39è CICLE DE CONFERÈNCIES 2023: 57-65 (2024)
Societat Andorrana de Ciències, Andorra
DOI: 10.2436/15.8060.35.5 ISBN: 978-99920-61-74-9

Els desafiaments de la intel·ligència artificial

· 21 d'abril del 2023 a les 7 del vespre
· Sala d'actes del Centre Cultural La Llacuna, Andorra la Vella
· Taula rodona
· Amb el suport de Creand© Crèdit Andorrà

Carles Gonzàlez i Mingorance

Professor de Filosofia al centre de batxillerat de l'Escola Andorrana

Es necessita alguna cosa més que intel·ligència per actuar intel·ligentment.
Fiódor Mikhailòvitx Dostoievski

No hi ha pràcticament res que no pugui canviar. Gairebé tot allò que és conegut està subjecte a processos de transformació i evolució. Aquesta és la idea que apuntala, com a principi metafísic, tot procés revolucionari. No obstant això, tot canvi comporta, també, l'aparició inevitable d'actituds contraposades. En aquest joc dialèctic, alguns abracen ferventment aquests canvis, uns altres els miren amb recel, preocupats davant la incertesa que sol generar la novetat. Actualment, el nostre món es troba immers en una nova revolució: *l'aparició* de la intel·ligència artificial. I de nou, davant la infinitud de possibilitats que la intel·ligència artificial presenta, el món s'ha tornat a dividir entre la sorpresa i la inquietud, entre l'entusiasme i l'alarmisme més visceral. A molts de nosaltres ens ha sorprès la capacitat que tenen per dur a terme tasques complexes com automatitzar processos, processar ingents quantitats de dades o generar imatges i textos de gran qualitat. Els més optimistes consideren fins i tot que quan arribi l'anomenada *singularitat tecnològica*[1] molts dels problemes que afecten la humanitat, com la pobresa, el canvi climàtic i les malalties, finalment se solucionaran. D'altres, en canvi, no són tan optimistes i adverteixen, com en el cas de Nick Bostrom, Elon Musk, Steve Wozniak o el ja difunt Stephen Hawking, que tots aquests avenços comporten un perill potencial per a la societat.

En les línies que segueixen em proposo, de manera breu, explorar i comprendre els desafiaments o problemes que la intel·ligència artificial té o pot produir en diferents àmbits, com el gnoseològic, l'econòmic, el jurídic i l'educatiu. A veure si així aconseguim aclarir tot això i evitem aquesta polarització tan maniquea, carrinclona i infundada que enfosqueix una qüestió extremament complexa.

El filòsof John Searle distingeix en el seu article *Minds, Brains, and Programs* entre intel·ligència artificial forta i intel·ligència artificial feble. La primera aspira a crear una màquina que pensi i

actuï com un humà. Aquesta és la creença fonamental que assumeixen molts dels especialistes i professionals del camp de la intel·ligència artificial quan asseguren que la intel·ligència humana pot ser formalitzada i reproduïda, i estableixen d'aquesta manera una analogia entre el funcionament del cervell humà i el d'una màquina. Aquesta idea la podem rastrejar en els inicis de la intel·ligència artificial quan el matemàtic Alan Turing va escriure el seu conegut article *Computing Machinery and Intelligence* i en la conferència de Dartmouth, esdeveniment organitzat per Marvin Minksy i John McCarthy l'estiu de 1956, on es va encunyar oficialment el terme *intel·ligència artificial*. No obstant això, igual que John Searle, considerem que el que tenim avui dia no són màquines amb una intel·ligència equivalent a la humana, capaç de comprendre i tenir consciència de si mateixes i del seu entorn, sinó sistemes d'intel·ligència artificial feble que malgrat realitzar amb resultats excel·lents tasques restringides a un àmbit molt concret, no pensen i ni actuen com l'ésser humà perquè no disposen d'intencionalitat, és a dir, no tenen una direccionalitat que es dirigeix als objectes del món exterior. Si li demanem com funciona, per exemple, a ChatGPT, un bot d'intel·ligència artificial especialitzat en la generació de textos, ens ofereix la resposta següent: *"GPT és un model de llenguatge basat en aprenentatge profund que s'entrena en grans quantitats de dades textuals. Durant el procés d'entrenament, el model aprèn patrons estadístics i relacions lingüístiques en el text. Després, quan se li presenta una nova entrada (una pregunta o una sol·licitud), utilitza aquests patrons i probabilitats per a generar una* resposta coherent i rellevant".

Aquesta resposta il·lustra perfectament la forma que té ChatGPT d'operar des d'un enfocament axiomàtic formal aliè al món exterior. Les paraules que selecciona per a la confecció d'aquesta resposta no es vinculen directament amb objectes del món exterior, sinó que se seleccionen basant-se en la probabilitat que unes certes combinacions de paraules són més comunes que unes altres. Això converteix ChatGPT en una mena de cec de Molyneux. En canvi, en el llenguatge humà les paraules sí que estan vinculades amb significats i objectes del món real, i és que el raonament humà depèn en bona mesura de la interacció que estableix amb el seu context (Dreyfus, 1972). Per tant, no pot reduir- se a regles formals i algorismes com en el cas de la intel·ligència artificial. Això és important d'assenyalar perquè, com veurem tot seguit, la falta d'interacció de la intel·ligència artificial, almenys de moment, amb l'entorn genera tota una sèrie de problemes que contribueixen a rebaixar una mica l'optimisme i les expectatives que alguns han dipositat en ella.

En els inicis la intel·ligència artificial es va basar en processos deductius per a la resolució de problemes lògics. Per exemple, Herbert Simon i Alan Newell van desenvolupar un programa pioner, anomenat "The Logic Theorist", que tenia l'habilitat de demostrar algun teorema lògic de manera més senzilla que la que havien aconseguit al seu moment Bertrand Russell i Alfred North Whitehead en la seva obra "Principia Mathematica". No obstant això, en la dècada dels 70, la intel·ligència artificial *lògica* es va enfrontar a un estancament (primer hivern) al no poder resoldre alguns problemes relacionats amb la traducció automàtica. La solució a aquests dificultats va passar pel desenvolupament d'un nou tipus d'intel·ligències artificials que aprenen mitjançant models d'aprenentatge que descansen en processos inductius com, per exemple, l'aprenentatge automàtic (de l'anglès, Machine Learning) i l'aprenentatge profund (Deep Learning).

L'ús de la inducció com a mètode té una llarga tradició en la història del pensament. No obstant això, també ha estat objecte de ressenyables objeccions per part de filòsofs notables com David Hume o Karl Popper. Val la pena, doncs, detenir-se uns instants i explicar aquestes limitacions que pot tenir l'aprenentatge automàtic i l'aprenentatge profund, a causa de l'ús de processos inductius en la realització de les seves funcions.

El primer problema és que tenen dificultats per manejar anomalies o dades atípiques. Per exemple, els algorismes de Machine Learning estan dissenyats per aprendre de les dades d'entrenament i generalitzar patrons que els permetin fer prediccions precises sobre noves dades. Aquests algorismes s'exerceixen bé quan les dades segueixen patrons clars que dibuixen tendències consistents, ja que poden ajustar els seus models en funció d'aquestes regularitats. No obstant això, quan es troben amb dades que són una *rara avis*, és a dir, que es desvien significativament de la majoria de les dades disponibles, els algorismes de Machine Learning poden presentar dificultats per integrar-los. I si alguna cosa ens va ensenyar Nassim Taleb en el seu llibre *El cigne negre*, és que el món és un lloc en constant canvi, on l'improbable –paradoxalment– és quelcom habitual i determinant, més del que suposem i estem disposats sovint a acceptar.

La segona limitació té a veure amb la fal·làcia de causa qüestionable. Els algorismes predictius de les intel·ligències artificials funcionen, com comentàvem abans, sense comprendre el context o les circumstàncies que envolten les dades amb els quals operen, i això dificulta la seva capacitat per identificar si la relació entre dos esdeveniments és realment causal o simplement una coincidència (Sema, Vincent i Grace, 2022). Per poder discriminar entre totes dues opcions, haurien de ser capaços de realitzar una exploració detallada del context i identificar possibles causes ocultes que poguessin estar influint en la relació entre aquests esdeveniments, però de moment aquesta tasca no poden fer-la de manera efectiva.

La tercera i última limitació té a veure amb el que comenta Erik J. Larson en la seva obra *El mite de la intel·ligència artificial*. Per a Larson, les intel·ligències artificials no pensen com l'ésser humà perquè, de moment, no tenen la capacitat de realitzar el que Charles Sander Peirce va anomenar *raonaments abductius*, un tipus d'inferència que, prenent com a base la descripció d'un esdeveniment o fenomen, estableix o arriba a una suposició explicativa (hipòtesi), que suggereix les possibles causes darrere del succés; i que nosaltres, els humans, realitzem constantment, cosa que és essencial, per exemple, per dur a terme la recerca científica.

Totes aquestes limitacions només mostren que, malgrat la seva capacitat per fer tasques complexes aparentment *intel·ligents*, les intel·ligències artificials encara no disposen d'un pensament conscient i d'una comprensió genuïna del que estan generant. Certament no coneixem què ens té preparat el futur, però de moment sembla que estem lluny d'aconseguir aquesta analogia amb la ment humana.

En un altre àmbit, les intel·ligències artificials també plantegen altres desafiaments addicionals relacionats amb qüestions ètiques i polítiques. És a dir, a més dels aspectes científics anteriorment analitzats, cal (pre)ocupar-se per les afectacions que poden tenir en la societat, veure fins a quin punt els temors associats amb les intel·ligències artificials estan justificats i presentar algunes de les solucions plantejades per fer-hi front.

Un ràpid tecleig en algun cercador web sobre les inquietuds i pors que provoquen les

intel·ligències artificials a escala social ens revela una ampli ventall de resultats. No obstant això, els resultats que més es repeteixen són la pèrdua generalitzada de llocs de treball i l'ús de dades esbiaixades o personals per a l'entrenament de les intel·ligències artificials.

L'arribada de les intel·ligències artificials ha instaurat un relat que augura importants i profundes transformacions en el mercat laboral. OpenAI, l'empresa responsable de ChatGPT, ha publicat un document en què analitza la incidència dels models de llenguatge gran (LLM) i els transformadors generatius preentrenats (GPT) en el mercat laboral dels Estats Units, i conclou que un 80% dels treballadors als Estats Units veurà almenys en un 10% les seves tasques influenciades per ChatGPT i tecnologies afins, mentre que al voltant del 19% dels treballadors podrien veure almenys un 50% de les seves tasques afectades (Eloundou, Manning, Mishkin, i Rock, 2023). De manera similar, un estudi realitzat per Goldman Sachs, un dels principals grups de banca d'inversió i de valors del món, afirma que podria afectar uns 300 milions d'ocupacions a tot el món, en automatitzar una quarta part del treball fet als Estats Units i Europa (Briggs i Kodnani, 2023). Aquest fenomen d'automatització és el que explica la preocupació de molts treballadors, que observen amb temor com moltes empreses, a la recerca de maximitzar resultats, puguin decidir reduir despeses prescindint dels seus serveis i condemnar-los d'aquesta manera irremeiablement a l'atur.

Aquest tipus de pors davant els avanços tecnològics no són cap novetat. Ja al segle XIX, els luddites es van oposar a les màquines, principalment telers industrials, que destruïen llocs de feina. Estaven justificades aquestes pors? Ho estan ara? Ho estan en part. Si bé és cert que els luddites del segle XIX van experimentar la pèrdua d'ocupació ocasionada per la mecanització de la indústria tèxtil, amb el temps l'adopció de noves tecnologies va portar a la creació de noves feines, un augment de la productivitat i una transformació econòmica i social sense precedents en la història de la humanitat. Actualment, fem front a un desafiament similar. Certament, l'impacte de les intel·ligències artificials en el mercat laboral està subjecte a la incertesa de predir el futur, per la qual cosa no podem afirmar de manera categòrica que pugui repetir-se l'esdevingut en el passat; no obstant això, tot apunta al fet que succeirà una cosa similar (Peng, Kalliamvakou, Cihon i Demirer, 2023) i (Noy i Zhang, 2023). En el curt termini, l'automatització de tasques repetitives i sovint tedioses podria resultar en la reducció d'ocupacions. Tanmateix, aquesta mateixa automatització té el potencial d'alliberar els treballadors de la realització d'aquestes tasques i permetre'ls, en canvi, concentrar-se en aquelles tasques més complexes i creatives, la qual cosa podria ser positiu si això acaba possibilitant la creació de nous llocs de treball altament especialitzats i de més valor agregat.

La creixent quantitat de serveis i prestacions oferts per institucions privades i públiques mitjançant l'ús d'intel·ligència artificial és ja una realitat difícil d'obviar. Aquests abasten des de predir nivells de criminalitat i reincidència (Angwin, Larson, Mattu i Kirchner, 2016), calcular els tipus d'interès en préstecs (Otero i Durán, 2018), fins a crear obres d'art (Navas, 2018), entre d'altres. No obstant això, a mesura que aquesta oferta de serveis s'amplia i es va normalitzant, apareixen com a contrapartida tota una sèrie de preocupacions vinculades amb l'ocupació i la selecció de dades per a l'entrenament d'aquests models d'intel·ligència artificial. Alguns d'aquests riscos a considerar són la possible conculcació de la privacitat dels usuaris, una exposició més gran a patir ciberatacs com el robatori d'identitat; la falta de consentiment

informat, ja que molts usuaris poden no ser del tot conscients que les seves dades personals són utilitzades per entrenar models d'intel·ligència artificial; la vulneració de drets de propietat intel·lectual, car les intel·ligències artificials no reconeixen l'autoria de les dades utilitzades en els seus entrenaments, i la sofisticació de les estratègies de manipulació, que poden arribar fins i tot a menyscabar la democràcia de manera substancial, especialment en escenaris electorals, tal com adverteix Connor Dunlop, responsable de polítiques públiques europees de l'institut Ada Lovelace. D'igual manera, l'ús que es fa d'algunes dades esbiaixades per a l'entrenament de les intel·ligències artificials pot donar lloc a la reproducció d'estereotips que impacten de manera negativa i generen danys morals en determinats col·lectius en condicions de vulnerabilitat (Suresh i Guttag, 2021), tal com ha succeït amb el programa informàtic Compas o l'algorisme SyRI (Ferrer, 2020).

Conscient tant d'aquestes problemàtiques exposades com dels diversos beneficis que, d'altra banda, les intel·ligències artificials poden oferir, la Unió Europea ha pres mesures concretes per regular l'ús de programes d'intel·ligència artificial que automatitzen la presa de decisions i que tenen el potencial d'afectar la vida i els drets de les persones (Roa i Sanabria, 2022), i exigeix "transparència algorítmica" i una major protecció de les dades personals a través del Reglament (UE) 2016/679 del Parlament Europeu i del Consell, de 27 d'abril de 2016, relatiu a la protecció de les persones físiques pel que fa al tractament de dades personals i a la lliure circulació d'aquestes dades, i la Resolució del Parlament Europeu, de 14 de març de 2017, sobre les implicacions de les macrodades en els drets fonamentals: privacitat, protecció de dades, no discriminació, seguretat i aplicació de la llei. No obstant això, és necessari assenyalar que una intel·ligència artificial *transparent* també podria generar certs problemes i tensions entre les normes de la legislació present i aquestes noves regulacions. Com, per exemple, fins a quin punt una empresa pot eximir-se d'oferir claredat sobre el funcionament d'un algorisme o els processos darrere de decisions automatitzades, emparant-se en la propietat intel·lectual?

Quan examinem els drets intel·lectuals, ens trobem davant el problema per determinar l'eventual autoria del que produeixen les intel·ligències artificials, perquè inevitablement sorgeix la qüestió si aquestes creacions alberguen algun indici d'activitat creativa que mereixi estar subjecta a resguard legal. Si es veiés confirmat això últim, sorgeix un segon problema entorn a l'absència de personalitat jurídica de l'agent identificat com a inventor. En aquest cas, no estem parlant d'un ésser humà, sinó d'una màquina, i la legislació actual no preveu l'opció d'atorgar drets d'autor a una entitat no humana. Fer-ho suposaria, sense cap dubte, un absurd en tota regla, ja que se li estaria pressuposant una autonomia que, de moment, li manca. A més, estaríem erròniament transferint la responsabilitat als sistemes d'intel·ligència artificial, quan en realitat aquesta continua sent dels agents humans. Per descomptat que podem delegar en una màquina unes certes decisions, però la responsabilitat ètica continuarà recaient en nosaltres. D'aquí ve que les paraules de l'historiador i pensador Yuval Noah Harari, quan afirma "que les conseqüències de cedir la decisió als algorismes poden ser terribles. Però si són una mica menys terribles que les nostres, els cedirem la decisió", siguin poc encertades, perquè, en realitat, el que estaríem transferint és la nostra sobirania a aquells que hi ha al darrere d'aquests sistemes d'intel·ligència artificial. Els algorismes no són neutrals ni infal·libles, sinó un reflex de les decisions i els interessos dels qui els creen i els instrueixen. Si en un futur

aquesta tecnologia planteja una amenaça a la humanitat, la responsabilitat haurà de caure sobre individus de carn i os. Per tant, en lloc de simplement cedir la decisió als algorismes, hem d'advocar per establir uns sòlids marcs reguladors que no eximeixin a ningú de les seves responsabilitats i garanteixin, en la mesura que sigui possible, la transparència en els processos de presa de decisions, a més d'assegurar-nos en tot moment que s'utilitzen en benefici de la humanitat i no en contra seva.

En el camp de l'educació, la irrupció de les intel·ligències artificials també ha dividit, en general, la comunitat en dos bàndols: els que pensen, com Emad Mostaque, CEO de Stability AI, que com a eina aplicada a l'educació aconseguirà resoldre tots els problemes que pateixen els sistemes educatius al voltant del món, i els que tenen una visió més catastrofista i consideren que la incorporació d'aquesta nova tecnologia a l'aula pot privar els alumnes de l'aprenentatge d'alguns processos cognitius importants. La meva posició sobre aquest tema és intermèdia. En funció de l'ús que li donarem, les conseqüències seran positives o negatives, i podran, fins i tot, com sovint succeeix, conviure totes dues opcions al mateix temps. Per tant, penso que és rellevant mantenir una actitud prudent quant a la inevitable integració de la intel·ligència artificial a les aules. Ens agradi més o menys, aquestes han arribat per quedar-se. L'oposició que generen és normal, la seva arribada ens obliga a repensar la nostra pràctica educativa de manera similar a com ho va fer al seu moment segurament l'arribada de la calculadora o d'internet. Llavors, igual que ara, es temia que perdéssim algunes habilitats, en deixar de sumar i restar a mà o de buscar informació en llibres i enciclopèdies. Però la veritat és que cada nova tecnologia ha portat reptes i beneficis propis, i la intel·ligència artificial en l'àmbit educatiu no en serà una excepció. Tancar els ulls no és una opció, més tenint en compte que aquest tipus d'eines seran presents en l'entorn laboral al qual s'incorporaran els nostres alumnes tard o d'hora. En aquest sentit, és essencial reconèixer que el que hem de fer és aprendre a conviure amb aquesta nova eina i guiar-los de tal manera perquè sàpiguen utilitzar-la de manera responsable i adequada, perquè, malgrat els desafiaments que planteja, són molts els beneficis que pot aportar.

Als docents, comptar amb aquesta mena d'eines ens pot ajudar, per exemple, simplificant la part més mecànica i burocràtica del nostre treball, possibilitant la inversió de temps addicional en aspectes crucials, com l'atenció als nostres estudiants. A més, a través de les interaccions quotidianes que estableixi l'estudiant amb la intel·ligència artificial, seria factible determinar-ne la velocitat d'aprenentatge i les àrees en les quals afronta desafiaments específics, cosa que facilitaria l'adaptació dels recursos didàctics segons les preferències i el nivell de desenvolupament competencial que cada estudiant mostri, la qual cosa generaria una experiència d'aprenentatge més efectiva. Això fins i tot obriria la possibilitat de prescindir dels exàmens tradicionals, perquè aquestes interacciones diàries permetrien portar un registre diari. D'altra banda, aquest nou enfocament també podria propiciar un canvi en l'organització dels espais escolars, potser permetria fer-ne agrupacions basades en el desenvolupament d'habilitats en lloc de per l'edat cronològica. Finalment, és important destacar que aquesta tecnologia també podria beneficiar els nostres estudiants, en proporcionar-los un assistent personal que els ajudés en la síntesi i resum de la informació, en la redacció de textos, així com en l'accés a una àmplia varietat de recursos i informació en línia només formulant una pregunta.

Ara bé, també existeixen alguns problemes sobre aquesta possible adopció de la intel·ligència

artificial a les aules. El principal podria ser la gradual supressió de processos cognitius, com la memòria, dins del context escolar. Internet i ara la intel·ligència artificial han estat encimbellats com a fonts de coneixement en substitució de la memòria. Però això suposa incórrer en un greu error: tenir accés a una gran quantitat d'informació, com la que es troba a internet o a través de la intel·ligència artificial, no garanteix que realment es comprengui o s'assimili tota aquesta informació sota la forma de coneixement. Perquè així sigui, la memòria, entre altres facultats, exerceix un paper fonamental en l'adquisició i construcció del coneixement.

En sintonia amb el que estic dient, el filòsof Plató relata en la seva obra *Fedre* l'arribada del déu egipci Theuth, que és conegut com el descobridor de conceptes com els números, el càlcul, la geometria i, especialment, les lletres. Un dia, el savi va mostrar a Thamus, rei d'Egipte, aquelles arts i li'n va exposar la utilitat amb la intenció de trobar l'aprovació del rei. Quan van arribar a les lletres, va dir Theuth: *"Aquest coneixement, oh Rei, farà més savis els egipcis i més memoriosos, perquè s'ha inventat com a fàrmac de la memòria i de la saviesa"*. Però el rei li va dir: *"Oh Theuth! A uns els és donat crear art, a uns altres jutjar què de mal o profit aporta als que pretenen fer ús d'ell. I ara tu, precisament, pare que ets de les lletres, per inclinació a elles, els atribueixes poders contraris als que tenen. Perquè és oblit el que produiran en les ànimes dels qui les aprenguin, en descurar la memòria, ja que, fiant-se de l'escrit, arribaran al record des de fora, a través de caràcters aliens, no des de dintre, des d'ells mateixos i per si mateixos. No és, doncs, un fàrmac de la memòria el que has trobat, sinó un simple recordatori. Aparença de saviesa és el que proporciones als teus alumnes, que no veritat. Perquè havent sentit moltes coses sense aprendre-les, semblarà que tenen molts coneixements, sent, al contrari, en la majoria dels casos, totalment ignorants, i difícils, a més, de tractar perquè han acabat per convertir-se en savis aparents en lloc de savis de veritat"* (traducció pròpia).

Però, per què és important la memòria i el coneixement per a l'ús de la intel·ligència artificial? La seva importància rau en el fet que els assoliments més destacats són aconseguits per aquells que posseeixen un profund domini del tema, la qual cosa permet formular preguntes adequades i discernir en les respostes rebudes entre allò que és vàlid i el que no ho és. Això, al seu torn, es tradueix en respostes cada vegada més pertinents i precises, amb la qual cosa es pot treure d'aquesta manera el màxim profit a una eina com ChatGPT.

Certament, amb l'arribada de la intel·ligència artificial es desplega davant nostre un futur incert, però és precisament aquesta incertesa la que fa que resulti fonamental adoptar una actitud prudent i evitar caure en comportaments excessivament polaritzats respecte a l'existència d'aquesta nova tecnologia. La intel·ligència artificial no ha de representar una amenaça per si mateixa, sinó una oportunitat per millorar la societat en conjunt, sempre que s'utilitzi de manera responsable.

Nota

1- Raymond Kurzweil (director d'enginyeria de Google) assegura, a tall de predicció, que cap al 2045 ocorrerà la singularitat tecnològica (moment en què la intel·ligència artificial superarà la intel·ligència humana). Però ja al 2029 les IA aconseguiran superar, per fi, el Test de Turing.

Referències

Algotive (2022, 29 de juliol). Historia de la Inteligencia Artificial, el Machine Learning y el Deep Learning. Algotive. https://www.algotive.ai/es-mx/blog/historia-de-la-inteligencia-artificial-el-machine-learning-y-el-deep-learning#ancla_3

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, 23 de març). Machine Bias.There's software used across the country to predict future criminals. And it's biased against blacks. Propublica,
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bostrom, N. (2017, 25 de febrer). *Superinteligencia: caminos, peligros, estrategias.* Teell Editorial, S.L.

Briggs, J. & Kodnani, D. (2023, 26 de març). The Potentially Large Effects of Artificial Intelligence on Economic Growth. Recuperat de https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf

Cortes, J. (2017, 9 d'octubre). Entrevista con Nick Bostrom. *El País.* https://elpais.com/retina/2017/12/02/tendencias/1512231406_905237.html

Dennis, M. A. (2023, 18 de setembre). Allen Newell American computer scientist. https://www.britannica.com/biography/Allen-Newell#ref733775

Dimensionia. (2023, 4 de maig). El impacto de la Inteligencia Artificial en la Educación: la Visión de Peter Diamandis. Recuperat de https://www.dimensionia.com/impacto-ia-educacion-segun-peter-diamandis/

Dreyfus, H. (1972), *What Computers Can't Do*. Recuperat de https://terrorgum.com/tfox/books/whatcomputersstillcantdo_acritiqueofartificialreason.pdf

Duffy, C. & Maruf, R. (2023, 18 d' abril). Elon Musk advierte que la IA podría causar la "destrucción de la civilización" aunque invierte en ella. CNN. https://cnnespanol.cnn.com/2023/04/18/elon-musk-advierte-ia-podria-causar-destruccion-civilizacion-trax/

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. Recuperat de https://arxiv.org/pdf/2303.10130.pdf

Ferrer, I. (2020, 12 de febrer). Países Bajos veta un algoritmo acusado de estigmatizar a los más desfavorecidos. *El País*. https://elpais.com/tecnologia/2020/02/12/actualidad/1581512850_757564.html

Fernández, J. (2023, 11 d'abril). Entrevista Connor Dunlop. La Expansión. https://www.expansion.com/economia-digital/protagonistas/2023/04/11/64351a6de5fdea60538b45e4.html

Harari, Y. (2019, 4 d'abril). Ei-Fei Li y Yuval Noah Harari en Conversación - La agitación de la IA que viene [Video]. YouTube. https://www.youtube.com/watch?v=d4rBh6DBHyw

iProfesional. (2023, 9 de Maig). Stephen Hawking advirtió que la inteligencia artificial puede derivar en el "fin de la humanidad". iProfesional. https://www.iprofesional.com/tecnologia/381541-stephen-hawking-predijo-los-riesgos-de-la-inteligencia-artificial

Kurzweil, R. (2005). *La singularidad está cerca: Cuando los humanos trascienden la biología.* Lola Books.

Larson, E.J. (2022, 17 d'octubre). *El mito de la Inteligencia Artificial: Por qué las máquinas no pueden pensar como nosotros lo hacemos*. Shackleton Books.

Locke, J. (2007). *Ensayo sobre el entendimiento humano.* Porrua.

López, M. (2023, 9 de Maig). Steve Wozniak, cofundador de Apple, tiene claro qué IA puede ser la que acabe con nosotros. Applesfera. https://www.applesfera.com/curiosidades/carrera-peligrosa-steve-wozniak-cofundador-apple-tiene-claro-que-ia-intenta-matarte

Mccorduck, P. (2004). *Machines Who Think. Recuperat de https://monoskop.org/images/1/1e/McCorduck_Pamela_Machines_Who_Think_2nd_ed.pdf*

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Recuperat de http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

Navas, S. (2018). "Obras generadas por algoritmos: En torno a su posible protección jurídica". *Revista de Derecho Civil, 5* (2018), 273-291.

Noy, S., & Zhang, W. (2023, 2 de març). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. Recuperat de: https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1.pdf

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (25 d'octubre). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. https://www.science.org/doi/10.1126/science.aax2342

Otero, L., & Durán, P. (2018). Fintech, blockchain y big data. En César García Novoa y Diana Santiago Iglesias (Dirs.), 4a Revolución industrial: impacto de la automatización y la Inteligencia artificial en la Sociedad y la economía digital (pp. 81-113). Thomson Reuters Aranzadi.

Parlamento Europeo. (2017, 14 de març). Resolución sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(ini)).
Peirce, C. S. (1878). *Deducción, inducción e hipótesis*. Recuperat de https://www.unav.es/gep/DeducInducHipotesis.html#nota1
Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023, 13 de febrer). The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. Recuperat de https://arxiv.org/pdf/2302.06590.pdf
Platón (2016, 4 de febrer) *Fedón / Fedro*. Alianza Editorial.
Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo. (2016). Relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).
Roa, M., & Sanabria, J. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. Revista Brasileira De Direito Processual Pena. V. 8 N. 1. Recuperat de https://revista.ibraspp.com.br/RBDPP/article/view/615/438
Searle, J. (1980). *Minds, brains, and programs. Behavioral and Brain Sciences*, 3(3), 417-457. Recuperat de https://home.csulb.edu/~cwallis/382/readings/482 searle.minds.brains.programs.bbs.1980.pdf
Sema K. S., Vincent, H., & Grace, Ch. (2022, 5 de maig). El caso de la Inteligencia Artificial causal. Stanford social innovation. https://ssires.tec.mx/es/noticia/el-caso-de-la-inteligencia-artificial-causal.
Suresh, H., & Guttag, J. (2021, 1 de desembre). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. Recuperat de https://arxiv.org/pdf/1901.10002.pdf
Taleb, N.N. (2011). *El Cisne negro*. Paidós Ibérica.
Turing, A. (1950, octubre). *Computing Machinery and Intelligence*. Mind, LIX (236), 433-460. doi:10.1093/mind/LIX.236.43