera menys problemes. I que produïx millors societats perquè compartim el coneixement, el repliquem i el reinventem, que és el que sabem fer els humans. Sempre han existit els defensors del coneixement tancat i propietari. Justament tots els sistemes de seguretat, per exemple, els criptocodis, tot allò que té a veure amb transaccions bancàries, el control de les nostres armes, etc., tot es basa en sistemes oberts. Un crea un codi i el primer que fa és publicar-lo i dir, així faig jo el meu codi, a veure qui el rebenta, a veure qui hi troba una errada. És així com es demostra si este codi és segur. Ningú que entenga una mica de tecnologia pot creure que el que és obert és pitjor. En cap sentit: ni en seguretat, ni com a desenvolupament, ni com a eficàcia. Tots els que defensen el codi privatiu és perquè tenen un interès econòmic i volen control, poder, monopoli. Ha estat així en tota la història d'internet”.

Arvind Narayanan, a qui ja hem citat diverses vegades, és professor de Ciències Computacionals de la Universitat de Princeton i director del Centre de Política i Tecnologies de la Informació d'esta universitat. La seua aproximació a la IA és molt interessant i convida a la reflexió: “[l'aprenentatge automàtic] és un tipus de tecnologia enganyosa en molts sentits, és molt fàcil enganyar-se un mateix. Així que molts dels que li estan fent bombo i platerets no només enganyen els altres, sinó que s'enganyen a si mateixos”. Des del seu departament fan una defensa fervent de totes les possibilitats que oferixen els models oberts, és a dir, aquells que en fan públic el codi font. Això s'enfronta frontalment a la defensa que es fa des de grans sectors de l'univers empresarial *tech*, per als quals la seguretat de la IA implica la restricció dels models oberts, i semblen oblidar que les eines, *per se*, amb codi publicat o privatiu, no són ni bones ni roïnes, sinó que depén de com les usem. El científic té una resposta que ja hem anomenat unes pàgines arrere: “Intentar crear un model d'intel·ligència artificial que no es puga utilitzar malament és com intentar crear un ordinador que no es puga utilitzar per a coses roïnes”.

## Intel·ligent, artificial i de codi obert

“No pots evitar solucions propietàries”, objecta Jordi Linares. I recorda la gran inversió que fan les empreses, de la mateixa manera que passa amb les farmacèutiques i les patents de medicaments. “Al marge dels sistemes propietaris, la salut dels sistemes oberts impulsats per universitats i institucions és excel·lent, fins al punt que són tan poderosos com els propietaris. Això és molt important perquè ens permet no haver de dependre dels sistemes propietaris i tindre una merescuda independència”. Este doctor en Informàtica de la UPV sosté que “és bo que hi haja els models oberts perquè garantixen que no estiguem a les mans de Microsoft, Google o Amazon, i que puguem tindre la llibertat d'utilitzar models per a aplicacions de qualsevol mena que no depenguen d'ells”. Però la batalla mai no ha estat equilibrada. Karlos G. Liberal pren

com a punt clau el moment del moviment en defensa del codi obert que canvia el programari lliure per l'*open source*. "Al programari lliure no hem aconseguit els espais d'emancipació que volíem. El que hem aconseguit és que Microsoft es quede amb tot el codi de GitHub perquè puguen entrenar el propi model d'IA amb el nostre codi".

Cory Doctorow<sup>108</sup>, pioner, agitador, escriptor futurista, va publicar un article en què retreia que no es poguera parlar de models lliures si no tenim la infraestructura, l'energia ni l'aigua que necessiten estes màquines per a funcionar. Sabem que estos grans models de llenguatge requereixen quantitats ingents de dades, energia i recursos naturals, però, com déiem, desconeixem la quantitat exacta de tot, perquè els processos d'ensinistrament són completament opacs: "Ningú ha tingut accés a la informació que se li ha donat, no hem participat en eixe procés, ni és gens *open*", prossegueix Linares. Una opció intermèdia a la realitat que planteja Doctorow és optar per aquells models pesats entrenats de manera més general que permeten un procés de *fine-tuning*, ajustament fi o personalitzat, del qual ja hem parlat, per tal d'instalar-lo al nostre equip i evitar regalar les nostres dades a les tecnològiques.

Això es pot fer tant amb alguns *foundation models* com amb models de llenguatge de molts tipus, des de clonació de veu fins a traducció, generació d'imatge, text, vídeo o 3D, etc. Pàgines com Hugging Face<sup>109</sup> compten amb més de mig milió de *foundation models* oberts amb tot el seu codi font publicat que es poden fer servir per a multitud de tasques i amb llicències més o menys restrictives. Entre les grans tecnològiques, Meta lidera el mercat dels models oberts fins i tot permetent llicència comercial en alguns casos. Per a la generació d'imatges, Stable Diffusion és un model lliure i obert que podem descarregar i *fine-tunejar* lliurement.

"Els models oberts, junt amb una fèrria comunitat activa de desenvolupadors, ens garantixen que hi haja alternativa als grans *foundation models* propietaris", aporta Linares. Les alternatives lliures, independentment del tipus de llicència que tinguen, ens oferixen transparència: en coneixem el codi, com i amb quines dades han estat entrenades i està disponible tota la documentació del model. A més, permeten la instal·lació en local, sense accés a internet, perquè les nostres dades no siguen compartides i perdem el control sobre elles. Però no són les opcions majoritàries ni moltes vegades les més senzilles a nivell usuari.

"Igual no tenim IA de codi obert, però sí que podem tindre IA de servei públic", opina Gala Pin. *Diners públics, codi públic*<sup>110</sup> és el lema d'una campanya de la Free Software Foundation. La diputada d'En Comú Podem porta a la conversa el Barcelona Supercomputing Center, un centre pioner en este camp que abasta des de la investigació fins a l'oferta de serveis de supercomputació, finançat amb diners públics. Com explica Gala Pin, oferix un servei de processament de dades i generació de procediments algorítmics que té vocació de servei públic i que permet a persones de l'acadèmia, científiques o a qualsevol, presentar-hi un projecte i, en cas que siga seleccionat, facilita la infraestructura tecnològica i les eines d'IA necessàries per a dur-lo a terme. "Des de la societat civil no hi ha possibilitat de recopilar totes les dades que tu necessites per a fer una intel·ligència artificial potent que tinga molta capacitat d'aprenentatge, ni màquines de computació de dades prou potents per a competir amb un Google o Amazon", però sí que es disposa d'eines institucionals per donar suport a la creació i acostar-la a altres sectors. "Jo crec que hi ha un debat no resolt, i

108 Doctorow, C. (2023, 12 d'agost). "Open" "AI" isn't. Medium. <https://doctorow.medium.com/open-ai-isnt-93c17a203aeb>

109 És la definició de ChatGPT: Hugging Face és una empresa i una plataforma de tecnologia especialitzada en eines i models d'intel·ligència artificial, amb un enfocament particular en el processament del llenguatge natural (NLP) i l'aprenentatge profund. Fundada el 2016, Hugging Face ha desenvolupat una comunitat activa i una extensa biblioteca de models i recursos d'IA que són àmpliament utilitzats tant per investigadors com per desenvolupadors. Compta amb un gran repositori de transformers i models preentrenats, multitud de datasets, i extensa documentació per a desenvolupadors, tot de codi obert i amb la premissa de compartir codi.

110 <https://publiccode.eu/es>



que la gent de l'*open source* no hem sabut resoldre, sobre com aplicar la filosofia des del principi fins al final a la intel·ligència artificial. Com la intel·ligència artificial amb xarxes neuronals que aprén per si mateixa, com això es fa *open source* del tot”.

En la línia de Gala Pin, hi ha propostes que van més enllà i parlen de la necessitat de convertir internet en un veritable servei públic. “La tecnologia mai no és neutral; sempre està condicionada per forces polítiques, socials i econòmiques, així com pels valors i les eleccions humanes. El naixement d’internet va estar molt influït per les filosofies llibertàries dels primers fundadors de Silicon Valley, i el nostre enfocament actual de la regulació tecnològica ha estat modelat predominantment pels desitjos neoliberals d’afavorir el creixement econòmic i el consumisme. Tot i això, estes ideologies no haurien de determinar els límits de la nostra imaginació. Tenint en compte tot allò que està en joc, és hora de preguntar-se si els models empresarials basats en el servei públic podrien oferir millors resultats per a la democràcia i els ciutadans”, escriu Helen Jay, directora de la secció de política de la cadena pública britànica Channel 4<sup>111</sup>.

L’escriptora i professora de Harvard Shoshana Zuboff va publicar el 2020 el llibre *La era del capitalisme de la vigilància. La lluita per un futur humà davant de les noves fronteres del poder*. Hi analitza com s’està produint una gran acumulació de riquesa i poder a Silicon Valley, on hi ha “mercats conductuals” en què es compren i es venen prediccions sobre el nostre comportament. Açò és, que utilitzen les nostres dades sense que ho sapiguem per a enganxar-nos als seus productes i serveis sense parar. Per a entendre la magnitud del que parlem, Zuboff aporta un exemple: un document filtrat de Facebook, empresa que ara es coneix com a Meta, va demostrar que el 2018 el seu centre d’intel·ligència artificial estava processant bilions de dades al dia per a produir al voltant de 6 milions de pronòstics de comportament.

Els algoritmes atresoren informació important sobre els nostres gustos i preferències i són capaços de *crear* necessitats de consum dels seus productes i serveis. Sense que a penes en siguem conscients, són els que trien la música que escoltem a Spotify, les sèries que veiem a Netflix i els productes que comprem a Amazon. L’enshittification d’internet té una artilleria poderosa darrere; el consum exacerbat que articula produïx grans beneficis econòmics a uns quants. I estos algoritmes corruptes controlen tota la xarxa, especialment les nostres xarxes socials. Com lamenta Marta Peirano en el seu llibre *El enemic coneix el sistema*, totes les ments acadèmiques i empresarials més brillants treballen en l’economia de l’atenció, en aconseguir el nostre *engagement*, el nostre temps i els nostres clics. Les notificacions push o l’*scroll* infinit i altres ocurrences tècniques de seducció són armes acuradament pensades per a retindre’ns el major temps possible en les seues aplicacions. I tenen efectes contradictoris en les *persones usuàries*: primer generen grans i addictius pics de dopamina, després acaben en quadres d’ansietat i addicció. Més adequat que persones usuàries seria, simplement, *consumidors*. Com diu la investigadora Paola Ricaurte, el món es convertix en una dada: “La colonialitat del poder a través de la tecnologia s’expressa mitjançant la captura algorítmica de la subjectivitat a través de l’extractivisme de dades i la mediació algorítmica que produïx models de món dels quals estem excloses”<sup>112</sup>.

111 Jay, H. (2024, 22 de març). *We Need a Public Service Internet to Free Us from Big Tech’s Grasp*. Scientific American. Helen Jay és doctoranda a la Universitat de Westminster, Anglaterra. [https://www.scientificamerican.com/article/we-need-a-public-service-internet-to-free-us-from-big-techs-grasp/?sma=tendencias\\_2024.03.26&utm\\_medium=email&utm\\_source=newsletter&utm\\_campaign=tendencias\\_2024.03.26](https://www.scientificamerican.com/article/we-need-a-public-service-internet-to-free-us-from-big-techs-grasp/?sma=tendencias_2024.03.26&utm_medium=email&utm_source=newsletter&utm_campaign=tendencias_2024.03.26)

112 La investigadora mexicana Paola Ricaurte va impartir una ponència al DecidimFest21 sobre extractivisme de dades, violència i colonialitat: [https://www.youtube.com/watch?v=LEIxHup6LQ&ab\\_channel=Decidim](https://www.youtube.com/watch?v=LEIxHup6LQ&ab_channel=Decidim)



# El bé comú, la privadesa i el capitalisme de plataformes

OpenAI es defineix com “una empresa de recerca i desplegament d’IA” la missió de la qual és “garantir que la intel·ligència artificial general beneficie tota la humanitat”. Segons les últimes estimacions<sup>113</sup>, els seus ingressos són de 1.000 milions de dòlars, amb una facturació de més de 80 milions de dòlars al mes. Cal destacar que segons alguns estudis<sup>114</sup>, un 23% d’adults nord-americans ha fet servir ChatGPT, que equival a uns 60 milions de persones, sobretot menors de 30 anys.

Les empreses que contracten la seua IA, els usuaris que paguen la tarifa –per exemple, la subscripció de pagament de ChatGPT4– i les dades personals que utilitza quan acceptem les seues condicions d’ús, són les principals fonts d’ingressos. Precisament OpenAI ha rebut una garrotada quant a gestió i extracció de dades personals a Espanya, on l’Agència Espanyola de Protecció de Dades (AEPD) paralitzà el març de 2024 l’activitat de Worldcoin, una empresa vinculada al seu CEO, Sam Altman, que portava des del juliol del 2023 escanejant l’iris a uns 400.000 espanyols, a canvi d’un lot de criptodivises que equivaldrien a uns 80 euros. Les dades recollides fins a la data per Worldcoin queden bloquejats, per la qual cosa no es podran tractar ni compartir en espera d’una investigació internacional que decidirà si és legal o no que una empresa privada aplegue este tipus d’informació biomètrica.

L’expert en dret digital Borja Aduara declarava la seua preocupació al diari El País<sup>115</sup>, ja que “no es posa el focus en el punt important: el problema no és que et donen diners pel teu iris, sinó si esta dada s’està tractant de forma correcta”. En total, i segons Marilín Gonzalo, ja porten 3,4 milions de dades recopilades. Però ¿per què vol Worldcoin i l’empresa matriu que hi ha darrere, Tools for Humanity Corporation, escanejar els nostres ulls? La resposta és que pretenen crear una identitat digital única per a cada persona al món utilitzant la tecnologia d’escaneig d’iris que permeta diferenciar la humana d’altres identitats creades per la IA. *No volem saber qui eres, només que eres únic*, diuen en l’apartat de privadesa de la seua pàgina web. També t’oferixen la possibilitat de guardar o no esta dada biomètrica, encara que advertixen que si tries l’opció d’esborrament de les dades, sí que es conserva un missatge que conté un codi del teu iris. Les denúncies sobre la legalitat dubtosa d’esta pràctica arriben també per l’opacitat i les poques explicacions que oferix sobre el com i el perquè del tractament d’unes dades tan sensibles. “OpenAI ara, d’open, no en té res. Fins al punt que desco-neixem de primera mà els detalls de l’arquitectura de xarxa neuronal que hi ha darrere de ChatGPT. El que sabem és per experimentació i per deducció”, critica Jordi Linares.

El nostre iris és un tret biomètric molt més precís que el reconeixement facial o que la petjada dactilar i que, mal usat, pot servir per a suplantar la identitat d’una persona. “A través de la mirada i de com dilata la pupiïlla una persona pots saber què li agrada, què li espanta, què li interessa i fins i tot certes característiques cognitives, com si té Parkinson”, indica a *El País* Carissa Véliz, professora de filosofia a la Universitat d’Oxford i autora del llibre *Privacidad es poder*.

113 Efrati, A. i Holmes, A. (2023, 29 d’agost). *OpenAI passes 1 billion revenue pace as big companies boost ai spending*. The Information. <https://www.theinformation.com/articles/openai-passes-1-billion-revenue-pace-as-big-companies-boost-ai-spending>

114 Tyson, A.; Kennedy, B. (2024). *Many Americans think generative AI programs should credit the sources they rely on*. Pew Research Center. [https://www.pewresearch.org/short-reads/2024/03/26/many-americans-think-generative-ai-programs-should-credit-the-sources-they-rely-on/?sma=newsletter\\_tecnologia&utm\\_medium=email&utm\\_source=newsletter&utm\\_campaign=tecnologia\\_2024.04.12\\_2](https://www.pewresearch.org/short-reads/2024/03/26/many-americans-think-generative-ai-programs-should-credit-the-sources-they-rely-on/?sma=newsletter_tecnologia&utm_medium=email&utm_source=newsletter&utm_campaign=tecnologia_2024.04.12_2)

115 G. Pascual, M. (2024, 8 de març). *Por qué el iris es el dato biométrico máspreciado*. El País. [https://elpais.com/tecnologia/2024-03-08/por-que-el-iris-es-el-dato-biometrico-mas-precioso.html?sma=tendencias\\_2024.03.12&utm\\_medium=email&utm\\_source=newsletter&utm\\_campaign=tendencias\\_2024.03.12](https://elpais.com/tecnologia/2024-03-08/por-que-el-iris-es-el-dato-biometrico-mas-precioso.html?sma=tendencias_2024.03.12&utm_medium=email&utm_source=newsletter&utm_campaign=tendencias_2024.03.12)



No és la primera vegada que l'AEPD actua així davant d'un cas de recopilació de dades biomètriques. El 2021 ja va multar Mercadona per instal·lar càmeres amb sistemes de reconeixement facial a 48 botigues sense el coneixement ni consentiment dels clients. Ana Valdivia parla de l'extensió dels "mecanismes de la datificació" segons els quals des que ens alcem fins que ens adormim, dins i fora de casa, estem produint dades, per exemple, amb els rellotges que van calculant les nostres pulsacions o comptant els nostres passos. "Tot això s'envia a bases de dades que alimenten algoritmes que van analitzant el nostre comportament". La privadesa ha sigut i continua sent una de les primeres damnificades des de la irrupció i l'extensió d'internet. Els dispositius d'emmagatzematge de dades estan molt estesos, però potser no totes les persones són conscients del que suposen. Carissa Véliz recorda que quan els nazis envaiïen un país, el primer que feien era apoderar-se dels registres locals com a primer pas per a controlar la població i, en particular, per a localitzar els jueus. La filòsofa proposa imaginar què passaria si arribara al govern un règim autoritari, i així ho explica a *El País*<sup>116</sup>: "Si poguera predir tots els nostres moviments, podria ser el començament d'un règim invencible. Perquè et faces una idea de com són de perilloses les dades personals, imagina't un règim com el nazi, però actualment, amb accés a dades en temps real sobre la teua ubicació, el teu perfil facial, la teua manera de caminar, la teua freqüència cardíaca, les teues idees polítiques, la teua afiliació religiosa i moltes coses més".

Com explica Mariano Sigman en el seu llibre i s'apuntava més amunt, l'avanç de la IA que veiem hui dia no seria possible si el suport tecnològic no haguera avançat alhora. Ací el nom clau és Nvidia, l'empresa tecnològica que ha estat definida com la firma més influent de Wall Street i el valor més important sobre el planeta Terra<sup>117</sup>. Nvidia anunciava al febrer del 2024 un augment del 265% en els ingressos trimestrals i uns beneficis de 12.285 milions de dòlars, xifres rècord per a la companyia. El 2023, el negoci d'IA va generar uns 15 mil milions de dòlars en ingressos, un 40% més que l'any anterior. De fet, va superar la divisió de jocs com la seua font d'ingressos més gran. En conseqüència, les accions s'han disparat més d'un 180% enguany. Els ingressos de Nvidia durant el conjunt de l'any creixen un 126%, fins a 60.922 milions de dòlars, segons els comptes comunicats a la Comissió de Valors i Borsa dels Estats Units (la SEC).

Nascuda el 1993, va començar fabricant xips gràfics i es va especialitzar com a proveïdora de targetes gràfiques per a videojocs fins que el 1999 va millorar notablement la tecnologia de les GPU, unitats de processament gràfic, cosa que va permetre l'impuls de les capacitats gràfiques dels ordinadors. El seu CEO, Hensen Juang, va optar el 2006 perquè els xips de Nvidia foren programables, fet que va permetre adaptar-los a usos diferents com la mineria de criptomonedes i, sobretot, aplicar-los als avanços en la investigació de la IA. Va ser Alphabet, empresa matriu de Google, la primera que va apostar per les GPU de Nvidia per a entrenar els grans models de llenguatge que desenvolupava. Este sofisticat maquinari necessita moltíssima energia per a funcionar, unes cinc vegades més que els processadors convencionals o CPU. L'entrenament dels LLM requereix desenes de milers de GPU operant dia i nit durant setmanes o mesos. Nvidia, sens dubte, lidera el camp del maquinari de desenvolupament d'IA i la seua col·laboració amb OpenAI ha estat, segons algunes fonts, fonamental per entendre l'avanç i el desenvolupament de models revolucionaris com ChatGPT, el gran exponent de la intel·ligència artificial generativa, màxima rival de professions creatives com la periodística.

116 Véliz, C. (2021, 12 de setembre). *Protejamos nuestros datos. No olvidemos cómo los usaban los nazis*. El País. <https://elpais.com/ideas/2021-09-12/protejamos-nuestros-datos-no-olvidemos-como-los-usaban-los-nazis.html>

117 Jiménez, M. (2024, 5 de febrer). *Nvidia supera las previsiones con unos resultados espectaculares gracias a la inteligencia artificial*. Cinco Días. <https://cincodias.elpais.com/companias/2024-02-21/nvidia-supera-las-previsiones-con-unos-resultados-espectaculares-gracias-a-la-inteligencia-artificial.html>



# Una manera de concloure

## Entre l'amença colossal i l'oportunitat

Potser encara estem a temps de decidir què volem que supose la irrupció de la intel·ligència artificial generativa tant en la nostra vida com en la professió periodística, si serà una amenaça massiva o una valuosa oportunitat per a la reinvençió. “Davant del periodisme de declaracions que tan en voga està, les màquines són imbatibles”. Antonio Lorenzo, periodista d'*El Economista* no en té cap dubte: “La IA generativa serà una amenaça colossal per als periodistes”, fins i tot estableix la data simbòlica de tres anys perquè realment estos models d'escriptura automàtica ocupen un seient a les redaccions. Els periodistes que escriuen com a màquines seran els primers a ser substituïts.

De nou és el moment de resignificar la professió, més enllà de l'arquetip de periodista-tertulià, per a ser realment el periodista especialista que persegueix la notícia, n'oferix una anàlisi encertada, aporta alguna cosa diferent i és capaç d'absorbir el context per a emmarcar-la d'una manera que resulte comprensible i desitjable per al públic. Lorenzo en diu “explicar eixa crònica que no siga capaç d'explicar la màquina”. Per a molts analistes, no queda tant perquè ni tan sols les notícies en exclusiva o les primícies importen massa. Allò que sí que romandrà és el to personal, les firmes que donen credibilitat i confiança. “El criteri, l'olfacte, l'astúcia i picardia professional, el diàleg amb la font, són les aportacions que fan irremplaçable el periodista”, afegí Flor Coelho.

QUI ACCIONA LA MÀQUINA?

## “Som intel·ligents perquè vivim”

La professora d'anglès del fill del neurocientífic Mariano Sigman va proposar este exercici a l'alumnat: escriuiu un paràgraf vosaltres mateixos, enfronteu-vos al desafiament de la pàgina en blanc, al repte de crear, de pensar, de discernir la millor manera de fer-ho. Després, passeu-ho a ChatGPT perquè ho millore. Quan tingueu els dos textos, compareu-los, vegeu quins canvis accepteu a ChatGPT i quins no. No té molt de sentit educar com si estos models no existiren o fins i tot prohibir-los, perquè ja sabem que no podem posar portes al camp, i a l'extens paratge de la IA molt menys, perquè internet i la tecnologia en general estan dissenyades per aprendre a botar-se les tanques, els tallafocs i quasi qualsevol classe de limitació. Això no lleva que el marc jurídic siga necessari perquè cada joc té les seues pròpies regles i les institucions públiques i els governs en conjunt han de ser garants i guardians dels drets de les persones al carrer i a la xarxa. Totes les persones entrevistades coincidixen en la importància d'educar, de formar-nos, de cultivar l'esperit crític, i més en el cas dels periodistes que ens dediquem a explicar, explicar o traduir el món que ens envolta.

El fort colp de la intel·ligència artificial i el seu impacte en quasi tots els àmbits de la nostra vida ens fa reflexionar i fer-nos preguntes. Com definim la intel·ligència, quines qualitats són pròpies del que és humà, com diferenciar-nos de les màquines, quin és el nostre paper i quin és i serà el seu. “La IA no pot emocionar. Les màquines només partixen del que ja estava clar, però únicament el creador, l'autor, és el que partix de la foscor, del dolor, i és capaç de desxifrar-lo amb esforç, és capaç d'emocionar”, reflexionava l'escriptor Lorenzo Silva, en eixe intent que es repetix per veure què ens diferencia, què ens fa millors. No queda tant perquè arribem a la definició que imagina Mariano Sigman: “Intelligent és tot allò que les màquines no poden fer”.

Judith Membrives concedix que és i serà una eina útil, però també posa nom als perills de la IA: “Diuen que busquen simular la intel·ligència humana, però és una visió de la intel·ligència molt estreta, no és contextual, no es té en compte tota la part física o material nostra que ens fa intelligents. Perquè nosaltres no només som intelligents perquè llegim, som intelligents perquè vivim en un context, ens relacionem amb persones. Tot això és aprenentatge, tot això no ho té la IA. Llavors es basa a codificar l'experiència humana, que és una cosa que segurament no és codificable, perquè no és només quantificable, també és qualitativa. Pensem que pot fer absolutament tot allò que fa un humà, li donem eixe estatus i a més permetem que les empreses que l'estan desenvolupant ho facen sense mecanismes de transparència consensuats, com ha estat i s'ha fet fins ara tota la producció del coneixement científic”.

El procés d'entrenament d'estos grans models de llenguatge (LLM) es fa a imatge i semblança del procés de l'aprenentatge del primer llenguatge, que és per inferència estadística, com ja s'ha dit. Els bebés no saben gramàtica ni tampoc els cal, no distingixen una preposició d'un substantiu, però creen una mena d'estructura estadística en què, per probabilitat i a base d'escoltar centenars o milers a vegades les frases de la seua família, van encertant, van millorant la sintaxi del seu llenguatge sense haver obert un llibre de text. “Aprenentatge estadístic”, en diu Mariano Sigman. El mateix que utilitza GPT.

L'ésser humà fa un segle que compartix la seua vida amb dispositius tecnològics més o menys sofisticats: l'*smartwatch*, el primer telèfon mòbil, l'Osborne 1 o primer ordinador portàtil d'11 kg de pes, la primera televisió en blanc i negre, la ràdio, els electrodomèstics. Abans d'això, les tecnologies primitives o iniciàtiques (el llenguatge, el foc, els estris de defensa, de caça o per a treballar el camp fets amb pedres i pells d'animals, la roda, la impremta). Tots ells han anat modelant el nostre dia a dia i cobrint buits, fins que en molts casos alguna cosa millor els ha anat substituïnt. Això ha suposat canvis en les nostres tasques i també ha implicat delegar-les als nous dispositius que anaven sorgint (ja no fem servir l'orientació, sinó que obrim Google Maps, ja no canviem les marxes si tenim un cotxe automàtic). La nostra història és la història de la substitució. Com més sofisticades són les tecnologies, més deleguem. Sigman dividix les tecnologies en dos grups: les que suplantem o substituïm les capacitats humanes i, per tant, atrofien les nostres pròpies capacitats fins a portar-nos al “sedentarisme cognitiu”, o aquelles que lleven fricció a les accions del nostre dia a dia i que multipliquen les nostres pròpies qualitats. La diferència entre optar per una bicicleta per a desplaçar-se o per un cotxe. El missatge de Sigman és apostar per una tecnologia “que no ens reemplace sinó que ens multiplique i ens permeta arribar a llocs on sols no arribaríem, però sent nosaltres els protagonistes i els qui prenem les decisions”.

El desafiament més gran de tota la intel·ligència artificial, segons escriu Sigman, és no deixar de ser qui som. Açò es pot aplicar fàcilment al periodisme: no perdre la nostra essència, la nostra capacitat crítica, continuar millorant les nostres qualitats, potenciar el valor (informatiu, veraç i també l'econòmic) de la nostra signatura. Que no trien ni la màquina ni l'algoritme, ni Netflix ni els reels d'Instagram, què vols llegir, escriure, comprar o veure a continuació. I sense oblidar que l'expansió voraç de la intel·ligència artificial no durarà eternament, perquè res no creix infinitament. Totes



les bombolles exploten, o com diuen Narayanan i Kapoor, “cap tendència exponencial no pot continuar indefinidament. Però pot ser difícil predir quan una tendència tecnològica es troba a punt d’estancar-se”.

Esta investigació només és un punt i seguit. Encara que, com hem vist, la història de la intel·ligència artificial es remunta diverses dècades arrere, el moment àlgid que protagonitzem comença tot just ara. Ens trobem en la cresta d’una ona que inevitablement trencarà en algun moment sobre les roques i no sabem bé quins trossos quedaran a la superfície. “George Orwell temia aquells que prohibien els llibres. Aldous Huxley temia que no hi haguera raons per a prohibir llibres perquè no quedava ningú que els volguera llegir”, va escriure Neil Postman el 1985 al seu llibre *Amusing Ourselves to Death*. La irrellevància del periodisme és una amenaça més gran que qualsevol supercomputadora *omnimodal* d’avançades xarxes neuronals sintètica, però de la professió dependrà diferenciar-se de l’*enshittification* o buscar intel·ligents excuses artificials amb què justificar la seua mediocritat ètica i informativa.



Internet no és més important que l'emergència climàtica, la justícia de gènere, la justícia racial, el genocidi o la desigualtat. Però Internet és el terreny on lliurarem estes lluites. Sense una Internet lliure, justa i oberta, la lluita està perduda abans de començar.







# Bibliografía

**Cano Orón, L., López Meri, A. (2024).** *Introducción al uso de la IA en periodismo: Guía de referencias y modos de uso.* Publicacions de la Universitat de València. <https://omp.uv.es/index.php/PUV/article/view/730>

**Andrés Pedreño Muñoz, A.; González Gosálbez, R.; Mora Illán, T.; del Mar Pérez Fernández, E.; Ruiz Sierra, J.; Torres Penalva, A. (2024).** *La IA que revolucionará el periodismo.* Grupo 1million bot.

**Peirano, M. (2019).** *El enemigo conoce el sistema: Manipulación de ideas, personas e influencias después de la economía de la atención.* Editorial Debate.

**Vázquez-Barrio, T.; Salazar García, I. (ed). (2023).** *Inteligencia artificial, periodismo y democracia.* Tirant Humanidades.

**Sigman, M., & Bilinkis, S. (2023).** *Artificial: La nueva inteligencia y el contorno de lo humano.* Editorial Debate.

**Diversos autores (2024).** *Propuestas para regular la Inteligencia Artificial desde una perspectiva de derechos.* Conclusiones II Jornades DAR - Democracia, Algoritmos y Resistencias.

**Diversos autores (2023).** *Biko Insights #6. Tendencias digitales para un mundo en constante cambio.* Jakala - Biko.

**Diversos autores (2020).** *Tecnológicas.* Revista d'Eldiario.es.

**Diversos autores (2023).** *Inteligencia artificial: Riesgos, verdades y mentiras.* Revista d'Eldiario.es.

**Cabrera, M. (2024).** *Tendencias: El año de las historIAS.* Story Baker.

# Glossari d'entrevistes

(fetes entre gener i març de 2024)

**Ujué Agudo**, doctora en psicologia i **Karlos G. Liberal**, data activista i *ciberpunk*, investigadors a BikoLabs.

**Miguel Arana**, investigador en Machine Learning Causal de la Universitat de Cranfield, al Regne Unit.

**Andrés Boix**, professor de Dret Administratiu de la Universitat de València.

**Mauricio Cabrera**, periodista mexicà i terapeuta de continguts creador de la *newsletter* Story Baker.

**Juan Candela**, responsable de Digital de *Las Provincias*.

**Flor Coelho**, periodista de *La Nación*, Argentina.

**Lorenzo Cotino**, catedràtic de Dret Constitucional de la Universitat de València.

**Iván Esteve**, director d'Informatius d'À Punt.

**Marilín Gonzalo**, periodista de *Neutral*.

**Javier Lifante**, cap de Digital d'À Punt.

**Jordi Linares**, doctor en Informàtica i professor de la UPV.

**Ricardo León**, advocada d'Algoritme Legal, especialitzada en Dret de les noves tecnologies i negocis.

**Antonio Lorenzo**, periodista d'*El Economista*.

**Judith Memvibres**, responsable de polítiques digitals a Lafede.cat i part d'Algorights, a més d'experta en l'*AI Act*.

**Pau Peiró**, responsable per part del SAMC del projecte 'Recerca i desenvolupament de sistemes intel·ligents avançats per a l'accessibilitat, catalogació i indexació de continguts textuais, radiofònics i audiovisuals' d'À Punt.

**Gala Pin**, diputada d'En Comú Podem al Congrés dels Diputats.

**Carmela Ríos**, periodista experta en xarxes socials, periodisme mòbil i desinformació.

**Íñigo Roy**, cap de Digital del *Levante-EMV*.

**Marta Suárez-Mansilla**, advocada d'Algoritme Legal, especialitzada en dret de les noves tecnologies i negocis.

**Ana Valdivia**, Professora investigadora en IA, Governança i Polítiques de l'Oxford Internet Institute de la Universitat d'Oxford.

# Glossari de termes

## A

### Algoritme

És un conjunt d'instruccions que un ordinador segueix per a fer una tasca. Els algorismes s'utilitzen en totes les àrees de la informàtica, inclosa la IA.

### Aprenentatge automàtic

Un subcamp de la IA que se centra en el desenvolupament d'algorismes que poden aprendre de les dades i millorar el seu rendiment amb el temps.

### Automatització

És l'ús de màquines per a fer tasques que abans es feien per part d'humans. L'automatització s'utilitza en molts sectors per a millorar l'eficiència i la productivitat.

## I

### Intel·ligència artificial (IA)

És un camp de la informàtica que busca crear màquines que puguin imitar les capacitats cognitives humanes, com ara l'aprenentatge, el raonament i la resolució de problemes.

### IA generativa

És un camp de la IA que se centra en el desenvolupament de models que poden generar contingut nou, com ara text, imatges o codi. La IA generativa es fa servir per a tasques com la creació d'art, la generació de música i la traducció automàtica.

### Intel·ligència Artificial General (AGI)

És un tipus hipotètic d'IA que seria capaç d'executar qualsevol tasca intel·lectual que pugui fer un humà. L'AGI encara no s'ha aconseguit, però és un objectiu de recerca a llarg termini per a molts científics de la IA. En la legislació europea (l'*AI Act*) es parla d'IA de propòsit general i IA de propòsit general amb risc sistèmic.

## L

### LLM

Large Language Models, models massius de llenguatge, són un tipus d'intel·ligència artificial, un programari avançat que s'ha entrenat amb una gran quantitat de text per a entendre i generar llenguatge humà. És la tecnologia que hi ha darrere dels xatbots, capaços de llegir, escriure, traduir i respondre preguntes en llenguatge natural, com ho faria una persona. Alguns exemples de LLM serien GPT d'OpenAI (Generative Pretrained Transformer), LLAMA (Large Language Model Meta) o PALM2 (Pathway Language Model 2) de Google AI).

## M

### Machine learning (ML)

És un subcamp de la IA que se centra en el desenvolupament d'algorismes que poden aprendre de les dades i millorar-ne el rendiment amb el temps.

### Màquina automatitzada

És una màquina que pot fer una tasca sense la intervenció humana. Les màquines automatitzades s'utilitzen en molts sectors, com ara la fabricació, l'agricultura i l'atenció mèdica.

### Models fundacionals

L'*AI Act* definix els *foundation models* com "un desenrotllament recent en què els models d'IA es desenvolupen a partir d'algorismes dissenyats per a optimitzar la generalitat i la versatilitat del resultat. Estos models sovint s'entrenen amb una àmplia varietat de fonts de dades i grans quantitats d'informació per a executar una àmplia gamma de tasques posteriors, incloses algunes per a les quals no van ser específicament desenvolupats ni entrenats. Estos sistemes poden ser unimodals



o multimodals, entrenats mitjançant diversos mètodes, com ara l'aprenentatge supervisat o l'aprenentatge per reforç". El terme *foundation models* va ser encunyat per investigadors de l'Institut d'Intelligència Artificial Centrada en l'ésser humà (HAI, per les sigles en anglés) de la Universitat de Stanford. La seua notorietat va arribar amb la publicació de l'article "On the Opportunities and Risks of Foundation Models" el 2021, liderat per Percy Liang i altres col·laboradors. Este treball va conceptualitzar i formalitzar l'enfocament d'estos models com una base unificadora per a diverses aplicacions d'intelligència artificial, tant pel seu potencial com pels riscos associats.

Alguns exemples serien:

- Gopher: desenvolupat per DeepMind, va ser un dels primers models a demostrar la capacitat de fer múltiples tasques.
- LaMDA (Language Model for Dialogue Applications): desenvolupat per Google AI, s'especialitza en la generació de diàlegs realistes i coherents.
- MUM (Multitask Unified Model): desenvolupat per Google AI, és un model multimodal que pot processar informació de text i imatges.

## P

### Processament del llenguatge natural (NLP, Natural Language Process)

És una branca de la intel·ligència artificial que permet a les màquines comprendre, interpretar i generar llenguatge humà. El seu objectiu és facilitar la interacció entre humans i ordinadors utilitzant el llenguatge natural, com quan parlem o escrivim. Les aplicacions de NLP inclouen assistents virtuals, traducció automàtica i anàlisi de textos.

### Processament natural del llenguatge (NLP)

És un camp de la IA que se centra en la interacció entre els ordinadors i el llenguatge humà. El NLP s'utilitza per a tasques com ara traducció automàtica, anàlisi de sentiment, resum de text i generació de llenguatge natural.

## S

### Self-supervised learning (SSL)

És un tipus d'aprenentatge automàtic en què els models s'entrenen amb dades sense etiquetar. L'SSL s'utilitza per a tasques com ara la representació de llenguatge, la detecció d'anomalies i la generació d'imatges.

### Sistemes algorítmics

Són sistemes que utilitzen algorismes per a prendre decisions. Els sistemes algorítmics s'utilitzen en moltes àrees, com ara la banca, la justícia penal i l'atenció mèdica.

## T

### Transformer

Un tipus d'arquitectura de xarxa neuronal dissenyada específicament per al processament del llenguatge natural (PLN). Va ser introduït en un document de recerca de Google AI el 2017 i ha revolucionat el camp del PLN. En essència, un *transformer* és un model basat en l'atenció. Això significa que, en lloc de processar les paraules d'una oració de forma seqüencial (una darrere l'altra), el *transformer* analitza les relacions entre totes les paraules de l'oració simultàniament. Com si poguera parar atenció en totes les paraules alhora per a entendre el context complet de l'oració. Per a aconseguir-ho, utilitza primer un codificador que processa l'entrada (per exemple, una oració) i genera una representació interna. Després, un descodificador usa eixa representació per a generar-ne l'eixida (per exemple, una traducció o un resum). A més, tant el codificador com el descodificador contenen capes d'atenció que permeten el *transformer* "parar atenció" a les parts més rellevants de l'entrada per a cada part de l'eixida.

## V

### Visió per computació

La visió artificial o visió per ordinador és una branca de la intel·ligència artificial que permet a les màquines interpretar i comprendre imatges o vídeos de manera similar a com ho fan els humans. Per això utilitza algorismes i tècniques avançades que permeten extraure informació útil de

dades visuals, com fotografies o seqüències de vídeo. Així poden identificar i localitzar objectes en una imatge i reconèixer patrons de les imatges o vídeos. És la tecnologia darrere del reconeixement facial i també dels vehicles autònoms, diagnòstics mèdics, realitat augmentada, etc.

X

**Xarxa neuronal**

Una xarxa neuronal és un model computacional inspirat en el cervell humà, compost per capes de nodes interconnectats (“neurons”) que processen dades en múltiples capes per a reconèixer patrons complexos. S'utilitza àmpliament en tasques com ara el reconeixement de veu i imatge. Hi ha dos tipus principals:

- Xarxa Neuronal Convolucional (CNN), un tipus de xarxa neuronal dissenyada per a processar dades estructurades en forma de quadrícula, com les imatges. Les CNN són efectives per a tasques de visió per ordinador com el reconeixement d'objectes i la classificació d'imatges.
- Xarxa Neuronal Recurrent (RNN), un tipus de xarxa neuronal que és efectiva per a processar dades seqüencials, com ara text o sèries temporals. Les RNN tenen “memòria” que els permet considerar informació prèvia a les seues prediccions actuals, fent-les adequades per a tasques com la traducció automàtica i la predicció de text.



