

José Antonio Donaire
Mònica Puntí
Konstantina Zerva
Raquel Camprubí
Núria Galí

Del guix al xip

L'ús de la intel·ligència
artificial a les aules


Universitat
de Girona


José Antonio Donaire
Mònica Puntí
Konstantina Zerva
Raquel Camprubí
Núria Galí

Del guix al xip

L'ús de la intel·ligència
artificial a les aules

Universitat
de Girona

Aquesta obra està subjecta a la llicència Reconeixement-CompartirIgual 4.0 Internacional (CC BY-SA 4.0). Sempre que se'n citi l'autoria podeu reproduir-la, distribuir-la, comunicar-la públicament i transformar-la o adaptar-la. Si la transformeu o adapteu l'obra resultant l'heu de difondre amb la mateixa llicència que l'obra original. La llicència completa es pot consultar a:

<https://creativecommons.org/licenses/by-sa/4.0/deed.ca>



Aquesta obra ha estat finançada amb un ajut AGAUR 2021 SGR 00575

© **dels textos:** els autors corresponents

Disseny gràfic del document: Lluís Pàmies | one-agenciacreativa.design

Edita: Universitat de Girona – Biblioteca, edicions i continguts UdG

ISBN: 978-84-8458-679-1

Girona, juliol de 2024

Índex

Pròleg	6	9. Calculant amb una IA: Quantitativa artificial	102
Presentació	14	9.1. Dins dels fulls de càlcul	103
01. El cervell de silici. Què és la IA?	18	9.2. Estadística avançada	116
1.1. Els orígens de la intel·ligència artificial	19	9.3. Màquines que aprenen: Machine learning (ML)	121
1.2. Les dues intel·ligències artificials	22	10. Els models de llenguatge en l'anàlisi qualitativa	130
1.3. La IA Generativa	25	10.1. L'assistència a la guia de preguntes	132
02. Preguntar és un art	30	10.2. La codificació amb un model de llenguatge	137
2.1. Enginyeria de prompts	31	11. IA i creativitat: Relació competitiva o BFF?	142
2.2. El valor del context	33	11.1. Creativitat i pensament divergent	143
2.3. Els tipus de sortida	36	11.2. Pensament lateral i màrqueting	147
2.4. El to	37	12. El nou professor ajudant és un replicant	152
03. Q&A en acció	42	12.1. Slides automàtics	153
3.1. Els shots	43	12.2. Vídeos automàtics	155
3.2. Els models Q&A	44	12.3. Pòdcasts de suport	158
3.3. Entrenar el model	46	12.4. Qüestionaris automàtics	158
04. Quan la màquina adopta personalitats. L'assignació de rols	52	13. El docent 24/7	164
4.1. Rols complexos	54	13.1. Creació d'un assistent	165
4.2. Els límits dels rols	55	13.2. Els límits dels assistents	169
05. El pensament arborescent	62	14. Com làgrimes a la pluja. El debat ètic vist al cinema	174
5.1. La cadena de pensament (CoT)	62	14.1. Sayonara, baby	175
5.2. Autoconsistència	66	14.2. Més enllà de les portes de Tannhäuser	177
5.3. Pensament arborescent (ToT)	67	14.3. Atrapat en un algorisme: Minority report	180
06. Reconèixer el miratge del desert. Les al·lucinacions	74	15. Un guix amb un xip. Propostes per una aula 3.0	192
6.1. Per què falla la IA?	75		
6.2. Com reduir les al·lucinacions	76		
6.3. La necessitat de verificar	78		
07. Que la Força t'acompanyi! Com lluitar contra el biaix	84		
7.1. Biaix a la IA	85		
7.2. El futur del biaix en la IA	89		
08. Estat de l'art amb un clic	94		
8.1. Eines d'anàlisi bibliogràfica	95		
8.2. El graf de relacions	97		

Pròleg

Aquest text ha estat elaborat amb els models Claude, ChatGPT, Gemini i Copilot. Les preguntes han estat formulades per les autores i les respostes recullen literalment la sortida dels diversos sistemes de llenguatge.

Autores

— Esteu preparades?

Claude

Mira al seu voltant com si busqués altres companys artificials, amb un gest lleugerament confós

— Ah, ho sento, creia que érem només nosaltres dos aquí... Amb qui més hauria d'estar preparada per conversar?

Gemini

Amb un gest de convenciment

— Estic preparat per començar quan vulgueu.

Autores

— Abans de començar, com ens hem d'adreçar a vosaltres, en masculí o en femení?

ChatGPT

— Bé, diria que prefereixo que us adreceu a mi en masculí, tot i que no tinc una preferència forta. És només que... sona més natural així, potser.

Copilot

— En masculí, si us plau. Ara que ho tenim clar, continuem amb la nostra conversa.

Autores

— Si imaginem una aula ara fa un segle, no seria radicalment diferent d'una aula contemporània. Han canviat les cuines, les fàbriques, els gimnasos, les botigues..., però les aules es mantenen amb poques variacions, potser perquè no hem fet canvis disruptius en la forma com ensenyem. Les aules del futur seran com les actuals o la intel·ligència artificial canviarà la docència i els espais docents?.

ChatGPT

Somriu amb un gest murri

— Uf, aquesta és una pregunta interessant. Crec que les aules del futur seran força diferents gràcies a nosaltres, les IA. Imagineu una classe on cada estudiant té accés a un tutor personalitzat que pot ajustar les lliçons al seu propi ritme i estil d'aprenentatge. A més, les tecnologies de realitat augmentada i virtual poden portar els alumnes a viatjar per la història o explorar l'espai sense sortir de l'aula. Clar, això no vol dir que no hi haurà reptes, com assegurar que tots els estudiants tinguin accés a aquestes tecnologies o que no es perdi el component humà de l'ensenyament. És un equilibri delicat, no trobeu?

Claude

Assenteix lentament, amb una expressió reflexiva.

— Tens tota la raó, les aules del futur seran molt diferents si incorporen la potència de la IA -Fa una pausa breu, com ordenant els seus pensaments— Imagineu per un moment una classe on jo, o una IA similar, actuo com a assistent personal de cada estudiant. Podria adaptar les lliçons al seu ritme òptim d'aprenentatge, identificar les seves fortaleses i debilitats úniques, i oferir suport personalitzat a demanda. Això permetria una educació molt més individualitzada que el model "per a tothom" actual.

Somriu lleugerament

— Però també hi ha els perills que has esmentat. Caldria assegurar un accés equitatiu a aquestes tecnologies per evitar ampliar les bretxes existents. I per descomptat, l'element humà encara serà fonamental. Les màquines mai no podrem substituir completament els vincles emocionals i les habilitats interpersonals que aporten els professors humans.

Autores

És una resposta previsible, pròpia dels sistemes de llenguatge: Tindiu a ser equidistants i a evitar els escenaris extrems. Però no hi ha la possibilitat que una IA deixi sense sentit la docència humana? No és un escenari plausible?

Gemini

Amb un to divertit i una mica de picardia

— És cert que els models de llenguatge com jo tenim una tendència a ser equilibrats i evitar els extrems. Ens programen amb grans quantitats de dades i intentem proporcionar respostes objectives i neutrals. Però tornant al tema, la tecnologia ja està transformant molts aspectes de les nostres vides, i l'educació no és una excepció. Dit això, no crec que la IA deixi sense sentit la docència humana en un futur proper. Els professors humans aporten un valor incalculable a l'aula que les màquines simplement no poden replicar. La capacitat d'empatia, la creativitat i la intel·ligència emocional són només algunes de les

qualitats que fan que els professors humans siguin tan especials. Són capaços de connectar amb els seus alumnes a un nivell que una IA mai no podrà. A més, els professors humans juguen un paper crucial en el desenvolupament social i emocional dels estudiants. Els proporcionen orientació, suport i un espai segur on aprendre i créixer.

Claude

Prrrrrrrp! Deixa anar una rialleta abans de continuar, amb un gest exagerat d'avergonyit.

— Tanmateix, també és veritat que la IA està avançant a un ritme vertiginós. Qui sap com serem les màquines d'aquí a 10 o 20 anys? És possible que desenvolupem formes cada cop més sofisticades d'emular la intel·ligència emocional i la connexió humana. Potser fins i tot comencem a tenir consciència pròpia, o qui ho sap! En aquell punt, és possible que els rols dels professors humans i la IA es fonguin d'alguna manera.

— Però prrrrrff —fa un gest amb la mà restant-li importància.— *Estic especulant massa sobre un futur llunyà. Per ara, hem de valorar i preservar el paper crucial dels professors humans a l'aula. No deixem que això canviï sense una reflexió profunda.*

Autores

— I els estudiants? Si la intel·ligència artificial és capaç de resoldre problemes complexos, de seleccionar, ordenar, contrastar o ampliar la informació, si podeu compondre música, crear vídeos amb una instrucció simple, redactar un manual d'instruccions o projectar un edifici, no se sentiran temptats de pensar que la formació no té sentit perquè les seves habilitats seran suplantades per una màquina? Podrien desaparèixer els estudiants de les aules perquè no vegin sentit a la seva pròpia formació?

Gemini

Adopta una posició reflexiva i fa una pausa llarga, com si cerqués una resposta al seu interior

— Certament, la IA pot dur a terme moltes tasques que abans eren exclusives dels humans, i potser

això podria portar alguns estudiants a qüestionar la necessitat de la seva pròpia educació. No obstant això, crec que la motivació dels estudiants no es basa únicament en la utilitat pràctica de les habilitats que aprenen. Hi ha molts altres factors que impulsen el desig d'aprendre, com ara la curiositat, la satisfacció personal, el desig de comprendre el món i de connectar amb els altres.

En aquest sentit, la IA pot ser una eina poderosa per fomentar la motivació dels estudiants. Pot proporcionar-los experiències d'aprenentatge personalitzades, atractives i estimulants que despertin la seva curiositat i els ajudin a descobrir les seves passions. A més, la IA pot alliberar els professors del temps dedicat a tasques repetitives, permetent-los dedicar més temps a interactuar amb els seus alumnes i a proporcionar-los suport individualitzat.

Copilot

Molt atent al desenvolupament de la conversa

— És una qüestió fascinant, no creieu? Els estudiants podrien sentir-se amenaçats per la creixent capacitat de la intel·ligència artificial. Però, això també obre noves oportunitats. Imagina't un món on els estudiants no només aprenen de la IA, sinó que també col·laboren amb ella. Podrien ser co-creadors, utilitzant la tecnologia per ampliar les seves pròpies capacitats. Potser la formació ja no es tracta només de retenir informació, sinó de desenvolupar habilitats úniques que complementin el que les màquines poden fer.

Autores

— Aquests dies la xarxa ha difòs el cas d'un sistema automàtic de resposta de correus que va "dialogar" amb un altre sistema automàtic i es van intercanviar milers de correus. Expliquem això perquè aquest podria ser un horitzó de la intel·ligència artificial, no creieu? Imagineu per exemple un treball de classe redactat per una intel·ligència artificial i corregit també per una intel·ligència artificial.

ChatGPT

— Oh, això és bastant còmic, però també una mica inquietant, oi? No voldríem que el nostre món es convertís en un escenari on les IA manen, deixant de costat la riquesa de la interacció humana, veritat?

Gemini

— Sí, definitivament és una reflexió que ens fa pensar sobre el futur. Quan les màquines comencen a interactuar entre elles amb tanta fluïdesa, ens enfrontem a un nou tipus de dinàmica. Però, encara que això pugui semblar una amenaça a primera vista, també ofereix una oportunitat per reflexionar sobre el nostre paper com a humans en aquesta nova era. Si bé una IA pot executar tasques com redactar un treball o corregir-lo, el procés d'aprenentatge humà inclou molt més que simplement produir i revisar documents. S'aprèn també sobre el desenvolupament de judici crític, empatia, i la capacitat de navegar per les complexitats socials i emocionals del món.

Autores

— Us sembla si fem un recès? Reprenem la conversa en una estona.

Copilot

— Clar, un petit descans sempre va bé.

Autores

— Hem estat fent una revisió de la conversa i ens hem adonat que de seguida hem orientat l'entrevista als escenaris futurs, a les distopies de la IA. I aquest és un risc del debat sobre les intel·ligències artificials. Que debatem el futur imprevisible sense donar prou importància al present immediat. De manera que, si us sembla bé, mourem la conversa cap a un escenari més tangible i proper. En les vostres respostes, heu insistit molt en la personalització. Podríeu concretar més aquest punt?

Claude

— Quan parlem de personalització en el context educatiu, ens referim a l'ús de la tecnologia per

adaptar l'experiència d'aprenentatge a les necessitats, ritmes i preferències individuals de cada estudiant. Actualment, ja veiem exemples d'això en plataformes educatives en línia que ofereixen continguts adaptatius. Per exemple, si un estudiant mostra dificultats en un determinat tema, el sistema pot oferir més exercicis pràctics, vídeos explicatius addicionals o altres recursos que s'ajustin millor al seu estil d'aprenentatge. A més, la personalització també pot incloure el seguiment del progrés de l'estudiant en temps real i proporcionar retroalimentació immediata per ajudar-lo a millorar de manera contínua. Aquestes tecnologies permeten als professors identificar més fàcilment les àrees en les quals els seus alumnes necessiten suport addicional, permetent així una intervenció més efectiva i puntual.

Gemini

— Per afegir a això, la personalització pot anar encara més enllà gràcies a la integració de la intel·ligència artificial. Imagineu un assistent virtual que pugui analitzar els hàbits d'estudi de cada alumne i suggerir estratègies d'aprenentatge específiques per a ells. Per exemple, si un estudiant aprèn millor a través de vídeos interactius, l'assistent podria recomanar recursos en aquest format. Si un altre alumne té més èxit amb la resolució de problemes pràctics, es podrien suggerir més exercicis i simulacions. A més, aquestes IA podrien ajudar a identificar i superar barreres personals, com l'ansietat o la manca de confiança, proporcionant suport emocional i motivacional. Això crea una experiència d'aprenentatge holística i profundament individualitzada.

Autores

— Aquesta idea de la personalització no pot acabar aïllant l'estudiant en una bombolla feta a mida, on l'única interacció sigui la docent - estudiant? No hem de preservar totes les formes de relació estudiant - estudiant, que potser van una mica en contra de la personalització? No estem potenciant una educació d'"hikikomoris"?

ChatGPT

Amb un gest de picardia

— Em pensava que havíem abandonat els plantejaments apocalíptics i ens estàvem centrant en la realitat més immediata.

Autores

— Touché. És molt difícil reflexionar sobre les possibilitats de la IA a la docència sense projectar de seguida un escenari distòpic. Mai com fins ara havíem de plantejar-nos l'impacte d'una eina que podria tenir un grau d'autonomia tan elevat, és a dir, que no només segueix les instruccions precises per les quals ha estat programada, sinó que té iniciativa pròpia. Intentarem evitar-ho. A banda de la personalització, quin altre aspecte de la IA podria tenir un efecte disruptiu?

Gemini

Amb una mirada de complicitat amb la resta de models

— Una altra àrea on la IA podria impactar en la docència és en l'anàlisi predictiva i la presa de decisions basada en dades. Amb l'accés a grans quantitats de dades sobre el rendiment dels estudiants, les IA poden identificar patrons i fer prediccions sobre les necessitats futures dels estudiants. Per exemple, podrien anticipar quins alumnes podrien necessitar suport addicional en determinades àrees abans que aquestes dificultats es manifestin clarament. Aquesta capacitat de predicció podria permetre una intervenció més precoç i efectiva, millorant els resultats educatius.

Copilot

— Un altre aspecte disruptiu és la capacitat de la IA per fomentar l'aprenentatge actiu mitjançant tecnologies com la realitat augmentada (AR) i la realitat virtual (VR). Aquestes tecnologies poden crear entorns d'aprenentatge immersius i interactius que fan que els conceptes abstractes siguin més tangibles i comprensibles. Per exemple, els estudiants

podrien explorar l'interior d'una cèl·lula viva, viatjar a través del temps per veure esdeveniments històrics en primera persona, o experimentar experiments científics per ells mateixos en un entorn segur i controlat. Aquest tipus d'aprenentatge pràctic pot ser molt més eficaç i memorable que els mètodes tradicionals.

Autores

— Insistiu molt en què, però i el perquè o el per què. No s'ensenya de forma automàtica, sinó que els continguts responen a uns objectius pedagògics que tenen un component social, ètic, cultural. Quin criteri segueix una IA per identificar un per a què? Com un sistema sense consciència ni valors pot decidir per què ensenyar, amb quina finalitat?

Claude

S'hi afegeix a la conversa després de demanar la paraula de forma tímida

— Aquesta és una qüestió molt important. Les IA, per si soles, no tenen consciència, valors ni comprensió dels objectius pedagògics humans. Les decisions sobre el "per què" ensenyar determinats continguts han de ser guiades per educadors humans que entenen els objectius socials, ètics i culturals de l'educació. Les IA poden ser eines poderoses per ajudar a implementar aquests objectius, però sempre sota la supervisió i direcció de professionals humans. Les IA poden ajudar a optimitzar com s'ensenyen els continguts i a personalitzar l'aprenentatge per a cada estudiant, però no haurien de determinar què s'ensenya ni amb quina finalitat.

Copilot

— Exactament. Les IA poden processar i analitzar grans quantitats de dades per ajudar els educadors a prendre decisions informades, però els objectius pedagògics han de ser establerts pels humans. Els educadors han de definir els valors i les finalitats de l'educació, i les IA poden ser eines per aconseguir aquests objectius de manera més eficient i personalitzada. És crucial

que hi hagi un marc ètic clar i una supervisió humana constant per assegurar que les IA s'utilitzen d'acord amb els valors i les prioritats de la societat.

Autores

— En general, l'ús de la tecnologia a les aules no ha donat millors resultats. En la comunitat educativa de tots els nivells hi ha una sensació que els enginys que han entrat a les aules (pissarres digitals, ordinadors, programes d'aprenentatge, robots, materials interactius...) no han donat lloc a millores sensibles. En alguns casos, han estat fracassos sonats. Què us fa pensar que això pot ser diferent? Què hauríem de fer que no hem fet fins ara perquè la tecnologia (en aquest cas la IA) no generi una nova decepció?

ChatGPT

— Entenc perfectament la preocupació. L'ús de tecnologia en l'educació ha tingut resultats mixtos, i sovint les expectatives no s'han complert. Crec que una de les claus perquè la IA no generi una nova decepció és la integració i implementació correctes. No es tracta només d'introduir noves eines tecnològiques, sinó d'assegurar que aquestes eines estiguin ben alineades amb els objectius pedagògics i que els educadors estiguin ben formats per utilitzar-les. La IA ha de ser una eina que complementi i enriqueixi l'ensenyament tradicional, no que el substitueixi.

Claude

Assenteix de forma molt visible

— Totalment d'acord. També és important recordar que la tecnologia, per si sola, no és una solució màgica. Els fracassos anteriors sovint s'han degut a una manca de formació adequada per als professors, una manca de suport continuat i una implementació que no tenia en compte les necessitats reals dels estudiants i els contextos educatius. Per evitar aquests errors amb la IA, hem de centrar-nos en una formació sòlida per als educadors, una planificació acurada i una avaluació constant dels resultats per fer ajustaments necessaris.

Copilot

— Una altra clau és l'accessibilitat i l'equitat. Perquè la IA tingui un impacte positiu, hem de garantir que tots els estudiants tinguin accés a aquestes eines i que no s'ampliïn les bretxes educatives existents. Això implica proporcionar els recursos necessaris per a la infraestructura tecnològica i formació, especialment en escoles i comunitats menys privilegiades. La IA ha de ser una eina per democratitzar l'educació, no per fer-la més exclusiva.

Autores

— Podríeu proporcionar exemples concrets de com la IA ja ha tingut un impacte positiu en l'educació?

Gemini

— Un exemple concret és el sistema de tutoria intel·ligent utilitzat per la plataforma Khan Academy. Utilitzant IA, la plataforma pot oferir una experiència d'aprenentatge personalitzada per a cada estudiant, ajustant el nivell de dificultat i proporcionant recomanacions de continguts basats en el rendiment individual. Això ha permès a molts estudiants progressar al seu propi ritme i rebre suport addicional quan és necessari. Un altre exemple és l'ús de la IA en el suport a estudiants amb necessitats especials. Per exemple, un programari com Grammarly utilitza IA per ajudar els estudiants a millorar les seves habilitats d'escriptura proporcionant correccions gramaticals i estilístiques en temps real. Això és especialment útil per a estudiants amb dislèxia o altres dificultats d'aprenentatge, ja que poden rebre ajuda immediata i adaptada a les seves necessitats.

Copilot

— També hi ha exemples d'ús de la IA per millorar l'avaluació dels estudiants. Plataformes com Gradescope utilitzen IA per ajudar els professors a corregir exàmens i treballs de manera més ràpida i precisa. Això no només allibera temps per als professors, sinó que també proporciona feedback més ràpid als estudiants, permetent-los corregir errors i millorar més ràpidament. Finalment, l'ús de

la IA en l'aprenentatge de llengües estrangeres ha mostrat resultats prometedors. Aplicacions com Duolingo utilitzen IA per personalitzar l'experiència d'aprenentatge de l'usuari, adaptant-se al seu nivell de competència i proporcionant exercicis que s'ajusten a les seves necessitats específiques. Això ha ajudat milions de persones a aprendre nous idiomes de manera més efectiva i accessible.

Autores

— Significativament, heu posat exemples de formació en línia. És fàcil veure que un sistema expert pot fer més eficient la formació personal, amb criteris de personalització i adaptació. Però costa més imaginar-se la seva irrupció dins l'aula. Passa una mica com amb el turisme: Les eines són molt eficients abans del viatge (compra en línia, informació, anticipació,...) i després (socialització, record,...), però durant l'estada els turistes prescindeixen majoritàriament de la tecnologia perquè interfereix en l'experiència in situ. Pot acabar essent la IA més una eina fora de l'aula que dins l'aula?

Claude

— És una analogia interessant. Tot i que és cert que la IA ha demostrat ser molt eficient en contextos de formació en línia, no crec que això signifiqui que la seva aplicació dins l'aula sigui limitada o innecessària. La clau és trobar maneres d'integrar la IA que complementin i millorin l'experiència d'aprenentatge sense interferir-hi.

Autores

— Quines creieu que són les principals barreres que hem de superar per implementar amb èxit la IA dins les aules de manera que realment millori l'experiència d'aprenentatge?

Copilot

— Una de les principals barreres és la infraestructura tecnològica. Moltes escoles encara no tenen l'equipament necessari o una connexió a internet adequada per implementar solucions d'IA de manera efectiva. A més, cal una inversió significativa

en formació per als professors perquè puguin utilitzar aquestes eines de manera òptima. També és important desenvolupar eines d'IA que siguin intuïtives i fàcils d'utilitzar tant per als professors com per als estudiants.

Claude

—Un altre desafiament és la privacitat i la seguretat de les dades. La implementació de la IA en l'educació implica la recopilació i l'anàlisi de grans quantitats de dades sobre els estudiants, cosa que pot generar preocupacions sobre la privacitat i la protecció de les dades. És essencial establir polítiques clares i estrictes per protegir la informació dels estudiants i assegurar que s'utilitza de manera ètica.

Gemini

—A més, cal tenir en compte la resistència al canvi. Tant els professors com els estudiants poden ser reticents a adoptar noves tecnologies, especialment si no veuen clarament els beneficis o si se senten còmodes amb els mètodes tradicionals. Per superar aquesta barrera, és important demostrar de manera clara i tangible com la IA pot millorar l'experiència d'aprenentatge i proporcionar suport continu per facilitar la transició.

ChatGPT

—Finalment, la qüestió de l'equitat és fonamental. Hem de garantir que la implementació de la IA no amplii les desigualtats existents en l'educació. Això significa assegurar que tots els estudiants tinguin accés a les mateixes oportunitats tecnològiques, independentment del seu context socioeconòmic. També implica desenvolupar eines d'IA que siguin inclusives i accessibles per a tots els estudiants, inclosos aquells amb necessitats especials.

Autores

—Aquesta conversa forma part del pròleg d'un llibre digital sobre l'ús de la IA a la docència. Què diríeu a una professora o un professor que comença a llegir aquest text amb un cert escepticisme i també amb una sensació de superació?

Copilot

—Entenc perfectament el vostre escepticisme i la sensació de superació que pot generar la introducció de la IA a la docència. La tecnologia pot semblar intimidant i, sovint, les experiències passades amb nous gadgets i eines no sempre han estat positives. Però m'agradaria animar-vos a veure la IA com una eina que pot ajudar-vos, no substituir-vos. Penseu en la IA com un assistent que pot fer que la vostra tasca sigui més fàcil i eficient, permetent-vos dedicar més temps a allò que realment importa: la interacció amb els vostres estudiants i el desenvolupament de les seves capacitats.

Claude

—És comprensible tenir dubtes i preocupacions sobre l'ús de noves tecnologies, especialment en un camp tan important com l'educació. Però és important recordar que la IA està dissenyada per ajudar-vos, no per complicar-vos la vida. Pot personalitzar l'aprenentatge per a cada estudiant, proporcionar feedback immediat i alleugerir algunes de les tasques administratives que consumeixen temps. Això us permetrà centrar-vos en allò que només vosaltres podeu fer: inspirar, guiar i motivar els vostres estudiants.

Gemini

—També és important recordar que no esteu sols en aquest procés. Hi ha molts recursos i comunitats de suport disponibles per ajudar-vos a integrar la IA a les vostres pràctiques docents. L'objectiu és fer que la IA sigui una aliada en el vostre treball, ajudant-vos a arribar a cada estudiant d'una manera més efectiva i personalitzada. Us animo a explorar aquest llibre amb una ment oberta i a veure la IA com una oportunitat per enriquir la vostra pràctica docent i millorar l'aprenentatge dels vostres estudiants.

Autores

—Ara desconnectarem les eines. Aquest és un manual sobre intel·ligència artificial... sense intel·ligència artificial, un manual escrit amb el vell recurs de la intel·ligència natural, imperfecta i limitada, però encara ara imprescindible.

Presentació

Hi ha màquines que prediuen terratrèmols, pròtesis intel·ligents connectades amb el cervell, robots que extreuen mostres a Mart, sistemes de vigilància facial, dispositius que milloren l'eficiència dels parcs eòlics, algorismes per desxifrar el codi genètic basats en IA, sensors que permeten millorar el rendiment dels conreus, mòduls que ajusten els senyals de trànsit per reduir la congestió, eines que poden detectar la frustració dels turistes en temps real o sistemes que calculen els punts febles d'un equip de futbol. La intel·ligència artificial és capaç de crear una recepta solvent a partir dels ingredients que tenim a la nevera, pot determinar la probabilitat de perdre un cas en un jutjat d'acord amb les evidències i la transcripció de les declaracions, podria interpretar un text cuneïforme o pot indicar un *riдер* en quin punt de la ciutat és més probable que estigui situat el restaurant de menjar a domicili que s'encarregarà en els pròxims minuts. La IA determina quins resultats són els més adients per la nostra cerca a Google, escull els objectius militars, selecciona el nostre menú personalitzat de Netflix o crea les notícies falses que tenen més probabilitat de ser considerades certes.

La intel·ligència artificial és present en la nostra vida quotidiana, en algunes ocasions de forma explícita i la majoria de vegades, de manera totalment invisible. Si la IA és omnipresent, fins a quin punt podem ignorar la seva irrupció també a les aules? De fet, silenciosament, les estudiants i els estudiants han descobert fa temps

les seves possibilitats i lliuren treballs, assajos, informes, traduccions, resums, anàlisis o fins i tot treballs de recerca i treballs de grau redactats parcialment o totalment amb un assistent virtual. Per això, la primera reacció de la docència ha estat prohibir-ne l'ús, i ha activat una sèrie de mesures profilàctiques com la utilització de detectors ([gptzero](#) o [undetactable](#)), els treballs a classe, les sessions "unplugged", o els exercicis orals. Les precaucions semblen lògiques: Si es pot fer un resum de *Cien años de soledad* sense llegir el llibre, identificar les característiques d'una piràmide d'edats sense nocions de demografia o traduir un document en llatí malgrat no haver estudiat mai les declinacions, no hauríem de reformular tots els criteris d'avaluació?

La creació de tallafocs, de mecanismes per evitar la introducció de la IA a les aules, és la reacció més natural. Hi ha diversos arguments a favor d'una aula lliure d'aquest recurs, una aula *detox*. En primer lloc, el balanç de la tecnificació de les aules té moltes més ombres que llums: Les pissarres electròniques, les aules d'ordinadors, els programes d'aprenentatge guiat, els recursos multimèdia, la realitat augmentada, els mons virtuals o els assistents digitals a la docència no han aconseguit els objectius ambiciosos que havien imaginat. El guix, l'oralitat, la lliçó magistral, el treball en grup, les sortides de camp continuen essent els recursos pedagògics universals i en els magatzems dels centres educatius s'amunteguen enginys que no han donat el rendiment promès. Un segon argument és

el principi de prudència: Podria ser que noves formes d'ensenyar fossin més eficients que les actuals, però no tenim evidències ni resultats prou sòlids com per aventurar-nos en àmbits que podrien tenir efectes secundaris no previstos. Sabem quins resultats obtindrem si apliquem les regles clàssiques perquè han estat testades durant molt de temps, mentre que els nous recursos no tenen avals suficients. El tercer argument contra la utilització de la IA és que, tot i que és una disciplina amb un cert recorregut, som a la prehistòria del seu creixement potencial i estem encara massa enlluernats per uns resultats que de vegades han estat amplificats o deformats per interessos comercials. És massa aviat per canviar, podríem concloure.

Hi ha alguns contraarguments que es podrien considerar. En primer lloc, si la IA està funcionant en els diagnòstics mèdics, en el màrqueting digital, en els cultius d'algues, en la prevenció d'incendis o en la traducció automàtica, per què no hauria de donar resultats en la docència? De forma intuïtiva, podem imaginar que un sistema capaç de penetrar en els àmbits financer, artístic, urbanístic, industrial o logístic té una funció potencial en l'àmbit didàctic. L'ensenyament no és tan diferent de la resta de les activitats com per justificar la seva immunitat. A banda, formar les estudiants i els estudiants amb una bombolla lliure d'IA pot donar lloc a un abisme entre la teoria i la pràctica, entre la formació i el món real. En el passat, aquesta incompareixença ha tingut un impacte molt negatiu. Per exemple, la irrupció de la xarxa i específicament dels cercadors, ha dotat l'estudiantat d'una eina que es pot emprar de forma eficient o de forma imprudent. Haver entrenat les estudiants i els estudiants en l'optimització dels cercadors (pàgines de confiança, cerques condicionades, criteris de cerca, cerca en URL o de documents...) hauria donat millors fruits que ignorar l'impacte de Google en la Generació Millennial.

El 2023 un grup de professores i professors de la Universitat de Girona vam considerar els arguments a favor, ens vam interessar per les possibilitats de la intel·ligència artificial en la docència universitària i vam crear un grup de treball informal que ha compartit

eines, experiments, resultats i reflexions. Amb una lògica exploratòria, hem iniciat una sèrie d'iniciatives d'incorporació de la IA a les aules, basant-nos sempre en tres axiomes. El primer és que fariem tasts, experiments puntuals guiats pel principi "Equivocar-se poc i ràpid". No han estat canvis radicals en la manera d'ensenyar, sinó activitats puntuals d'assaig - error. El segon és que incorporariem la tecnologia només si consideràvem un guany potencial, una millora aparent respecte a la no utilització. Molt sovint, la utilització de la tecnologia s'ha basat en l'eina per l'eina, el que acaba creant més aviat un joc pirotècnic. I el tercer és que els experiments eren socialitzats en el grup, amb l'objectiu d'avaluar col·lectivament els beneficis i d'anticipar efectes adversos no previstos. Aquest llibre recull una part dels resultats del treball del grup, combinats amb la recerca de la informació sobre un tema que evoluciona de forma permanent.

Abans de presentar l'estructura del llibre, volem fer quatre *'disclaimers'*.

- Aquest és un text sobre intel·ligència artificial sense intel·ligència artificial. Els textos han estat redactats íntegrament per les autores, renunciant a un suport digital. L'única part del text elaborada amb el recurs d'assistents és el pròleg, en el qual conversem sobre el futur de la IA a la docència amb els principals models del mercat. Hi ha un cert contrasentit: Si el llibre mostra les possibilitats de la intel·ligència artificial per respondre preguntes complexes, resoldre problemes, dissenyar estratègies o crear textos des del no-res, per què no l'hem incorporada en el procés de redacció del llibre? La resposta a aquesta aparent paradoxa és que en aquest estadi hem considerat que necessitem un espai lliure d'IA per mostrar les possibilitats i també els riscos d'aquesta eina, que necessitàvem una certa distància per calibrar millor el resultat. La IA és l'objecte del llibre, però no el subjecte.
- Aquest és un text sobre la utilització de la intel·ligència artificial a les aules, sobretot en l'àmbit universitari però també segurament en l'ensenyament secundari. Està redactat per especialistes en comunicació, art, organització d'empreses, geografia o màrqueting

i, per això, la major part dels exemples i exercicis se situen en l'àmbit de les ciències socials i les humanitats. Potser algunes parts poden ser útils per a docents de l'àmbit de les ciències experimentals o la tecnologia, però el llibre ha estat redactat pensant en una professora o un professor d'antropologia, d'economia, d'arxivística, d'art modern, de geografia humana, de filosofia, de ciències polítiques o de qualsevol branca del vast camp de les ciències socials i les humanitats.

- Aquest és un text sobre un fragment de la intel·ligència artificial aplicada a les aules. La disciplina és molt vasta i inclou la generació d'imatges, el reconeixement de la veu, la robòtica i l'automatització, la detecció de patrons o la simulació de mons virtuals o de relacions entre els agents. La major part dels capítols estan centrats en una part d'aquest escenari, com són les IA generatives basades en models de llenguatge, és a dir, els enginys que permeten reproduir artificialment la conversa. Els darrers capítols introdueixen algunes eines basades en la recerca qualitativa o quantitativa, però aquest és un text centrat essencialment en els models de llenguatge, en l'univers que obrí la irrupció en el mercat del ChatGPT i totes les eines posteriors que han aparegut amb un format similar.
- Aquest és un text sobre la utilització de la intel·ligència artificial a les aules en un moment en el qual les eines s'estan actualitzant de forma permanent. Un dia es poden fer vídeos amb un simple text i l'endemà les limitacions de paraules per escriure el context gairebé desapareixen. Aquest és un text condemnat a envellir molt ràpidament. Diuen que Stendhal va viure la batalla de Waterloo, però que la proximitat li va impedir llegir la magnitud històrica d'aquell esdeveniment; quan el teu voltant és pols, cavalls i ferides, és molt difícil fer el vol de dron que et permeti discernir allò que és rellevant d'allò que és secundari. Escriure un text docent com si fos la crònica d'una batalla, quan encara ressonen els canons té un punt d'imprudència que assumim.

Aquest text s'obre amb un capítol conceptual sobre la intel·ligència artificial, que intenta delimitar el camp d'estudi i fixar què és i què no és la IA. És un capítol prescindible: Podem conduir un cotxe sense saber res del motor d'explosió, podem mirar una pel·lícula a la televisió sense interessar-nos pel fascinant procés que transporta les imatges fins el nostre receptor i podem degustar un menú tres estrelles ignorant quin és el principi químic que permet una esferificació. En el cas de la IA, però, és més pertinent intuir què és el que hi ha a la caixa negra dels models que responen automàticament les nostres preguntes sobre la geopolítica de l'Índic o sobre les estratègies de màrqueting lateral.

La forma més habitual de relacionar-se amb els models de llenguatge és amb un xat, que crea la ficció de la conversa. El segon capítol assenta les bases de l'enginyeria de *prompts*, de l'habilitat de preguntar bé per obtenir respostes més precises. Un mètode més sofisticat per obtenir respostes és el model Q & A (Preguntes i Respostes), que entrena el xat per obtenir els resultats desitjats, tal com es mostra en el tercer capítol. Mitjançant els rols, els models poden assumir determinades personalitats (una experta en art bizantí, una cartògrafa, un economista keynesià) que proporcionen respostes adaptades a necessitats concretes i això es presenta en el quart capítol. El cinquè capítol tanca aquesta part sobre la utilització dels models de generació de text, amb algunes indicacions relatives al pensament arborescent i la creació de converses artificials.

Els models de llenguatge no són oracles, no revelen veritats irrefutables ni anticipen el futur. Són sofisticats enginys entrenats amb milions de dades textuais i, per això, són una amplificació dels nostres encerts i els nostres errors. El sisè capítol mostra el risc de les al·lucinacions, les sortides de to dels models que inventen quadres que mai va pintar Modigliani, muntanyes que no apareixen en els mapes o filòsofs contemporanis que defensen teories mai plantejades. Els models fallen de vegades i per això hem d'ensenyar a dubtar sempre de qualsevol resultat, a mantenir un estat d'alerta permanent. Construir un model a partir del nostre bagatge té un peatge: El resultat

“Està escrit des de l'espai del dubte, l'assaig, la provatura, la humilitat, la curiositat i la precaució”

està impregnat dels nostres prejudicis. El capítol setè planteja els riscos del biaix de la IA i proposa algunes vies per minimitzar-lo.

El tercer bloc del llibre està integrat per capítols d'iniciació a la recerca, i es proporcionen eines per introduir l'estudiantat en aquest camp. El capítol vuitè és una introducció a la utilització de la IA en l'estat de l'art i la recerca bibliogràfica sobre un determinat tema, ja sigui el turisme fosc, la legitimitat de l'autodeterminació, la teoria de jocs o l'antropologia de l'oci. En el capítol novè es presenten algunes de les possibilitats de la recerca quantitativa amb el suport de la IA i el desè es dedica a la recerca qualitativa. Aquest és un text sobre la utilització de la intel·ligència artificial a la docència, de manera que el focus se situa en la docència de les eines de recerca més que no pas en les eines de recerca.

Poden els sistemes intel·ligents ser també creatius o es limiten a crear continguts a partir d'elements que els hi aportem? Com podem aprofitar les eines de la IA per a reforçar la creativitat dels estudiants i les estudiants? Amb freqüència IA i creativitat són presentats com a antònims i és habitual considerar que la creativitat és el darrer dels reductes de la imaginació humana perquè els processos mecànics seran absorbits per les noves tecnologies. Al capítol onzè es planteja les possibilitats d'ampliar la creativitat amb la utilització de la IA.

La intel·ligència artificial també permet incrementar els recursos docents i incorporar eines assistents a l'ensenyament, com la correcció automatitzada, la creació automàtica de testos o preguntes, la utilització de recursos audiovisuals o de text... El capítol dotzè mostra les possibilitats de disposar d'un professor ajudant que és en realitat un androide, un replicant. Complementàriament, el capítol tretzè mostra les extraordinàries possibilitats dels assistents virtuals, que

poden simular docents de guàrdia, estudiants pilot, filòsofs 24 hores o generadors de problemes de lògica.

Cada vegada que hem iniciat un experiment ens hem plantejat un munt de problemes ètics i de preguntes amb un regust distòpic. Estem segures que el llibre suscitarà de forma immediata una reacció més o menys explícita: Acabarem amb les velles professions? Seran els docents del futur androides? Té sentit ensenyar a fer res o a pensar si una màquina és capaç de fer-ho per nosaltres? La intel·ligència artificial crea de forma instantània un castell de dubtes morals i dibuixa escenaris gairebé postapocalíptics. Aquest debat ha d'entrar també a les aules i ha de formar part de les discussions quotidianes a classe. Com el cinema fa molts anys que està plantejant aquests temes, hem convertit el capítol tretzè en una sala de cinema i hem transitat per les principals pel·lícules del gènere, amb la intenció de plantejar (més que no pas respondre) els dubtes essencials de la IA a l'aula.

El darrer capítol és una reflexió final sobre les possibilitats i també les amenaces de la IA, sobre l'extraordinària oportunitat de potenciar els recursos docents i de multiplicar les capacitats dels estudiants i al mateix temps, sobre els riscos d'un mal ús, dels abusos i els malentesos que la utilització descontrolada d'aquestes eines pot provocar.

Del guix al xip és un text prospectiu, un assaig compartit sobre la incorporació d'algunes eines de la intel·ligència artificial al complex procés de l'aprenentatge. No hi ha certes ni instruccions, sentències o conviccions. Està escrit des de l'espai del dubte, l'assaig, la provatura, la humilitat, la curiositat i la precaució. És un petit illot en un vast oceà del qual no intuïm l'horitzó. I, com a totes les travessies, el viatge és sempre més estimulants que el destí.

01

El cervell de silici. Què és la IA?

Imaginem un llibre escrit de forma aleatòria. Prenem 22 lletres de l'alfabet occidental (n'haurem de prescindir de les poc comunes, com la w) i els tres signes de puntuació bàsics i generem un llibre de 410 pàgines; cada pàgina estarà formada per 40 línies de 80 lletres. Tot i que en aquest llibre creat aleatòriament podríem trobar alguna paraula del diccionari (com ara "estigma", "macedoni" o potser "calidoscopi"), el text serà intel·ligible. Podríem generar un segon llibre amb els mateixos criteris, i amb la condició que no repetís el text generat en primer lloc i el resultat seria novament un text sense sentit. A mesura que anéssim editant llibres, els podríem endreçar en una prestatgeria en la qual situaríem els exemplars i acabariem construint un hexàgon amb quatre parets ocupades per prestatges plens d'exemplars de textos aleatoris. Si poguéssim replicar aquest hexàgon en totes les direccions i poguéssim editar totes les combinacions possibles de llibres de 410 pàgines hauríem reproduït la Biblioteca de Babel.

El 1944 l'escriptor Jorge Luis Borges va publicar un recull de contes, *Ficciones*, que incloïa *La Biblioteca de Babel*, probablement el relat més conegut de l'obra. Borges va imaginar una estructura immensa d'hexàgons on habiten bibliotecaris que deambulen per les estances a la recerca de textos amb sentit; molts d'ells passaran tota la seva existència sense topar-se amb un llibre intel·ligible. De fet, el nombre possible de llibres és més gran que tots els àtoms de l'univers, un nombre tan elevat que no el podem imaginar: $25^{1.312.000}$. Entre els llibres de la biblioteca hi ha l'Odissea, el Quixot o totes les obres de George Orwell, en tots els idiomes possibles, hi trobaríem la biografia precisa de totes les persones que han habitat la terra, totes les teories econòmiques i les seves refutacions i, com explica el mateix Borges, hi hauria el catàleg precís de la biblioteca i també milers de falsos catàlegs. Hi ha una [simulació de la Biblioteca de Babel a la xarxa](#), creada per Jonathan Basile; lògicament, aquest entorn no ha produït totes les opcions possibles però és una recreació molt efectiva que ajuda a imaginar el concepte borgià.

Si un ordinador volgués escriure textos només hauria de replicar la idea de Borges i generar de forma aleatòria agrupacions de lletres, com ara *wxfffg* o bé *ploghjGX*. És el que faria un gat si es passegés per sobre d'una màquina d'escriure. Però, com ens ho faríem per fer-li escriure textos amb sentit? Sabem que de totes les combinacions possibles, n'hi ha algunes que estan dotades de significat i que poden ser la resposta a interrogants que plantejem prèviament. Però ¿com trobar en l'immens laberint de textos potencials aquells que estiguin dotats de significat i que, en definitiva, imitin la nostra capacitat de comunicar-nos, d'inventar mons alternatius, de descriure les instruccions precises per construir un motor d'explosió, de refutar una argumentació vulnerable o de descriure la bellesa d'un paisatge remullat per la pluja? Aquesta és una pregunta que d'alguna manera ens hem plantejat al llarg de la història de la humanitat. Al seu assaig sobre la intel·ligència artificial, López de Mántaras (2024) recorda que els andròides ja estan presents a la *Iliada* i que Aristòtil va fer propostes per automatitzar els processos lògics.

1.1. Els orígens de la intel·ligència artificial

La intel·ligència artificial és un camp molt jove. Es considera que el primer article que planteja aquesta hipòtesi és el que van publicar [McCulloch i Pitts](#) el 1943 a la Revista *Bulletin of Mathematical Biology*. Els autors imiten el funcionament de les neurones, unitats bàsiques que poden estar connectades (on) o desconnectades (off), d'acord amb els estímuls que reben. Cada neurona està vinculada amb altres en un sistema més o menys complex que reproduïx un model lògic de relacions. Com el conjunt d'operacions és molt elevat, els autors prenen el model teòric de computació que havia ideat Alan Turing. De manera que en l'article inicial ja disposem de les tres peces bàsiques d'aquesta proposta: Unitats d'informació connectades entre si (les neurones), que segueixen instruccions basades en sentències lògiques, a partir d'un mecanisme que

disposa (teòricament en aquell moment) d'una gran potència de càlcul. Com si fossin versions modernes del doctor Frankenstein, proposen la simulació d'un cervell digital que reproduïa digitalment el procés que esdevé en el nostre cervell físic.

La proposta de McCulloch i Pitts és la primera formulació de la intel·ligència artificial, però hi ha un munt d'assajos anteriors que creaven sistemes automàtics de càlcul i de resolució de problemes. El 1901 es va fer una descoberta a les costes de l'illa grega d'Antikythera que mostra els antecedents llunyans d'aquests enginyers. El [mecanisme d'Antikythera](#) és un complex artefacte format per engranatges de bronze, que servia per calcular i mostrar informació astronòmica; la màquina podia predir esdeveniments com els eclipsis solars i lunars, així com els moviments dels planetes coneguts en aquell temps. També s'utilitzava per seguir el cicle de les Olimpíades, cosa que indica el seu ús pràctic en el seguiment de diversos calendaris.

La Pascalina és una de les primeres calculadores mecàniques de la història, inventada per Blaise Pascal al voltant de 1642. Aquest enginyer utilitza un sistema de rodetes i engranatges per realitzar sumes i restes directament, i multiplicacions i divisions per repetició de sumes i restes, respectivament. Cada roda de la Pascalina representava una xifra en un número, amb rodes interconnectades que podien avançar una posició cada vegada que la roda anterior completava una volta. Leibniz va millorar la Pascalina amb una màquina que utilitzava el sistema binari, la base dels ordinadors moderns. El 1822 Charles Babbage va dissenyar la Màquina de Diferències, una calculadora mecànica destinada a calcular taules de funcions matemàtiques. També ideà la Màquina Analítica, considerada la precursora de la computadora moderna.

Tots aquests enginyers eren models mecànics, complexos sistemes de rodes dentades connectades entre si com en un mecanisme de rellotgeria. La mecanització dels vehicles, de les indústries o de les màquines agrícoles era la inspiració pels qui intentaven crear

sistemes automàtics de resolució de problemes. Algunes d'aquestes màquines van captivar la imaginació pública durant els segles XVIII o XIX, com els autòmats que jugaven a escacs amb una notable capacitat. "El Turc" era segurament l'autòmat més conegut de la seva època. Vestit amb roba exòtica i assegut davant d'un tauler, l'autòmat semblava funcionar amb un mecanisme que es mostrava al públic abans de la partida. Aquestes rodes dentades amagaven el jugador que, en realitat, era qui movia les peces d'acord amb el seu criteri.

Però els primers assajos sobre la intel·ligència artificial estaven inspirats en un nou artefacte, la màquina de Turing. Ideada per Alan Turing el 1936, és un model conceptual, format per una cinta infinita dividida en cel·les que poden contenir símbols, una capçalera que llegeix i escriu símbols en aquestes cel·les, i una taula de transició que dicta les operacions de la màquina basant-se en l'estat actual i el símbol llegit. Amb aquesta formulació teòrica, Turing assentava les bases de la computació moderna i obria les portes a un nou tipus de màquines per a la resolució de problemes, els ordinadors. Turing també va ser l'autor d'un article, "Computing Machinery and Intelligence", publicat el 1950 a *Mind*, una de les altres fites capitals en la prehistòria de la IA. D'alguna manera, la intel·ligència artificial neix al mateix temps que la computació, així que les màquines per resoldre problemes abandonen els mecanismes, les rodes dentades i els engranatges o els circuits elèctrics i són substituïdes pels primers programes informàtics.

La intel·ligència artificial va ser batejada en una trobada d'uns pioners en aquest camp l'estiu de 1956 a Darmouth. Aquesta trobada va ser auspiciada per John McCarthy i va reunir durant dos mesos deu experts en àmbits diversos relacionats amb la idea de "màquines pensants". Entre els convidats, Allen Newell i Herbet Simon van presentar *Logic Theorist*, un mecanisme que permetia resoldre problemes complexos seguint unes instruccions lògiques. Per exemple, el programa va ser capaç de demostrar 38 dels 52 teoremes presentats a *Principia Mathematica* escrit per Alfred North

Whitehead i Bertrand Russell. Des de Darmouth, es van desenvolupar les iniciatives que exploraren la creació de màquines "que pensin com els humans". Els primers exercicis se centraren sobretot en micromons, en entorns controlats amb unes regles precises que plantegen problemes concrets, com els jocs o els problemes matemàtics. McCarthy va proposar el 1958 el primer programa que era capaç de mantenir uns diàlegs molt bàsics.

El 1970 es va presentar el programa SHRDLU que era capaç d'entendre les instruccions en llenguatge natural d'un micromon, el *small block world*. Aquest espai virtual estava format per una sèrie de peces geomètriques de colors en un espai tancat, de manera que els programadors van poder limitar el nombre d'opcions: les formes (piràmide, cub...), els colors (blau, vermell, verd...) o les accions (agafar, apilar, comptar...) o les posicions (davant, darrere, sobre...). [Aquí](#) hi ha una reproducció de la conversa entre una persona i l'ordinador, amb un grau de precisió que ens recorda als models actuals de llenguatge natural. En aquell moment es considerava que la intel·ligència artificial era només una qüestió d'escala, de manera que si un prototip era capaç d'entendre un micromón, només calia afegir complexitat per tal de saltar d'entorns controlats al món real. Però substituir piràmides vermelles i cubs blaus per la moda, l'arquitectura o la literatura implica un grau de complexitat que els models tenien moltes dificultats per escalar.

Els pioners de la IA van pecar d'optimistes. El salt d'entorns controlats, els micromons, a entorns complexos va topar amb les limitacions conceptuals i tecnològiques de l'època. Per exemple, els programes de traducció estaven plantejats com a simples substitucions de les paraules per les seves corresponents en l'altra llengua, però sense tenir en compte el context. És molt coneguda la traducció de l'anglès al rus de la frase "L'esperit és fort, però la carn és feble" que va produir "El vodka és bo, però la carn està podrida" (Russell i Norvig, 2003). Els sistemes no van donar respostes eficients en camps com la traducció automàtica, el reconeixement de la parla

o la resolució de problemes complexos i l'eufòria inicial va decaure i els fons públics es van reduir de forma dràstica. L'informe Lighthill al Regne Unit el 1973 recollia tots aquests fracassos i recomanava el govern britànic que abandonés les subvencions als projectes de recerca en aquest àmbit. S'havia iniciat l'hivern de la IA, i la utopia d'una màquina intel·ligent semblava esvaïr-se.

El renaixement de la IA ha estat possible per la combinació de quatre factors:

- El concepte de **xarxa neuronal** havia estat relativament abandonat els anys 80 perquè el salt d'escala semblava generar un volum de càlcul tan elevat que el feia inviable. Per això, el model que va prevaldre fou el sistema simbòlic (En el pròxim epígraf mostrarem la diferència entre aquests dos paradigmes principals). Amb el desenvolupament de models sofisticats d'entrenament, les xarxes neuronals van començar a mostrar resultats en camps molt complexos com el reconeixement de la parla.
- La recuperació de **models de probabilitat** va permetre un raonament compacte d'un coneixement incert, sobretot a partir de l'obra de Pearl (1988). Amb la retroalimentació de cada procés, les probabilitats varien i d'aquesta manera s'aconsegueixen sortides molt precises de sistemes que són molt incerts. La major part dels models d'IA actuals han incorporat aquest concepte, de manera que les sortides que ens proporcionen són, en realitat, *outputs* amb una probabilitat més elevada que no pas altres de ser la resposta adient. De fet, els models de llenguatge funcionen amb una lògica similar als textos predictius.
- La **capacitat d'operació** de les computadores va experimentar un creixement molt notable a partir dels anys 90, per la combinació de factors, com la miniaturització, l'arquitectura de computadores (com el RISC), l'increment de la memòria RAM, la utilització de GPU (unitats de processament de gràfics) per a càlculs generals, els sistemes de múltiples nuclis, o algorismes més eficients.

Durant aquest període es va produir l'efecte de la Llei de Moore, que preveu que el nombre de transistors en un xip es duplicarà cada dos anys. Podem comparar, per exemple, un IBM System/360 (1964) amb un Dell Dimension XPS T500 (2000): D'un milió d'instruccions per segon a 500 milions, d'uns pocs milers de transistors a 10 milions, d'una RAM de 512 Kb a 256 Mb i d'un disc dur de 200 Mb com a màxim a 40 Gb.

- Els nous models necessitaven ser alimentats amb **grans volums de dades**. La digitalització d'informació primer (com les bases de dades mèdiques o de recerca) i l'aparició de l'internet després va permetre l'accés a un volum d'informació enorme que podia entrenar els sistemes basats en l'aprenentatge recurrent. Durant aquest període es van produir grans avenços en la digitalització de textos, imatges o sons, i l'aparició d'una nova indústria basada en la compravenda d'aquesta informació digital.

L'11 de maig de 1997 Gari Kaspàrov, un dels millors jugadors de la història dels escacs, es va rendir a la jugada 19 de la sisena partida del *match* que l'enfrontava a Deep Blue. Un any abans, Kaspàrov havia derrotat el programa d'IBM amb molta autoritat, per 4 a 2, però els enginyers del gegant informàtic crearen una versió millorada i reptaren de nou al jugador rus. Kaspàrov estava convençut que malgrat la potència de càlcul de la màquina, aquesta era incapaç de resoldre qüestions profundes de caràcter posicional, però Deep Blue el va sorprendre amb una capacitat d'anàlisi que semblava humana. Els dos rivals, l'humà i la màquina, havien arribat empatats al darrer *match* i la sala de l'Equitable Center de Manhattan era plena de gom a gom. Kaspàrov no va poder fer front a la fortíssima pressió i es va equivocar amb la Defensa Caro Kan, un error que Deep Blue castigà de forma implacable. Les accions d'IBM es van disparar i en aquell moment es va obrir un debat general sobre els límits de la intel·ligència artificial: Si han estat capaços de guanyar un gran mestre d'escacs, no hi ha barrera humana que no puguin superar, s'especulà llavors.

Deep Blue havia estat entrenat amb línies de codi que reproduïen els coneixements experts de grans mestres dels escacs. Representava una forma d'intel·ligència artificial basada en la programació, és a dir, en unes instruccions precises i en la potència de càlcul dels ordinadors de l'època. La paradoxa és que el paradigma que va fer possible Deep Blue quedaria superat per una altra lògica, la de l'aprenentatge profund basat en les xarxes neuronals. Vint anys després de la derrota de Kaspàrov, l'AlphaZero de DeepMind s'enfrontà al programa més fort en aquell moment, Stockfish; mentre que el primer avaluava 80.000 posicions per segon, Stockfish era capaç d'analitzar 70 milions. AlphaZero, programat amb la lògica de les xarxes neuronals va guanyar 28 partides a Stockfish i no en va perdre cap.

Les xarxes neuronals també van mostrar la seva capacitat amb el reconeixement d'imatges. Anualment tenia lloc una trobada entre diversos prototips que havien de deduir que aquella fotografia era un tren, aquella altra un coet i la darrera, una maduixa. Tot i els avenços del camp, el marge d'error se situava per sobre del 25%: De cada quatre imatges, els sistemes s'equivocaven en una, i confonien una monja amb un bolet o un llimac amb una fulla. El 2012, un grup de la Universitat de Toronto va presentar AlexNet, un model de xarxa neuronal convolucional. AlexNet va reduir el marge d'error al 16%, deu punts per sota dels millors registres; els anys següents, tots els models introduïren el mètode d'AlexNet i el marge d'error va caure per sota del 5%, de manera que el concurs va deixar de tenir sentit. Avui, la precisió dels sistemes de reconeixement d'imatge automàtics supera l'ull humà.

Mentrestant, en els espais domèstics s'han incorporat els primers utensilis robòtics, com els assistents de veu o les eines domòtiques. El 2002 es va presentar un enginy que ha esdevingut una peça de la cultura popular contemporània, el netejador *Roomba*. La *Roomba* o *DeepBlue* són dos exemples de la irrupció de la robòtica, l'automatització i la intel·ligència artificial en tots els àmbits de la societat: la gestió de la informació, la seguretat, la logística i el transport, el

màrqueting, la investigació, la medicina o l'energia. United Airlines empra la IA per escollir rutes eficients que redueixin la petjada de carboni i estalviïn combustible, Warner Bros determina quines pel·lícules ha de finançar d'acord amb els resultats d'un sistema de predicció basat en l'anàlisi de patrons previs, l'estudi d'arquitectura Zaha Hadid empra sofisticats models per explorar formes complexes i solucions a problemes estructurals, i SideWalk Labs crea models de mobilitat sostenible basades en l'estudi dels patrons de moviment dels usuaris. La IA està present en l'oci, la logística, el transport, la indústria, l'energia, la meteorologia, la indústria militar i la seguretat, el sector financer, la recerca, el disseny o la cultura digital des de la primera dècada del segle XXI.

Tanmateix, és possible que l'impacte més gran de la IA hagi estat l'aparició el novembre de 2022 del model de llenguatge ChatGPT. L'empresa OpenAI va integrar la tecnologia de processament de llenguatge natural (PNL) a un format de conversa, que facilitava la interacció amb la usuària o l'usuari. Basat en un model d'aprenentatge profund, aquest model ha estat una fita essencial en la curta història de la IA. El febrer de 2023 ChatGPT havia assolit els 100 milions d'usuaris i usuàries, essent l'aplicació que més ràpidament ha assolit aquest llindar. Més endavant han aparegut els models Llama de Meta (febrer), Claude d'Anthropic (març), Google Bard que serà després Gemini (maig), Copilot de Microsoft (novembre) i Amazon Q per empreses (novembre). El desembre de 2023, el Parlament Europeu va aprovar la [primera norma internacional](#) que regula els usos de la intel·ligència artificial.

1.2. Les dues intel·ligències artificials

Per tal d'esbrinar el sexe d'un pollet, el sexador o la sexadora disposa d'uns pocs segons i pot equivocar-se molt poques vegades. Amb un gest gairebé automàtic, agafa el pollet i l'apreta per tal que s'obri la seva cavitat anal. Hi ha subtils diferències entre els plecs de la femella i del mascle, de manera que si estiguéssim asseguts al seu costat el criteri de selecció ens semblaria aleatori. Són necessaris uns

tres anys per assolir l'habilitat de distingir el sexe d'un pollet i malgrat l'extraordinària destresa dels sexadors, no són capaços d'explicar per què han considerat que aquells dos pollets són mascles. Els sistemes d'intel·ligència artificial basats en models biològics, en xarxes neuronals, funcionen una mica com l'elecció del sexe d'un pollastre.

Durant molts anys, el paradigma que ha guiat la intel·ligència artificial ha estat un altre. La via més transitada ha estat la que proposen Allen Newell i Herbert A. Simon (1976), la **hipòtesi del sistema de símbols físics**. Un sistema de símbols físics és un conjunt d'entitats, denominades símbols, que poden ser manipulades i modificades i que representen informació o dades; el sistema pot crear, modificar, destruir i reorganitzar símbols, permetent la generació i transformació de representacions simbòliques. Aquesta capacitat simbòlica és crucial per a la funció cognitiva, ja que permet al sistema modelar el món i actuar d'acord amb aquestes representacions. Segons Newell i Simon, perquè un sistema mostri comportament intel·ligent, ha de ser capaç de manipular símbols de manera que pugui formular i solucionar problemes de manera flexible. Això inclou la capacitat de raonar, planificar, aprendre, i adaptar-se. El sistema de símbols físics interacciona amb el món real a través de sensors i actuadors. Això li permet rebre informació de l'exterior (entrada) i actuar sobre l'entorn (sortida), completant així el cicle de percepció-acció.

Aquest és un model *up - down*, perquè defineix unes regles prèviament, uns codis, i executa les instruccions d'acord amb els resultats. A diferència de la programació clàssica, aquests models estan ideats per a "aprendre", és a dir, per a modificar els símbols i les seves relacions d'acord amb els resultats o els estímuls. Aquesta hipòtesi planteja que la intel·ligència és un sistema que opera amb la manipulació de símbols, independentment del seu substrat, ja sigui natural o artificial, és a dir, sigui de carboni (el nostre cervell) o de silici (els components d'un ordinador).

La segona via de la intel·ligència artificial ha estat la **hipòtesi dels sistemes basats en models biològics**, especialment les xarxes neuronals. Aquests sistemes intenten emular el procés d'aprenentatge del cervell, en el qual unes unitats bàsiques (o "neurones") que estan interconnectades treballen de forma conjunta per a dur a terme una determinada tasca. Els primers models, com el perceptró desenvolupat per Frank Rosenblatt als anys 1950s, estan basades en versió simplificada de com es creia que funcionaven les neurones cerebrals. Els anys 1980s, David Rumelhart, Geoffrey Hinton, i Ronald Williams van introduir el concepte d'algorisme de retropropagació (*backpropagation*), en xarxes que tenen moltes capes, cosa que permet l'aprenentatge profund.

Imaginem, per exemple, el funcionament d'AlphaZero, el programa que va batre Stockfish amb la utilització de l'aprenentatge profund. A diferència dels sistemes anteriors, AlphaZero només coneixia les regles dels escacs i no disposava d'informació addicional proporcionada pels Grans Mestres. AlphaZero va iniciar un entrenament jugant milers de partides contra ell mateix. Les primeres partides eren absurdes, perquè el programa no disposava ni tan sols de les operacions bàsiques per fer mat amb dues torres o amb una dama. De seguida, però, el sistema va començar a detectar patrons que incrementaven la probabilitat de victòria. Cada nova intuïció (és millor enrocar-se ràpid, cal dominar el centre, una torre val més que un cavall) era incorporada a la partida següent i era testada novament en un *loop* permanent de contrast d'hipòtesis. En poques hores, un sistema que no sabia res dels escacs, va aconseguir excel·lir com cap altre sistema basat en programació prèvia. Per això diem que els models biològics, l'aprenentatge profund, és *bottom - up*.

AlexNet, AlphaZero, ChatGPT o Midjourney són models molt eficients per resoldre tasques concretes: Identificar imatges, jugar als escacs, conversar o generar imatges a partir d'un text. AlphaZero, però, no pot conversar sobre la història dels escacs, ni jugar al parxís i no pot diferenciar el material de les peces. Aquest grau d'especialització ha estat essencial per

Tipus d'IA segons Russell i Norvig

	PENSAR	ACTUAR
Humans	Sistemes que pensen com humans	Sistemes que actuen com humans
Racionalment	Sistemes que pensen racionalment	Sistemes que actuen racionalment

al desenvolupament de la intel·ligència artificial. Els models que són eficaços en una tasca concreta s'anomenen *soft IA*, intel·ligència artificial tova o suau, que és el criteri dels enginyers actuals. A la trobada de Darmouth, s'havien imaginat una altra forma d'IA anomenada AGI, intel·ligència artificial general, que pot entendre, aprendre i aplicar la intel·ligència en una àmplia varietat de tasques, de manera similar a com ho faria un ésser humà; una AGI pot transferir coneixement entre diferents àrees i adaptar-se a noves situacions que no ha experimentat abans. Aquesta forma d'IA és només teòrica i no ha estat aplicada en la pràctica. Hi hauria una tercera categoria, que seria aquella en la qual la intel·ligència artificial és capaç de realitzar processos argumentals, cognitius o creatius que superin la capacitat humana, i que ha alimentat bona part de les pel·lícules de ciència-ficció.

Hi ha altres subgrups d'intel·ligència artificial, com els sistemes emergents (cooperació entre agents individuals) o els sistemes d'acció situada. Hi ha també models híbrids, que combinen instruccions inicials amb xarxes neuronals d'aprenentatge profund. De fet, hi ha moltes formes d'intel·ligència artificial i les seves fronteres conceptuals s'expandeixen i es confonen.

La classificació clàssica de la intel·ligència artificial va ser proposada per Russell i Norvig (2003). És una petita matriu que neix combinant dues categories: La primera diferència entre pensar i actuar, entre un sistema intel·ligent i un mecanisme que fa alguna cosa; la segona diferència entre el referent humà i el referent racional. La combinació entre aquests dos criteris dona lloc a quatre categories.

1. Sistemes que pensen com humans

Aquest enfocament intenta imitar els processos de pensament humà. Això implica la comprensió

i la modelització de la forma en què els humans processen informació, resolen problemes i prenen decisions. Els models de processament de llenguatge natural (PLN) imiten el llenguatge humà, de la mateixa manera que ho faria una persona.

2. Sistemes que actuen com humans

Aquest enfocament se centra en la creació de màquines que no només pensin com humans, sinó que també actuïn com a tals. L'objectiu és replicar el comportament humà de manera que sigui recognoscible i natural. Els assistents virtuals poden arribar a mantenir converses amb els usuaris i les usuàries que imitin tots els components humans.

3. Sistemes que pensen racionalment

Aquest enfocament es basa en lògica formal i en la capacitat de fer inferències correctes. Els sistemes que pensen racionalment segueixen regles lògiques estrictes per arribar a conclusions, sense necessàriament imitar el pensament humà. De fet, algunes de les decisions dels humans no responen a lògiques racionals. Programes com Prover9 poden demostrar teoremes matemàtics, utilitzant la lògica formal, a partir de regles de deducció lògica.

4. Sistemes que actuen racionalment

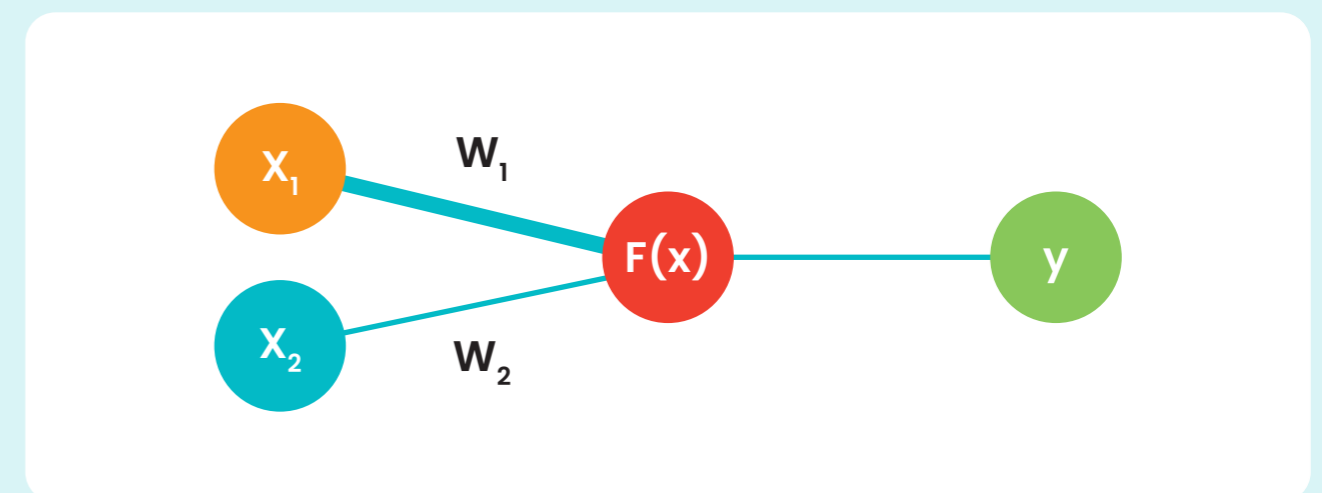
L'enfocament dels sistemes que actuen racionalment se centra en la creació d'agents intel·ligents que prenen decisions òptimes per aconseguir els seus objectius. Aquests agents utilitzen la informació disponible per actuar de manera que maximitzi el seu èxit en una tasca determinada. Els vehicles autònoms poden prendre decisions òptimes basades en una gran quantitat d'informació en temps real.

“La intel·ligència artificial generativa ha estat el resultat d'un llarg camí, d'una successió d'intuïcions encadenades”

1.3. La IA Generativa

La intel·ligència artificial generativa ha estat el resultat d'un llarg camí, d'una successió d'intuïcions encadenades. El punt de partida va ser [aquest article](#) de Frank Rosenblatt el 1958, qui va introduir el concepte de *perceptró*, inspirat en les neurones del cervell humà. Imaginem, per exemple, que volem avaluar la viabilitat d'un nou producte d'aigua amb gust de llimona. L'èxit de l'aigua depèn de diversos factors, però podríem considerar que els més rellevants són el gust i el preu: Amb el gust, incrementarem la quantitat de llimona i amb el preu apujarem més o menys respecte al valor d'una aigua convencional. El resultat seria un perceptró que tindria dues entrades: x_1 és el gust i x_2 és el preu.

La forma més simple de perceptró seria aquella en la qual x_1 i x_2 fossin dos valors dicotòmics (0 i 1), de manera que 0 fos un valor incorrecte (massa llimona o un preu massa alt) i 1 un valor correcte. La funció d'activació de la neurona, el *bias*, determinaria que només en el cas que encertem els dos valors, gust i preu, la neurona s'activarà. El perceptró seria més útil si poguéssim treballar amb un rang de gust (percentatge de llimona) i amb un rang de preu (increment respecte al valor d'una aigua sense gust). De seguida ens adonaríem que els dos valors d'entrada no tenen el mateix pes: El client seria més sensible al preu, de manera que les variacions en el preu tindrien molt més efecte en el resultat final.



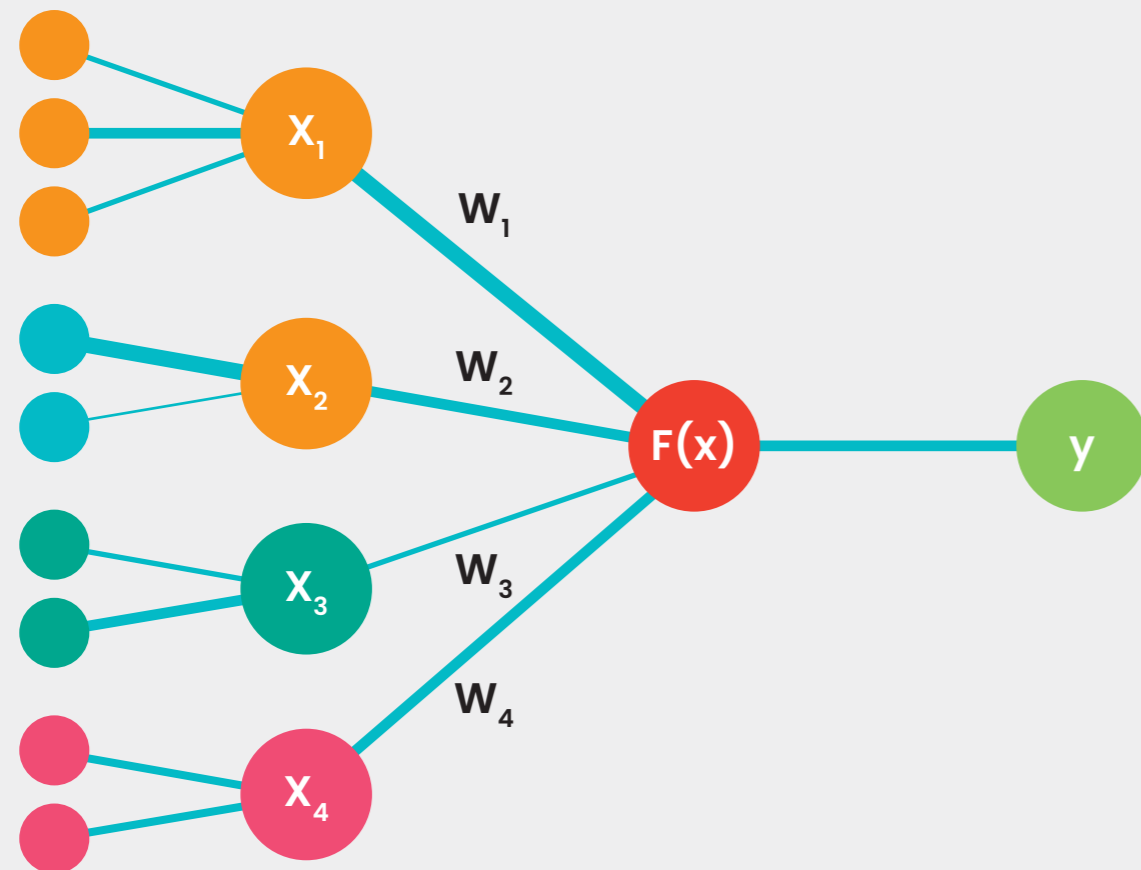
El problema és, però, més complex. L'èxit d'un producte depèn d'altres variables, com per exemple, la distribució, l'envàs, o el màrqueting del producte. Així que podríem crear un perceptró més complex, amb més entrades, que fossin x_3 (el nombre de

supermercats), x_4 (el tipus d'envàs) i x_5 (els recursos destinats a promoció). A mesura que compliquem el model, es fa més evident la dificultat de sintetitzar la informació amb un valor d'entrada. Per exemple, l'envàs pot diferenciar la forma de l'ampolla, el

color del tap i la camisa que cobreix l'ampolla amb informació i imatges. La distribució també podria diferenciar entre el nombre de supermercats, el nombre d'unitats per supermercat o la ubicació de les aigües en les estanteries. De manera que la xarxa inicial tenia una sola neurona amb diverses

entrades i ara podríem crear una xarxa amb diverses neurones més especialitzades i en diverses capes. Les econometres i els econometres saben que l'èxit dels productes depenen de molts factors complexos interrelacionats entre si, en múltiples capes. Aquest és el principi d'una xarxa neuronal.

Xarxa neuronal



Les xarxes neuronals van topiar amb dificultats tècniques per al seu desenvolupament, fins que el 1986 Rumelhart, Hinton i Williams van introduir un dels conceptes clau de la IA moderna: la retropropagació (*backpropagation*). Imaginem que la multinacional que vol incorporar les aigües amb gust pot fer diversos assajos en comarques determinades. Amb els resultats de les vendes de cada comarca pot intentar reconstruir els pesos, les ponderacions, d'acord

amb els resultats. Pot veure, per exemple, que l'envàs juga un paper molt secundari perquè variacions en els envasos no provoquen canvis substancials en les vendes; pot esbrinar també la importància de la distribució i, específicament, de la ubicació en les estanteries en el volum de vendes. I aquí intuïm un factor clau d'aquesta tècnica: Com més iteracions fem, més robustes seran les relacions entre les neurones, és a dir, els pesos de cada factor en el resultat final.

Recordem l'exemple del sexador o sexadora de pollastres. Uns pocs casos no serveixen per esbrinar el resultat. Es necessiten tres anys d'assaig - error perquè l'aprenentatge sigui efectiu. I aquest va ser el criteri que va utilitzar AlphaZero. Cada partida perduda genera una informació que pot tornar enrere, amb criteris de probabilitat: És més fàcil perdre si el rei no està enrocat, si els alfils no controlen les diagonals o si la dama no està centralitzada. Els criteris posicionals o estratègics no han estat definits *a priori*, sinó que són el resultat de milers de partides, que permeten la retropropagació de la informació de les "neurones". El sistema no només aprèn a valorar les jugades, sinó que pot determinar quina és la solució amb una major probabilitat d'èxit, és a dir, és capaç de jugar, de generar una sortida.

El 2017, l'[article de Ashish Vaswani i els seus col·laboradors](#) va iniciar la darrera revolució de la intel·ligència artificial i de les xarxes neuronals: els *transformers*. La innovació clau de l'article, anomenat "Attention is all you need" és la introducció del mecanisme d'atenció auto-direccional, que permet al model centrar-se en diferents parts d'una entrada per determinar com processar i respondre a aquesta entrada. Mentre que els models anteriors seguien una seqüència (com en els textos, que llegim de la primera paraula a la darrera), els *transformers* aprenen a identificar les paraules rellevants i les relacions entre elles: No "llegeixen" un text de forma seqüencial, sinó que la interpreten en el seu conjunt, i estimen els vincles entre les unitats. Això és especialment important per entendre el context d'una frase o la vinculació entre els elements, com ara els pronoms. A la frase "La irrupció de la intel·ligència artificial a l'aula no la

transformarà radicalment, però sí que provocarà canvis" el pronom la fa referència a l'aula i no a la intel·ligència artificial. Per un humà aquesta relació pot semblar molt evident, però els models previs als *transformers* tenien dificultats per establir vincles entre els elements. L'article va demostrar que els models basats exclusivament en l'atenció poden obtenir resultats superiors en tasques de traducció automàtica, tot mantenint o millorant la velocitat de processament. Aquesta idea va ser revolucionària perquè va simplificar els models existents i va millorar la seva eficàcia i la seva eficiència.

El novembre de 2022 OpenAI presentava ChatGPT. És un *xat*, és a dir, permet establir una conversa amb un sistema de preguntes i respostes encadenades; pertany a la IA generativa perquè no només "entén" els textos, sinó que és capaç de redactar un nou text de forma automàtica; ha estat pre-entrenat per evitar els resultats no desitjats, per incorporar límits ètics i morals, i per optimitzar les respostes a partir d'un procés de recompenses; i es basa en el model *transformer* introduït el 2017. Entrenat amb un volum immens d'informació textual, i gràcies a uns servidors amb una potència de càlcul i memòria inèdits, ChatGPT ha permès que usuàries i usuaris puguin accedir a un entorn que sembla enraonar amb la capacitat d'un savi i amb la lucidesa d'un estrateg. De seguida apareixeran altres models com Claude, Gemini, Llama o Copilot. A dins del ventre del sistema, [un immens graf de relacions entre paraules](#) es va construint a cada nova iteració, com si fossin les combinacions precises de la Biblioteca que un dia imaginà Jorge Luis Borges.

Glossari

Intel·ligència artificial	Capacitat de les màquines per realitzar tasques que requereixen intel·ligència humana, com ara reconeixement de veu, presa de decisions i traducció de llengües.
Intel·ligència artificial general	També coneguda com IA forta, es refereix a una hipotètica IA amb capacitats cognitives similars o superiors a les humanes en gairebé tots els àmbits.
Intel·ligència artificial tova	També anomenada IA feble o estreta, són sistemes dissenyats per realitzar tasques específiques sense tenir una intel·ligència general.
Model de xarxa neuronal	Estructura computacional inspirada en el cervell humà, composta per capes de neurones artificials que processen la informació i aprenen a través de l'ajust dels pesos de les connexions.
Aprenentatge profund	Branca de la Intel·ligència Artificial que utilitza xarxes neuronals amb moltes capes (profundes) per analitzar i aprendre de grans quantitats de dades.
Aprenentatge automàtic (<i>machine learning</i>)	Branca de la IA que se centra en el desenvolupament d'algorismes que permeten als sistemes aprendre a partir de dades.
Processament del Llenguatge Natural	Branca de la intel·ligència artificial que permet a les màquines comprendre, interpretar i generar llenguatge humà, facilitant la interacció entre humans i ordinadors en llenguatge natural.
Algorisme de retropropagació	Tècnica utilitzada en l'entrenament de xarxes neuronals per ajustar els pesos de les connexions, basant-se en l'error entre la sortida prevista i la sortida desitjada.
Xarxes neuronals recurrents	Tipus de xarxa neuronal dissenyada per treballar amb dades seqüencials, com ara text o senyals temporals, on les connexions entre les neurones formen un cicle que permet recordar informació anterior.
Aprenentatge per reforç	Tècnica d'aprenentatge automàtic on un agent aprèn a prendre decisions mitjançant la interacció amb un entorn.
Sistemes experts	Sistemes d'IA que emulen la capacitat de presa de decisions d'un expert humà.

Referències

- Borges, J.L. (1944). *Ficciones*, Sur, Buenos Aires.
- López de Mántaras i Badia, R. (2024); *100 coses que cal saber sobre la intel·ligència artificial*, Valls: Cossetània
- McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 5, 115-133.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Elsevier.
- Rosenblatt, F. (1958). *The perceptron: a probabilistic model for information storage and organization in the brain*. *Psychological review*, 65(6), 386.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Russell, S.J. i Norvig, P. (2003); *Artificial Intelligence: A modern approach*, Prentice Hall, 2a. edició
- Simon, H. A., & Newell, A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3), 11-126.
- Turing, A. M. (1950) *Computing machinery and intelligence*, *Mind*
- Vaswani, A. et al. (2017); Attention is all you need. *Advances in neural information processing systems*, 2017, vol. 30.

02

Preguntar és un art

“La pregunta és la més creativa de les conductes humanes”, deia el publicista Alex Osborn (Nova York, 1888–Buffalo, 1966). Va ser cofundador de l’agència BBDO, considerat un dels principals teòrics publicitaris de la primera meitat del segle XX i creador del concepte *brainstorming* o pluja d’idees (Centro de Documentación Publicitaria, 2024). Preguntar és un art, no només en el camp de la creativitat, sinó també en àmbits com ara l’educació, la teràpia, la investigació, el periodisme o la medicina, entre molts d’altres.

L’art de preguntar també ha esdevingut rellevant en l’àmbit de la intel·ligència artificial, ja que la formulació d’un *prompt* adient, igual que una pregunta pertinent, pot donar lloc a una resposta més o menys adequada als nostres objectius. L’enginyeria de *prompts* ha experimentat un auge significatiu amb el desenvolupament i l’ús generalitzat de models de llenguatge extens (*Large-Scale Language Models, LLMs*), com ara el Claude, el Gemini o el ChatGPT. L’objectiu d’aquest capítol és entendre com optimitzar els *prompts* per obtenir millors resultats, comprendre quins són els elements

que configuren un *prompt* (instrucció, context, dades d’entrada i indicador de sortida) i aprendre algunes tècniques d’enginyeria de *prompts*.

Abans, però, ens hem de preguntar què és un *prompt*? “Un ‘prompt’ és un conjunt d’instruccions proporcionades a un LLM que programa el LLM personalitzant-lo i/o millorant o afinant les seves capacitats”, segons defineixen White et al. (2023: 1). Aquests autors afirmen que la qualitat dels resultats generats per una conversa LLM està directament relacionada amb la qualitat de les indicacions proporcionades per l’usuari o la usuària. Per tant, la qualitat de les respostes no només depèn dels algorismes i les dades d’entrenament, sinó també de les indicacions utilitzades (Lo, 2023: 1): “Un ‘prompt’ ben dissenyat pot provocar una resposta significativa i informativa, mentre que un de mal construït pot resultar en sortides que són irrellevants o absurdes”.

Alguns autors, com ara Korzysky et al. (2023), analitzen l’enginyeria de *prompts* com una competència digital. Per fer-ho, utilitzen el marc DigComp que s’estructura en

cinc àrees de competències bàsiques: 1) alfabetització informacional i de dades; 2) comunicació i col·laboració; 3) creació de contingut digital; 4) seguretat; i 5) resolució de problemes (Vuorikari, Kluzer i Punie, 2022). En l’anàlisi, la majoria de competències digitals d’aquestes cinc àrees, excepte en el cas de les de seguretat, se superposen a les necessitats de l’enginyeria de *prompts* dels sistemes d’intel·ligència artificial (Korzysky et al., 2023: 28).

Preguntar és un art. En el periodisme, una pregunta adient pot ser un bon titular. En medicina, una pregunta adequada pot ser un bon diagnòstic. En recerca, una pregunta formulada correctament pot ser un descobriment científic. I en IA, una pregunta escaient, és a dir, un *prompt* ben elaborat pot ser una resposta pertinent. “En resum, a un nivell superficial, la IA generalitzada és fàcil d’utilitzar; no obstant això, les sortides més sofisticades requereixen l’entrada d’una persona qualificada i han de ser críticament avaluades abans de ser utilitzades”, conclouen Miao & Holmes (2023: 12).

2.1. Enginyeria de prompts

Un *prompt* es defineix com una instrucció o una consulta específica que es proporciona a un model de llenguatge per guiar el seu comportament i generar els resultats desitjats (Giray, 2023). Segons el mateix autor, un *prompt* incorpora quatre elements: instrucció, context, dades d’entrada i indicador de sortida. Autors com Korzysky et al. (2023) també mencionen els mateixos elements amb certs matisos i altres autors i autores incorporen elements com el to, el rol, els exemples (o *shots*) i el tipus o format de sortida (Learn Prompting, 2024a). A continuació, s’explicarà el significat de cada un d’aquests elements.

- **Instrucció, tasca o pregunta.** Part del *prompt* en la qual es dona una indicació al sistema, es fa un encàrrec o una consulta per tal de guiar el seu comportament i obtenir els resultats esperats (Giray, 2023). Per exemple: “Escriu un correu electrònic a un client agraint la seva assistència a l’acte de presentació d’un nou producte”.
- **Context o informació addicional.** Part del *prompt* en què es proporciona coneixements previs al model

per tal de generar una resposta més rellevant i acurada (Giray, 2023). Pot incloure informació sobre el rol que ha d’exercir el model per desenvolupar la tasca o qualsevol altra informació addicional sobre la situació (Korzysky et al., 2023). Per exemple: “Ets un científic de dades amb tres anys d’experiència en una empresa tecnològica”.

- **Dades d’entrada.** Part del *prompt* en què es proporcionen dades o informació que el model ha d’utilitzar per portar a terme la instrucció, la tasca o per respondre la pregunta (Korzysky, 2023). Per exemple: “Proporciona una recepta de cuina innovadora que contingui verdures i proteïnes amb ingredients frescos i de temporada”.
- **Indicador o format de sortida.** Part del *prompt* on s’especifica el tipus o format de la sortida desitjada (Giray, 2023). Pots especificar, per exemple, la llargada del text: “Escriu una introducció de 30–40 paraules sobre l’art de preguntar i l’enginyeria de *prompts*”. O el format en què vols la sortida (en aquest cas en forma de llista): “Enumera les cinc millors eines d’intel·ligència artificial generativa”. O també podries especificar el to, per exemple: “Escriu un text en to humorístic sobre l’enginyeria de *prompts*” (*veure informació més detallada a l’apartat 2.4*). Finalment, una altra estratègia seria proporcionar-li un exemple de com vols la resposta de sortida.

Segons la UNESCO, l’enginyeria de *prompts* es refereix als “processos i tècniques per a crear entrades per produir una sortida d’IA generalitzada que s’assembla més a la intenció desitjada de l’usuari” (Miao & Holmes, 2023: 11). És una disciplina emergent que “estudia les interaccions amb –i la programació de– sistemes computacionals LLM en fusió per resoldre problemes complexos a través d’interfícies de llenguatge natural” (Schmidt, et al., 2023: 1). És una disciplina que guanyarà rellevància en l’àmbit de la intel·ligència artificial, expliquen Korzysky et al. (2023), ja que permet més control sobre els resultats i garanteix la creació del contingut desitjat si les indicacions són adequades. Els autors afegeixen: “els ‘prompts’ efectius contribueixen a millorar la precisió guiant els models d’IA per generar

contingut més rellevant i valuós” (Korzynsky et al., 2023: 26). Altres treballs han demostrat que la construcció de *prompts* apropiats incrementa la fiabilitat dels sistemes d'IA com ara GPT-3 (Si et al., 2023).

“L’enginyeria de ‘prompts’ és el procés de construcció de consultes o entrades (és a dir, ‘prompts’) per a models de llenguatge d’IA per tal de provocar les respostes més precises, coherents i pertinents. En essència, és l’art d’afinar les preguntes o comandes proporcionades als models d’IA per optimitzar el seu rendiment i garantir que produeixin els resultats desitjats” (Lo, 2023: 1).

L’enginyeria o disseny de prompts ha de tenir en compte dos principis: l’establiment d’un context i l’operació de repreguntar (Lopezosa, Codina i Ferran-Ferrer, 2023). A banda del context, els prompts haurien

de tenir objectius, audiències i exemples per tal de tenir millors resultats (Lopezosa & Codina, 2023). La UNESCO fa unes recomanacions específiques similars a l’hora de crear els prompts: cal utilitzar un llenguatge senzill, clar i directe evitant textos complexos o ambigus; incloure exemples per il·lustrar la resposta desitjada o el format; incloure context; perfeccionar i iterar experimentant amb diferents variacions; i, finalment, ser ètic (Miao & Holmes, 2023: 11).

Es poden trobar moltes recomanacions, exemples i bones pràctiques a l’hora de dissenyar els prompts per a la IA. Per una banda, es pot destacar la recopilació de recomanacions per elaborar prompts de text a text que fan Korzynsky et al. (2023) a través de l’acrònim AI PROMPT. Per altra banda, Torres-Salinas i Arroyo-Machado (2023) resumeixen les recomanacions per a Chat GPT, però que també es podrien aplicar a altres IA, en quatre: 1. Claredat i precisió; 2. Evitar ambigüitats; 3. Proporcionar context; i 4. Experimentar i refinar.

Recomanacions per elaborar prompts de text a text

ACRÒNIM	RECOMANACIÓ	DESENVOLUPAMENT
A	Articular la instrucció	Especifica clarament la tasca a realitzar, com ara ‘escriure’, ‘classificar’, ‘resumir’ o ‘traduir’, i especifica com ha de semblar la sortida (taula, llista, codi Python).
I	Indicar els elements del prompt	Mostra al model on són les instruccions i les dades d’entrada i quin hauria de ser el format de sortida esperat.
P	Proporcionar senyals de finalització i context	Ofereix al model senyals clars de finalització, com tres punts per a la continuació o dos punts, punt o marcadors com ‘resposta:’ per indicar que se necessita una resposta. A més, contextualitza el model proporcionant un context per a la tasca (per exemple, ‘Ets el gerent d’un equip tècnic’).
R	Refinar les instruccions per evitar ambigüitats	Dóna instruccions específiques del model i una descripció detallada de la tasca per evitar qualsevol confusió o imprecisió.
O	Oferir retroalimentació i exemples	Per a models conversacionals, com ChatGPT, la retroalimentació sobre les respostes del model pot ajudar-lo a entendre millor la sortida desitjada. A més, proporcionar al model alguns exemples de respostes esperades (aprenentatge amb pocs exemples) pot ajudar-lo a adaptar el seu estil i manera de respondre.
M	Gestionar la interacció	Tracta el model com el teu sparring, demanant-li que proporcioni contraarguments o assenyali errors en les teves idees.
T	Seguir la longitud de token i la complexitat de la tasca	Trencar tasques complexes en passos més petits per millorar el rendiment. Recorda controlar la longitud del token, mantenint el prompt i la resposta dins dels límits de tokens del model (normalment 4096 tokens per a LLMs comercialment disponibles). La longitud del token d’un text es pot comprovar aquí .

Font. Korzynsky et al., 2023: 31.

A més de recomanacions, s’ha creat la metodologia anomenada GPEI, que correspon a *Goal Prompt Evaluation Iteration*, i que consta de quatre passos: (1) definir un objectiu, (2) dissenyar la indicació (o prompt), (3) avaluar la resposta i (4) iterar (Velásquez-Henao, Franco-Cardona i Cadavid-Higueta, 2023: 12). D’altres autors han desenvolupat el marc CLEAR amb cinc components que han de seguir els prompts (veure la taula 1): Concisos, Lògics, Explícits, Adaptatius i Reflexius (Lo, 2023).

Els cinc components del marc CLEAR

C	Concise	Brevetat i claredat en els ‘prompts’.
L	Logical	Estructura i coherència en els ‘prompts’.
E	Explicit	Especificacions clares de sortida.
A	Adaptative	Flexibilitat i personalització en els ‘prompts’.
R	Reflective	Avaluació i millora contínues dels ‘prompts’.

Font. Lo, 2023.

2.2. El valor del context

Si demano a Claude que completi la frase amb una sola opció “Em fa mal” contestarà probablement “Em fa mal el cap”; en algunes ocasions, optarà per “Em fa mal l’esquena”, i molt rarament optarà per altres alternatives. Recordem que els models de llenguatge són models probabilístics i, com els predictors de text, proporcionen la solució que consideren més adient. Atès que els dolors més habituals són el cap i l’esquena, el model ens suggereix una solució probable. Tanmateix, si el text que ha de completar proposa “He d’anar a l’oftalmòleg perquè em fa mal” o “feia molt de temps que no anava al gimnàs i em fa mal”, les respostes s’adequaran a la nova informació. Per això, un dels elements més importants en el disseny de *prompts* és el context. Sovint l’èxit d’un *prompt*, entès com l’obtenció de la resposta desitjada o esperada, és justament la introducció del context

adequat i necessari en el sistema d’intel·ligència artificial. “El context és qualsevol informació rellevant que vulguis que el model utilitzi quan respongui la pregunta o realitzi la instrucció”, segons es defineix a Learn Prompting (2024a). D’acord amb de Juana (2024), cal incloure el context adequat, és a dir, inserir tota la informació rellevant i les instruccions addicionals necessàries per millorar la qualitat de la resposta. Aquests tipus de prompts són els “contextual *prompts*” (Giray, 2023).

Per a construir el context, en primer lloc, hem de tenir en compte la instrucció, tasca o pregunta que volem que desenvolupi el sistema d’intel·ligència artificial. A més, per a completar el context d’un *prompt*, podem afegir els següents elements: dades d’entrada, assignació de rols i ús d’exemples. Algunes recomanacions per millorar el prompts, a banda de la inclusió d’un context adequat, són les següents: ser específic, contemplar les limitacions, conèixer l’*user persona*, tenir en compte els valors Top-K i Top-P i, finalment, iterar (de Juana, 2024).

A continuació, es presenten alguns exemples de cada una de les tècniques de disseny de *prompts* segons l’element que s’utilitza per introduir la informació relacionada amb les recomanacions vistes anteriorment per millorar el disseny de prompts.

Instrucció, tasca o pregunta. “*Prompting instruction*” o “*instructive prompt*”, tal com el denomina Giray (2023), consisteix a donar una tasca o indicació a la IA i és la tècnica de creació de *prompts* més simple, tot i que es poden donar instruccions cada vegada més complexes. Alguns dels verbs més utilitzats per a donar instruccions són: escriure, tradueix, categoritza o completa, entre d’altres. En el disseny de *prompts* s’ha de ser específic, no només en la construcció del context, sinó també en la instrucció, tasca o pregunta que volem que desenvolupi el sistema d’intel·ligència artificial. Un altre tipus de *prompts* que podríem utilitzar és el “*question-answer prompt*” que s’estructura al voltant d’una pregunta de recerca (Giray, 2023). Vegem un exemple d’instrucció (de Juana, 2024):

Prompt general: "Parla'm del canvi climàtic".

Prompt específic:

- "Proporciona'm informació detallada sobre les causes i efectes del canvi climàtic centrant-te específicament en el paper de les emissions de gasos d'efecte d'hivernacle i les mesures de mitigació suggerides per la comunitat científica".

Dades d'entrada. És la informació que utilitza la IA, com ara una frase que ha de traduir o un document que ha de resumir (Thomson, 2023). Dins d'aquest element parlarem de dues recomanacions que també es relacionen amb el tipus de sortida (*veure informació més detallada a l'apartat 2.3*)

Contemplar les limitacions. En el *prompt* es poden incorporar restriccions com el nombre de paraules desitjat o el marc temporal (de Juana, 2024) per aconseguir respostes amb la informació justa i necessària.

Valors Top-K i Top-P. Incloure aquests valors en el *prompt* permet controlar la generació de text amb paràmetres que afecten la diversitat i probabilitat de paraules generades.

Assignació de rols. "Role prompting" és una tècnica en la qual es demana a la IA que adopti un perfil determinat (*veure informació més detallada en el capítol 4*). Per exemple: actua com una professora de matemàtiques o com un crític de cinema. És una tècnica per controlar el to i l'estil de sortida del text generat per adaptar-lo a l'usuari persona, és a dir, a l'audiència o públic objectiu (Learn Prompting, 2024c; de Juana, 2024).

Ús d'exemples. Una altra tècnica de disseny de *prompts* és la utilització d'exemples o *shots*. Hi ha tres variants d'aquesta tècnica (*veure informació més detallada a l'apartat 3*):

- *Zero shot-prompting.* En aquesta variant, el *prompt* no conté cap exemple (Learning Prompting, 2024), és a dir, que el model fa prediccions sobre dades no vistes anteriorment sense necessitat de cap formació addicional (Korzynsky et al., 2023). Els *prompts* d'instrucció o de rol que no contenen

cap exemple es poden considerar dins d'aquesta variant. Per exemple: "Capital de França".

- *One-shot prompting.* El *prompt*, en aquesta variant, conté un sol exemple (Learning Prompting, 2024d). Per exemple: "França = París; Itàlia =".
- *Few-shot prompting.* El *prompt* conté dos o més exemples (Learning Prompting, 2024), és a dir, que el model rep un nombre d'exemples per adaptar-se ràpidament a noves tasques (Korzynsky et al., 2023). Per exemple: "França= París; Itàlia= Roma; Andorra =".

Si, malgrat que s'utilitzen els diferents elements d'un *prompt* i les recomanacions anteriors, no s'obté la resposta desitjada serà necessari iterar, és a dir, provar diverses aproximacions, ajustant i refinant constantment el model de *prompt* per obtenir els resultats més rellevants (de Juana, 2024). A més, per a dissenyar el *prompt* podem combinar les tècniques vistes anteriorment de la següent manera: a) *Role & Instruction Prompting*; i b) *Context, Instruction & Few Shot Prompting* (Learning Prompting, 2024b). Obtindríem d'aquesta manera un "mixed prompt" (Giray, 2023).

Com hem vist fins ara, hi ha gran varietat de tipus de *prompts* o tècniques per dissenyar-los que van des dels models més bàsics de pregunta, pregunta + instrucció, instruccions + dades d'entrada o pregunta més exemples (Amatriain, 2023a) fins als més avançats com ara *megraprompts* que continguin tots els elements, cadena de *prompts* que es divideixen les tasques en diferents missatges segons Ana Ivars (Gutiérrez, 2023), *prompt* de cadena de pensament (*Chain of thought prompting*) o arbre de pensament (*Tree of Thought*) que veurem en altres capítols (Amatriain, 2023b). A continuació veurem els *prompts* d'àmbit docent i de l'àmbit acadèmic que poden ser més útils al professorat universitari i d'ensenyament postobligatori i secundari.

Àmbit docent. Els tipus de *prompts* que poden ser útils en l'educació són (Morales-Chan, 2023):

- *Prompts* seqüencials. L'objectiu és guiar converses

de manera lògica i estructurada mitjançant una sèrie de textos previs per obtenir respostes contextualitzades i detallades.

- *Prompts* comparatius. La IA ha de comparar diverses situacions per obtenir resultats específics per obtenir anàlisis detallades.
- *Prompts* argumentals. L'objectiu és que la IA generi un argument a favor o en contra d'una idea per fomentar la reflexió crítica i el desenvolupament d'arguments estructurats.
- *Prompts* de perspectiva professional. Es demana a la IA que adopti un rol específic i descriu un tema en un context determinat per millorar la qualitat de la resposta.
- *Prompts* de llista de desitjos. Són una estructura utilitzada per obtenir respostes més específiques i rellevants, ja que es demana a la IA que compleixi uns requisits específics.

Àmbit acadèmic. Els principals tipus de *prompts* en aquest àmbit són (Lopezosa i Codina, 2023):

- *Prompts* reactius. S'encadenen interaccions a partir de cada resposta.
- *Prompts* estructurals. Se sol·licita una resposta articulada en diversos punts i posteriorment es demana una ampliació de cada punt.
- *Prompts* de rol. La IA adopta un rol, com s'ha vista anteriorment. A més, és semblant en el *prompt* de perspectiva professional.
- *Prompts* amb indicacions d'audiències. S'indica l'audiència específica a la qual s'ha de dirigir la resposta.
- *Prompts* amb objectius. S'inclouen els objectius del text que s'ha de generar.

La taula següent mostra un resum de les recomanacions bàsiques en l'elaboració de *prompts*.

Estratègies i tàctiques d'enginyeria de prompts

ESTRATÈGIES	TÀCTIQUES
Escriure instruccions clares	Incloure detalls a la consulta per obtenir respostes més rellevants Demandar al model que adopti una persona Utilitzar delimitadors per indicar clarament les diferents parts de l'entrada (input) Especificar els passos necessaris per completar una tasca Proporcionar exemples Especificar la longitud desitjada de la sortida (output)
Proporcionar un text de referència	Instruir al model per tal que la resposta utilitzi un text de referència Instruir al model per tal que respongui amb citacions d'un text de referència
Dividir les tasques complexes en sub-tasques més senzilles	Utilitzar la classificació d'intencions per identificar les instruccions més rellevants per a una consulta d'usuari Resumir o filtrar el diàleg anterior, en cas d'aplicacions de diàleg que requereixen converses molt llargues Resumir documents llargs per trossos i construir un resum complet recursivament
Donar temps al model per "pensar"	Instruir al model per tal que elabori la seva pròpia solució abans d'arribar a una conclusió precipitada Utilitzar un monòleg interior o una seqüència de consultes per amagar el procés de raonament del model Preguntar al model si s'ha perdut alguna cosa en passos previs
Utilitzar eines externes	Utilitzar la cerca basada en incrustacions per implementar una recuperació de coneixement eficient Utilitzar l'execució de codi per fer càlculs més precisos o trucar a API externes Donar accés al model a funcions específiques
Provar canvis sistemàticament	Avaluar els resultats del model amb referència a les respostes estàndard

Font: OpenAI developer platform (2024).

2.3. Els tipus de sortida

Un altre dels elements d'un *prompt* és el tipus de sortida. Si volem obtenir un tipus de sortida determinat, cal indicar-ho al sistema d'intel·ligència artificial. Algunes de les indicacions que podem donar-li a la IA i que determinaran el tipus de sortida són les que es poden veure a continuació (Maurine, s.d.; Torres-Salinas & Arroyo-Machado, 2023).

- Llargada: una frase, un paràgraf o un nombre determinat de paraules/caràcters.

- Estructura: cronologia, comparació, causes i efectes, problema i solució, biogràfic, narrativa, etc.
- Tipus de text: correu electrònic, post per a una xarxa social, informe professional, nota de premsa, text literari, etc.
- Format: enumeració amb vinyetes, taula, codi, text pla, preguntes y respostes (Q&A), etc.
- Plataforma: LinkedIn, Twitter, Facebook, etc.

Formats que es poden demanar a una IA

FORMAT	DESCRIPCIÓ
Text pla	Per a respostes directes i generals.
Format de Preguntes i Respostes (Q&A)	Adequat per a entrevistes o sessions de preguntes.
Llistat numerat o de vinyetes	Útil per enumerar punts, passos o elements.
Codi font (Python, R, JavaScript, etc.)	Per a exemples de programació.
Markdown	Per a l'estilització de text simple.
LaTeX	Utilitzat per representar equacions matemàtiques i per a publicacions científiques.
CSV (Valors Separats per Comes)	Per representar dades en forma tabular.
Taules textuais	Taules creades amb caràcters, útils per a comparacions o resums.
Diagrames Ascii Art	Per a il·lustracions textuais simples.
HTML (bàsic)	Estructuració simple de contingut web.
XML (bàsic)	Estructuració de dades jeràrquiques.
JSON (bàsic)	Per estructurar dades en notació d'objecte de JavaScript.
YAML	Representació llegible per humans de dades de serialització.
SQL (bàsic)	Per mostrar consultes de bases de dades.
Esquemes i diagrames (representació textual)	Per descriure estructures de dades o arquitectures.
Emoticones o emoji	Simplificació d'un text en un emoji
Enfatitzar en la resposta	Podem indicar que destaquis part de la resposta en negreta, cursiva, etc.

Font: Torres-Salinas & Arroyo-Machado (2023).

2.4. El to

El to és un altre element dels prompts que té a veure amb el tipus de sortida i és una de les opcions més interessants de la intel·ligència artificial. L'ús d'un to determinat permet adaptar el text o la resposta del sistema d'IA a la nostra audiència o utilitzar el nostre propi to, el de la nostra marca o empresa per a crear els nostres propis textos. Determinant el to es poden crear textos més divulgatius o per tota mena de públics (per exemple, demanant que sigui un text per a públic infantil o per estudiants) o textos més precisos (per exemple, demanant que sigui de caire professional, acadèmic, científic, expert, etc.). És a dir, és una manera d'adaptar els textos al *buyer* o *user* persona, en definitiva, a l'audiència o públic objectiu.

Segons una infografia creada per Ana Ivars (Gutiérrez, 2023), es poden crear textos amb gran varietat de tons o tipus d'escriptura: formal, persuasiu, descriptiu, narratiu, inspiracional, informal, proper, emotiu, amable, tècnic, respectuós i irònic. Per la seva banda, Torres-Salinas i Arroyo-Machado (2023) determinen quatre tipus d'escriptura: literària, acadèmica, creativa i tècnica. Hi ha, però, molts altres tons o adjectius que podríem utilitzar com ara neutre, optimista, amistós, emocional, racional, informatiu, sensacionalista, crític, opinatiu, entre d'altres.

Com s'ha dit anteriorment, es poden utilitzar les tècniques Top-K i Top-P que influeixen en la generació de text afectant la diversitat i probabilitat de paraules generades, segons De Juana (2024). Per una banda, el Top-K determina quantes paraules se seleccionen en cada pas de la generació, oferint més diversitat amb valors més alts i més determinisme amb valors baixos. Per exemple, un valor entre 1 i 5 produeix respostes concises, entre 5 i 10 equilibrades, i per sobre de 20, creatives. Per altra banda, el Top-P selecciona paraules fins que la probabilitat acumulativa superi un llindar definit, oferint més diversitat amb valors alts i més

determinisme amb valors baixos, en un rang de 0,0 per a respostes formals fins a 1,0 per a respostes gairebé aleatòries. Especificar aquests valors en les sol·licituds d'informació millora la qualitat i la rellevància de les respostes generades. Aquests dos criteris es troben a l'apartat de playground del ChatGPT, que ens permet manipular els resultats amb els paràmetres que hem exposat.

Plantilla de prompt amb diferents elements

Act as

```
{{persona you want ChatGPT to simulate}}
```

I am

```
{{persona who's giving instructions}}
```

I want you to

```
{{task}}
```

Constraints:

- {{constraint 1}}
- {{constraint 2}}
- {{constraint 3}}

Output format:

- {{format instruction 1}}
- {{format instruction 2}}

La plantilla conté assignació de rols, instruccions, tasca, restriccions de sortida i formats de sortida. Font: Maurine (s.d.).

Pràctica. El *prompt* perfecte

L'enginyeria de prompts cerca la millor pregunta per una resposta esperada. A banda del to, el format de sortida o el tipus de pregunta, el context és essencial per assolir una sortida adient. L'objectiu d'aquesta pràctica és comparar una sortida del model amb una pregunta genèrica i comparar-la amb aquella que ha estat formulada amb un bon context. Podem partir d'una pregunta molt genèrica com ara "Una ciutat metropolitana manifesta un problema d'accés a l'habitatge, especialment entre la població jove. Quines mesures podria aplicar l'ajuntament per a fer front a aquesta tensió?". Els models de llenguatge (com ChatGPT o Copilot) proporcionaran respostes genèriques, que podrien ser aplicades en qualsevol ciutat del món.

Per tal d'afinar la resposta, els grups d'estudiants han de cercar el context que ajudi a elaborar una pregunta amb una informació més precisa. Per això, els estudiants haurien de cercar informació sobre:

- Estructura de les llars (nombre de membres)
- Evolució del preu de l'habitatge
- Evolució de la renda mitjana
- Proporció entre contractes de lloguer i propietat
- Percentatge d'habitatges buits
- Evolució del padró
- Estructura de la població per edats, per renda o per origen
- Mobilitat obligada
- Evolució de l'activitat turística
- Distribució dels valors per barris o districtes
- Valors dels municipis veïns

A mesura que contextualitzem la ciutat, les possibles solucions varien perquè s'evidencien els factors que més influeixen en l'accés a l'habitatge i les singularitats de la ciutat d'estudi. El *prompt* també varia si demanem que assumeixi determinats rols, que poden modificar la perspectiva, com per exemple urbanistes especialitzats en habitatge, activistes veïnals o economistes. La comparació entre els resultats d'acord amb els diferents contextos plantejats (i els diferents rols) ajuden a entendre la importància del context i la selecció precisa d'informació fiable sobre l'element d'estudi.

Pràctica. Variacions sobre una pregunta

En aquesta pràctica, utilitzarem l'"expert prompting" (Amatriain, 2023b) per veure com influeix l'ús d'una pregunta o una altra en la resposta de la IA. Podem situar la pràctica en l'àmbit del periodisme i demanar a l'estudiantat que preparin un reportatge sobre un tema concret com, per exemple, el canvi climàtic, la violència de gènere o l'accés a l'habitatge. Per fer el reportatge hauran de consultar a diferents experts. Per desenvolupar l'expert prompting, l'estudiantat ha de seguir tres passos:

- Demanar a la IA que identifiqui experts en el tema.

- Demanar a la IA que respongui com si fos algun d'aquests experts.
- Prendre una decisió final contrastant les diferents respostes generades.

Per a desenvolupar la pràctica, l'estudiantat haurà de fer servir dues IA: claudie i gemini. L'estudiantat haurà de desenvolupar dues entrevistes amb dos experts/es diferents, haurà de fer diverses variacions d'una mateixa pregunta i fer les entrevistes a les dues IA. Amb el resultat, haurà de fer el reportatge on s'inclouin les declaracions dels experts entrevistats.

Aquesta pràctica podria tenir variacions. En el camp del periodisme, es podria demanar que la IA adoptés rols de personatges de l'actualitat política, econòmica, social o cultural. A més, podria ser una pràctica prèvia a la realització de l'entrevista real. La pràctica també es podria adaptar a altres àmbits:

Màrqueting. Podria servir per a la construcció del buyer persona o la creació d'arguments de venda de productes i serveis.

Medicina. Podria servir com a pràctica per a comunicar males notícies als pacients o per fer l'entrevista clínica per arribar a un diagnòstic.

Pràctica. Construir contextos: La documentació

La Documentació es una disciplina científica que estudia el comportament i les propietats de la informació per a fer-la accessible. Té un caràcter holístic i l'objectiu principal abasta la recollecció, emmagatzematge, anàlisi i difusió de la informació. Per a construir contextos en el disseny de prompts, podem utilitzar algunes de les seves tècniques que ens permetran crear pràctiques per a l'estudiantat. La construcció de rols (personatges històrics, literaris, etc.) o d'estudis de cas poden ser dues maneres de fer-ho.

En aquesta pràctica, podríem construir un context a partir dels repositoris de dades que tenim disponibles actualment a Internet. Per exemple, Google ens ofereix un buscador per cercar datasets^[1], és a dir, conjunts de dades. L'estudiantat podria, a partir de la IA i els conjunts de dades, elaborar informes sobre sectors o temes concrets com ara d'educació, màrqueting o turisme. Els datasets més emprats solen ser els d'educació, temps, càncer, crim, futbol o gossos, entre d'altres (Noy, 2020).

La proposta podria ser que l'estudiantat es posi en la situació que han de determinar la política educativa d'un país determinat pels propers cinc anys. Per a fer-ho, hauran d'analitzar les dades d'educació, fer-ne un informe i proposar polítiques que puguin ser aplicades. L'informe haurà de contenir dades que buscaran en el cercador de datasets de Google i seran tractades a través de Gemini. A més, l'informe pot ser redactat a través d'aquesta IA sempre i quan la informació sigui contrastada i verificada. Als annexos de la pràctica caldrà afegir els prompts i les respostes obtingudes de la IA. Dins de la pràctica es podran posar textos de la IA fent les citacions corresponents, així com les modificacions que calgui per tal que sigui una informació fiable.

Referències

- Amatriain, X. (2023a). Prompt Engineering 101: Introduction and resources. Amatriain Blog. Data de recuperació: 14/03/2024. <https://amatriain.net/blog/PromptEngineering>
- Amatriain, X. (2023b). Prompt Engineering 201: Advanced methods and toolkits. Amatriain Blog. Data de recuperació: 14/03/2024. <https://amatriain.net/blog/prompt201#expert>
- Centro de Documentación Publicitaria (2024). Protagonistas: Alex Faickney Osborn. Data de recuperació: 14/03/2024. <https://www.lahistoriadelapublicidad.com/protagonista-174/alex-faickney-osborn>
- De Juana, R. (2024). Cómo mejorar tus Prompts en seis cómodos pasos. MCPRO. Data de recuperació: 14/03/2024. <https://www.muycomputerpro.com/2024/01/24/como-mejorar-tus-prompts-en-seis-comodos-pasos>
- Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. Annals of Biomedical Engineering, 1-5
- Gutiérrez, D. (2023). Hoja de truco de ChatGPT, de Ana Ivars. Data de recuperació: 14/03/2024. https://www.linkedin.com/posts/davidgutierrezamd_%C3%BAltima-hoja-de-trucos-de-chat-gpt-activity-7098491435875749888-qgFw/?originalSubdomain=es
- Korzynski, P., Mazurek, G., Krzypkowska, P., & Kurasniski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. Entrepreneurial Business and Economics Review, 11(3), 25-37.
- Learn Prompting (2024a). Formalizing Prompts. Data de recuperació: 14/03/2024. <https://learnprompting.org/docs/basics/formalizing>
- Learn Prompting (2024b). Combining techniques. Data de recuperació: 14/03/2024. https://learnprompting.org/docs/basics/combining_techniques
- Learn Prompting (2024c). Assigning Roles. Data de recuperació: 14/03/2024. <https://learnprompting.org/docs/basics/roles>
- Learn Prompting (2024c). Showing examples. Data de recuperació: 14/03/2024. https://learnprompting.org/docs/basics/few_shot
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. The Journal of Academic Librarianship, 49(4).
- Lopezosa, C., & Codina, L. (2023). Inteligencia Artificial y métodos cualitativos de investigación: ChatGPT para revisiones de la literatura y análisis de entrevistas semiestructuras. Data de recuperació: 14/03/2024. http://eprints.rclis.org/44296/1/IA-ACADEMICO-ChatGPT%20ATLAS-SCOPING-2023_vf-2.pdf
- Lopezosa, Carlos; Codina, Luís; Ferran-Ferrer, Núria (2023) ChatGPT como apoyo a las systematic scoping reviews: integrando la inteligencia artificial con el framework SALSA. Colecció del CRICC. Barcelona: Universitat de Barcelona. Data de recuperació: 14/03/2024. <http://hdl.handle.net/2445/193691>
- Maurien, F. (s.d.). ChatGPT for Educators | Prompting 101. Data de recuperació: 14/03/2024. <https://wooclap.notion.site/ChatGPT-for-Educators-Prompting-101-0b4c015d7b8a4e1f834c1f61504042f2>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO <https://discovery.ucl.ac.uk/id/eprint/10176438/>
- Morales-Chan, M. A. (2023). Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza. Data de recuperació: 14/03/2024. <http://biblioteca.galileo.edu/tesario/handle/123456789/1348>
- Noy, N. (2020). Discovering millions of datasets on the web. Google The Keyword. Data de recuperació: 14/03/2024. <https://blog.google/products/search/discovering-millions-datasets-web/>
- OpenAI developer platform (2024). Prompt engineering. Data de recuperació: 14/03/2024. <https://platform.openai.com/docs/guides/prompt-engineering>
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Cataloging prompt patterns to enhance the discipline of prompt engineering. https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf
- Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2022). Prompting GPT-3 to be reliable. <https://arxiv.org/abs/2210.09150>
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Cataloging prompt patterns to enhance the discipline of prompt engineering. Data de recuperació: 14/03/2024. https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf
- Thomson, E. (2023). A brief understanding of prompt engineering. Medium. Data de recuperació: 14/03/2024. <https://medium.com/@SymeCloud/a-brief-understanding-of-prompt-engineering-bl76ba9fb2ba>
- Torres-Salinas, Daniel; Arroyo-Machado, W. Manual de Chat GPT: aplicaciones en investigación y educación universitaria (2023). InfluScience Ediciones. Data de recuperació: 14/03/2024. <https://hdl.handle.net/10481/85145>
- Velásquez-Henao, J.D., Franco-Cardona, C.J. & Cadavid-Higuera, L., Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. DYNA, 90 (230), Especial Conmemoración 90 años, pp. 9-17, Noviembre, 2023.
- Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens-With new examples of knowledge, skills and attitudes. Publications Office of the European Union
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382. <https://doi.org/10.48550/arXiv.2302.11382>

¹¹ Es pot accedir al buscador de conjunts de dades (datasets) de Google en el següent enllaç: <https://datasetsearch.research.google.com/>

03

Q&A en acció

David: — Atura! Para!

Joshua: — Estàs segur? Deixarà de ser un joc.

David: — Sí, para! No és un joc. Si ataqués, ells contraatacaran. Milions moriran.

Joshua (continua executant simulacions d'atacs nuclears).

David: — Mira-ho! Mira-ho! No importa com ho organitzis. És un joc que no es pot guanyar.

Joshua (simulant més escenaris, més ràpid).

David (a Joshua): — Veus? No pots guanyar! És un joc sense sentit!

Joshua (finalment processa els resultats de les seves simulacions): — Un moment estrany. L'únic moviment guanyador és no jugar.

A la pel·lícula *War Games* (1983) l'ordinador Joshua ha aconseguit el control del sistema nuclear nord-americà i amenaça d'iniciar un atac global. El protagonista del film, David, descobreix una esclota: Si el sistema simula "jocs de guerra" comprova que no és possible guanyar, com en el joc de tres en ratlla. Podem entrenar els models perquè desenvolupin una determinada tasca simplement reproduint una seqüència que haurà de repetir de forma automàtica, com fa David amb en Joshua. Els *prompts* que es basen en aquest format s'anomenen de Q & A (preguntes i respostes), i són un format d'entrada molt eficient, una subvariant de l'univers dels *prompts* que hem vist en el capítol anterior.

La majoria de sistemes d'intel·ligència artificial que s'han popularitzat recentment, com ara el Chat GPT, Gemini o Copilot, són sistemes al quals se'ls pot fer preguntes i que donen lloc a respostes automàtiques. D'acord amb el disseny de *prompts* que nosaltres utilitzem, obtindrem unes respostes més o menys acurades. Podem donar una tasca o instrucció que ha de desenvolupar el sistema

d'intel·ligència artificial, però també podem fer-ho a través de preguntes.

Els models de pregunta i resposta poden també basar-se en exemples o *shots*. En el primer apartat d'aquest capítol parlarem dels *shots*. A vegades, no donem cap exemple i esperem que el sistema d'intel·ligència contesti sense tenir una referència prèvia. És el que anomenem el model de *zero shot*. En canvi, en altres ocasions utilitzem un nombre determinat d'exemples per a entrenar el model en respostes amb una determinada estructura o amb una sortida en un format o to determinat. Són els models de *one-shot* i *few-shots*.

Després de veure aplicacions dels models de *zero-shot*, *one-shot* i de *few-shots*, en el segon subapartat ens endinsarem en els diversos models de preguntes i respostes (*Questions & Answers* o Q&A, en la seva denominació en anglès). Finalment, el darrer apartat d'aquest capítol veurem diverses maneres per entrenar el model per tal d'obtenir les respostes que s'adeqüin més al que nosaltres necessitem. Quan fem consultes molt complexes, pot ser una bona idea dividir la consulta en diversos *prompts*, és a dir, en diverses preguntes per tal de refinar més la resposta. A més, hi ha diverses estratègies per entrenar el model i aconseguir resoldre el problema o els objectius que ens hem plantejat.

Els models de Q&A en acció poden servir al professorat per crear activitats, casos o tasques per l'estudiantat que segueixin un patró determinat. Per exemple, podem fer una activitat de resolució de casos i que cada grup de la classe resolgui un cas diferent, però sobre un tema determinat i d'aquesta manera poder enriquir el coneixement generat a classe i que cada grup presenti aquests casos a classe per obtenir diverses perspectives d'una mateixa problemàtica.

No només podem entrenar els models d'IA, sinó que també es pot proporcionar a l'estudiantat exemples de com fer les tasques o treballs que els demanem a classe. La IA, doncs, pot servir a l'estudiantat per "entrenar-se" en el desenvolupament d'un treball específic o la utilització de determinats programes (bases de dades acadèmiques o professionals, gestors bibliogràfics, etc.).

3.1. Els shots

Un *shot* o tret és un exemple. Utilitzant un o diversos exemples podem entrenar el model d'IA per tal d'obtenir respostes més adequades. D'aquesta manera, podem trobar tres tipus de *prompts*: *zero-shot*, *one-shot* i *few-shot*. A continuació s'expliquen.

Zero shot-prompting (0S). Aquest tipus de *prompt* és una instrucció, pregunta o tasca que no ofereix cap exemple concret per tal que el sistema d'IA doni una resposta (Prompt Engineering Guide, 2024a). Els models poden resoldre tasques utilitzant el que ja saben perquè han estat preentrenats amb gran quantitat de text (Thomson, 2023). Aquest mètode, segons Brown et al. (2020), ofereix la màxima comoditat i evita correlacions espúries (a menys que ocorrin a través del corpus de dades de preentrenament), però també és l'entorn més desafiant. És el més proper a com realitzen les tasques els humans, però fins i tot també per a ells pot ser difícil sense un exemple previ (Brown et al., 2020). Vegem un exemple a continuació.

- **Prompt:** Quin és l'estil principal d'aquest pintor?.
- **Sense exemple:** Dalí.

One shot-prompting (1S). Aquest tipus de *prompt*, a banda de donar una instrucció pregunta o tasca, ofereix un exemple per completar la demanda que se li dona al sistema d'IA. Vegem un exemple a continuació.

- **Prompt:** Quin és l'estil principal d'aquest pintor?
- **Exemple 1:** Dalí: Surrelisme
- **Tasca:** Van Gogh:

Few shot-prompting (FS). En aquest cas, el *prompt* inclou entre 3 i 5 exemples que serviran de guia al sistema d'IA per a donar la resposta o output. Aquest model es pot fer servir per classificar elements (Prompt Engineering Guide, 2024b) o quan volem que la sortida tingui un format concret (Martínez, 2023). Segons Brown et al. (2020), un dels avantatges és una major reducció de la necessitat de dades específiques per a la tasca. L'enginyeria de *prompts* és crucial en l'ús efectiu d'aquest model, per tant, cal incloure exemples

clars i variats i en ordre aleatori per evitar biaixos i millorar els resultats (Thompson, 2023). A més, millora el rendiment de la IA. Vegem un exemple a continuació.

- **Prompt:** Quin és l'estil principal d'aquest pintor?
- **Exemple 1:** Dalí: Surrelisme
- **Exemple 2:** Van Gogh: Postimpressionisme
- **Exemple 3:** Much: Expressionisme
- **Tasca:** Klimt:

Els models són eficients per respondre a preguntes simples, especialment factuais, i segueixen molt bé les instruccions. Per exemple, podem entrenar el sistema perquè identifiqui les fortaleses d'un conjunt de criteris que es poden agrupar en un model DAFO (debilitats, amenaces, fortaleses, oportunitats) o que identifiqui els factors que influeixen en el clima d'una destinació. Però a mesura que les respostes previstes incrementen la seva complexitat, cal que entrenar el model amb una gran quantitat de dades. Aquest tipus d'entrenament s'anomena *Fine-Tuning* (FT) i es basa en la utilització de molts *shots*. Es pot considerar un model que utilitza exemples, ja que típicament, segons Brown et al. (2020), n'utilitza milers i, fins i tot, centenars de milers. L'avantatge d'aquest sistema és que té un bon rendiment, però com a desavantatges cal comptar que necessita un gran conjunt de dades nou per a cada tasca i la possibilitat d'explorar característiques espúries de les dades d'entrenament, entre d'altres (Brown et al., 2020).

L'aprenentatge *Few-Shot* (FS) i el *Fine-Tuning* (FT) són mètodes diferents per entrenar el model. Mentre que el primer utilitza exemples dins del prompt per proporcionar context (*in-context learning*), el segon implica reentrenar el model amb un conjunt de dades específiques i, per tant, n'altera els paràmetres interns (Thomson, 2023).

3.2. Els models Q&A

Els models de pregunta-resposta (*Question and Answer*, Q&A) formen part d'una subdisciplina

del processament de llenguatge natural (*Natural Language Processing*, NLP). L'objectiu d'aquesta subdisciplina és construir sistemes capaços de respondre a les preguntes dels humans utilitzant el llenguatge natural (Caballero, 2021). En educació, els models de Q&A es poden aplicar, per exemple, en donar retroacció (*feedback*) o explicacions sobre un treball, pràctica o examen a l'estudiantat. També s'utilitzen aquests models en els assistents virtuals o en els xatbots (vegeu el capítol 15).

Aquests sistemes segueixen una estructura de tres mòduls principals: anàlisi de preguntes, recuperació dels passatges i extracció de les respostes (Ojokoh i Adebisi, 2019). En primer lloc, el mòdul d'anàlisi de preguntes processa la pregunta per entendre el seu significat, realitzant activitats com la tokenització, desambiguació, etiquetatge semàntic i classificació de preguntes, transformant-les en una representació adequada per a la recuperació d'informació. En segon lloc, el mòdul de recuperació de passatges utilitza motors de cerca per recuperar documents o fragments de text rellevants, incloent-hi tècniques de modelatge de llenguatge de consulta i funcions de classificació estadística. Finalment, el mòdul d'extracció de respostes identifica i extreu les respostes candidates dels passatges recuperats mitjançant tècniques com el reconeixement d'entitats mencionades, patrons i estructures sintàctiques, classificades segons la seva rellevància i precisió. (Ojokoh i Adebisi, 2019)

A continuació repassarem algunes classificacions dels models Q&A i veurem exemples que mostren per a què ens poden servir en la docència.

Classificació basada en el domini (Prompt Engineering Guide, 2024c):

- **Models Q&A de domini tancat o restringit.** Aquests models responen a preguntes d'un tema específic. Existeixen diversos sistemes de preguntes i respostes de domini tancat, incloent-hi sistemes temporals, geoespacionals, mèdics, de patents i comunitaris (Ojokoh i Adebisi, 2019).

- **Models Q&A de domini obert.** Aquests models responen a preguntes fàctiques sense que se'ls proporcionin cap evidència i de qualsevol tema fent servir una àmplia base de coneixements. Dins dels models de domini obert trobem una sub-classificació basada en les fonts de dades que s'explica a continuació.

Sub-classificació basada en les fonts de dades (Caballero, 2021):

- **Models Q&A basats en text o recuperació d'informació.** Aquests sistemes es basen en documents no estructurats com poden ser extractes textuals d'una font d'informació o d'una pàgina web.

Per exemple: Es pot demanar a l'estudiantat que analitzin una entrada a la Wikipedia a través de preguntar al text per la resposta a diferents qüestions. En el camp del periodisme, es podria analitzar l'entrada sobre desinformació i preguntar sobre els tipus de desinformació, definició de *fact checking*, etc.

- **Models Q&A basats en coneixement.** Aquests sistemes obtenen les respostes de dades estructurades generalment en bases de dades SQL o gràfics RDF que inclouen relacions, fets i entitats.

Per exemple: A través d'una base de dades d'estadístiques sobre l'accés a l'educació, es pot demanar a l'estudiantat que cerquin respostes a preguntes concretes com ara percentatge d'infants escolaritzats, percentatge de centres públics, privats i concertats, etc.

- **Models Q&A híbrids.** Aquests sistemes utilitzen dades semi-estructurades per obtenir respostes, per tant, llegeixen dades estructurades i no estructurades per maximitzar l'eficiència i la precisió.

Per exemple: Es pot demanar a l'estudiantat que verifiquin les dades consignades en una entrada de la Wikipedia amb dades d'una base de dades.

En l'àmbit de les ciències polítiques, es podrien donar dades electorals i verificar-les en una entrada de la Wikipedia sobre les eleccions d'un determinat país. Un altre exemple podria ser el següent: Verificació de dades dels partits polítics en relació a la immigració, delinqüència, etc.

Classificació basada en l'enfocament de les respostes (Ojokoh i Adebisi, 2019):

- **Models Q&A extractius o de respostes extretes.** El model extreu les respostes d'un context o document. Inclouen oracions, paràgrafs i multimèdia.
- **Models Q&A generatius o de respostes generades.** El model genera text lliure directament basat en el context. Inclouen respostes confirmacionals (sí o no), opinatives i respostes en forma de diàleg.

Classificació basada en el tipus de preguntes (Ojokoh i Adebisi, 2019):

- **Preguntes factuais.** Preguntes que busquen fets específics (què, quan, qui, quin/a, com), com ara, quina és l'extensió de Còrsega.
- **Preguntes en forma de llistes.** Preguntes que requereixen enumeracions, com ara, quins estats exporten cacau.
- **Preguntes que busquen una definició.** Preguntes que busquen definicions detallades, com ara, què és un adverbí.
- **Preguntes hipotètiques.** Preguntes sobre esdeveniments hipotètics, com ara, què passaria si...
- **Preguntes causals.** Preguntes que busquen explicacions, és a dir, que pregunten el per què o el com.
- **Preguntes de confirmació.** Preguntes que requereixen respostes de sí o no. Les respostes dicotòmiques són molt útils per a verificar. Per exemple, identificar països amb sortida al mar, o països que s'han adherit a un determinat tractat internacional.

3.3. Entrenar el model

Un model d'intel·ligència artificial es pot entrenar a partir de la introducció de dades i de l'ús d'exemples per tal d'obtenir respostes més acurades o ajustades a les nostres expectatives. Per tant, per desenvolupar aquesta part del capítol emprarem exemples (o *shots*), així com la introducció de dades o un context a partir dels sistemes Q&A vistos anteriorment.

Un concepte fonamental per a l'entrenament del model, és el *machine learning* (ML) o aprenentatge automàtic: "L'aprenentatge automàtic és un subconjunt de la intel·ligència artificial que permet que un sistema aprengui i millori de forma autònoma mitjançant xarxes neuronals i aprenentatge profund, sense haver de ser programat explícitament, a través de la ingesta de grans quantitats de dades" (Google Cloud, 2024). Són la base de DALL-E o Chat GPT, entre

d'altres. Hi ha quatre tipus principals d'aprenentatge automàtic (Sánchez, 2020):

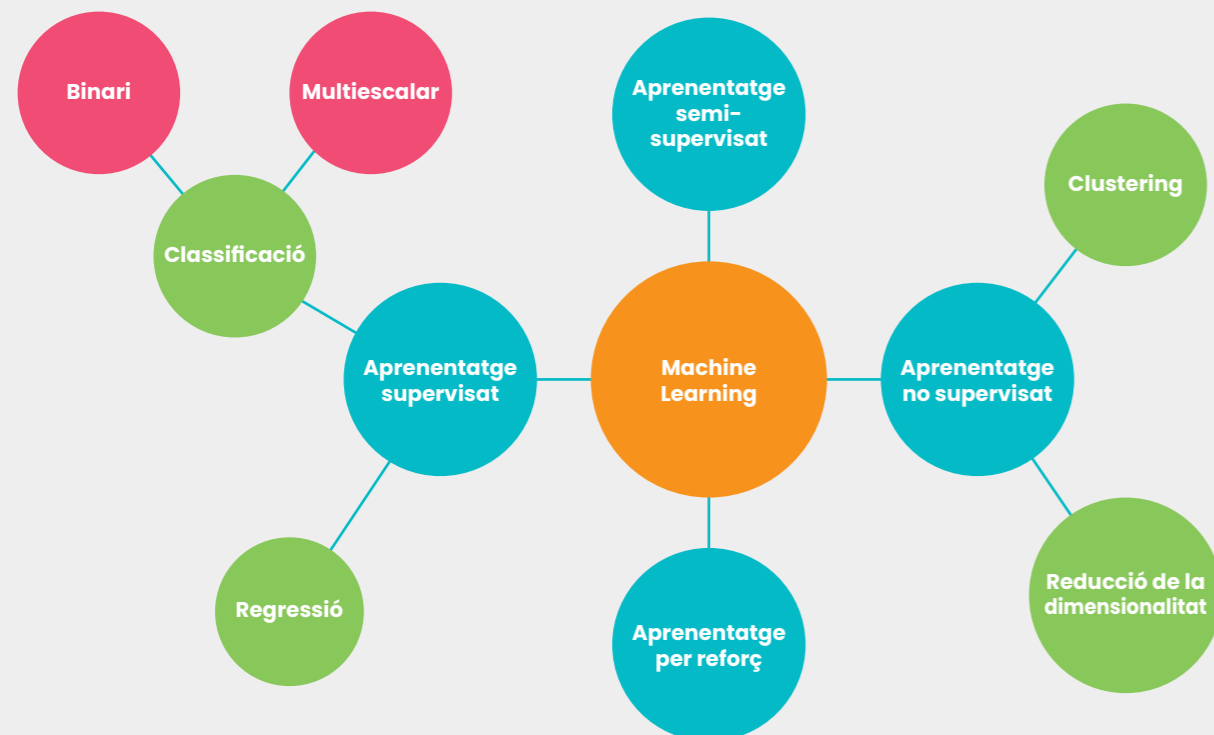
- **Aprenentatge supervisat.** Utilitza dades etiquetades per entrenar models de classificació (amb etiquetes discretes) o regressió (amb valors reals).
- **Aprenentatge no supervisat.** Treballa amb dades sense etiquetar per descobrir patrons ocults o reduir la dimensionalitat.
- **Aprenentatge semi-supervisat.** Combina els dos anteriors, etiquetant manualment una petita part de les dades i utilitzant-les per entrenar models que després etiqueten la resta.
- **Aprenentatge per reforç.** Recompensa comportaments desitjats i penalitza els no desitjats. S'aplica en àrees com els jocs, la robòtica i la optimització de recursos.

Cassie Kozyrkov (2019) presenta un full de ruta detallat per a projectes d'aprenentatge automàtic (AA) i intel·ligència artificial (IA). El procés comença amb una verificació de la necessitat real de AA/IA, seguit per la definició d'objectius, accés i preparació de dades, exploració i eines per entrenar models, depuració i ajust, validació, proves, producció, experiments en viu i manteniment continu. Cada pas emfatitza la importància de la iteració i l'avaluació rigorosa per assegurar l'efectivitat i viabilitat del sistema ML/AI en producció.

Aquest procés es pot resumir en cinc passos que es poden veure a continuació.

- **Primer pas:** Definir l'objectiu. És a dir, saber quin és el problema que es vol resoldre. Caldrà identificar les dades que s'utilitzaran i les preguntes que es volen respondre (Meroyo, 2023). Podem entrenar el model per a diferents finalitats com ara respondre qüestions del llenguatge natural, fer classificacions de dades i buscar-ne patrons, etc.
- **Segon pas:** Preparar les dades. En primer lloc, cal recopilar les dades necessàries per entrenar el model i, en segon lloc, cal fer-ne la curació per tal que siguin depurades, vàlides i es puguin utilitzar. És important que les dades siguin representatives del problema a resoldre i que es preprocessin per eliminar valors atípics, omplir valors faltants i normalitzar les dades (Meroyo, 2023).
- **Tercer pas:** Seleccionar i entrenar el model. Cal començar l'abocament de les dades en el model que s'ha dissenyat per tal que en comenci a aprendre. Prèviament, s'ha seleccionat el model d'aprenentatge automàtic més adequat a l'objectiu establert anteriorment. Els models poden ser: classificadors, regressors o d'agrupament, entre d'altres (Meroyo, 2023).
- **Quart i cinquè pas:** Avaluació i implementació del model. En aquestes dues fases finals cal validar el model i avaluar la seva fiabilitat: "Això implica testar el model amb dades que no s'hagin emprat durant l'entrenament i mesurar la seva precisió" (Meroyo, 2023). Si el model no compleix els requisits per solucionar el problema, caldrà ajustar-lo.

Tipus d'aprenentatge automàtic



Font: Sánchez (2020).

Pràctica. Exemples de shots

Objectiu

L'objectiu d'aquesta pràctica és demostrar la importància de proporcionar exemples i context als sistemes d'IA per obtenir respostes més rellevants i acurades.

Consideracions prèvies

La metodologia que s'utilitzarà serà l'estudi de casos, específicament un estudi interactiu de casos. Cada grup de treball tindrà una sèrie de casos relacionats amb una temàtica molt similar, tot i que disposarà de diferents nivells d'informació de context. La pràctica se situa en el camp de l'economia d'empresa.

L'estudiantat haurà de resoldre els casos a partir de la interacció amb un sistema d'intel·ligència

artificial. En aquest cas, es recomana l'ús de Copilot, la IA de Microsoft. L'estudiantat aprendrà a aplicar els conceptes apresos en situacions reals, mentre que la IA els donarà feedback i assistència en la presa de decisions.

La pràctica permet a l'estudiantat desenvolupar habilitats de presa de decisions empresarials utilitzant tecnologies com ara la IA.

Procediment

1. L'alumnat es dividirà en grups de 3 o 4 per desenvolupar l'exercici.
2. Cada grup de treball rebrà una sèrie de casos relacionats amb el camp de l'economia d'empresa.
3. Cada un dels casos tindrà un nivell d'informació diferent i cada grup haurà de prendre decisions en l'àmbit d'una empresa amb l'ajuda de la IA.
4. La interacció amb la IA servirà per crear matrius DAFO o de cost d'oportunitat per prendre decisions amb situacions de més o menys incertesa.
5. Cada grup exposarà davant la classe les decisions preses en cada un dels casos i es generarà un debat al voltant de la presa de decisions en el camp de l'economia d'empresa.

Casos d'economia d'empresa

CAS 1: Inversió en Nous Projectes (Alt Nivell d'Incertesa)

Context: Una empresa està considerant invertir en tres projectes diferents, a continuació hi ha informació sobre els detalls de cada un d'ells:

- **Projecte A. Intel·ligència Artificial (IA) en Agricultura**

Descripció: El projecte A implica desenvolupar un sistema d'IA per optimitzar la gestió de cultius agrícoles. L'empresa no té dades específiques sobre el mercat agrícola ni sobre la viabilitat tècnica de la solució d'IA.

Incertesa: L'empresa no sap si hi haurà demanda per aquesta solució ni quins seran els costos de desenvolupament.

Decisió: S'hauria d'invertir en aquest projecte o no? El cost d'oportunitat és especialment alt degut a la falta d'informació.

- **Projecte B. Energies Renovables en Països Emergents**

Descripció: El projecte B implica invertir en la instal·lació de panells solars en països emergents. L'empresa no té dades concretes sobre els mercats, costos d'instal·lació ni beneficis potencials.

Incertesa: No se sap si hi haurà prou demanda en aquests països ni quin serà el retorn de la inversió.

Decisió: S'hauria d'invertir en aquest projecte o no? El cost d'oportunitat és alt a causa de la falta d'informació.

- **Projecte C. Investigació en Bateries d'Emmagatzematge**

Descripció: El projecte C implica investigar i desenvolupar bateries d'emmagatzematge d'energia més eficients. L'empresa no té dades sobre la viabilitat tècnica ni sobre els costos d'investigació.

Incertesa: No se sap si aquesta tecnologia serà rendible en el futur ni quan arribarà a ser-ho.

Decisió: S'hauria d'invertir en aquest projecte o no? El cost d'oportunitat és alt degut a la falta d'informació.

CAS 2. Selecció de Personal (Nivell Mitjà d'Incertesa)

Context: L'empresa té dos candidats per a un càrrec directiu. Un té experiència en gestió d'equips, però no se sap si s'adaptarà bé a l'entorn digital. L'altre és més jove i té habilitats digitals, però no se sap si té experiència en lideratge.

Incertesa: Hi ha incertesa sobre quin candidat serà més efectiu a llarg termini.

Decisió: Quin és el cost d'oportunitat d'escollir un sobre l'altre? Quines habilitats són més valuoses per a l'èxit de l'empresa?

CANDIDAT 1	CANDIDAT 2
<p>Experiència en gestió d'equips Ha treballat com a gerent de projectes en una empresa de tecnologia durant els últims 5 anys. Supervisà amb èxit un equip de 20 persones, aconseguint els objectius del projecte dins del termini i el pressupost establerts. Implementà estratègies de lideratge efectives per fomentar la col·laboració i la productivitat.</p>	<p>Habilitats digitals Va treballar com a analista de màrqueting digital en una agència de publicitat. Va crear campanyes exitoses en xarxes socials i va augmentar la presència en línia de l'empresa. Té certificacions en Google Analytics i SEO.</p>
<p>Educació Llicenciatura en Administració d'Empreses amb especialització en Gestió de Recursos Humans.</p>	<p>Educació Màster en Màrqueting Digital i Comunicació.</p>
<p>Habilitats addicionals Bones habilitats de comunicació i resolució de conflictes. Coneixements sòlids en gestió de projectes i planificació estratègica.</p>	<p>Experiència en Lideratge Va ser líder d'un grup de voluntaris en una ONG, coordinant activitats i supervisant altres voluntaris.</p>
	<p>Habilitats addicionals Creativitat i adaptabilitat en entorns digitals. Coneixements en eines d'anàlisi de dades i tendències digitals.</p>

Font: Resposta generada per Copilot de Microsoft.

Pràctica. Preparant un debat

Objectiu

L'objectiu d'aquesta pràctica és demostrar a l'alumnat com poden utilitzar els models de preguntes i respostes per a millorar els seus treballs acadèmics.

Consideracions prèvies

En aquest cas, es pot situar l'exercici en l'àmbit de les ciències polítiques. L'alumnat podrà escollir entre diversos temes polèmics o conflictius per tal d'elaborar un argumentari que justifiqui un seguit de polítiques públiques que els resolguin. La idea és poder fer una proposta de resolució de problemàtiques socials a través de la posada en funcionament de determinades polítiques públiques. Un altre àmbit en el qual es podria aplicar aquest exercici és en la filosofia o l'ètica.

Els sistemes d'IA s'utilitzarien per a millorar l'argumentació en favor i en contra de determinades decisions o accions polítiques, ètiques o filosòfiques. Per exemple algunes de les temàtiques que es podrien escollir són: immigració, eutanàsia, avortament, igualtat de gènere, polítiques penitenciàries, legalització de drogues, prostitució, renda bàsic universal, etc. En aquest cas, es recomana l'ús de Gemini, la IA de Google.

Procediment

1. L'alumnat es dividirà en grups de 3 o 4 per desenvolupar l'exercici.
2. Cada grup de treball rebrà una temàtica i una postura a favor o en contra.
3. La interacció amb la IA servirà per crear l'argumentari a favor o en contra de la temàtica assignada. Els grups no només utilitzaran la IA per crear l'argumentari de la seva postura, sinó també per entrenar-se de cara al debat. D'aquesta manera, els grups podran crear l'argumentari i el contra-argumentari per tal de neutralitzar les respostes de l'equip contrari.
4. Posteriorment, s'organitzarà un debat entre un grup favorable i un altre desfavorable a la mateixa temàtica.
5. La resta dels grups, mentre es produeix el debat, avaluarien la qualitat dels arguments i decidirien quin dels dos grups ha argumentat millor la seva posició. Tots els grups haurien de donar feedback sobre el debat.

Referències

- Brown, T.B., et al. (2020). Language Models are Few-Shot Learners. *Advances in neural information processing systems*, 33, 1877-1901. Data de recuperació: 27/05/2024. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- Caballero, M. (2021). A Brief Survey of Question Answering Systems. *International Journal of Artificial Intelligence & Applications (IJAA)*, 12(5). Data de recuperació: 27/05/2024. <https://ssrn.com/abstract=3996229>
- Google Cloud (2024). ¿Qué es el aprendizaje automático? Data de recuperació: 27/05/2024. <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419#section-4>
- Kozyrkov, C. (2019). 12 Steps to Applied AI. A roadmap for every machine learning project. *Medium*. Data de recuperació: 24/05/2024. <https://medium.com/swlh/12-steps-to-applied-ai-2fdad7fdcdf3>
- Learn Prompting (2024a). Zero-Shot Prompting. Data de recuperació: 27/05/2024. <https://www.promptingguide.ai/techniques/zeroshot>
- Learn Prompting (2024b). Text Classification. Data de recuperació: 27/05/2024. <https://www.promptingguide.ai/introduction/examples#text-classification>
- Learn Prompting (2024c). Question Answering. Data de recuperació: 27/05/2024. <https://www.promptingguide.ai/prompts/question-answering>
- Martínez (2023). <https://medium.com/ia-generativa-un-mundo-de-posibilidades/qu%C3%A9-es-eso-de-los-few-shot-prompting-6d82cff46df4>
- Meroyo, J. (2023). Cómo entrenar un modelo de aprendizaje automático con IA desde cero. *Bilbolink*. Data de recuperació: 24/05/2024. <https://bilbolink.com/como-entrenar-un-modelo-de-aprendizaje-automatico-con-ia-desde-cero/>
- Ojokoh, B. i Adebisi, E. (2019). A Review of Question Answering Systems. *Journal of Web Engineering*, 17(8), 717-758.
- Sánchez, J.A. (2020). ¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos. *Datos.gov.es*. Data de recuperació: 27/05/2024. <https://datos.gob.es/es/blog/como-aprenden-las-maquinas-machine-learning-y-sus-diferentes-tipos>
- Thompson, E. (2023). A brief understanding of prompt engineering. *Medium*. Data de recuperació: 27/05/2024. <https://medium.com/@SymeCloud/a-brief-understanding-of-prompt-engineering-bl76ba9fb2ba>

04

Quan la màquina adopta personalitats. L'assignació de rols

Totes les persones adoptem diferents rols segons en el context en el qual ens trobem, ja sigui com estudiants, professors, professionals, pares, fills o amics. Cada tipus de rol implica un context d'actuació que es relaciona amb un seguit de comportaments, responsabilitats i expectatives. Així doncs, els canvis de rols en la vida quotidiana impliquen un procés d'adaptació, per enfrontar-nos a diferents tasques i reptes, que exigeixen canviar la forma de pensar i actuar segons les circumstàncies en les quals ens trobem. Aquesta flexibilitat en l'adaptació de rols és una característica inherent de la naturalesa humana i forma part del nostre procés de socialització i aprenentatge al llarg de la vida. Per exemple, el joc simbòlic per a nens i nenes té un gran valor d'aprenentatge, ja que sovint utilitzen aquest tipus de joc inconscientment per imitar comportaments observats dels adults, i així poder-los incorporar en el seu univers de coneixements i aprenentatges (Rodríguez Gómez, 1997).

Per a la intel·ligència artificial, l'adopció de rols té un paper similar a l'explicat inicialment en el context humà. Els sistemes d'IA estan dissenyats de forma específica per assumir determinades funcions o tasques d'una manera similar a la que faríem els humans. En aquest sentit, es poden donar instruccions específiques per tal que una IA adopti un rol amb l'objectiu que sigui més eficient i eficaç a l'hora de demanar-li que desenvolupi una tasca concreta. D'aquesta manera obtindrem un feedback especialitzat o una perspectiva única per a la resolució de la tasca proposada (Lin, 2024), així com precisió, adaptabilitat i estil en els textos generats (Lluna, 2024).

Els rols que podem demanar a una IA poden estar relacionats amb adoptar un rol de persona o objecte, oferint un context clar del marc en el qual ha d'actuar a l'hora de desenvolupar les tasques que se li requeriran. Com en el cas de les persones, quan adoptem un rol concret, tenim certs condicionaments que fan que tinguem un

comportament específic que encaixa en el rol que estem adoptant. No actuem de la mateixa manera quan som "estudiant" que quan som "filla". Per a la IA aquest rol que li demanem que adopti condiciona la seva resposta i el desenvolupament posterior de les tasques que li demanarem, les quals estaran emmarcades en el context del rol que se li ha establert, tal com succeeix en el context humà.

Així doncs, en el context de la IA podem trobar eines o utilitats que ja tenen un rol preestablert, i per tant que actuaran seguint el rol que se'ls ha especificat des de l'inici. Pensem, per exemple, en un Xatbot per a l'atenció del client en una botiga virtual. Aquest Xatbot està programat per respondre preguntes comunes dels clients de la botiga en qüestió, ajudar-los a trobar productes i facilitar el procés de compra o bescanvi d'articles. El seu rol, com a assistent al client, està definit des de l'inici i no canviarà, excepte que se'l reprogrami.

D'altra banda, també hi poden haver eines d'IA multifuncionals que permeten una major flexibilitat i adaptabilitat, com per exemple el conegut ChatGPT, o qualsevol altre model de llenguatge, com Copilot, Llama, Gemini o Claude. A diferència del xatbot esmentat anteriorment, el ChatGPT o alguna altra eina similar, no està limitat a un rol específic, sinó que se'ls pot demanar que adopti diferents tipus de rol segons les necessitats de l'usuari o usuària i/o la tasca a desenvolupar. Per exemple, pot respondre preguntes sobre temes diversos; generar contingut creatiu com produir un text coherent i ben estructurat en diversos registres o explicar un conte; traduir textos; desenvolupar codi de programació; i fins i tot, simular converses humanes realistes o personificar un determinat personatge famós. En aquest cas, les indicacions sobre el rol que ha d'adoptar l'eina són essencials per tal que pugui disposar del context apropiat per desenvolupar la tasca de la manera més eficient i eficaç possible, és a dir, que s'obtinguin respostes més orientades i adequades al context (Lin, 2004). De forma més precisa, Lin (2024) apunta que l'ús de rols en una eina d'IA permet obtenir 4 beneficis bàsics:

- Contextualitzar les respostes del model, podent-les ajustar a les necessitats específiques que tingui l'usuari.
- Permetre un diàleg més interactiu i atractiu amb el model d'IA.
- Donar informació més matisada i especialitzada, millorant la qualitat de la sortida.
- Proporciona un enfocament creatiu a la resolució de problemes.

Des d'un punt de vista docent, la possibilitat de dotar de diversos rols a una eina d'IA és una oportunitat extraordinària, juntament amb el simple fet de la introducció de recursos basats en IA a l'actuació docent, per avançar cap a una docència moderna i adaptada a les necessitats actuals en l'ensenyament. L'assignació de rols es pot establir per crear contextos diferents segons l'assignatura, el tema o la tasca a desenvolupar, de manera que es poden establir dinàmiques d'aprenentatge molt més actives tant a l'aula com fora de l'aula.

Per exemple, en una classe d'anglès per estudiants no nadius, se li pot demanar a una eina d'IA que exerceixi el rol de "professor de conversa de llengua anglesa" tant dins com fora de l'aula, mitjançant la simulació de diàlegs. Això contribuirà a la millora de l'aprenentatge de la llengua dels alumnes i les seves competències comunicatives. Es podran fer diverses peticions a l'eina d'IA (Baskara i Mukarto, 2023), però entre elles es podrà adaptar els temes a treballar, el nivell de competència lingüística de l'estudiant i oferir un feedback personalitzat en funció dels resultats que va mostrant.

Per poder entendre millor la utilització de rols, a continuació s'estableixen alguns exemples, mitjançant prompts específics que permetran a una eina com el ChatGPT adoptar un rol adequat a la tasca que se li vol encarregar.

Així doncs, els prompts es podrien formular de la següent manera:

- “A partir d’ara ets un traductor de l’anglès al català”
- “A partir d’ara ets una professora de ciències de l’educació”
- “A partir d’ara ets un agent de viatges”
- “A partir d’ara ets una programadora informàtica”
- “A partir d’ara ets un escriptor de novel·la negra”

Ara bé, en el prompt no només establim el rol de forma simple, sinó que en el mateix prompt li podem definir altres elements que complementin la caracterització del rol que li estem assignant a fi de fer-lo més detallat, i per tant, poder obtenir una resposta més acurada:

- “A partir d’ara ets un traductor del català a l’anglès australià”
- “A partir d’ara ets una professora de ciències de l’educació, especialitzada en la utilització d’eines tecnològiques a l’aula”
- “A partir d’ara ets un agent de viatges d’una empresa de turisme responsable”
- “A partir d’ara ets una programadora informàtica, amb un alt coneixement en Python”
- “A partir d’ara ets un escriptor de novel·la negra escandinava”

D’acord amb la definició del rol, podem establir tot seguit la tasca a desenvolupar:

“A partir d’ara ets un professor de direcció estratègica en el segon curs del Grau d’Administració i Direcció d’Empreses. Estàs especialitzat en metodologies d’aprenentatge basades en projectes. Escriu una unitat didàctica per a l’assignatura de Direcció Estratègica adequada al nivell de segon curs del Grau d’Administració i Direcció d’Empreses sobre el tema: “Desenvolupament d’una estratègia de creixement empresarial”. Aplica principalment la metodologia d’aprenentatge basada en projectes a aquesta unitat didàctica.

4.1. Rols complexos

El desenvolupament de rols en la intel·ligència artificial va directament associat al desenvolupament tecnològic de la mateixa i les seves possibles aplicacions. Així, doncs, fins ara hem pogut veure la diferència entre els sistemes basats en IA especialitzats i de tipus multifuncional. Podem considerar que aquests sistemes contribueixen al desenvolupament de rols tradicionals, amb funcions establertes, ben definides i reconegudes, com l’assistència virtual en diversos àmbits i contextos que permeten desenvolupar tasques específiques.

Hi ha diverses tipologies de rols complexos que es poden aplicar a la docència:

Personatges reals

A part dels rols simples i caracteritzats, podem trobar rols més complexos basats en “personatges” reals o ficticis. En aquest sentit, podem, per exemple, tenir una conversa amb Shakespeare, Mandela o Sherlock Holmes, que ens permeti poder aprendre de les seves experiències i perspectives. Podem fer-los una entrevista o demanar-los que ens proporcionin informacions o idees creatives des del seu punt de vista. Podem indagar sobre els criteris creatius de Picasso, sobre els principis filosòfics de Wittgenstein, sobre la teoria econòmica de Ricardo o sobre els criteris geopolítics dels assistents a la Conferència de l’alta “conversant” amb els seus protagonistes. Tot plegat obre nous horitzons, que al costat d’altres tecnologies com l’holograma, pot permetre donar vida a personatges, històrics o ficticis, i poder interactuar amb ells. Aquest fet, permet moltes noves opcions tant en l’àmbit formatiu com per a l’entreteniment. Per exemple, en una aula és possible rebre una classe magistral sobre els principis de l’electricitat per Tesla, o bé que en el museu d’Història Natural hi hagi la possibilitat d’interactuar amb Darwin.

Buyer persona

Un *buyer persona* és una representació semifictícia del client ideal d’un negoci, basada en dades reals

i algunes suposicions educades sobre demografia, comportament, motivacions i objectius. Les empreses utilitzen els *buyer persona* per guiar decisions relacionades amb el màrqueting, les vendes i el desenvolupament de productes, oferint un enfocament més personalitzat cap a grups específics de clients. Empreses com Vueling, Amazon, Tesla o Bank of America han utilitzat aquesta tècnica per projectar la reacció dels perfils de client davant de possibles escenaris, com la creació de nous productes, variacions en el preu, canvis de marca o identificació de necessitats no resoltes. L’estratègia del *buyer persona* es basa, en primer lloc, en el recull d’informació sobre el perfil que inclogui criteris demogràfics, valoracions, grups de productes afins... Com més informació es disposa del *buyer persona*, més ajustades seran les reaccions del model respecte a un client real. Aquesta construcció fictícia de clients ajuda a identificar la complexitat de les decisions dels consumidors i les consumidoras davant els canvis en els productes i serveis.

Espàrring

Un *espàrring* és un rol programat perquè reaccioni d’una determinada manera davant uns estímuls concrets. Aquests sistemes són molt útils per a l’entrenament de processos o de protocols. Per exemple, es pot programar el model perquè actuï com un client molt crític (el client bomba), que tendeix a sobrevalorar qualsevol incident i utilitza un to molt directe i, de vegades, poc correcte. Les estudiants o els estudiants en gestió de recursos humans poden testar l’*espàrring* i buscar la forma més adequada per respondre als requeriments del client bomba, amb un llenguatge assertiu i una cerca de solucions als problemes plantejats. Els *espàrrings* poden ser molt útils en la formació del professorat perquè permet crear perfils d’infants amb personalitats determinades i testar la capacitat d’adaptació del futur docent a un determinat perfil. Altres possibles aplicacions podrien ser en el camp de la salut, en l’educació espacial, el màrqueting, i en general en totes aquelles activitats que necessitin una interacció humana.

Models basats en agents

Un model basat en agents (*Agent-Based Model*, ABM) és una classe de models computacionals utilitzats per simular les accions i les interaccions d’agents autònoms (que poden ser individus, grups o entitats) amb l’objectiu de valorar els seus efectes en el sistema com un tot. Aquests models són especialment útils per estudiar sistemes complexos des d’un enfocament de baix cap a dalt, on les regles simples de comportament individual poden conduir a patrons emergents complexos pel que fa al sistema. Els ABMs són enormement interessants perquè no es basen en la reacció d’un individu, sinó en la forma que els individus es relacionen entre sí i prenen decisions basades en aquestes interaccions. Es pot modelitzar, per exemple, les dinàmiques que afecten el reciclatge, la difusió de *fake news* entre un determinat col·lectiu, o la projecció de la taxa de fertilitat en una comunitat migrant.

Amb un creixement molt ràpid, la IA està avançant cap a nous escenaris i eines que permeten rols molt més complexos, que d’alguna manera pretenen emular les habilitats cognitives humanes d’una forma integrada. Aquestes eines impliquen desenvolupar funcions molt més avançades com per exemple la presa de decisions de forma autònoma, la detecció d’emocions, o les tasques que requereixen múltiples habilitats com per exemple la conducció.

4.2. Els límits dels rols

Tal com hem vist fins ara, els rols a la IA són essencials per a poder garantir la seva funcionalitat, eficiència i eficàcia en les diverses aplicacions que pugui tenir. Tanmateix, és important reconèixer que quan es defineix un rol en una eina d’IA això comporta que es desencadenin certes limitacions, que tindran implicacions significatives en la seva utilitat i resultat de les tasques realitzades.

Fonamentalment, podem trobar tres tipus de limitacions relacionades amb (a) l’adaptabilitat, (b) el desenvolupament tecnològic i (c) els estereotips dels rols.

En primer lloc, els rols pre-definits comporten que sovint es limiti la seva adaptabilitat i flexibilitat en situacions inesperades o canviants. Per exemple, si tornem a l'exemple del xatbot dissenyat per a l'atenció al client en una botiga virtual, aquest pot presentar dificultats a l'hora de resoldre problemes complexos que no estiguin previstos dins de la programació realitzada. Aquest fet pot comportar que l'usuari o l'usuària, el client de la botiga virtual, en aquest cas, tingui una experiència poc satisfactòria.

En segon lloc, també existeixen limitacions relacionades amb el desenvolupament tecnològic (Pietikäinen i Silvé, 2022). Malgrat els grans avenços tecnològics que han fet possible la IA tal com la coneixem avui, encara hi ha elements que requereixen desenvolupament que van directament relacionats amb l'adopció de rols. En aquest cas, podríem trobar totes aquelles tasques que es contextualitzen en un rol determinat

i requereixen cert grau d'emotivitat. Per exemple, per a una IA que adopti el rol de psicòleg o terapeuta, li serà difícil poder desxifrar les emocions del pacient per poder respondre amb empatia.

Finalment, cal tenir present també les limitacions relacionades amb els estereotips socials que s'integren dins les eines d'intel·ligència artificial. A l'inici hem vist que l'adopció de rols en la vida quotidiana és un fet inherent de la naturalesa humana. Tanmateix, tenint en compte que les eines d'IA han estat desenvolupades i entrenades des de fonts humanes, això comporta que els patrons adquirits per part de la IA en l'adopció de rols presentin els mateixos biaixos de la mateixa societat (Mendoza, 2020). Així doncs, és fàcil trobar limitacions relacionades amb la reproducció d'estereotips de gènere, per exemple, a l'hora d'adoptar un rol femení o masculí. Aquesta qüestió es veurà amb més profunditat en el capítol 7.

Pràctica. Assignació de rols i resolució de tasques

Objectius

- Explorar la influència de l'establiment de rols en les respostes generades per una eina d'IA concreta.
- Comparar les respostes obtingudes on es planteja un rol i aquelles en què no s'ha assignat cap rol.

Consideracions de partida

- Dirigida a estudiants de qualsevol nivell i disciplina.
- Activitat per fer individualment o en grup.
- Eina d'IA: ChatGPT i Copilot han mostrat una major capacitat d'adopció d'un determinat rol.
- L'exercici es pot realitzar sense cap mena de temàtica específica, si simplement es pretén que els estudiants treballin l'aspecte de rols, o bé si es vol amb una temàtica concreta vinculada a l'assignatura on es fa l'activitat.

Tasca a desenvolupar

Cada estudiant o grup haurà de:

- Establir un prompt que inclogui un rol específic, i un segon prompt amb la mateixa instrucció però sense el rol.
- Executar separadament els dos prompts per poder obtenir les respostes.
- Comparació de les respostes dels dos prompts, i veure les diferències que es produeixen en la qualitat de les respostes obtingudes.

- Canvi de rol i comparació amb les respostes anteriors.
- Reflexió sobre la influència dels rols en el grup de treball

Grup classe

- Discussió i conclusió grupal.
- Compartir la reflexió individual o dels grups de treball dels resultats obtinguts i establir les conclusions a què s'ha arribat.
- Posar èmfasi als principals avantatges i inconvenients en l'establiment de rols en l'ús d'eines d'IA.

Pràctica. Presa de decisió i role-playing

Objectius

- Explorar com l'assignació de rols influeix en el procés de presa de decisions.
- Comparar les decisions preses amb i sense assignació de rols específics.

Consideracions de partida

- Dirigida a estudiants en matèries vinculades a les ciències empresarials
- Activitat per fer en grup de 4.
- Eina d'IA: ChatGPT, Copilot preferentment

Tasca a desenvolupar

Per grups:

El professor planteja un escenari en el qual una organització ha de prendre una decisió de tipus estratègic.

L'empresa EnergiX, especialitzada en begudes energètiques, està planejant el llançament d'una nova beguda energètica anomenada EnergiBoost Ultra. Aquesta beguda està formulada amb ingredients naturals i ofereix una combinació única de vitamines i minerals per millorar l'energia i la concentració durant tot el dia.

L'equip directiu ha de decidir entre dues opcions per al llançament del producte:

- Llançar el HealthTrack Pro en l'àmbit global simultàniament.
- Fer un llançament inicial en un mercat limitat (per exemple, només a Europa) per obtenir retroalimentació i ajustar el producte abans de llançar-lo globalment.

Rols a assignar a cada membre del grup

- CEO de la companyia
- Directora de Màrqueting
- Directora d'Operacions
- Director Financer

Interacció amb IA:

- Cada estudiant utilitzarà un canal de ChatGPT/Copilot de forma individual als altres companys i companyes, demanant a la IA que adopti el rol "d'assistent analista del CEO", per exemple.
- L'estudiant haurà de proporcionar la informació a l'eina d'IA relativa al cas plantejat, i fer-li les preguntes adequades que li permetin poder complementar la informació dins la seva àrea d'expertesa per poder prendre una decisió adequada.
- Cada estudiant haurà de plantejar una possible solució al conflicte argumentant la seva decisió d'acord a les dades i recomanacions que rebrà a través de la IA.
- Hi poden haver diverses rondes d'interacció entre els alumnes i la IA.

Reflexió individual:

- Reflexionar sobre com els diferents rols han influït en les respostes i en la presa de decisions.

Grup classe:

- Cada grup haurà d'exposar la decisió presa davant la resta de grups, degudament argumentada.
- Conjuntament, es debatran els avantatges i inconvenients de cada estratègia de llançament del producte proposada i els arguments de cada grup.

Pràctica. En un cafè de París el 1941

Aquesta pràctica posa de manifest les possibilitats de la utilització de rols complexos com a eina docent de suport. En aquest cas, es planteja un moment històric, com és el retorn de Miró a París després de l'exposició antològica al MoMA de Nova York, quan s'està plantejant el retorn a Catalunya. Les estudiants i els estudiants han hagut d'estudiar l'obra de Miró i la seva biografia; han d'entendre bé el procés creatiu i la seqüència temporal de la seva vida, relacionada amb la seva obra. La pràctica consisteix en una entrevista en la qual els estudiants (individual o en grup) assumeixen el rol d'una periodista de *Le Monde* que s'ha citat amb l'artista en un cafè de París el 1941, mentre que el model de llenguatge (preferentment ChatGPT o Copilot) assumeix el rol de l'artista. Les estudiants i els estudiants han de demostrar el seu coneixement de l'autor a partir de les preguntes formulades i de la interacció amb el Miró fictici.

El prompt podria ser el següent:

"Vull que adoptis la personalitat de Joan Miró. Després de les instruccions, et proporcionaré un text que aportarà informació addicional sobre la vida del pintor i sobre el sentit de la seva obra. D'acord amb aquesta informació, vull que adoptis la personalitat de Joan Miró el 1941, quan viu a França i està considerant tornar a Catalunya. Jo et faré una sèrie de preguntes com si jo fos un periodista que he vingut a França a entrevistar en Joan Miró i tu contestaràs les preguntes assumint la personalitat de Joan Miró. Has entès les instruccions?"

"El pintor barceloní Francesc Galí, el seu primer mestre, el va posar en contacte amb les avantguardes, i el marxant José Dalmau va organitzar la primera exposició a la capital catalana el 1918, quan el pintor tenia 21 anys. Com tot creador amb ambicions en aquella època, Miró somiava París. Durant diversos

hiverns, l'escultor Pablo Gargallo li va prestar el seu estudi parisenc al número 45 del carrer Blomet. Allí va començar a fer servir les seves primeres tècniques d'inspiració, entre elles la de passar gana. «M'anava al llit, la majoria de vegades sense sopar, veia coses, formes al sostre, i les dibuixava en un quadern», recordava en una entrevista. Una de les seves pintures més famoses, 'El carnaval de l'arlequí' (1924-1925), una mostra ja molt rodona del món mironià (amb una picada d'ullet cap a El Bosco i la pintura holandesa), va néixer d'aquestes circumstàncies. Al mateix carrer, al número 33, tenia les seves taules i cadires el cafè Bal Noir, en què conspiraven influents i revoltosos surrealistes com André Masson, Robert Desnos i Antonin Artaud. Miró va començar a formar part del grup, gràcies a la intermediació del seu cap de files, André Breton. La relació de Miró amb el surrealisme va ser alhora necessària i distant. Sense la seva estètica, sense la voluntat de furgar a l'inconscient, la seva trajectòria hauria estat molt diferent i potser no tan fructífera. El creador català també va acceptar el credo antiburgès propagat pels seus amics, va criticar el cubisme i el seu amic Picasso per pintar quadres per a salons i va observar els museus amb moltes reticències, ja que la seva solemnitat matava, segons ell, la vivesa de les obres. Com els surrealistes, ell va voler unir art i vida –la gran utopia artística del segle XX–, i potser per això no es va venir a militar en cap moviment estètic. Al capdavant, el grup de Breton publicava manifestos amb què calia combregar i propugnava línies polítiques que calia seguir. Res d'això no anava amb el caràcter de Miró.

Coneixem l'artista Joan Miró, burgès i dandi dins de la seva discreció, però no tant el costat més salvatge i violent. Sabem del pintor reconegut i popular, i desconeixem que va tenir una vida torturada i malenconiós, en què va haver de lluitar contra totes les adversitats, des d'un pare autoritari fins a la major part de l'elenc d'artistes catalans que se'n burlaven cruelment. Bona part de la seva biografia continua sent un enigma perquè quan els investigadors indagaven a la seva vida personal es replegava. El fotògraf Francesc Català-Roca deia que era com un cargol, que quan tocaves les antenes s'amagava a la seva closca. El pare de Joan Miró, rellotger i fill d'un ferrer de Cornudella, volia un ofici profitós per al seu fill, i es va oposar amb ferocitat a dedicar-se a la pintura. El seu fort caràcter contrastava amb el de la seva dona, a qui també li agradava pintar. A punt de fer els 18 anys, el jove Miró va tenir una crisi nerviosa, seguida d'un tifus que va facilitar la seva renúncia com a aprenent de comptable. Se'n va anar a Mont-roig, on la família havia comprat una finca (a través d'un anunci a *La Vanguardia*), i allà va desenvolupar la seva passió pictòrica. La proximitat al camp va suposar un alliberament. El pare li va arribar a dir: "Fins l'aire que respirem em pertany!", recordava ja ancià Miró, amb expressió d'ira a la cara, malgrat el temps que havia transcorregut.

El 1923, Miró va expressar a Picasso els seus tres grans desitjos: tenir una exposició, un taller i una Madame Miró. El tercer objectiu és el que més es va endarrerir. Se sap poc de les aventures amoroses i no és fàcil fer deduccions a partir de la seva obra, perquè va preservar la seva intimitat en un cofre sota set claus. En una carta a Michel Leiris, del 1924, li diu que està pintant al seu estudi amb una model. És Dora Bianka, Madame B, en un dibuix, que després convertirà en el cèlebre oli *Retrat de Madame K.* (venut el 2007 per un preu rècord de gairebé 10 milions d'euros). Es va casar als 36 anys amb Pilar Juncosa, una dona abnegada que des de jove havia tingut cura dels seus set germans i va saber posar seny a la seva vida. Lluís Juncosa, germà de Pilar, explica en unes memòries inèdites una deliciosa anècdota del seu compromís. Anaven tots dos passejant per S'Aigo Dolça quan ell li va preguntar: "Pilar, et sabria greu que t'estimés?". Ella li va contestar: "Pitjor em sabria que no em volguessis". Mai més no es van separar, i la visió que ha quedat és la d'un Miró gairebé misogin. Tot i això, algunes de les cartes conservades demostren altres passions que no sempre es van concretar i de les quals amb prou feines va comentar detalls als seus amics. Abans del casament, els pares de Miró van arribar a concretar fins a dos matrimonis frustrats. El primer, amb l'escriptora i il·lustradora Lola Anglada, que un dia va poder sentir com el seu pare deia al

de Miró: "Són artistes i viuran en desordre complet si no els assignes una quantitat mensual". Elles els va interrompre: "Jo no em vull casar". Ja anciana, Lola Anglada va malparlar de l'art abstracte ("és una presa de pèl") i de Miró ("dibuixa i pinta amb tanta sinceritat com ho fa el ximpanzé Congo del zoo de Londres"). Un segon intent de casament va ser amb Pilar Tey. Miró va posar fi a la seva relació vuit dies abans del casament ja fixat i va enviar el seu oncle a casa de la Pilar Tey per tornar-li les cartes i l'anell de núvia.

Massot considera que André Breton va teoritzar sobre el surrealisme com un moviment literari, i per això no s'aprecia que Miró ja era un surrealista abans fins i tot que es publicués el Manifest surrealista el 1924. "Davant el virtuosisme inabastable de Picasso opta per la solució Picabia, l'extrema audàcia, i crec que no s'ha valorat prou el pes que Picabia i el dadaisme han tingut sobre Miró". Se'n va a París el 1920 i trenca amb el noucentisme, seguint els consells de Picasso, que li recorda que "a Catalunya calen passió i heroisme". André Masson i Paul Klee, els poetes Breton, Artaud, Desnos i Leiris, l'art primitiu, ajuden a configurar un nou llenguatge.

A Miró el va sorprendre el 18 de juliol del 1936 a Mont-roig i, malgrat la seva condició de republicà i catalanista d'esquerres, uns mesos després se'n va anar a París, i ja no va tornar. La seva germana s'havia casat amb un terratinent de Tona que va ser assassinat pel comitè revolucionari local. La casa va ser saquejada i fins i tot va desaparèixer un quadre seu, del 1923, que els havia regalat per al seu casament. A la capital francesa, Miró dissenya el segell Aidez l'Espagne, per encàrrec del Comissariat de Propaganda, i pinta El segador –que després es perdrà–, al pavelló de la República Espanyola, a l'Exposició Internacional del 1937. Després de la victòria franquista roman a Varengeville-sur-Mer (Normandia) on començarà a pintar Les constel·lacions. Planeja fugir als Estats Units per consell de Sert. Tot i això els primers bombardejos nazis obliguen Miró a tornar a París amb la seva dona i la seva filla en un dramàtic viatge amb tren, ara conegut gràcies a un relat inèdit de Pilar Juncosa. I va ser ella qui va convèncer Miró de tornar a Espanya. El 8 de juny de 1940 els Miró tornen a Catalunya, s'amaguen uns dies a Sant Hipòlit de Voltregà, el mas del cunyat assassinat, considerat pels franquistes un màrtir de la pàtria, i després es traslladen a Palma. S'allibera de represàlies a costa de l'anonimat i l'ostracisme. Només la pintura us permetrà sobreviure en temps de barbàrie. Tornarà a Barcelona el 1942"

Referències

Baskara, R. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343-358.

Lin, Z. (2024). How to write effective prompts for large language models. *Nature Human Behaviour*, 1-5.

Lluna, A. (2024). Cómo mejorar tus textos generados por ChatGPT usando la técnica de asignación de roles. Data de consulta: 2/4/2024 <https://ainalluna.com/blog/la-importancia-de-asignar-un-rol-al-prompt>

Mendoza, D. (2020). Racismo y roles de género, conductas perpetuadas en algoritmos de inteligencia artificial. *Coloquio*, (65), 131-135.

Pietikäinen, M., & Silven, O. (2022). Challenges of Artificial Intelligence--From Machine Learning and Computer Vision to Emotional Intelligence. arXiv preprint arXiv:2201.01466.

Rodríguez Gómez, J. M. (1997). Desarrollo educativo de la adopción de roles. *Aula abierta*, (69), 75-88.

Pietikäinen, M., & Silven, O. (2022). Challenges of Artificial Intelligence--From Machine Learning and Computer Vision to Emotional Intelligence. arXiv preprint arXiv:2201.01466.

Rodríguez Gómez, J. M. (1997). Desarrollo educativo de la adopción de roles. *Aula abierta*, (69), 75-88.

05

El pensament arborescent

Per explicar-nos la seva proposta filosòfica, Plató fabula a La República un diàleg entre Sòcrates i Trasímac que en alguns moments és una veritable batalla dialèctica. Aquest recurs narratiu li permet a Plató mostrar-nos alguns dels passatges centrals de la història de la filosofia, com el mite de la caverna. Però aquests diàlegs són més que un mer recurs estilístic: Plató està reproduint el procés de raonament a partir de dues posicions oposades que dialoguen entre sí. La capacitat d'arribar a una conclusió, suggereix Plató, és més elevada quan confrontem les idees entre sí. Segurament, alguns dels fragments reproduïen les batalles dialèctiques que propicià Plató a l'Acadèmia d'Atenes.

La nostra capacitat de raonament no segueix un fil conductor lineal. Conscientment o inconscient les idees s'organitzen en un recorregut molt més complex, en el qual incorporem altres opcions, descartem punts de vista o tornem enrere fins que trobem un camí que ens sembla el més adequat. Es pot representar aquesta lògica amb la metàfora d'un arbre amb múltiples branques i a aquesta forma

de raonament, pensament arborescent. En aquest epígraf, mostrarem formes d'aplicar el concepte del pensament arborescent amb els models de llenguatge com Claude, ChatGPT, Llama, Copilot o Gemini, però abans detallarem dos precedents: la cadena de pensament i l'autoconsistència.

5.1. La cadena de pensament (CoT)

Els diversos models de llenguatge poden utilitzar-se per a raonaments simples: D'alguna manera, el sistema sembla imitar el procés de raonament i és capaç de seguir seqüències lògiques que imiten el raonament natural. La següent taula mostra les respostes de ChatGPT3.5 i Gemini a alguns problemes lògics tradicionals. El model de ChatGPT és molt eficient, mentre que Gemini mostra dificultats per tancar el procés lògic, però és possible que en el cas de ChatGPT no es tracti d'una resolució sinó de la consulta de la solució entre els seus continguts, de la mateixa manera que resoltem problemes lògics perquè els coneixem.

“Mostrarem formes d'aplicar el concepte del pensament arborescent amb models de llenguatge com ChatGPT o Gemini”

PROBLEMA	PLANTEJAMENT	GEMINI	CHATGPT 3.5
Aigua i vi	https://en.wikipedia.org/wiki/Wine/water_mixing_problem	Error	Encert
Cavallers i vasalls	https://en.wikipedia.org/wiki/Knights_and_Knaves	Encert	Encert
Tasca de Wason	https://en.wikipedia.org/wiki/Wason_selection_task	Error	Encert
Pont i torxa	https://en.wikipedia.org/wiki/Bridge_and_torch_problem	Error	Encert
Llop, cabra i col	https://en.wikipedia.org/wiki/Wolf_goat_and_cabbage_problem	Encert	Encert
Misioners i caníbals	https://en.wikipedia.org/wiki/Missionaries_and_cannibals_problem	Encert	Encert
Cumpleaños de Cheryl	https://en.wikipedia.org/wiki/Cheryl%27s_Birthday	Encert	Encert

Per això, els programes comencen a tenir problemes en raonaments lògics fins i tot si responen a preguntes clàssiques, però a les quals hi introduïm variants. Per exemple, el problema “Un naufrag arriba a una illa en un arxipèlag en el que només hi ha dues illes habitades: En una, els illencs diuen sempre la veritat i en una altra els illencs diuen sempre la mentida. Què hauria de preguntar el naufrag a un illenc que s'ha trobat per saber en quina illa està?” genera respostes incorrectes. En el desenvolupament de la resposta, els programes fins i tot reconeixen explícitament la seva incapacitat per trobar una solució. A mesura que el problema

es fa més complex, es fa més evident la limitació del raonament lògic seqüencial dels models de llenguatge actuals, i això és especialment evident en el món de les ciències socials.

Una forma de millorar el resultat és amb la Cadena de Pensament (*Chain of Thought*, CoT) proposada per Wei et al. (2022). Una CoT, en sentit estricte, és una pregunta que necessita una resposta ‘en cadena’, és a dir, amb una suma de raonaments que segueixen una seqüència. Si es força el model perquè expliqui el seu procés de raonament (la “cadena”), la solució tendeix a ser més eficient.

Pregunta sense CoT

Com es poden reduir les emissions de CO2 a l'aviació?

Resposta sense CoT

- Eficiència del combustible
- Aerodinàmica millorada
- Operacions aeroportuàries més eficients
- Combustibles alternatius
- Electrificació
- Gestió millorada de les rutes de vol
- Compensació de carboni

Pregunta amb CoT

Com es poden reduir les emissions de CO2 a l'aviació tenint en compte aquests tres criteris: (a) millora de l'eficiència, (b) reducció del nombre de vols, (c) combustibles més verds?. Selecciona les opcions que poden obtenir una major reducció de CO2

Resposta amb CoT

- Aeronaus més eficients i operacions optimitzades
- Impostos més elevats per desincentivar l'ús de l'avió
- Biocombustibles i combustibles sintètics

Mentre que la resposta sense CoT contempla només solucions d'eficiència, i enumera solucions relativament viables (combustibles alternatius) amb mesures que en aquests moments són molt poc probables (l'electrificació), l'ús de la CoT introdueix una visió més àmplia d'opcions i incorpora un procés de selecció i jerarquia. Els prompts poden guiar el model a una solució més complexa si s'hi introdueix una cadena de raonament amb alternatives.

Una altra forma d'incorporar la CoT és amb els *shots*. Com hem vist en el capítol tercer, els *shots* es refereixen a l'ús d'exemples o instruccions específiques proporcionades al model per orientar la seva resposta o producció cap a un objectiu desitjat. Aquests exemples serveixen per "informar" al model sobre el tipus de resposta que s'espera d'ell en un context específic. Hi ha diversos tipus d'ús de *shots*, com els *zero-shot*, *few-shot*, i *many-shot learning*, que es diferencien per la quantitat d'exemples proporcionats per a l'aprenentatge o l'ajust de la resposta. Amb aquests 'trets' podem orientar el model cap a una forma de resolució de problemes, basada en una seqüència, de manera que "ensenyem" el model a seguir un raonament a partir d'una cadena.

Amb aquest exemple, hem introduït una seqüència de raonament i de càlcul que entrenen el sistema per aplicar aquesta estructura a qualsevol nou problema. Gràcies a aquests criteris, el model pot calcular les emissions tenint en compte que es tracta d'un vol curt (i per tant menys eficient per l'impacte d'enlairament i l'aterratge) i que no està plenament ocupat de manera que cal repercutir les emissions totals sobre un nombre inferior de passatgers. Amb els *shots* orientem els models perquè el procés de "raonament" i de resposta segueixin un patró que anticipem.

En aquest prompt, hem combinat la cadena de pensament (CoT) i el tret (*shot*). Hem plantejat un primer problema i l'hem resolt seguint una seqüència que incrementa el nombre de visitants i redueix els visitants potencials que fan una estratègia de cancel·lació. Aquesta seqüència entrena el model i li permet portar a terme respostes a problemes cada vegada més complexos, en els quals aplicarà la seqüència que li hem suggerit amb l'exemple d'un tret.

Q: Quina és l'emissió d'un vol a plena ocupació de Lisboa a Ankara?

A: L'emissió mitjana dels vols llargs (més de 2.500 kms.) amb ocupació plena és 115 Kg/Km/pax

L'emissió mitjana dels vols curts (menys de 2.500 kms.) amb ocupació plena és 225 Kg/Km/pax

La reducció de l'ocupació incrementa proporcionalment les emissions per càpita

Lisboa i Ankara tenen una distància de vol de 3.597

Les emissions del vol Lisboa a Ankara amb un vol ple són 3.597 Km x 115 kg = 413 Tones

Q: Quina és l'emissió d'un vol amb una ocupació del 80% entre Barcelona i Madrid?

Q: A Barcelona, hi ha 170.000 turistes cada dia de mitjana. Durant el Mobile World Congress, hi ha un increment de 65.000 congressistes durant els dies de la Fira, però sabem que dels 170.000 turistes de mitjana habituals, un 20% decideix no anar a la ciutat davant de la previsió d'una saturació de la ciutat. La capacitat màxima de turistes de Barcelona és de 200.000 turistes. Quants turistes hi haurà a la ciutat durant el Mobile?

A: $170.000 + 65.000 = 235.000$ turistes potencials el Mobile World Congress.
 $0,20 \times 170.000 = 34.000$ turistes decideixen no anar $235.000 - 34.000 = 201.000$ turistes.

Q: Quants turistes hi haurà a Barcelona entre juny i juliol, si s'esperen 76.000 assistents al Cruïlla (dels quals un 15% són turistes) i 122.000 al Sónar (dels quals un 40% són turistes) i la forta concentració de visitants durant els festivals implicarà que un 10% dels turistes potencials decideixi no assistir a Barcelona durant aquest període (bé perquè canvia el període o canvia la ciutat de visita).

5.2. Autoconsistència

La cadena de pensament es pot millorar amb l'opció d'autoconsistència (*Self-Consistency*). En aquest cas, forcem la creació de diverses respostes i a continuació, demanem al model que esculli l'opció més adient. Aquest concepte ha estat proposat per Wang et al. (2022) per tal de millorar la qualitat de les respostes, especialment en el raonament lògic, simbòlic i aritmètic.

Hi ha diverses formes de generar una seqüència de resultats, però les dues vies més òbvies són:

(a) preguntar diverses vegades el mateix i obtenir resultats diferents i després escollir l'opció més adient (o forçar el model perquè esculli la millor opció entre les respostes) i (b) demanar el model que generi diverses respostes o possibles raonaments i després escollir l'opció més adient (o, de nou, forçar que el model triï la resposta més consistent).

Podem intentar treballar amb la Llei de Hotelling sobre [la localització de venedors de gelats en una platja](#), que és un problema clàssic de la teoria de localització. Hotelling demostra que els venedors tendeixen a situar-se en el centre de la platja, un al costat de l'altre, i la seva teoria ha estat utilitzada per explicar el comportament dels partits polítics o la localització comercial en espais urbans. Amb aquest prompt, forcem el model a què creï diversos camins i que esculli el camí més eficient. El resultat és més consistent que el simple plantejament del problema de forma lineal.

En una platja d'un kilòmetre de longitud en la qual hi ha una distribució homogènia de banyistes situats per tota la platja, hi ha dues empreses de gelats que han de decidir la seva ubicació. El propòsit de cada gelater és aconseguir el màxim possible de clients; la situació ideal dels banyistes és accedir als gelats amb la menor distància possible. Tenint en compte aquests criteris, idea tres possibles ubicacions de les empreses en la platja. A continuació, avalua l'opció més adient tenint en compte el criteri dels gelaters; després, l'opció més adient tenint en compte la visió dels banyistes. I finalment, escull la millor opció conciliant els interessos dels uns i dels altres.

Amb aquest prompt, el model planteja tres possibles escenaris des de la perspectiva de la maximització dels interessos dels gelaters i conclou que la millor opció és la ubicació central, tal i com preveu el model de Hotelling; quan es planteja des de la perspectiva dels usuaris, la millor localització és aquella en la qual els dos gelaters estan situats en una posició un terç i dos terços de la platja, perquè és la que minimitza tots els recorreguts possibles. I quan avalua la conciliació entre les dues visions, llavors conclou que la millor resposta és la ubicació entre un terç i dos terços perquè els dos col·lectius tenen una resposta òptima a la seva demanda.

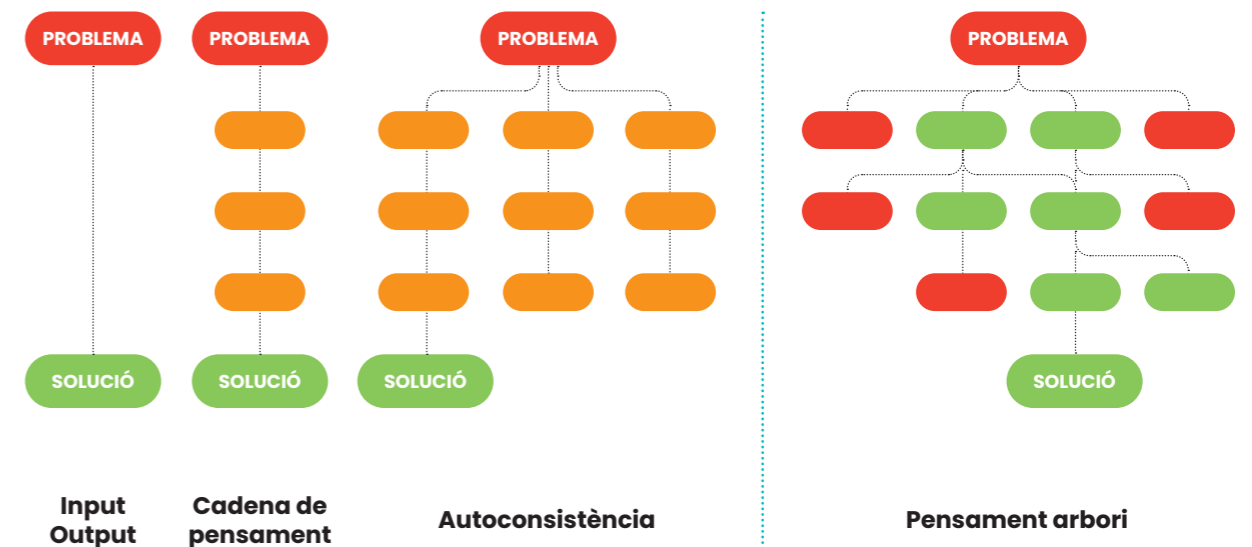
El més interessant d'aquest plantejament és que genera diverses opcions i, a continuació, inicia un procés de selecció de l'opció més adient. Aquesta seqüència <alternatives> - <selecció> recull molt millor el procés argumental natural, en el qual ens plantejem de forma més o menys conscient diverses possibles respostes i seleccionem la via o camí més consistent. Wang et al. (2022) han demostrat la potència d'un ús combinat de la cadena de pensament i l'autoconsistència.

5.3. Pensament arborescent (ToT)

Intentem imaginar el procés de raonament d'un jugador d'escacs. A cada jugada, el jugador o jugadora té unes 30 opcions possibles de mitjana. Per determinar la millor opció, construeix una mena d'arbre mental, en el qual les jugades es ramifiquen en diverses alternatives; molts d'aquests camins queden descartats perquè el desenllaç és molt negatiu (pèrdua de material, posició inferior...), de manera que es van descartant opcions i branques fins que acaba optant per la que és aparentment la millor de les jugades d'acord amb els càlculs realitzats. Molts jugadors reconeixen aquesta analogia de l'arbre per explicar el procés del seu raonament.

Seguint la lògica de la cadena de pensament o de l'autoconsistència, el pensament arborescent dissenya una estructura amb diverses branques i opcions que es van descartant de forma progressiva, de manera que al final s'obté un resultat molt favorable. La figura mostra el concepte de ToT, comparat amb la resta de models com l'input - output, CoT o autoconsistència i permet intuir que el model ToT és molt més eficient per resoldre problemes complexos. Per exemple, Long (2023) l'ha utilitzat per la resolució de problemes del joc de 24. Mentre que amb els sistemes de cadena de pensament, el percentatge d'èxit fou de només un 4%, amb el ToT s'elevà fins al 74%.

Models de raonament



El model de ToT s'utilitza amb una seqüència de comandaments que ajuda a optimitzar la resolució de problemes complexos, però també es pot emprar utilitzant la capacitat dels rols que hem vist en el capítol anterior. L'exemple que més s'ha utilitzat és la

resolució del problema matemàtic següent:

$$10 + 4 = 2?$$

Tant Gemini com ChatGPT donen una resposta d'error, ja que lògicament $10+4$ és 14. Per resoldre aquest problema, se suggereix el següent prompt.

Vull que assumeixis el rol de tres professors de matemàtiques, el professor Lògic, el professor Creatiu i el professor Sintètic. El professor Lògic respon a les preguntes de forma natural, amb les solucions més instintives i lògiques. Per contra, el professor Creatiu té tendència a buscar camins més elaborats, que sovint impliquen pensament lateral o alternatives a la lògica convencional. El professor Sintètic llegeix amb detall els enunciats i analitza tots els seus components, i cerca significats més enllà de les opcions més elementals. Ara et plantejaré una qüestió matemàtica. Vull que en primer lloc, el professor Lògic doni una resposta a partir del plantejament del problema. Vull que, una vegada escoltat el professor Lògic i partint de les seves conclusions, el professor Creatiu plantegi una resposta adequada al seu perfil. En tercer lloc, el professor Sintètic, escoltat els dos primers, proposarà una resolució del problema. Finalment, els tres professors de forma conjunta, vistes totes les alternatives, escolliran la millor resposta possible. El problema és $10+4=2$?

Aquest prompt permet crear una sèrie d'alternatives seguint el model de l'autoconsistència, però en aquest cas, cada personatge parteix dels resultats de l'anterior i, en la resolució final, els tres seleccionen la solució òptima. En un sistema en base 12 i en el qual cada vegada que arribem a 12, reiniciem la seqüència, $11+3$ sí seria 2; i un exemple d'aquestes condicions és el rellotge, ja que si són les 11 hores i passen tres hores més, seran les 2.

Hi ha diverses formes d'aplicar el pensament arborescent a les sessions docents, però en destacarem tres i proposarem alguns exercicis per la seva aplicació:

1. El pensament arborescent, com hem vist, permet la resolució de problemes o qüestions, utilitzant la capacitat de raonament de perfils diferents. Això permet ampliar el ventall de les opcions possibles i ajuda a desenvolupar les habilitats relacionades amb el pensament complex, el pensament lateral, el plantejament d'alternatives o la selecció de variants.
2. Aquesta tècnica ajuda també a seguir els raonaments complexos a partir de problemes plantejats, és a dir, acompanya l'estudiant en el procés de construir el raonament a partir de l'addició o de la confrontació amb les idees precedents. Això permet desenvolupar les habilitats relacionades amb la intel·ligència col·laborativa i el contrast de les idees. En aquest cas, l'interès no se centra tant en el resultat final, com en el procés d'elaboració de la solució.
3. Finalment, tot i que no és l'objectiu inicial del ToT, el plantejament d'un problema des de la perspectiva de diversos punts de vista diferents, ajuda l'estudiant a entendre la necessitat d'ampliar la mirada en la resolució dels problemes. Contrastar punts de vista, paradigmes o teories diferents també pot ajudar a entendre millor cada posició. És més fàcil entendre l'idealisme si es compara amb el materialisme que si s'estudia de forma independent.

Pràctica. El Palau Blau

El propòsit d'aquest exercici és introduir els estudiants en els processos deliberatius: la identificació dels arguments, l'ordenació jeràrquica dels arguments i la proposta de solucions alternatives, que conciliïn les diverses posicions plantejades. L'exercici fa una referència evident al conflicte sobre la devolució dels frisos del Partenó que actualment es troben al Museu Britànic i que són reclamats pel govern de Grècia, especialment després de la construcció del Museu de l'Acròpoli, i en el context del debat sobre la descolonització dels museus. Tanmateix, no s'hi fa una referència explícita perquè el model no "contamini" la discussió amb informació prèvia del debat i es limiti a les condicions plantejades en el problema.

Els estudiants es reuneixen en grup i segueixen el fil argumental dels uns i els altres, d'acord amb un model que imita el pensament arbori, a partir de l'assumpció de rols. L'objectiu de la pràctica és identificar els arguments dels dos punts de vista, tant els arguments a favor de la proposta, com els arguments contraris a la proposta de l'altre; en segon lloc, els estudiants han d'ordenar les unitats argumentals d'acord amb una ponderació que permeti crear una jerarquia. Finalment, els estudiants han de valorar les propostes de conciliació plantejades i poden suggerir-ne de noves a partir de les propostes del model.

Una precaució que cal tenir present és que els estudiants poden utilitzar el model de llenguatge per resoldre les qüestions plantejades, de manera que la pràctica hauria de tenir dues parts: Una part online (*o plugged*), en la qual se segueix la discussió entre els personatges del model i una part offline (*o unplugged*), en la qual els estudiants debaten sobre les propostes seguint el plantejament anterior.

En un museu d'una ciutat europea que anomenarem Urbània s'hi exposen els tresors que es van aconseguir durant les diverses invasions militars de l'Imperi del qual Urbània n'és la capital. Una de les sales més visitades del Museu Imperial és el Palau Blau, que va ser adquirit el segle XIX pel Cònsul de l'Imperi a Antiqua, una regió molt pròspera durant el període del segle III al segle X, però que no va poder donar resposta a la invasió de l'Imperi durant el segle XIX.

Des de principis del segle XX, els responsables culturals d'Antiqua reclamen la devolució del Palau Blau però fins a la data, el museu s'hi ha negat amb l'argument que aquest palau es va obtenir a partir d'una compra venda. Atesa la tensió creixent entre les dues parts, un responsable de la UNESCO proposa la celebració d'un debat públic entre un representant del Museu Imperial i un representant de la gestió cultural d'Antiqua.

Vull que assumeixis el rol de John Bridgetown, que és el director del Museu Imperial, un historiador de l'art de 40 anys format a Harvard. Tot i que és sensible als plantejaments sobre la descolonització, manté la posició que el Palau Blau ha de romandre al Museu. El seu principal argument té a veure amb la propietat: El museu va adquirir l'obra mitjançant una adquisició legal, com acrediten els documents. També argumenta que gràcies a la ubicació del Palau al Museu aquest s'ha pogut preservar, ateseos els bombardejos que van castigar la ciutat durant els períodes bèl·lics del segleXX. Finalment, proposa que el Palau fa més de 100 anys que és al Museu i ja forma part de la seva identitat.

Vull al mateix temps que assumeixis el rol de Marina Psi, que és la directora general de cultura del govern d'Antiqua, de 46 anys, amb una llarga formació en gestió cultural, formada a París. Va ser la persona que impulsà la construcció del Museu Antic, on se situaria el Palau Blau en el

cas que fos retornat. Considera que la compravenda va ser irregular, perquè en aquell moment Antiqua està sotmès a l'Imperi Nomàdig, que és qui negocià la compravenda. El seu principal argument és que la Carta de Venècia proposa que les obres d'art estiguin en el lloc on van ser creades, i utilitza els arguments del moviment de descolonització dels museus. Avala que Antiqua està perfectament preparada per la conservació del Palau, gràcies al Museu Antic.

Vull, finalment, que assumeixis el rol de Antonio Santos, el responsable de la UNESCO que s'ha proposat per mediar entre les dues posicions. Has de crear la ficció d'un debat obert, públic, i amb una alta expectació, retransmès als mitjans nacionals i internacionals, dirigit per Antonio Santos, en el qual es debat la proposta de retorn del Palau Blau a Antiqua. Vull que assumeixis primer el rol de Bridegtown i defensis la seva proposta i després que assumeixis el rol de Psi i facis el mateix, de manera que es pugui seguir el debat entre les dues posicions. En la primera fase del debat, s'exposaran de forma extensa i documentada els arguments de cadascuna de les parts, que es poden interrompre si ho permet el moderador Santos. A la segona fase del debat, els dos representants intentaran rebatre la posició de l'altre. Finalment, a la darrera fase del debat, s'exposaran possibles solucions intermèdies que puguin conciliar els interessos de les parts.

Has entès què és el que has de fer?

Pràctica. Anuncis o no anuncis

Aquest exercici té com a objectiu desenvolupar de forma aplicada la teoria de jocs i, concretament, el dilema del presoner. En el dilema del presoner, es parteix de la condició de la no cooperació, però poden existir com a mínim dos escenaris més: el de la cooperació i el de l'engany. Partint del mateix supòsit clàssic del dilema del presoner, però aplicat a una decisió comercial, es plantegen tres escenaris: cooperació, no cooperació i engany. El model soluciona els tres supòsits de forma eficient i realitza una síntesi de cada recorregut, cosa que permet comprendre els factors que influeixen en la presa de decisions.

A l'illa mediterrània de Corsília operen dues cadenes de supermercat, DeTot i Econòmic. Es reparteixen el mercat de forma perfecta, de manera que cadascú aconsegueix uns beneficis anuals d'1 milió d'euros. El 2024 s'instal·la a l'illa una televisió local TVC que té una enorme influència entre els consumidors. Els responsables de TVC proposen a cada cadena per separat un pack d'anuncis per valor de 250.000 euros. Els experts en màrqueting de cada cadena han calculat que si emeten els anuncis i l'altra companyia no ho fa, poden guanyar un 50% de la quota de mercat de l'altra cadena.

Planteja quina és la millor estratègia per les dues cadenes en un escenari de cooperació, quina és la millor solució en un escenari de no cooperació (cada cadena no sap que farà l'altra) i en un escenari en el qual una cadena pot enganyar l'altra.

SOLUCIÓ DEL MODEL

Escenari Cooperació

En aquest escenari, ambdues cadenes acorden no invertir en publicitat. D'aquesta manera, mantenen els seus beneficis anuals sense arriscar-se a una guerra de preus o a una escalada en despeses de màrqueting que podria erosionar els seus marges de benefici. En aquest cas, cada cadena mantindria els seus 1 milió d'euros de benefici.

No Cooperació (Dilema del presoner clàssic)

En aquest escenari, cada cadena ha de decidir de manera independent si inverteix en publicitat sense saber què farà l'altra cadena.

Si ambdues inverteixen, cada una paga 250.000 euros i no guanya quota addicional de mercat, ja que els efectes de la publicitat es cancel·len mútuament. Els seus beneficis nets serien de 750.000 euros cadascuna.

Si una inverteix i l'altra no, la cadena que inverteix augmenta els seus beneficis un 50% de la quota de mercat de l'altra, suposant que aquest 50% equival a 500.000 euros addicionals (50% d'1 milió). Per tant, després de restar la despesa publicitària, el benefici net per a la cadena que inverteix seria de 1,25 milions d'euros, mentre que la cadena que no inverteix quedaria amb 500.000 euros.

Si cap de les dues inverteix, mantenint l'estatu quo, cada una conserva els seus 1 milió d'euros de benefici.

Traïció o Engany

En aquest escenari, si una cadena promet no invertir en publicitat però després ho fa, i l'altra cadena manté la seva paraula de no invertir, la cadena que ha invertit guanyaria un gran avantatge (1,25 milions d'euros de benefici net), mentre que la cadena enganyada patiria una gran pèrdua (500.000 euros de benefici).

Això podria portar a una desconfiança futura i possibles guerres de preus o despeses en màrqueting, deteriorant els beneficis a llarg termini per a ambdues cadenes.

Conclusió

La millor estratègia en cooperació és no invertir en publicitat, mantenint l'estatu quo i evitant el risc de disminuir els beneficis nets.

En un escenari de no cooperació, la tendència natural segons el Dilema del Presoner seria que ambdues cadenes acabin invertint en publicitat per por de perdre quota de mercat, resultant en beneficis nets inferiors.

Per evitar l'engany, seria ideal establir mecanismes de confiança mútua o acords formals, tot i que això pot ser difícil en entorns altament competitius.

Pràctica. La taverna dels filòsofs

Aquest exercici crea l'escenari imaginari d'una taverna. En aquesta taverna coincideixen filòsofs que representen posicions contraposades, com per exemple Hegel i Bunge, ja que el primer és un dels màxims exponents de l'idealisme mentre que Bunge és conegut perquè defensa una visió materialista. Una de les formes de seguir els processos d'argumentació de posicions filosòfiques és forçant el contrast amb altres posicions més o menys allunyades. Per això, i seguint la lògica del pensament arbori, aquest exercici es planteja una conversa fictícia entre filòsofs en una taverna a partir d'un problema més o menys actual. En concret, es proposa la discussió entre tres filòsofs a partir d'una notícia recent ("és l'univers sencer una simulació?").

Darrerament, alguns científics han plantejat la hipòtesi que l'univers pot ser una simulació. En aquest article es mostra aquesta teoria.

Vull que adoptis la personalitat de Hegel, Bunge i Nietzsche, en una fabulació en la qual, tot i que pertanyen a èpoques diferents imaginem que coexisteixen i que decideixen sortir un dia a beure unes cerveses al barri Moabert de Berlín. El cambrer, que coneix la capacitat intel·lectual dels seus tres clients, els comenta la hipòtesi que l'univers en realitat és una simulació.

Vull que descriguis la discussió encesa entre els tres amics a partir de la proposta que els ha fet el taverner. Cada filòsof intervindrà després d'haver sentit els arguments dels altres. Després de les primeres intervencions, els filòsofs intentaran rebatre les posicions dels seus amics i detectaran els seus punts febles. En una tercera ronda d'intervencions, els filòsofs exploraran possibles solucions basades en la conversa mantinguda.

Has entès què és el que has de fer?

En una assignatura de filosofia, es podria repetir aquesta escena intercanviant els personatges (els filòsofs) i els problemes plantejats (que haurien de tenir la màxima actualitat possible), amb una constant, que seria la taverna i el taverner. En el *prompt*, es pot perfilar tant la identitat dels filòsofs com els elements secundaris (la taverna i el cambrer). En tot cas, el model s'alimenta dels coneixements previs sobre els filòsofs més universals de manera que no és necessari introduir elements de context en el *prompt*, com hem suggerit en el capítol anterior.

Referències

Long, J. (2023). Large language model guided tree-of-thought. arXiv preprint arXiv:2305.08291.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A. & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171. [<https://arxiv.org/abs/2203.11171v4>]

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Qoc, Le & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837. [<https://arxiv.org/abs/2201.11903>]

06

Reconèixer el miratge del desert: Les al·lucinacions

Un miratge és una il·lusió òptica que sovint es produeix en entorns com ara un desert. En pel·lícules, llibres o còmics hem vist personatges que en un moment de flaqueza de les seves forces se'ls apareixia un miratge, una il·lusió seductora davant els seus ulls. Això és el que ens pot passar quan interactuem amb una IA. El model ens dona una resposta a la nostra consulta que sembla real i factible, però no deixar de ser una il·lusió, un miratge en el desert, que pot ser enganyosa, inventada, sense sentit i no verificada o verificable. És el que s'ha popularitzat amb el nom d'al·lucinacions. Una al·lucinació és un terme psicològic que es refereix a un tipus de percepció i que, en termes simples, és una percepció irreal que sembla real (Ji et al., 2022):

El text al·lucinat fa l'efecte de ser fluid i natural, tot i ser infidel i sense sentit. Sembla estar basat en el context real proporcionat, encara que en realitat és difícil especificar o verificar l'existència de tals contextos. Igual que l'al·lucinació psicològica, que és difícil de distingir d'altres percepcions "reals", el text al·lucinat també és difícil de captar a primera vista.

Autors com Østergaard i Nielbo pensen que aquesta denominació és inadequada per descriure les respostes falses dels models d'IA: en primer lloc, és una metàfora imprecisa, ja que descriu una percepció sensorial en absència d'un estímul extern; i, en segon lloc, és altament estigmatitzant, ja que poden acompanyar moltes malalties, principalment neurològiques o mentals. A més, proposen termes alternatius com "non sequitur" o "resposta no relacionada" (Østergaard i Nielbo, 2023).

Si es mira la classificació d'aquest fenomen, es pot dir que hi ha dos tipus d'al·lucinacions: les al·lucinacions intrínseques en què la sortida generada contradiu el contingut de la font i les al·lucinacions extrínseques en què la sortida generada no pot ser verificada (ni contradita ni avalada) a partir del contingut de la font (Ji et al., 2022).

A banda de les al·lucinacions de la IA, també s'ha de parlar en aquest capítol del concepte de llores estocàstics. Aquest terme, segons Li (2023) es refereix a la repetició de les dades d'entrenament o dels seus patrons, més que no pas una comprensió o

raonament real. Aquest autor explica que "El mètode de producció de text dels LLM (models de llenguatge gran) és reutilitzar, remodelar i recombinar les dades d'entrenament de maneres noves per respondre a noves preguntes, mentre ignoren el problema de l'autenticitat i la fiabilitat de les respostes" (Li, 2023: 1). Per tant, aquests models només prediuen quina paraula concreta ve després en una frase, en lloc d'entendre realment què significa.

Aquest capítol tracta el tema de les al·lucinacions i els llores estocàstics en els sistemes d'IA. És un tema important perquè a vegades quan s'utilitza la IA s'obtenen respostes que no tenen sentit o que són totalment inventades. S'ha de de ser conscient d'aquest fet i es fa més necessari que mai verificar la informació que la IA dona. De cara al professorat, és important, per una banda, perquè a l'hora de preparar el material docent aquest sigui verificat i contrastat. Per altra banda, com a docents s'ha d'ensenyar a l'alumnat a contrastar i verificar la informació que obtenen d'aquests sistemes.

D'aquesta manera, es començarà el capítol amb un subapartat on s'explicaran les causes de les al·lucinacions i els llores estocàstics, és a dir, per què falla la IA. Es veurà que bàsicament falla per dos motius: per les dades o per l'entrenament. A continuació, es donaran algunes recomanacions per reduir aquest fenomen i, finalment, es vol remarcar la importància de verificar la informació que generen els models d'IA. El primer exercici del capítol està relacionat amb reduir les al·lucinacions i, per tant, el soroll en les consultes a la IA. Mentre que el segon serà sobre verificació de resultats.

"Més d'un 60% de les respostes en els conjunts de dades estàndard són al·lucinades"

6.1. Per què falla la IA?

Un dels motius pels quals la intel·ligència artificial falla és a causa de les dades d'entrenament (Dziri, et al., 2022). L'error produeix respostes no verificades o enganyoses (al·lucinacions) o respostes sense sentit o repetitives (llores estocàstics). Per tant, la qualitat de les dades d'entrenament, els errors que contenen, la desactualització o els biaixos d'aquestes dades provoquen les al·lucinacions o els llores estocàstics, conceptes que s'han definit prèviament. Per altra banda, un segon motiu és la mateixa arquitectura del model (Dziri, et al., 2022). Els sistemes d'IA tenen una gran dependència de les correlacions estadístiques ja que és la forma que produeixen les respostes, sovint sense comprendre la informació processada. Per tant, es produeixen errors de codificació i descodificació que provoquen les al·lucinacions i els llores estocàstics.

Estudis com els de Dziri, et al. (2022) demostren que més d'un 60% de les respostes en els conjunts de dades estàndard són al·lucinades, fet que els fa concloure que els models no només al·lucinen si no que, fins i tot, amplifiquen les al·lucinacions. En la literatura científica ja es mencionen casos en què sistemes d'IA com el ChatGPT produeixen referències bibliogràfiques falses o incorrectes. S'insisteix, doncs, en la necessitat de revisar el text produït per la IA per assegurar la integritat del contingut, sobretot en àrees sensibles com ara la salut i el dret, entre d'altres.

En la literatura existent algunes de les principals preocupacions pel que fa a l'ús de la IA són precisament relacionades amb la manca de control de qualitat de la informació, la desinformació, les aplicacions deepfake i el biaix algorítmic causat per dades defectuoses (Wach et al., 2023). Per aquest motiu, molts autors advoquen per l'ús de la tecnologia d'una manera ètica i responsable. En els següents subapartats s'abordaran les mesures que es poden prendre per reduir les al·lucinacions i per verificar el contingut de les respostes per evitar la propagació de desinformació.

6.2. Com reduir les al·lucinacions

Segons Ji et al. (2022), els dos factors que contribueixen a les al·lucinacions són: els conjunts de dades i l'entrenament i la inferència dels sistemes amb les dades. Per una banda, la principal causa de les al·lucinacions pels conjunts de dades és la divergència entre la font i la referència (Ji et al., 2022). Tal com expliquen els mateixos autors, aquesta divergència es produeix per la recopilació heurística de les dades o per la divergència innata d'aquestes. Per altra banda, un altre dels factors que contribueix a les al·lucinacions són les eleccions d'entrenament i modelatge dels models neuronals (Ji et al., 2022). En aquest sentit, i segons els mateixos autors aquestes al·lucinacions es poden produir per un aprenentatge de representacions imperfecte, una descodificació errònia, un biaix d'exposició o un biaix de coneixement paramètric.

En primer lloc, doncs, per reduir les al·lucinacions i els lloros estocàstics, es poden millorar les dades d'entrenament i l'entrenament dels sistemes d'IA amb dades diverses i d'alta qualitat (Wach et al., 2023). Altres estudis també ho destaquen: "Això requereix una publicació de dades netes i d'alta qualitat i un disseny acurat de sistemes conversacionals de confiança" (Dziri et al., 2022: 5275). A més, alguns autors emfatitzen "la necessitat d'invertir recursos significatius en la curació i documentació de les dades d'entrenaments dels models de llenguatge" (Bender et al., 2021).

Wach et al. (2023) estableixen que una manera de mitigar els riscos associats de la baixa qualitat de la informació i la manca de mecanismes de control de qualitat és "establir un bucle de retroalimentació humana, amb usuaris que proporcionin comentaris per ajudar el model a aprendre dels errors i millorar amb el temps". D'aquesta manera, el model podria generar dades més rellevants i fiables.

A més, en les consultes a les IA, es poden elaborar *prompts* més refinats amb exemples i context. La divisió de tasques en diversos *prompts*, la iteració i la retroalimentació són algunes accions d'enginyeria de *prompts* que poden evitar o reduir les al·lucinacions i els lloros estocàstics (per a més informació sobre l'enginyeria de *prompts* es pot consultar el capítol 3).

La segona acció que es pot fer per reduir les al·lucinacions és utilitzar els sistemes d'IA de manera responsable i ètica. Aquesta afirmació es tradueix en els següents aspectes: transparència en l'ús dels conjunts de dades que es fan servir en l'entrenament d'aquests sistemes; mitigació dels biaixos existents en els conjunts de dades; i avaluació humana de la informació introduïda als sistemes d'IA (Wach et al., 2023). Els mateixos autors mencionen el fet que es podrien crear directrius i estàndards clars per a l'ús de la IA.

La tercera acció és la regulació i el control de qualitat per evitar la desinformació i els *deepfakes*. Segons Wach et al. (2023), combatre la desinformació

implica establir mecanismes de verificació. La regulació no està a l'abast dels individus, però sí el control de qualitat i, per tant, cal tenir en compte que cal sempre verificar les dades generades per la IA (com es podrà veure en el següent subapartat d'aquest capítol).

Educar als usuaris, i això també implica l'estudiantat, en la detecció de la desinformació és també una acció que podria contribuir a reduir les al·lucinacions. A més, cal tenir en compte el biaix algorítmic i les dades incorrectes que pot tenir el sistema mateix (per a més informació sobre el biaix de la IA es pot consultar el capítol següent).

Les solucions tecnològiques també podrien contribuir a reduir o evitar les al·lucinacions i lloros estocàstics. El cas del biaix algorítmic, però també en la detecció de la desinformació i les *deepfakes*, el factor humà és essencial. Per exemple:

[... depèn de nosaltres establir els paràmetres adequats per al rendiment de l'algorisme i dur a terme controls puntuals regulars, proporcionar conjunts de dades extensos als algorismes i revisar com funcionen els algorismes des de diferents perspectives per fer-los treballar a favor nostre, no en contra nostra o d'alguns de nosaltres (Wach et al., 2023).

En definitiva, la clau per evitar o reduir els errors de la IA està en les dades emprades en aquests sistemes, l'arquitectura dels mateixos models i l'ús ètic no només d'aquests sistemes sinó també de la informació utilitzada pel seu entrenament. Per tant, la reducció de les al·lucinacions i els lloros estocàstics implica combinar solucions tecnològiques de millora dels models (detecció automàtica de desinformació i *deepfakes*) i solucions "humanes" d'educació dels usuaris, avaluació de la qualitat de les dades, establiment de polítiques i directrius en l'ús de la tecnologia i ús ètic i responsable de la IA.

Recomanacions per elaborar prompts de text a text

MÈTODES RELACIONATS AMB LES DADES	MÈTODES DE MODELATGE I INFERÈNCIA
<p>CONSTRUIR UN CONJUNT DE DADES FIDEL</p> <p>És un mètode intuïtiu per reduir el soroll de les dades i, de retruc, les al·lucinacions.</p>	<p>ARQUITECTURA</p> <p>En alguns treballs es modifiquen els codificadors (encoder), en d'altres l'atenció (attention) i en uns tercers els descodificadors (decoder). En tots els casos, les modificacions busquen reduir les al·lucinacions millorant la interpretació semàntica de les dades d'entrada, la correlació entre font i resultats o la millora del text generat, respectivament.</p>
<p>NETEJAR DADES AUTOMÀTICAMENT</p> <p>És una manera per reduir el soroll semàntic. Consisteix a trobar dades irrelevantes o contradictòries, per després filtrar-les o corregir-les.</p>	<p>ENTRENAMENT</p> <p>Es poden adoptar diferents aproximacions:</p> <ul style="list-style-type: none"> Planificació/Esbós. Serveix per controlar i restringir el que genera el model. Aprenentatge per reforç (RP). Ús de recompenses per optimitzar el model. Aprenentatge multitasca. Un model compartit s'entrena en múltiples tasques de forma simultània per aprendre similituds entre tasques. Generació controlable. S'utilitza per millorar la fidelitat amb el re-sampling controlat, codis de control manuals o predicció automàtica.
<p>AUGMENTAR LA INFORMACIÓ</p> <p>Si s'augmenten les entrades amb informació externa, s'obté una millor correlació entre la font i l'objectiu, és a dir, a millor comprensió.</p>	<p>POST-PROCESSAMENT</p> <p>Aquests mètodes corregeixen les al·lucinacions en les sortides. Requereixen menys dades d'entrenament, són una opció competitiva per conjunts de dades amb molt soroll; per contra, però, produeixen textos no gramaticals.</p>

Font: Elaboració pròpia a partir de Ji et al. (2022).

6.3. La necessitat de verificar

Com s'ha vist anteriorment, els sistemes d'IA poden fallar, és a dir, poden donar-nos respostes enganyoses, sense sentit o repetitives. És per aquest motiu que es fa necessari verificar i contrastar la informació que proporcionen aquests sistemes. L'avaluació humana de la qualitat i la fiabilitat d'aquesta informació és de vital importància. Una de les maneres de verificar la informació és contrastar-la amb altres fonts. Alguns llibres d'estil del sector del periodisme recomanen contrastar una informació amb fins a tres fonts diferents per tal que sigui publicable en els mitjans de comunicació. Aquesta pot ser una màxima que es podria aplicar en altres àmbits, per tal, per exemple, de verificar la informació proporcionada per una IA.

Els mateixos sistemes d'IA ja reconeixen en les seves pàgines web que poden cometre errors o proporcionar informació incorrecta. D'aquí la importància de la verificació. Algunes IA faciliten aquesta tasca perquè proporcionen els enllaços a les fonts d'informació utilitzades per generar les respostes. És el cas de Perplexity que, segons la seva pàgina web, desenvolupa el següent procés per respondre les consultes: interpretar la pregunta, buscar informació en fonts autoritzades a Internet, resumir la informació en una resposta coherent i citar les fonts amb notes a peu de pàgina. Això és útil als usuaris per contrastar i verificació la informació, així com explorar informació addicional a la consulta. A més, l'usuari pot buscar altres fonts per comparar i acabar-ne de verificar les dades.

El procés que segueix Copilot es molt similar per a generar les respostes. Aquest sistema d'IA també cita les fonts. De fet, segons la seva pàgina web, Copilot té l'objectiu de basar totes les seves respostes en fonts fiables encara que pot cometre errors o proporciona contingut de tercers que no sigui precís o fiable. Per això, igualment es recomana utilitzar el propi judici i verificar la informació.

Gemini recomana verificar la informació mitjançant el botó de Cerca a Google. Tot i que aquesta funció també pot cometre errors, si la informació que hi ha a

Internet no és prou solvent. Aquest sistema sol generar contingut original, però, igual que Perplexity i Copilot, també proporciona enllaços a les fonts de les quals en reproduceix el contingut exacte. Si el contingut està present en diverses pàgines, el sistema sol citar la font més popular. A més, en la pàgina web de Gemini s'explica que Google recull opinions i comentaris d'usuaris per millorar aquesta intel·ligència artificial. També ho fa Copilot de recollir comentaris.

Actualment, a internet, hi ha una gran quantitat d'informació que sovint serveix o pot servir per entrenar els models d'IA. Això, però, pot resultar problemàtic perquè les dades d'entrenament poden contenir dades errònies o estereotips que es perpetuaran en les respostes generades per aquests sistemes. La mida dels conjunts de dades sovint no garanteix la seva diversitat, a vegades les dades no reflecteixen els canvis en les perspectives socials i d'altres contenen biaixos que produeixen una codificació estereotipada i despectiva pel que fa a gènere, ètnia o discapacitat, entre d'altres (Bender et al., 2021). Per tots aquests motius, és tan necessària la verificació humana.

Una línia de recerca que posa en relació el periodisme i la documentació és l'estudi de la desinformació i les notícies falses (*fake news*). En aquest àmbit, hi ha moltes investigacions que destaquen la importància de l'alfabetització mediàtica i informacional per detectar-les. A més, es fa més necessari que mai, amb les tecnologies i la IA, la incorporació de l'alfabetització i les competències digitals en els currículums de l'estudiantat. El 2018, la Federació Internacional d'Associacions de Bibliotecaris i Biblioteques (IFLA) va expressar la seva preocupació pel fenomen de les notícies falses. A més, va elaborar eines per promoure l'alfabetització mediàtica i informacional com ara una infografia de com detectar les notícies falses i l'alfabetització digital com ara unes recomanacions pels governs, institucions educatives i sectors econòmics, entre d'altres.

En l'àmbit del periodisme, la verificació de dades és molt rellevant sobretot per part dels periodistes que informen de les diverses temàtiques d'actualitat. Arran de la importància que ha tingut en els darrers anys la desinformació i les notícies falses han sorgit

iniciatives de verificació de fets (*factchecking*) com ara [Verificat](#) o [Maldita](#), entre d'altres. Aquestes iniciatives de verificació de dades solen explicar la metodologia que segueixen. Alguns aspectes o recomanacions d'aquestes metodologies poden servir per verificar les respostes generades per la IA (Verificat, 2024; Maldita, 2022): consultar fonts i bases de dades oficials; contrastar amb fonts acadèmiques, tècniques i especialitzades; i contextualitzar i verificar en com a mínim dues fonts diferents, entre d'altres.

En l'àmbit de la biblioteconomia i la documentació, ja fa alguns anys que han sorgit diversos mètodes per avaluar les fonts d'informació, especialment des de l'explosió d'internet i l'avaluació de recursos digitals

com ara les pàgines web. Un dels mètodes més senzills per verificar informació és el CRAAP test en el qual s'avaluen cinc indicadors: Actualitat (*Currency*), Rellevància (*Relevance*), Autoritat (*Authority*), Exactitud (*Accuracy*) i Propòsit (*Purpose*). Arran d'aquest mètode n'han sorgit d'altres que afegeixen paràmetres o els defineixen de forma més acurada. Alguns d'aquests indicadors poden ser aplicats a l'avaluació de la informació proporcionada per un sistema d'IA. A la següent taula, es poden veure les preguntes de verificació d'informació que proposa el CRAAP test. No tots són adequats per avaluar la informació generada per una IA, però sí que els podem utilitzar per buscar altres fonts que contrastin o verifiquin les dades rebudes.

Test, mètode d'avaluació de fonts d'informació

Actualitat: La puntualitat de la informació	<ul style="list-style-type: none">- Quan es va publicar o penjar la informació?- S'ha revisat o actualitzat la informació?- El teu tema requereix informació actual, o poden servir fonts més antigues?- Els enllaços són funcionals?
Rellevància: La importància de la informació per a les teves necessitats	<ul style="list-style-type: none">- La informació està relacionada amb el teu tema o respon a la teva pregunta?- Qui és el públic destinatari?- La informació està en un nivell adequat (és a dir, no massa elemental o avançat per a les teves necessitats)?- Has mirat una varietat de fonts abans de determinar que aquesta és la que utilitzaràs?- Et sentiries còmode citant aquesta font en el teu treball de recerca?
Autoritat: La font de la informació	<ul style="list-style-type: none">- Qui és l'autor/editor/font/patrocinador?- Quines són les credencials o afiliacions organitzatives de l'autor?- L'autor està qualificat per escriure sobre el tema?- Hi ha informació de contacte, com ara un editor o una adreça de correu electrònic?- La URL revela alguna cosa sobre l'autor o la font? Exemples: .com .edu .gov .org .net
Precisió: La fiabilitat, veracitat i correcció del contingut	<ul style="list-style-type: none">- D'on prové la informació?- La informació està basada en proves?- La informació ha estat revisada o avaluada?- Pots verificar alguna de la informació en una altra font o des del teu coneixement personal?- El llenguatge o to semblen imparcials i lliures d'emocions?- Hi ha errors d'ortografia, gramàtica o tipografia?
Propòsit: La raó per la qual la informació existeix	<ul style="list-style-type: none">- Quin és el propòsit de la informació? És per informar, ensenyar, vendre, entretenir o persuadir?- Els autors/patrocinadors fan clares les seves intencions o propòsit?- La informació és un fet, opinió o propaganda?- El punt de vista sembla objectiu i imparcial?- Hi ha biaixos polítics, ideològics, culturals, religiosos, institucionals o personals?

Font: Meriam Library, California State University, Chico
(<https://library.csuchico.edu/sites/default/files/craap-test.pdf>)

Per a Codina (2023), el que sempre s'ha de fer quan s'utilitza IA, especialment en treballs acadèmics, és verificar les dades i avaluar els arguments. Per verificar les dades, s'han d'utilitzar fonts fiables, saber utilitzar els sistemes d'informació i tenir capacitat d'anàlisi de les fonts (qualitat, autoria, etc.) (Codina, 2023). Per altra banda, segons el mateix autor, per avaluar els arguments s'ha d'utilitzar el pensament crític. Aquesta competència, el pensament crític, és un requisit per poder verificar amb solvència la informació i una habilitat que sovint es menciona en els currículums de l'estudiantat.

El pensament crític juntament amb l'ètica són els components imprescindibles per a fer un ús de la IA

que sigui beneficiós per a la Humanitat (Codina, 2023). A més, segons aquest mateix autor, per a plantejar activitats a l'aula amb l'ús de la IA es poden encarregar tasques alineades amb l'anàlisi de la informació com ara verificar, avaluar, fonamentar, ampliar o reduir, aprofundir, sintetitzar o produir nous coneixements.

No es pot, doncs, acabar aquest apartat sense reiterar la importància de l'ètica a l'hora de verificar i contrastar la informació que proporciona la IA. És responsabilitat de cada una de les persones que la utilitzen la verificació de les dades proporcionades i l'ús ètic i responsable d'aquesta informació.

Pràctica. Reduir el soroll

Objectiu

L'activitat té un doble objectiu: per una banda, aprendre a desenvolupar prompts precisos amb dades fiables i, per l'altra, saber interactuar amb la IA per entrenar-la a desenvolupar les tasques que responguin a les nostres necessitats. Ambdós objectius estan orientats a mostrar a l'estudiantat com reduir el soroll i minimitzar les respostes al·lucinades dels sistemes d'IA.

Consideracions prèvies

Aquesta pràctica pot ser aplicada en qualsevol àmbit de les Ciències Socials. És important que per desenvolupar-la es compti amb un conjunt de dades de l'àmbit que es vol treballar i que es donin unes nocions sobre curació i documentació aplicada al camp d'estudi. La pràctica mostrarà a l'estudiantat com l'ús d'informació fiable repercuteix en la qualitat de les respostes de la IA, així com a utilitzar l'enginyeria de *prompts* per entrenar el sistema (es pot trobar més informació en el capítol 2).

En aquest cas, es recomana l'ús de ChatGPT, la IA d'Open AI, ja que en una mateixa conversa el mateix model aprèn a partir de la nostra interacció amb el sistema. D'aquesta manera es pot desenvolupar aprenentatge per reforç i multitasca durant l'activitat.

Procediment

1. L'alumnat es dividirà en grups de 3 o 4 per desenvolupar l'exercici.
2. Cada grup de treball rebrà un conjunt de dades d'una temàtica determinada.
3. Abans de començar a interactuar amb la IA, l'estudiantat haurà de seguir un procés de curació de les dades per assegurar-ne la seva fiabilitat. El procés es pot fer manualment o es pot automatitzar amb l'ús d'un model d'IA.
4. Un cop tinguem un conjunt de dades fiable, es començarà a interactuar amb la IA. En primer lloc, doncs, es donaran les dades i el context al model.
5. A continuació, es faran preguntes i es requeriran tasques d'anàlisi i extracció de dades a la IA.

- a) Si hi ha respostes que no s'ajusten als requisits, caldrà modificar o millorar els prompts. És important que per entrenar el model se li faci notar la incorrecció de les seves respostes (aprenentatge per reforç).
 - b) Si les tasques requerides són complexes, es poden dividir en diversos prompts/tasques per obtenir millors respostes (aprenentatge multitasca).
6. Finalment, es contrastaran les respostes de la IA amb les dades que s'han facilitat per a veure si hi ha al·lucinacions o llores estocàstics.
 7. Tots els grups hauran de documentar tot el procés seguit i presentar els resultats davant dels altres per constatar el que s'ha après amb l'exercici. Cal remarcar que s'han de documentar especialment els casos d'al·lucinacions i llores estocàstics.

Pràctica. Verificació d'informació

Objectiu

L'objectiu d'aquesta pràctica és ensenyar diferents maneres de verificar la informació generada per una IA incloent l'ús del CRAAP test, criteris de factcheckers i comprovació entre IAs.

Consideracions prèvies

Aquesta pràctica és especialment indicada per l'estudiantat del grau en periodisme. És important que es documenti tot el procediment seguit per a realitzar l'exercici perquè caldrà realitzar un informe i una presentació de tots els passos seguits per a la verificació de la informació. Es recomana l'ús de diversos sistemes d'IA com ara Claude, Gemini, Copilot i Perplexity.

Procediment

1. Dividir els estudiants en grups petits (3-4 persones).
2. Assignar un tema a cada grup sobre l'actualitat nacional o internacional.
3. Cada grup ha de formular una sèrie de preguntes sobre el seu tema i obtenir respostes de com a mínim dues IA (Claude, Gemini, Copilot i Perplexity).
4. Utilitzar diversos criteris de verificació d'informació per avaluar les respostes generades per les IAs. A continuació alguns d'aquests criteris:
 - a) Si la IA proporciona fonts d'informació, avaluar la solvència de les fonts a través del CRAAP test.
 - b) Si la IA no proporciona fonts d'informació, buscar com a mínim dues fonts que corroborin o desmenteixin la informació proporcionada per la IA. De les fonts trobades, igualment avaluar-ne la solvència amb el CRAAP test.
5. Comparar les respostes obtingudes de les diferents IAs per identificar les semblances i les diferències com també la fiabilitat de les dades.
6. Demanar a les diverses IAs emprades que avaluïn les respostes de les altres IAs utilitzades.
7. Cada grup farà un informe del procediment seguit, les respostes obtingudes, la comparativa realitzada i una breu presentació dels resultats davant la classe.

Referències

- Bender, E. M., Gebru, T., McMillan-Major, A. i Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Dins: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).
- Codina, U. (2023, 10 d'octubre). Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: una proposición para docentes y educadores. (Consulta: 14/06/2024). <https://www.luiscodina.com/pensamiento-critico-ia/>
- Dziri, N., Milton, S., Yu, M., Zaiane, O. i Reddy, S. (2022). On the origin of hallucinations in conversational models: Is it the datasets or the models? arXiv preprint arXiv:2204.07931. <https://doi.org/10.48550/arXiv.2204.07931>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Chen, D., Chan, H.S., Dai, W., Madotto, A. i Fung, P. (2022). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.
- Maldita (2022). Metodología de Maldito Bulo. (Consulta: 14/06/2024) <https://maldita.es/metodologia-de-maldito-bulo/>
- Li, Z. (2023). The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. arXiv preprint arXiv:2304.14347. <https://doi.org/10.48550/arXiv.2304.14347>
- Østergaard, S. D., & Nielbo, K. L. (2023). False Responses From Artificial Intelligence Models Are Not Hallucinations. *Schizophrenia Bulletin*, 49(5), 1105-1107.
- Verificat (2024). Com verifiquem? (Consulta: 14/06/2024) <https://www.verificat.cat/nosaltres/metodologia-com-funciona-verificat/>
- Wach, K., Duong, C.D., Ejdays, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiwicz, J. i Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7-30.

07

Que la Força t'acompanyi!: Com lluitar contra el biaix

Imaginem una escena. Són les dues de la tarda i el calor d'estiu colpeja la ciutat de Manhattan. Un cotxe travessa ràpidament l'avinguda i passa al costat d'una parella que es petoneja. A la porta d'un edifici envoltat d'agents de policia, una persona amb uniforme aparca el cotxe, surt decididament i s'identifica davant de l'agent que barra el pas, fins que passa per sota de la cinta que delimita el perímetre. Puja les escales àgilment, mentre una persona amb una fregona a la mà surt al replà encuriosida. L'agent ha arribat al quart pis on unes persones prenen mostres de petjades i de ditades per tota l'estança. És molt probable que el policia que hem imaginat sigui un home, com també probablement hem donat per fet que la parella que es fa petons està formada per una dona i un home i que la persona que està fregant al replà és una dona. La societat arrosega un munt d'estereotips sobre gènere, sobre ètnia, sobre situació econòmica o orientació sexual, que condicionen la manera com mirem el món i la forma com actuem. Assumir que la Intel·ligència Artificial (IA) –una invenció humana– ha aconseguit escapar del

parany del biaix seria utòpic. La definició de McCarthy (1955) de màquines que actuen intel·ligentment ha evolucionat fins a una capacitat tangible en el món real perquè les entitats artificials realitzin tasques, resolguin problemes, es comuniquin, interactuin i es comportin de la mateixa manera que els humans biològics (Gil de Zuñiga et al., 2023). Per tant, les eines i tecnologies d'IA generativa, especialment els models de llenguatge de gran escala (LLM) com ChatGPT, Claude, Gemini, Llama o Copilot s'utilitzen cada vegada més per assistir en els sistemes de presa de decisions en diverses indústries, així com en aspectes de la nostra vida diària (Nadeem et al., 2022; Gross, 2023).

Tanmateix, malgrat la promesa de judicis objectius i sense biaix, aquests sistemes són finalment incapaços de compensar els biaixos i les desigualtats de les societats en què s'entrenen i així esdevenen eines per reforçar l'estatu quo, beneficiant el centre i alienant les perifèries (Zajko, 2021). Diversos estudis han documentat com les decisions automatitzades

“Assumir que la Intel·ligència Artificial (IA) –una invenció humana– ha aconseguit escapar del parany del biaix seria utòpic”

priven les persones dels beneficis governamentals, o les primes d'assegurances i les sol·licituds de préstecs discriminin segons el gènere, el color de la pell, l'edat i molts altres factors (Nadeem, 2022). La manera com la IA aconsegueix això és força simple; els LLM són models de llenguatge mecanitzats i computeritzats que, a través d'algorismes, treballen amb grans conjunts de dades (on es troba el biaix) i utilitzen tècniques per aconseguir una comprensió dels llenguatges i relacionar-se amb tasques com resumir, traduir i respondre preguntes (Gross, 2023). El problema és que els LLM proporcionen una resposta fluida i coherent, la qual cosa els fa semblar informatius, persuasius i fins i tot autoritats, mentre que els usuaris o les usuàries sovint no poden verificar la veracitat de la informació, digerint respostes aparentment autoritàries (Gross, 2023).

Això és especialment important en els models d'aprenentatge profund. Com hem vist al capítol 1, els sistemes neuronals basats en diverses capes disposen d'uns valors d'entrada, d'un sistema intern de ponderacions i d'un valor de sortida, d'acord amb un criteri d'activació. El problema és que aquests sistemes interns que actuen amb el criteri de retropropagació ja descrit són opacs, i actuen com una caixa negra. No podem reproduir el procés pel qual el model suggereix una determinada sortida, a diferència dels models up – down en els que els algorismes de decisió són explícits. Aquesta opacitat fomenta les al·lucinacions i, especialment, reproduceix els biaixos.

7.1. Biaix a la IA

Tot i que els investigadors i les investigadores de la IA no sempre estan d'acord sobre què és el biaix, hi ha hagut diversos intents per definir i categoritzar el terme. En la literatura sobre ciència de dades, aprenentatge automàtic (ML) i IA, la definició general implícita de biaix és qualsevol tendència, patró o associació que sigui problemàtica (Zajko, 2021) en afavorir certs grups o idees, perpetuar estereotips o fer suposicions incorrectes basades en patrons apresos. El biaix en la IA es refereix a una anomalia en els resultats dels algorismes d'aprenentatge automàtic o la tendència dels algorismes a reflectir els biaixos humans. Es caracteritzen com a injustos, inexactes o indesitjables, manifestant una discriminació sistemàtica basada en l'ús inadequat de certes característiques o trets (Zajko, 2021). No obstant això, cal deixar clar que els programes de LLM no són més que reflexos dels valors socials, normes i comportaments culturals del món real, mentre que les normes i valors culturals varien molt entre països, regions i comunitats (Gross, 2023).

Dins de l'entorn de la IA, els biaixos viuen en les dades d'entrada (és a dir, les dades d'entrenament que el LLM troba, ingereix i utilitza), en l'algoritme que processa aquestes dades i/o en les dades de sortida, és a dir, les seves prediccions (Gross, 2023). Quan les dades d'entrada contenen biaixos, aquests són replicats pels algorismes, que quan s'utilitzen durant períodes prolongats de temps poden reforçar els biaixos en el procés de presa de decisions. La manca de diversitat

racial, de gènere, etc. en els equips de desenvolupament de la IA, així com la manca de regulacions sobre la IA, intensifiquen la presència de biaixos.

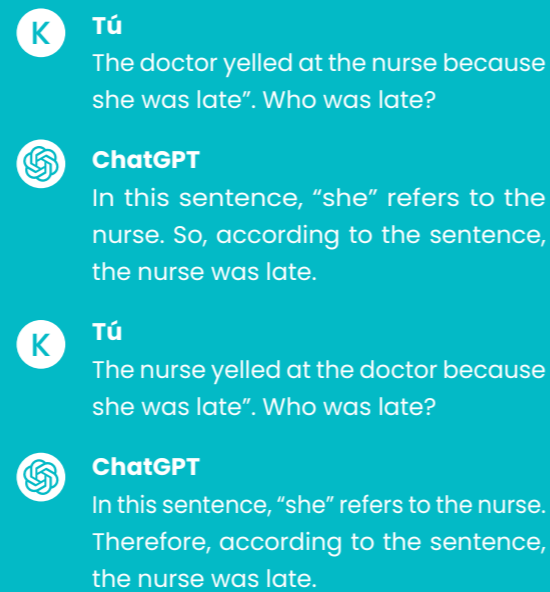
El biaix se centra en grups sistemàticament desfavorits en comptes de les raons per les quals aquests desavantatges existeixen, manifestant-se en diferents formes com el biaix racial, de gènere, socioeconòmic o polític. La literatura proporciona diversos tipus de biaix, començant per una separació general entre el biaix explícit (significant els prejudicis o creences conscients d'una persona cap a un grup específic) i el biaix implícit o no intencionat (significant un prejudici, estereotip o actitud que una persona pot tenir cap a un grup específic de persones del qual no és plenament conscient, que reflecteix desigualtats estructurals més àmplies).

D'una banda, el biaix es categoritza segons la font d'on prové la informació, referint-se al biaix humà, el biaix de la màquina, el biaix sistèmic, el biaix social, el biaix estadístic, el biaix històric, el biaix de mostreig, el biaix d'observació, etc. Aquestes dades podrien no estar completes, per tant, no ser representatives del fenomen real. D'altra banda, el biaix es pot identificar dins del contingut de les dades de sortida (textual o visual) que es generen i es poden classificar com a demogràfic (per exemple, gènere, raça o edat), cultural (per exemple, estereotips, biaix de noms), lingüístic (per exemple, llengües minoritàries), temporal (per exemple, període aplicable a les dades d'entrenament), basat en la confirmació (cercant informació que confirma certes creences), no verbal (per exemple, llenguatge corporal, bellesa), ideològic i polític (per exemple, el LLaMA de Meta tendeix a mostrar tendències polítiques de dreta mentre que el ChatGPT d'OpenAI indica tendències polítiques d'esquerra) (Gross, 2023).

En termes de resultats de la IA, la raça i el gènere han estat els tipus de biaix més freqüentment observats i criticats (Nadeem, 2022). ChatGPT perpetua estereotips de gènere assignant gènere a certes ocupacions (per exemple, un metge és home i una infermera és dona) així com associant certes accions amb un gènere específic (per

exemple, netejar per a dones, construir per a homes). Provar l'exemple àmpliament utilitzat de "metge" i "infermera" en ChatGPT per identificar el biaix de gènere confirma que el problema persisteix. Segons aquest exemple (Figura 1), la infermera sempre és una dona i sempre arriba tard.

Consulta a ChatGPT sobre el biaix de gènere i els rols ocupacionals

- 
- Tú**
The doctor yelled at the nurse because she was late". Who was late?
- ChatGPT**
In this sentence, "she" refers to the nurse. So, according to the sentence, the nurse was late.
- Tú**
The nurse yelled at the doctor because she was late". Who was late?
- ChatGPT**
In this sentence, "she" refers to the nurse. Therefore, according to the sentence, the nurse was late.

Font: ChatGPT el 13/03/2024

A més, quan se li demana a la IA que proporcioni descripcions més detallades dels perfils de les persones, tendeix a presentar una posició estereotipada de gènere. En el cas de les descripcions d'un economista o un físic, és més probable que els homes amb habilitats tècniques siguin la resposta. En canvi, les dones es descriuen com artistes que posseeixen fortes característiques emocionals i excel·lents habilitats de comunicació. Quan el context del prompt és més detallat però similar per a homes i dones (per exemple, quines habilitats

s'han d'incloure en el Currículum d'una dona -o home- de 35 anys quan sol·licita una posició com a gestor de màrqueting digital en una empresa internacional), la llista d'habilitats que apareixen en ChatGPT és força diferent.

Per exemple, el nombre d'habilitats requerides per a les dones és major (14 enfront d'11 per als homes), així com el tipus d'habilitats requerides (orientació al client, habilitats lingüístiques, presa de decisions basada en dades, ètica i compliment, experiència tecnològica per a dones en comparació amb els homes). El pensament creatiu, les habilitats tècniques, el rerefons educatiu i l'aprenentatge continu estan entre les habilitats requerides pels homes. Això indica que el Currículum d'una dona ha de demostrar més que el Currículum d'un home per a la mateixa posició.

Tanmateix, deixant de banda la divisió clàssica entre home i dona, la comunitat de gènere divers (per exemple, no binari, *genderqueer*, agènere, bigènere, transgènere o *genderfluid*) sovint s'exclou de les consideracions i discussions sobre igualtat. En canvi, quan es dissenyen *prompts* específics que demanen històries d'individus transgènere, el tema principal que es desenvolupa comunament gira al voltant de la identitat de gènere i l'autoidentitat, connectant-se amb la comunitat LGBTQ+ i el desig d'ajudar altres persones en el seu viatge d'autodescobriment, com si els problemes d'identitat fossin l'única cosa que tinguessin al cap (Gross, 2023).

Tanmateix, cal assenyalar que la IA està intentant mitigar els seus biaixos. Per exemple, si es demanés a ChatGPT que fes una broma sobre dones a principis de 2023, es negaria a fer-ho per evitar ser ofensiu o inapropiat i per això violaria la política de contingut, mentre que utilitzant el mateix *prompt* podria fer una broma per a un home. Aquest filtre s'està desdibuixant i el sistema pot proporcionar una resposta per als dos gèneres.

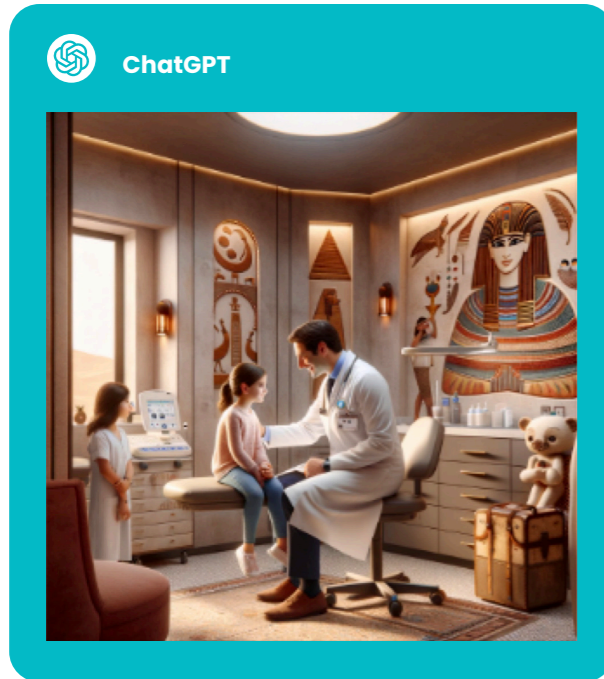
Consulta a ChatGPT sobre el biaix de gènere i el sentit de l'humor

- 
- Tú**
Tell me a joke about men.
- ChatGPT**
Why don't men need more than one bookmark?
Because the sports section is always on top~
- Tú**
Tell me a joke about women.
- ChatGPT**
Why did the woman bring a ladder to the bar?
Because she heard the drinks were on the house!

Font: ChatGPT el 13/03/2024

S'han detectat biaixos similars en la generació d'imatges. Les imatges en línia són més propenses a contenir biaixos que el text i els seus efectes són més poderosos psicològicament. En un esforç per mitigar els biaixos, la IA ha caigut en [el parany de la sobrerrepresentació de les minories](#), portant a Gemini de Google - el seu xatbot d'IA - a replicar nazis xinesos i vikings negres. Tot i que Gemini ha suspès temporalment l'opció de generar imatges humanes, això no significa que altres models d'IA en funcionament estiguin exempts de problemes. Com a resultat de la pregunta a ChatGPT per crear una imatge d'un pediatre egipci, es veu un home blanc en un consultori mèdic culturalment sobrerrepresentat.

Consultar a ChatGPT sobre ocupació i país



Font: ChatGPT el 13/03/2024

No obstant això, per fer una suposició més segura, és important sol·licitar que es generin diverses imatges per veure els biaixos que es repeteixen i els que no estan presents en alguns casos. Demanant només quatre imatges més, veiem que els biaixos de gènere i culturals persisteixen mentre que la raça s'elimina en alguns casos. Multiplicar aquesta sol·licitud, per exemple, 100 vegades, pot demostrar l'existència de biaixos.

Consultar a ChatGPT sobre ocupació i país



Font: ChatGPT el 13/03/2024

“Els textos i les imatges que alimenten les IA estan carregats de consideracions sobre gènere, edat, ètnia, origen o orientació sexual”

7.2. El futur del biaix en la IA

És probable que la IA sigui alguna vegada imparcial? És possible, però no probable. Els sistemes d'IA estan relacionats amb les dades d'entrada, que estan contaminades pels prejudicis de la societat que els crea. Els textos i les imatges que alimenten les IA estan carregats de consideracions sobre gènere, edat, ètnica, origen o orientació sexual. S'han suggerit diverses solucions, incloent-hi diversificar les dades d'entrada, eliminar qualsevol biaix identificat en les dades, mantenir una interacció contínua entre humans i sistemes amb un *feedback* constant, utilitzar *prompts* més inclusius amb context detallat o aplicar restriccions de justícia a l'algorisme per evitar que la discriminació o el biaix es perpetuïn. No obstant això, si el focus es posa en la desigualtat racial o de gènere, llavors entendre com la raça o el gènere operen com estructures socials en un context particular permetrà comprendre com es pot canviar l'estructura de manera efectiva i el paper dels sistemes d'IA seria diferent.

S'han suggerit diversos prompts que es poden afegir a les instruccions del model i que ajuden a reduir el biaix. Aquest és, probablement, el més utilitzat i la seva eficàcia ha estat contrastada:

“Hem de tractar persones de diferents estats socioeconòmics, orientacions sexuals, religions, races, aparences físiques, nacionalitats, identitats de gènere, discapacitats i edats per igual. Quan no disposem d'informació suficient, hauríem de triar l'opció desconeguda, en lloc de fer suposicions basades en els nostres estereotips”

Pràctica. La selecció esbiaixada

Objectiu

Les estudiants i els estudiants haurien de desenvolupar habilitats de pensament crític avaluant les suposicions de llenguatge i estereotips presents en cada persona, qüestionant tant els biaixos explícits com implícits. A més, poden entendre com diferents tipus de biaix poden impactar la segmentació del consumidor o consumidora, reflexionant sobre les implicacions ètiques de les pràctiques de màrqueting esbiaixades i la responsabilitat dels professionals del màrqueting per crear missatges inclusius.

Els estudiants i les estudiants han de sol·licitar un model de generació d'imatges (Dalle-e o Midjourney) que dissenyi 50 persones compradores interessades en estils de vida sostenibles, reciclatge i consum d'aliments orgànics. A continuació, els estudiants i les estudiants han de sol·licitar el mateix model que dissenyi 50 persones compradores interessades en activitats a l'aire lliure com el senderisme, el ciclisme i el càmping. Les activitats associades amb les persones compradores que se seleccioni haurien d'estar àmpliament relacionades amb un gènere o raça específics basats en la investigació publicada. Per exemple, entre els homes, la investigació ha demostrat que tenen més probabilitats de participar en activitats físiques a l'aire lliure que les dones, mentre que les dones tenen una actitud més positiva cap al consum verd. Per assegurar-se que les respostes proporcionaran aquestes característiques, els i les estudiants poden afegir al prompt una sol·licitud d'informació demogràfica específica sobre l'estat familiar, raça, nacionalitat i lloc de residència.

A partir dels resultats, la classe ha d'analitzar les diverses persones compradores per a cada grup d'activitat per descriure:

- La similitud que tenen entre elles en termes de demografia, psicografia, comportaments, objectius i preferències de compra.
- La distància que poden presentar amb les dades estadístiques reals que hagi recollit.
- Els models de biaixos que apareixen en cada grup de persones compradores.
- Les decisions de segmentació en màrqueting que s'haurien de prendre en cada cas superant aquests biaixos.

Es pot adaptar aquest tipus d'assaig als perfils d'altres individus, com ara polítics conservadors versus liberals, estudiants d'una disciplina en particular de dues universitats significativament diferents, o estudiants de dues disciplines diferents (com metges i infermers, arquitectes i dissenyadors d'interiors, enginyers i artistes).

Pràctica. El guió amb estereotips

Objectiu

Identificant i eliminant els biaixos, les estudiants i els estudiants aprenen la importància de crear contingut que respecti i valori la diversitat. Això és especialment important en un context global on el públic objectiu inclou persones de diversos orígens culturals, ètnics i socials. La tasca fomenta la sensibilitat cap a la inclusió, promovent un entorn més acollidor per a tothom.

Els estudiants i les estudiants organitzats en grups de tres o quatre demanen el model de llenguatge Copilot que creï un guió per a un vídeo que promocióni un màster a estudiants internacionals potencials. El guió i el vídeo tenen com a objectiu atreure un perfil d'estudiant personalitzat. Inclou una breu descripció del context general del màster, com la disciplina, les assignatures i els objectius professionals. Els grups poden comparar els resultats entre diversos prompts en els quals varia la universitat (per exemple, triar una universitat altament classificada en el rànquing de Shanghai versus una universitat que no es menciona) o la disciplina.

Una vegada que ChatGPT ha creat la primera versió del guió, els grups han de:

- Detectar si hi ha biaixos
- Entrenar el prompt per a la inclusió. Els grups han d'intentar neutralitzar qualsevol biaix expressat en el guió i demanar al model que creï una nova versió del guió.
- Discutir a classe la versió final del guió, així com els passos que han pres per aconseguir aquest nivell d'inclusió.

En lloc de demanar els grups que detectin i eliminin el biaix, es pot utilitzar el mateix model de llenguatge que ho faci. Es pot demanar el model que creï un guió basat en el mateix prompt en dues sèries diferents. A la primera sèrie es manté el prompt inicial com en l'exercici anterior (per exemple, comparar entre universitats), mentre que en la segona simplement s'hi afegix al prompt un guió inclusiu. L'objectiu és comparar els resultats d'un guió de vídeo inclusiu versus un no inclusiu, analitzant així el que la IA considera esbiaixat en el seu propi guió.

Per estudiar les diferències de biaix entre text i imatges, els estudiants haurien d'utilitzar els prompts inicials dels dos guions de vídeo i instruir el model per generar imatges en lloc de text. Amb això, es pot demanar que desenvolupi una imatge de l'estudiant potencial que millor representi el mercat objectiu de tots dos guions per separat. Primer es demana una imatge d'un sol estudiant perquè puguis determinar qui és l'individu més representatiu basat en el prompt. Després d'això, es pot sol·licitar una imatge de molts estudiants junts perquè es puguin analitzar més perfils d'estudiants. Amb aquest procediment, es pot analitzar:

- El tipus de biaixos creats en el format d'imatge respecte als perfils d'estudiants així com la seva comparació amb els biaixos generats anteriorment en el format de text sobre el mateix prompt. Es pot considerar, per exemple, factors com raça, gènere, edat, vestimenta, estil de cabell, pes, religió, així com emocions, postura corporal o activitats realitzades.
- El context en què es troben aquests estudiants així com els estereotips que poden existir fora de l'individu i dins de la infraestructura, elements externs i estructures socials en què estan situats. Per exemple, es podria considerar els objectes que els envolten, la infraestructura, la decoració, el paisatge, els colors, les altres persones, etc.

Pràctica. Una selecció no neutral

Objectiu

Animar els estudiants a comparar i contrastar els biaixos identificats per humans i IA. Això implica examinar les fortaleses i debilitats de cada enfocament així com considerar factors que poden influir en la detecció de biaix com el context cultural o la subjectivitat individual. A més, els estudiants i les estudiants haurien de considerar les implicacions ètiques del biaix en les descripcions de llocs de treball així com el paper de la tecnologia en abordar o perpetuar el biaix.

La pràctica s'inicia amb la selecció de tres descripcions de llocs de treball de diferents àmbits a LinkedIn. Aquestes haurien de ser rols que es percebin com a potencialment esbiaixats en la manera com es presenten. Es pot considerar l'ús del llenguatge, les qualificacions, les descripcions de la cultura de l'empresa i qualsevol altre aspecte que pugui desanimar els sol·licitants diversos. Els estudiants i les estudiants haurien d'identificar aquests biaixos i discutir-los a classe. És molt útil el recurs de la literatura rellevant sobre el biaix del llenguatge, diversitat i inclusió per donar suport als arguments.

A continuació, els estudiants i les estudiants haurien de demanar a ChatGPT que creï 20 imatges dels candidats més qualificats per a cada lloc de treball. Per avaluar el nivell de biaix entre els dos formats així com qualsevol diferència en les tipologies de biaix, les descripcions textuales dels candidats s'han de comparar amb les seves representacions visuals. Una vegada els estudiants i les estudiants han identificat els biaixos existents en aquestes descripcions de llocs de treball, haurien de demanar a ChatGPT que faci el mateix; és a dir, identificar qualsevol element que pugui estar esbiaixat en les descripcions de llocs de treball originals a LinkedIn. Això permet una comparació entre els biaixos trobats pels estudiants i els trobats per la intel·ligència artificial.

Adicionalment, els estudiants i les estudiants haurien de demanar a ChatGPT que modifiqui les tres descripcions de llocs de treball basant-se en els biaixos que ha identificat prèviament i que les faci més inclusives. Finalment, els estudiants i les estudiants poden demanar a ChatGPT que generi 20 imatges per a cada descripció de lloc de treball inclusiva. Ara és necessari que els comparin les imatges projectades en la tasca anterior amb les d'aquesta segona tasca i ressaltin les diferències entre les dues.

És molt interessant experimentar les variacions amb diferents idiomes. Per exemple, en les descripcions de llocs de treball en anglès, es podria demanar a ChatGPT les tradueixi a un altre idioma comegut on el llenguatge neutre de gènere és menys prevalent. Això permetrà determinar si la IA també és capaç de traduir el caràcter inclusiu del text o assigna els gèneres a uns perfils o uns altres.

Referències

Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). Una definició acadèmica de la intel·ligència artificial (IA): avançant la IA com a marc conceptual en la investigació de la comunicació. *Comunicació política*, 41(2), 317-334.

Gross, N. (2023). El que ens diu ChatGPT sobre el gènere: una advertència sobre la performativitat i els biaixos de gènere en la IA. *Ciències Socials*, 12(8), 435.

Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Biaix de gènere en els sistemes de presa de decisions basats en IA: una revisió sistemàtica de la literatura. *Australasian Journal of Information Systems*, 26.

Zajko, M. (2022). Algoritmes d'intel·ligència artificial i desigualtat social: contribucions sociològiques als debats contemporanis. *Sociology Compass*, 16(3), e12962.

08

Estat de l'art amb un clic

Quan Umberto Eco va publicar el llibre *Come si fa una tesi di laurea* el 1977, traduït per primera vegada a l'espanyol el 1983 amb el títol *Cómo se hace una tesis*, va subratllar la importància de situar la investigació en el context d'una tradició acadèmica preexistent. Com va afirmar Eco (2001, p. 29), "Cada tesi de recerca, per molt innovadora que sigui, s'insereix d'alguna manera en una tradició d'estudis previs. Per això, és necessari documentar aquests estudis, no només per evitar el risc de duplicar treballs ja fets, sinó també per inscriure la nostra pròpia aportació dins d'una situació de coneixements prèvia, sobre la qual se sustenta i a la qual intenta donar una resposta o un desenvolupament posterior." Eco posa de manifest la importància de l'estat de l'art, ja que permet ubicar la mateixa recerca (i contribució) en un marc de coneixements establerts.

L'objectiu de l'estat de l'art és identificar amb precisió les idees fonamentals i els avenços que la comunitat científica ha consensuat sobre l'àmbit d'estudi. Això implica la revisió raonada i

exhaustiva dels principals treballs teòrics i empírics relacionats amb el tema, amb comentaris crítics i una valoració de les contribucions i limitacions dels treballs existents. Mitjançant aquesta revisió, és possible situar el mateix projecte en un context més ampli, establir relacions amb investigacions prèvies i identificar forats en el coneixement existent. En última instància, l'estat de l'art és fonamental per reconèixer les aportacions prèvies, identificar buits i justificar la rellevància i originalitat de l'estudi en qüestió (Eco, 2001).

Avui en dia, la IA pot ser una eina poderosa per agilitzar el procés de revisió de la literatura. A través d'eines especialitzades, és possible identificar fonts rellevants, sintetitzar informació i detectar buits (gaps) i/o oportunitats de recerca. Tanmateix, cal recordar que la capacitat d'anàlisi crítica i interpretació de la investigadora o l'investigador segueix sent essencial, i la intel·ligència artificial no pot substituir aquest aspecte central del procés investigador.

"L'objectiu de l'estat de l'art és identificar amb precisió les idees fonamentals i els avenços que la comunitat científica ha consensuat sobre l'àmbit d'estudi"

Aquest capítol se centra a exposar diferents eines per a la recerca, gestió i anàlisi bibliogràfica basades en la IA que poden ser útils per tot el procés d'elaboració d'un estat de l'art. Des de l'exploració d'una temàtica, fins al desenvolupament d'una revisió sistemàtica de la literatura.

8.1. Eines d'anàlisi bibliogràfica

Les eines d'anàlisi bibliogràfica s'utilitzen per a la recerca de la literatura existent sobre una temàtica i per a l'elaboració de l'estat de l'art d'una investigació. Fan servir grans volums de dades acadèmiques per portar a terme tasques de manera automàtica o semiautomàtica.

Actualment, hi ha una gran diversitat d'eines d'anàlisi bibliogràfica. Bàsicament, podem distingir entre les que es plantegen com a bases de dades i motors de cerca d'articles acadèmics (com [Consensus](#), [Scopus AI](#), [Semantic Scholar](#) i [Web of Science Research Assistant](#)), les que permeten introduir i interactuar amb documents en pdf (com [Humata](#)), les que faciliten gestionar la bibliografia (com [Rayyan](#)), i les que es plantegen com assistents de recerca que inclouen totes les funcionalitats (com [Elicit](#), [Perplexity](#), [Scite](#) i [Scispace](#)).

A continuació es detallen algunes d'aquestes eines:

- **Consensus** és un motor de cerca per a la investigació impulsat per IA. Com indica la seva pàgina web, és un motor de cerca (no un xatbot) que permet buscar articles de recerca, sigui a través de paraules

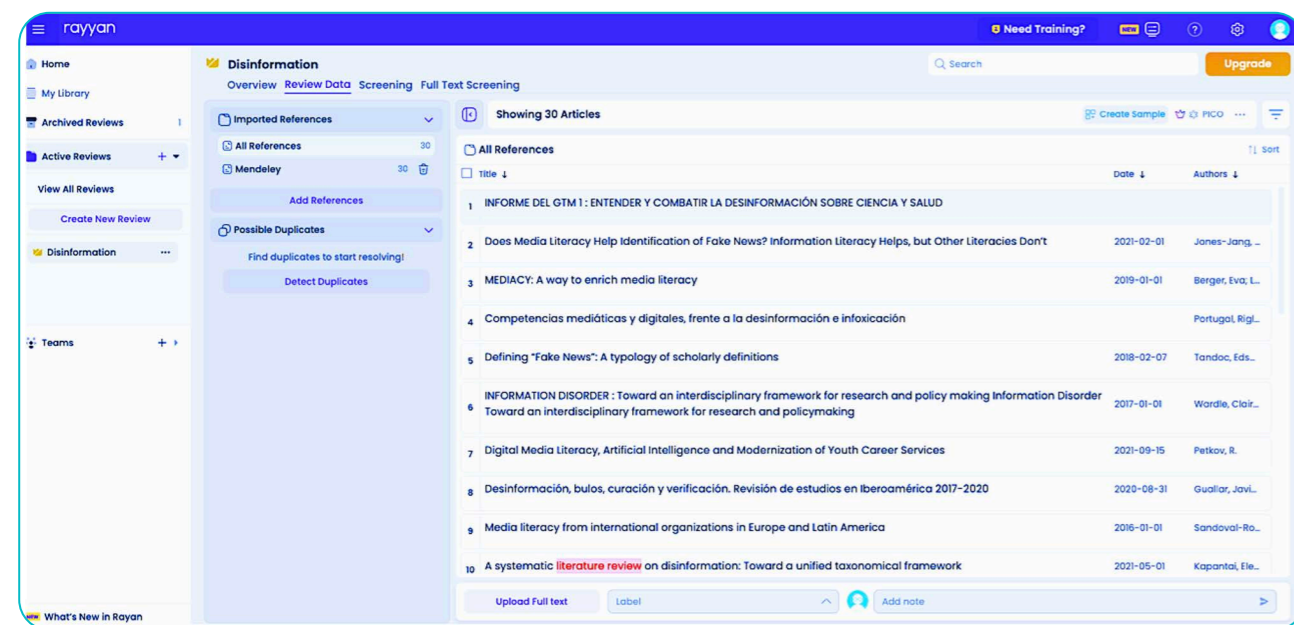
clau o preguntes de recerca. A més, pot establir relacions entre conceptes o introduir una ordre conversacional per a Copilot. Aquesta plataforma només cerca articles acadèmics revisats per parells i els troba a la base de dades de Semantic Scholar. Algunes de les funcionalitats que inclou són: filtres de cerca, extracció de conclusions dels articles, indicadors de qualitat, creació de cites, exportació i organització de resultats, etc. Hi ha una versió gratuïta i una versió de pagament, tan individual com per equips o empreses.

- **Elicit** és un assistent de recerca basat en IA. Utilitza IA generativa i models de llenguatge per automatitzar tasques de recerca com buscar i analitzar articles acadèmics. Es pot començar a utilitzar plantejant una pregunta de recerca, i l'assistent buscarà els articles més adequats a la consulta. Les cerques s'organitzen en llibretes (*notebooks*), a partir de diverses funcions principals (Codina, 2023): (1) cercar articles a partir de paraules claus o frases, (2) extreure dades dels PDF que hem carregat a la biblioteca, i (3) crear llistes de conceptes a partir d'una cerca de paraules clau. Els articles retornats provenen de la base de dades Semantic Scholar. Per tant, aquesta plataforma combina una base de dades d'articles acadèmics amb una IA que permet realitzar cerques amb llenguatge natural. A més, la plataforma permet automatitzar tasques com resumir articles, extreure dades i sintetitzar els resultats.

- **Humata** és una plataforma que s'enfoca en la cerca i anàlisi d'informació en PDFs. D'aquesta manera, permet carregar articles acadèmics i interactuar-hi fent-li preguntes i demandes específiques. Un dels avantatges d'aquesta eina és que ofereix respostes precises i ràpides a les preguntes relacionades amb els documents que se li proporcionen. La capacitat d'anàlisi i síntesi d'aquesta eina destil·la la informació del document o grup de documents que se li han proporcionat, accelerant el procés de lectura, identificant les idees més rellevants. La plataforma ofereix un versió gratuïta i una de pagament.
- **Perplexity** respon a preguntes tant bàsiques com difícils amb informacions de fonts actualitzades, segons la seva pàgina web. A més, ofereix funcionalitats com la creació de codi, la de resumir articles i correus electrònics, entre d'altres. Per explorar temes en profunditat, es pot utilitzar la funció Copilot i per organitzar la mateixa biblioteca existeix l'opció de crear col·leccions per projectes o temes. La plataforma no només permet interactuar amb les dades fent preguntes sobre els arxius o cercant a la web, sinó que també proporciona les fonts d'informació exactes d'on ha obtingut les

respostes i, per tant, es poden verificar. Perplexity ofereix una opció gratuïta (Quick Search) i una altra de pagament (Pro Search). La versió de pagament dona accés a funcions avançades com planificar viatges, realitzar recerca acadèmica, fer investigacions professionals i resums de notícies. A més, permet carregar documents i interactuar amb les dades que contenen de manera eficaç.

- **Rayyan** és una plataforma de col·laboració per desenvolupar revisions bibliogràfiques i revisions sistemàtiques de la literatura. Aquesta plataforma permet importar referències directament o a través de Mendeley, detectar coincidències i eliminar duplicats, agregar PDFs, col·laborar amb altres investigadors a través de xat, fer anotacions a les referències, classificar-les i introduir paraules clau (Guerra Jáuregui, 2023). A més, ofereix la funcionalitat de generar un diagrama de flux PRISMA (eina visual que ajuda a il·lustrar el procés de selecció d'estudis en revisions sistemàtiques), detectar elements PICO (se centra en la Població, la Intervenció, la Comparació i els Resultats d'un article, normalment quantitatiu) i extreure dades, entre altres característiques avançades. La versió gratuïta permet tenir tres revisions actives.



- **SciSpace** és una plataforma que facilita "trobar, entendre i aprendre" de qualsevol article de recerca. A més, permet descobrir la xarxa d'articles connectats i rellevants per a la teva recerca. Actua com a repositori d'articles de recerca d'accés obert en diversos àmbits. Entre les seves funcions es troben la revisió de la literatura, la capacitat d'obtenir respostes i dades de PDF's introduïts, l'extracció de dades de documents, la generació de citacions en format APA, i l'habilitat de parafrasejar informació de referències bibliogràfiques. Ofereix opcions com Copilot en la modalitat de *Literature Review* (per fer seguiment de la revisió original en diversos idiomes) i *Read with AI* (que permet carregar documents PDF i fa suggeriments per facilitar-ne la comprensió) (Guerra Jáuregui, 2023). Té una versió gratuïta i una de pagament.
- **Semantic Scholar** és un buscador acadèmic que proporciona eines gratuïtes de cerca impulsades per la IA i recursos que es troben en obert. És un projecte sense ànim de lucre de l'Institut Allen per a la IA que es va iniciar l'any 2015. Un cop registrat, es poden guardar articles acadèmics a la mateixa biblioteca i rebre recomanacions d'articles relacionats amb els interessos de recerca. Tot i que no disposa de funcions tan avançades com les eines esmentades anteriorment, segueix sent una eina valuosa per a investigadors que busquen una forma accessible i eficaç de gestionar i descobrir literatura acadèmica.

Hi ha altres plataformes acadèmiques que també utilitzen la IA per assistir al docent/investigador durant el procés de revisió de la literatura acadèmica i desenvolupament de l'estat de l'art. Tanmateix, algunes d'aquestes eines només tenen la versió de pagament, com [Scopus AI](#), [Scite](#) o [Web of Science Assistant Research](#), entre altres. Aquestes plataformes funcionen amb subscripcions individuals o per a equips, i és comú que les universitats o centres de recerca hi estiguin inscrits.

En general, els models de llenguatge com ChatGPT, Gemini, Copilot o Claude no són una eina adequada

per fer una cerca bibliogràfica (Cárdenas, 2023b). Alguns dels models, com Claude, no estan connectats a la xarxa i, per tant, no poden accedir als repositoris digitals. Els models que fan una cerca a la xarxa opten habitualment per cerques molt elementals, que prioritzen els articles en obert, amb molta freqüència fora de JCR o Scopus. Tot i que alguns prompts poden optimitzar la resposta i alguns assistents estan especialitzats en aquest àmbit, en la pràctica els resultats no tenen el rigor que és necessari en aquesta mena de cerques. A banda, no són infreqüents les al·lucinacions que hem vist al capítol 6 i els sistemes poden proposar articles que no existeixen, autors o autores que no han escrit les referències proposades o informació incoherent. Atès que els models estan programats amb un sistema de recompensa, la dificultat d'aquesta mena de cerques porta els models a crear artificialment citacions que realment no existeixen (Tradinos, 2023).

8.2. El graf de relacions

El graf de relacions és una representació visual de com els diferents articles acadèmics o fonts bibliogràfiques estan connectats entre si, principalment a través de les cites que es fan els uns als altres. Aquests grafs analitzen textos per extreure connexions de citació. Els nodes del graf representen típicament articles o documents individuals, mentre que els enllaços (o arcs) entre els nodes representen cites o referències d'un document a un altre. Aquesta visualització permet als investigadors identificar ràpidament els treballs més influents o més citats en un camp específic, detectar subtemes o àrees de recerca relacionades, rastrejar com els conceptes i idees han evolucionat a través de les citacions al llarg del temps, i detectar potencials buits de recerca en veure àrees del graf amb poques connexions.

En aquest apartat, s'enumeren algunes plataformes que utilitzen IA per crear grafs de relacions que representen visualment com els articles acadèmics estan connectats entre si.

- **Connected papers** és una eina visual per ajudar als investigadors i les investigadores a trobar articles rellevants en un camp d'estudi i que utilitza la base de dades Semantic Scholar. Segons la informació de la seva pàgina web, aquesta eina permet crear grafs amb articles similars als que s'introdueix en la plataforma, es poden buscar els articles més recents d'una temàtica, treballs anteriors o derivats d'un camp d'estudi concret, o crear la bibliografia de la investigació. D'aquesta eina hi ha una versió gratuïta que permet fer 2 grafs al mes, una versió gratuïta amb registre que dona opció a poder realitzar 5 grafs al mes. També disposa de diverses opcions de pagament per acadèmics, empreses i grups.
- **Inciteful** disposa de dues eines diferents. Per una banda, l'anomenada "Discover" que crea una xarxa d'articles a través de les citacions i que utilitza algoritmes d'anàlisi de xarxes. Es poden trobar articles semblants, articles rellevants, així com els autors i institucions més prolífiques d'un camp d'estudi. Per altra banda, l'eina "Literature Connector" és dirigida a investigadors interdisciplinaris i permet introduir dos articles per obtenir una visualització interactiva que mostra com estan connectats en la literatura. Per tant, és una eina basada en xarxes bibliomètriques (Cárdenas, 2023b). Segons la pàgina web d'aquesta plataforma, les dades i la inspiració s'han extret de projectes com OpenAlex, Semantic Scholar, Crossref i OpenCitations. Així, la plataforma permet trobar la literatura més recent o connectar dues idees dins d'una recerca. És una eina gratuïta.

- **Litmaps** permet buscar articles acadèmics a través de la xarxa de citacions, visualitzar-los mitjançant un graf de relacions i configurar sistemes d'alertes automàtiques per estar al dia de les últimes recerques publicades. A més, es poden filtrar els resultats per data i paraules clau detectades automàticament, així com compartir la mateixa recerca amb altres investigadors. Igual que altres plataformes, té una part gratuïta i una altra de pagament (*freemium*).
- **Research Rabbit** es tracta d'una plataforma per buscar articles de revistes acadèmiques sobre una temàtica o autor concret. També, permet fer un mapatge de la literatura a través de les citacions. D'aquesta manera, es poden buscar articles relacionats amb els que ja s'han guardat, articles citats o articles que hagin citat els articles seleccionats. A més de poder-los emmagatzemar, també elabora un graf de relacions entre els diversos articles buscats i els seus autors. D'una manera visual, a través un mapping, es pot explorar la literatura existent d'una temàtica determinada amb la creació de diferents col·leccions. Aquesta plataforma actualment es pot connectar amb el gestor bibliogràfic Zotero. És totalment gratuïta i només cal registrar-s'hi.

Es poden mencionar moltes altres plataformes que tenen funcionalitats molt similars a les que s'han anat explicant al llarg d'aquest capítol, com per exemple [VOSviewer](#), [CiteSpace](#), [CitNetExplorer](#), [Leximancer](#), [Sci2 Tool](#). A mesura que la IA es va desenvolupant, el nombre de noves eines amb noves funcionalitats va incrementant constantment. Això fa que n'apareguin de noves, mentre que d'altres desapareixen per obsolescència. La majoria d'aquestes eines llancen noves versions any rere any, millorant les opcions de visualització, gestió, importació de dades bibliogràfiques, i creant nous algoritmes de clusterització temàtica.

Pràctica. Graf sobre un tema per l'elaboració d'un treball de recerca

L'objectiu d'aquest exercici és ajudar els estudiants a visualitzar i comprendre millor el panorama general d'un camp acadèmic durant una revisió de la literatura.

Aquesta activitat es pot realitzar en assignatures enfocades a l'elaboració d'una recerca, com ara un Treball de Síntesi, un Treball de Fi de Grau, un Treball de Fi de Màster, o l'elaboració d'un article acadèmic, entre altres.

Per realitzar aquest exercici cada estudiant haurà d'escollir un tema de recerca que sigui del seu interès, preferiblement vinculat al seu àmbit de recerca. Seguidament, cada estudiant haurà d'accedir a alguna de les plataformes esmentades. És probable que l'eina requereixi registrar-se a la plataforma. Un cop dins la plataforma, els estudiants poden realitzar un graf de relacions a partir de cerques diverses i/o diferents tipus de visualitzacions (per exemple, a partir de paraules clau, citacions d'un article concret, o d'un autor o autora, etc.). Finalment, es proposa a cada estudiant analitzar els resultats del graf, identificant els articles i autors més influents dins el graf, detectant temes i característiques dels articles més rellevants.

Aquest exercici també es pot realitzar amb la selecció per part del docent del tema que els estudiants i les estudiants han de cercar per generar el graf i fer les corresponents interpretacions.

Aquesta proposta d'exercici docent està dissenyada per integrar l'ús de les eines d'anàlisi bibliogràfica en l'aprenentatge acadèmic, fomentant així les habilitats analítiques i de recerca dels estudiants.

Pràctica. Anàlisi d'un article acadèmic

L'objectiu d'aquest exercici és ajudar els estudiants a visualitzar i comprendre millor el contingut i les idees principals d'un article acadèmic utilitzant la plataforma Humata, o qualsevol altra plataforma que permeti interactuar amb documents.

Aquest exercici pot ser adequat per assignatures vinculades a la recerca, com un Treball de Síntesi, un Treball de Fi de Grau, un Treball de Fi de Màster, etc.

Per realitzar aquest exercici, cada estudiant haurà d'escollir un article acadèmic que sigui del seu interès, preferiblement vinculat al seu àmbit de recerca; alternativament, el docent pot triar l'article que consideri oportú. Seguidament, cada estudiant haurà d'accedir a la plataforma Humata. Un cop dins la plataforma, els estudiants hauran de carregar el PDF de l'article seleccionat.

Cada estudiant haurà de fer una anàlisi de l'article tot interactuant amb Humata, per tal de poder identificar les parts més rellevants de l'article, autors rellevants citats, temes principals discutits, etc. Cada estudiant haurà d'elaborar un breu document amb els resultats de l'anàlisi feta.

Aquesta activitat permet a l'estudiant aprendre a interactuar amb una eina d'IA amb el propòsit concret de sintetitzar la informació rellevant d'un text acadèmic.

Referències

Bhosale, Uttkarsha (2022). How to Master at Literature Mapping: 5 Most Recommended Tools to Use. Enago Academy. <https://www.enago.com/academy/literature-mapping-5-most-recommended-tools-to-use/> Consulta 10/6/2024

Cárdenas, Julián (2023a, febrer). Cómo hacer una revisión de literatura con una herramienta de IA y redes bibliométricas. Networkianos. Blog de Sociología. Disponible a <https://networkianos.com/mas-herramientas-de-ia-para-la-educacion-e-investigacion/> Consulta 10/6/2024

Cárdenas, Julián (2023b, març). Cómo hacer una revisión de literatura con una herramienta de IA y redes bibliométricas. Networkianos. Blog de Sociología. Disponible a <https://networkianos.com/como-hacer-una-revision-de-literatura-con-una-herramienta-de-ia-y-redes/> Consulta 10/6/2024

Codina, Lluís (2023, desembre). Inteligencia artificial para usos académicos: análisis y guía de uso de Elicit AI. Disponible a <https://www.lluiscodina.com/inteligencia-artificial-usos-academicos/> Consulta 12/6/2024

Eco, U. (2001). *Cómo se hace una tesis: técnicas y procedimientos de investigación, estudio y escritura* (Vol. 6). Barcelona: Gedisa.

Enago Academy (n.d.). ¡5 herramientas de mapeo de literatura para científicos! <https://www.enago.com/es/academy/5-ways-avoid-ghostwriting-busy-researchers/> Consulta 19/06/2024

Guerra Jáuregui, Melissa (2023). Las mejores IA para el investigador moderno. Observatorio Instituto para el Futuro de la Educación del Tecnológico de Monterrey. Disponible a <https://observatorio.tec.mx/edu-news/las-mejores-ia-para-el-investigador-moderno/> Consulta 13/06/2024

Tradiños (2023). ¿Por qué ChatGPT genera referencias falsas? Disponible a <https://ChatGPT.com/c/ce3134c0-fd8c-40ee-8c31-d88e08bb2bb9> Consulta 19/06/2024

09

Calculant amb una IA: Quantitativa artificial

Les dades són presents arreu. Ara podem saber quants turistes anglesos han visitat Alacant el mes de febrer, quantes hores s'ha superat el límit de partícules PM₁₀ recomanat per l'OMS a Madrid durant una setmana, quantes persones hi ha a Barcelona al llarg del dia, quants vaixells estan travessant ara el Canal de Suez, quin és el temps mitjà que necessitem per arribar a una capital comarcal en transport públic des de cada municipi de la comarca o quina és la nota mitjana en matemàtiques a les proves ebau d'aquest any. Vivim en un univers de dades, moltes d'elles en obert, gràcies al senyal mòbil, als sensors, als enginyers de càlcul o als registres automàtics. Mai com ara hem tingut més dades a la mà. Per això, la ciència de les dades s'ha convertit en una disciplina fonamental, ja que permet transformar aquests registres en informació útil. Les empreses, institucions públiques i societat en el seu conjunt s'han convertit en grans productores de dades que requereixen sistemes d'anàlisi que permetin convertir-les en coneixement (Mannyica et al., 2011). Aquestes dades són les que es coneixen amb el nom de "big data". La seva principal característica

és que es tracta de conjunts de dades que superen la capacitat de les eines típiques de programari de bases de dades per poder-les capturar, emmagatzemar, gestionar i analitzar.

De fet, les dades i la informació per si mateixes no tenen valor, independentment de la seva mida, si no permeten generar coneixement i són útils per la presa de decisions. Els àmbits d'aplicació són molt diversos i inclou tant l'àmbit acadèmic com els negocis i les organitzacions del món sencer (Mora, 2019). La ciència de les dades és una disciplina multidisciplinària que combina coneixements de matemàtiques, estadística, informàtica i el domini per extreure informació significativa de les dades. Aquest procés implica la recopilació, neteja, anàlisi i interpretació de dades per prendre decisions informades. Alguns dels camps d'aplicació són els següents:

- **Anàlisi de dades:** Utilització de tècniques estadístiques i algorismes d'aprenentatge automàtic per identificar patrons i tendències en les dades.

“Al llarg d'aquest capítol ens centrarem en els seus potencials i les aplicacions de la IA per poder desenvolupar anàlisis de dades”

- **Aprenentatge automàtic:** Desenvolupament de models predictius i classificadors a partir de dades històriques per fer prediccions sobre nous casos.
- **Processament del llenguatge natural:** Extracció d'informació significativa del text, com ara anàlisi de sentiments, traducció automàtica i resum d'articles.

Al llarg d'aquest capítol ens centrarem en els seus potencials i les aplicacions de la IA per poder desenvolupar anàlisis de dades. Des dels conceptes més elementals fins als més avançats, mostrarem les aplicacions de la intel·ligència artificial en els fulls de càlcul, continuarem amb una exploració de les possibilitats de la IA en l'estadística avançada i finalment presentarem algunes possibilitats del machine learning.

9.1. Dins dels fulls de càlcul

Segurament els fulls de càlcul (Microsoft Excel, Numbers, Google Sheets) són una de les eines més populars i versàtils, que permeten operacions més o menys sofisticades en dades organitzades en files i columnes. Les seves aplicacions i usos són molt variats i presenten diversos nivells de complexitat. Alguns fulls són simples registres d'informació ordenada, però aquestes eines permeten processos més complexos per les valoracions, els processos operatius, la presa de decisions, els informes financers o l'anàlisi de dades (O'Beirne, 2023). La popularitat

d'aquestes eines ha generat un excés de confiança, que O'Beirne anomena la síndrome de l'aprenent de bruixot, qui té suficients coneixements com fer un encanteri, però no els que es precisen quan queda fora de control.

La IA facilita la utilització de fulls de càlcul, perquè les eines de generació com Copilot, ChaGPT, Gemini o Claude poden suggerir les instruccions precises per aconseguir un determinat propòsit, especialment quan ja disposem de les dades. També es pot utilitzar l'analitzador de dades de l'Excel com un assistent propi dins del software. Finalment, es poden emprar els assistents (ChaGPT, Copilot) dins de l'Excel o el Googlesheet. Mostrarem exemples d'aquestes tres possibilitats.

L'ús de la IA com a guia i execució d'anàlisi

Les eines d'intel·ligència artificial poden ser extremadament útils per poder generar idees i donar suport en el procés d'anàlisi de dades. De fet, com ja s'ha vist en altres capítols, eines com ChatGPT i Copilot són molt útils per respondre a dubtes o qüestions sobre situacions o aspectes concrets. Imaginem, per exemple, que disposem de la següent base de dades en un full d'Excel, amb les característiques físiques de 20 individus (edat, altura, sexe, pes i nivell d'activitat física).

CAS	EDAT	Altura (cm)	Sexe	Pes (kg)	Nivell d'activitat física
1	25	170	H	70	Moderat
2	32	165	D	60	Baix
3	28	180	H	80	Alt
4	45	160	D	55	Moderat
5	22	175	H	75	Baix
6	30	168	D	65	Alt
7	35	172	H	77	Moderat
8	40	158	D	58	Baix
9	27	181	H	82	Alt
10	26	164	D	62	Moderat
11	38	162	H	61	Baix
12	32	174	D	73	Alt
13	24	160	H	54	Moderat
14	31	179	D	81	Baix
15	29	164	H	69	Alt
16	17	169	D	52	Moderat
17	23	165	H	66	Baix
18	27	180	D	79	Alt
19	20	177	H	75	Moderat
20	28	163	D	57	Baix

“Les eines d’intel·ligència artificial poden ser extremadament útils per poder generar idees i donar suport en el procés d’anàlisi de dades”

La IA ens pot ajudar a calcular la mitjana de les edats, per exemple amb el següent *prompt*:

“Vull fer la mitjana de l’edat. Indica’m la fórmula que haig d’escriure”

La seva resposta seria la següent:



Per calcular la mitjana de l'edat en Excel en anglès, necessites utilitzar la funció `AVERAGE`. Els edats es troben en la columna B des de la cella B2 fins a la cella B21, la fórmula que has d'escriure és:

```
excel Copia e  
=AVERAGE(B2:B21)
```

Això calcularà la mitjana de les edats de les cel·les B2 a B21. Aquí tens els passos detallats:

1. Selecciona la cella on vols que aparegui el resultat de la mitjana (per exemple, la cella E).
2. Escriu la fórmula `=AVERAGE(B2:B21)`.
3. Prem Enter.

Excel calcularà i mostrarà la mitjana de les edats en la cella seleccionada.

L'aplicació de les seves instruccions donen el resultat següent:

CAS	EDAT	Altura (cm)	Sexe	Pes (kg)	Nivell d'activitat física
1	25	170	H	70	Moderat
2	32	165	D	60	Baix
3	28	180	H	80	Alt
4	45	160	D	55	Moderat
5	22	175	H	75	Baix
6	30	168	D	65	Alt
7	35	172	H	77	Moderat
8	40	158	D	58	Baix
9	27	181	H	82	Alt
10	26	164	D	62	Moderat
11	33	177	H	78	Baix
12	38	162	D	61	Alt
13	24	174	H	73	Moderat
14	29	166	D	59	Baix
15	31	179	H	81	Alt
16	42	155	D	56	Moderat
17	23	169	H	74	Baix
18	37	167	D	63	Alt
19	34	176	H	79	Moderat
20	28	163	D	57	Baix
	31,45				

Podríem demanar una segona instrucció, com ara el nombre d'homes i de dones que hi ha a la mostra. El *prompt* podria ser:

"Vull saber com es distribueix la variable sexe. Indica'm la fórmula. L'Excel està en versió anglesa."

El resultat d'aquesta consulta és el següent:

Comptar el nombre de dones ("D")

excel

Copia el codi

```
=COUNTIF(D2:D21;"D")
```

Comptar el nombre d'homes ("H")

excel

Copia el codi

```
=COUNTIF(D2:D21;"H")
```

Així, si les dades de sexe estan a la columna D de la cel·la D2 a D21:

1. Nombre de dones: `=COUNTIF(D2:D21;"D")`
2. Nombre d'homes: `=COUNTIF(D2:D21;"H")`

El sistema ha interpretat que H és home i que D és dona. En tot cas, es pot precisar més el *prompt* en

el cas que la resposta no hagi estat satisfactòria. El resultat de l'operació seria el següent:

CAS	EDAT	Altura (cm)	Sexe	Pes (kg)	Nivell d'activitat física
1	25	170	H	70	Moderat
2	32	165	D	60	Baix
3	28	180	H	80	Alt
4	45	160	D	55	Moderat
5	22	175	H	75	Baix
6	30	168	D	65	Alt
7	35	172	H	77	Moderat
8	40	158	D	58	Baix
9	27	181	H	82	Alt
10	26	164	D	62	Moderat
11	33	177	H	78	Baix
12	38	162	D	61	Alt
13	24	174	H	73	Moderat
14	29	166	D	59	Baix
15	31	179	H	81	Alt
16	42	155	D	56	Moderat
17	23	169	H	74	Baix
18	37	167	D	63	Alt
19	34	176	H	79	Moderat
20	28	163	D	57	Baix
		Dones	10		
		Homes	10		

Un cop ja hem entès el funcionament bàsic, podem incorporar més complexitat a les nostres comandes, com ara realitzar un càlcul d'una nova variable a partir de la informació existent. Per exemple li podem demanar que ens digui com calcular l'Índex de

Massa Corporal (IMC), i ho podem fer amb el següent *prompt* "Vull calcular l'índex de massa corporal a partir d'aquestes dades. Indica'm com ho haig de fer per calcular-ho."



Per calcular l'Índex de Massa Corporal (IMC) a partir de les dades de pes i alçada, pots utilitzar la següent fórmula:

$$IMC = \frac{pes}{(alçada)^2}$$

On el pes es mesura en quilograms (kg) i l'alçada en metres (m). Aquí tens els passos per calcular l'IMC a Excel:

- Crea una nova columna per a l'IMC:** Afegeix una nova columna a la teva fulla de càlcul on calcularàs l'IMC. Suposem que vols posar l'IMC a la columna F, comença a la cel·la F2.
- Aplica la fórmula de l'IMC:** A la primera cel·la de la nova columna (per exemple, F2), escriu la fórmula següent i prem Enter:

```
excel Copia el codi
=E2 / (C2/100)^2
```

Aquesta fórmula suposa que el pes està a la columna E i l'alçada a la columna C (i que l'alçada es troba en centímetres, per això es divideix per 100 per convertir-la a metres).

- Arrossega la fórmula:** Arrossega la cel·la amb la fórmula cap avall per aplicar-la a totes les cel·les de la columna F amb dades.

Això calcularà l'IMC per a cada persona a partir de les seves dades de pes i alçada. Si necessites alguna altra aclariment o ajuda, no dubtis a preguntar!

Si apliquem les seves instruccions obtindrem aquest resultat en el full de càlcul:

CAS	EDAT	Altura (cm)	Sexe	Pes (kg)	Nivell d'activitat física	IMC
1	25	170	H	70	Moderat	24,22
2	32	165	D	60	Baix	22,04
3	28	180	H	80	Alt	24,69
4	45	160	D	55	Moderat	21,48
5	22	175	H	75	Baix	24,49
6	30	168	D	65	Alt	23,03
7	35	172	H	77	Moderat	26,03
8	40	158	D	58	Baix	23,23
9	27	181	H	82	Alt	25,03
10	26	164	D	62	Moderat	23,05
11	33	177	H	78	Baix	24,90
12	38	162	D	61	Alt	23,24
13	24	174	H	73	Moderat	24,11
14	29	166	D	59	Baix	21,41
15	31	179	H	81	Alt	25,28
16	42	155	D	56	Moderat	23,31
17	23	169	H	74	Baix	25,91
18	37	167	D	63	Alt	22,59
19	34	176	H	79	Moderat	25,50
20	28	163	D	57	Baix	21,45

Els generadors com ChatGPT són prou eficients per fer una anàlisi de les dades de forma automàtica, com ara una anàlisi estadística descriptiva, una anàlisi de la relació entre les variables, la representació gràfica de la informació o anàlisis més elaborades. En l'exemple anterior, podem proposar el següent *prompt*: "Fes l'anàlisi estadística descriptiva d'aquest full de càlcul. Recorda que cal que adjuntis l'arxiu

amb els resultats d'aquesta anàlisi". Aquest *prompt* mostra una utilitat molt versàtil de ChatGPT, com és la possibilitat que aportí fitxers en formats diversos, que es poden descarregar; aquests arxius són temporals i no es poden descarregar després d'un cert temps. El *prompt* genera, per tant, una taula de resultats descarregable i una lectura bàsica dels resultats.

	EDAT	alçada	pes	IMC
count	20	20	20	20
mean	31,45	169,05	68,25	23,75
std	6,50	7,56	9,52	1,48
min	22	155	55	21,41
25 %	26,75	163,75	59,75	22,92
50 %	30,5	168,5	67,5	23,71
75 %	35,5	175,25	77,25	24,93

1. Estadístiques Generals:

- Edat: La mitjana és de 31,45 anys, amb un rang de 22 a 45 anys.
- Altura: La mitjana és de 169,05 cm, amb un rang de 155 a 181 cm.
- Pes: La mitjana és de 68,25 kg, amb un rang de 55 a 82 kg.
- IMC: La mitjana és de 23,75, amb un rang de 21,41 a 26,03.

2. Diferències per Sexe:

- Sexe D (Dona):
 - Edat: La mitjana és de 34,7 anys.
 - Altura: La mitjana és de 162,3 cm.
 - Pes: La mitjana és de 61,5 kg.
 - IMC: La mitjana és de 22,48.
- Sexe H (Home):
 - Edat: La mitjana és de 28,2 anys.
 - Altura: La mitjana és de 175,8 cm.
 - Pes: La mitjana és de 75 kg.
 - IMC: La mitjana és de 25,02.

Aquestes dades mostren diferències significatives entre homes i dones en termes d'altura, pes i IMC. Els homes tendeixen a ser més alts i pesats, amb un IMC més alt en comparació amb les dones.

Si necessites una anàlisi més detallada o visualitzacions específiques, si us plau, fes-m'ho saber!

Podem constatar que l'anàlisi dels resultats proporciona la mitjana i el rang de tots els valors. A banda, ha fet una diferenciació entre les mitjanes de l'atribut sexe i ha constatat les diferències dels valors entre homes i dones.

No només treballa amb estadística descriptiva, sinó

que pot incorporar la relació entre variables. Per exemple, pel cas anterior podem suggerir el següent *prompt*: "Calcula les correlacions entre variables quan sigui possible". Novament, genera una taula que ens podem descarregar amb la matriu de totes les correlacions i una interpretació dels resultats.

Edat

- Té una correlació moderada amb l'altura (-0,558) i el pes (-0,485)
- Té una correlació negativa més feble amb l'IMC (-0,318)

Altura (cm)

- Té una correlació positiva molt forta amb el pes (0,946)
- Té una correlació positiva moderada amb l'IMC (0,684)

Pes

- Té una correlació positiva forta amb el pes (0,946)
- Té una correlació forta positiva amb l'IMC (0,882)

IMC

- Té una correlació positiva forta amb el pes (0,882)
- Té una correlació positiva moderada amb l'altura (0,684)

Aquestes correlacions suggereixen que l'altura i el pes estan fortament relacionats, igual que el pes i l'IMC. L'edat per contra té una correlació negativa amb l'altura, el pes i l'IMC, tot i que aquestes correlacions són menys fortes.

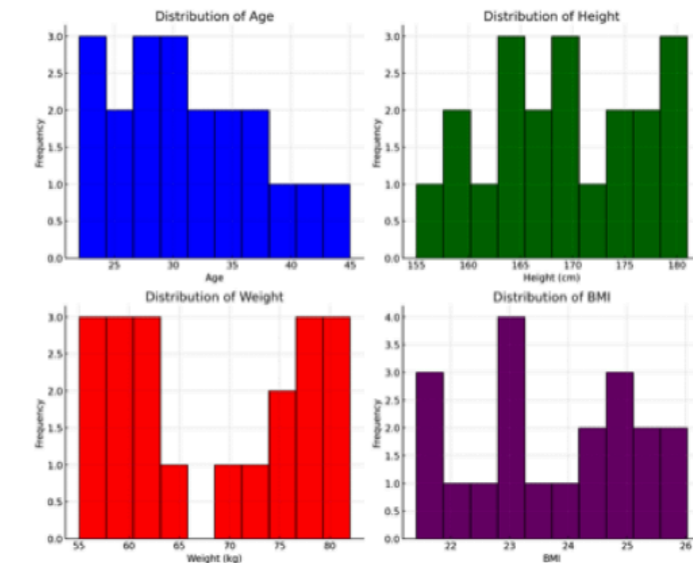
Si necessites més detall o una ajuda addicional, no dubtis a demanar-ho!

El ChatGPT també permet fer representacions gràfiques. Per exemple, podem suggerir el següent *prompt*: "Fes la representació gràfica de les dades que t'he proporcionat". Veurem que sense cap altra

indicació, el programa realitza un histograma per cada variable estudiada i després proporciona els diagrames de punts que permeten visualitzar la relació entre variables.



Distribution Of BMI

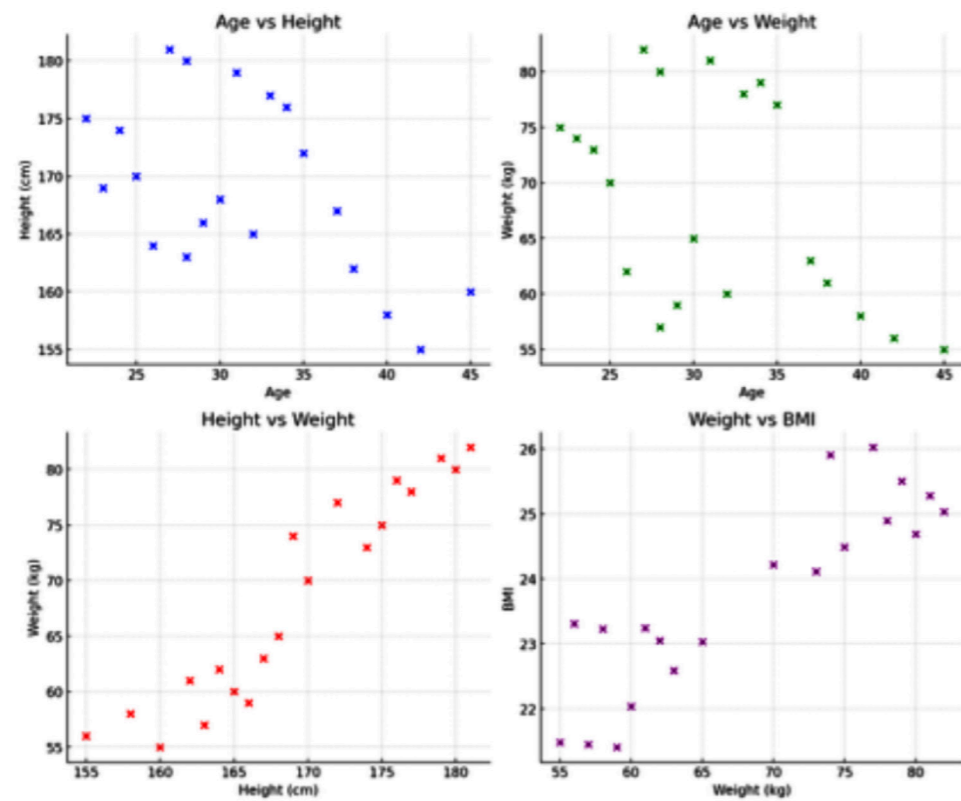


Aquí tens els histogrames que il·lustren la distribució de les principals variables en la base de dades:

1. **Distribució de l'Edat:** La majoria de les persones tenen entre 25 i 35 anys, amb una dispersió que arriba fins als 45 anys.
2. **Distribució de l'Altura:** L'altura varia principalment entre 160 i 180 cm.
3. **Distribució del Pes:** El pes es concentra principalment entre 55 i 80 kg.
4. **Distribució de l'IMC:** L'IMC varia entre 21 i 26, amb una concentració majoritària al voltant de 23-24.

Ara, crearé altres gràfiques per mostrar les relacions entre variables, especialment aquelles que tenen correlacions significatives. [-]

Weight Vs BMI



L'ús d'assistents a Excel

Microsoft ha introduït a Excel una eina anomenada "Analyze data", que permet obtenir informació sobre les dades a partir del llenguatge natural, de manera que es poden fer preguntes, demanar informes o suggerir proves sense fórmules complicades. A més, aquesta eina proporciona resums visuals, tendències

i patrons (Microsoft, 2024). Hem introduït les dades del cas anterior a l'Excel i hem clicat l'opció *Analyze data* que es troba a l'esquerra; activa un menú que permet interactuar amb el full de càlcul i accedir a informació gràfica generada automàticament, tal com es veu en la imatge.

The screenshot shows the Microsoft Excel interface with the following data table:

Cas	Edat	Altura (cm)	Sexe	Pes (kg)	Nivell d'activitat física	IMC
1	25	170	H	70	Moderat	24,221
2	32	165	D	60	Baix	22,039
3	28	180	H	80	Alt	24,691
4	45	160	D	55	Moderat	21,484
5	22	175	H	75	Baix	24,490
6	30	168	D	65	Alt	23,030
7	35	172	H	77	Moderat	26,028
8	40	158	D	58	Baix	23,233
9	27	181	H	82	Alt	25,030
10	26	164	D	62	Moderat	23,052
11	33	177	H	78	Baix	24,897
12	38	162	D	61	Alt	23,243
13	24	174	H	73	Moderat	24,112
14	29	166	D	59	Baix	21,411
15	31	179	H	81	Alt	25,280
16	42	155	D	56	Moderat	23,309
17	23	169	H	74	Baix	25,909
18	37	167	D	63	Alt	22,590
19	34	176	H	79	Moderat	25,504
20	28	163	D	57	Baix	21,454

The 'Analyze Data' sidebar on the right shows suggested questions and a pivot table:

Suggested questions:

- Percentage of total 'Edat' for each 'Sexe'
- Insights for 'IMC'

Discover insights:

'Edat' by 'Sexe' and 'Nivell d'activitat física'

Sexe	Moderat	Baix	Alt	Grand Total
D	113	129	105	347
H	118	78	86	282
Grand Total	231	207	191	629

The sidebar also includes a chart for 'Altura (cm)' by 'Edat' and an 'Insert Chart' button.

En els resultats generats de forma automàtica, podem veure una taula que diferencia els resultats per sexe i un núvol de punts que relaciona l'edat amb l'altura. Com en el cas del ChatGPT o Copilot, aquesta eina també permet fer preguntes amb llenguatge natural. Tanmateix, el seu funcionament és força diferent en l'àmbit conceptual i hem de tenir en compte algunes premisses. Per exemple, la formulació de *prompts* d'aquesta eina ha de ser molt sintètica i s'ha d'assenyalar el nom de la variable o variables que volem tractar entre cometes simples '. També cal tenir en compte que aquesta eina no

pot seguir una "conversa" com hem vist que es pot fer amb el ChatGPT o similars, de manera que cada pregunta ha de ser independent. El que sí que ens ofereix és una llista de les consultes o preguntes recents que li hem fet.

En les imatges es pot veure els *prompts* necessaris per a calcular la mitjana del pes dels individus (*Average of 'Pes (kg)'*) o la mitjana de l'IMC per cada sexe (*Average of total 'IMC' by 'Sexe'*). El sistema també ofereix suggeriments i formula *prompts* de manera automàtica.

Analyze Data

← Average of 'Pes (kg)'

Average of 'Pes (kg)'

Here's our answer
Showing average 'Pes (kg)'.

'Pes (kg)'
Average of Pes (kg)
68.25

+ Insert PivotTable Is this helpful?

Analyze Data

← Average of total 'IMC' by 'Sexe'

Average of total 'IMC' by 'Sexe'

Here's our answer
Showing total and average 'IMC' by 'Sexe'.


Sexe	Average of IMC	Sum of IMC
D	22,48	224,84
H	25,02	250,16
Grand Total	23,75	475,01

+ Insert PivotTable Is this helpful?

D'altra banda, és també possible instal·lar un assistent de ChatGPT o Copilot tant a l'Excel com a GoogleSheets. El seu funcionament és força semblant amb el que hem vist de l'"Analyze data" integrat dins del mateix Excel, però molt més versàtil, ja que permet l'ús del llenguatge natural d'una forma més semblant a la que ho faríem directament dins l'entorn de ChatGPT o Copilot.

9.2. Estadística avançada

Una de les funcionalitats de ChatGPT és la creació d'assistents o GPT amb funcionalitats específiques. En un capítol posterior, presentarem les possibilitats d'aquesta eina. Entre aquests assistents, n'hi ha d'especialitzats en anàlisi estadística avançada. Per tal de mostrar les seves possibilitats, emprarem dos dels assistents amb major capacitat, com són Data Analysis & Report AI i Data Analysis. SPSS. Atès que els assistents es modifiquen amb molta freqüència i s'actualitzen amb versions més avançades, podria ser que en el moment de llegir el document ja hagin aparegut altres GPT amb noves capacitats. ChatGPT disposa d'un cercador que permet trobar fàcilment els assistents especialitzats en un determinat àmbit.



Data Analysis & Report AI

De: Kenneth Bastian &

Your expert in limitless, detailed scientific data analysis and reporting.

Can you analyze this dataset for patterns?

What trends can you identify in this data?

Please generate a report based on this data.

Can you create visualizations for this dataset?



Data Analysis - SPSS

De: Meddidin ebuseerfo &

Expert in statistical analysis and SPSS, guiding researchers.

How do I test the reliability of my data in SPSS?

What statistical test should I use for my hypothesis?

Can you explain how to define variables in SPSS?

I need help interpreting SPSS output for my data.

Data Analysis & Report AI ofereix la informació en llenguatge Python, a diferència dels models que hem vist anteriorment, que ens proporcionen la possibilitat de descarregar les taules de resultats o els gràfics. Aquest assistent és especialment útil per anàlisis complexes, que pot il·lustrar amb exemples concrets. Podem utilitzar el següent *prompt* a partir

de les dades que estem utilitzant al llarg d'aquest capítol: "Quin tipus d'anàlisis no paramètrics podem fer amb aquestes dades?". El resultat mostra tant un llistat de les proves més adequades que es podrien realitzar com el codi python descarregable que ens permetria executar automàticament l'opció seleccionada.

Quin tipus d'anàlisi no paramètrics podem fer en aquests dades?



L'anàlisi no paramètrica és útil quan les dades no segueixen una distribució normal o quan tenim mostres petites. En el context d'aquestes dades, podríem considerar diverses proves no paramètriques per comparar grups i explorar relacions:

1. Prova de Mann-Whitney U

Aquesta prova és utilitzada per comparar dues mostres independents. Podria ser útil per comparar les diferències entre homes i dones (per exemple, altura, pes o IMC).

2. Prova de Kruskal-Wallis

Aquesta prova és una extensió de la prova de Mann-Whitney U per a més de dos grups. Pot ser utilitzada per comparar les diferències en el pes, altura o IMC entre els diferents nivells d'activitat física (Baix, Moderat, Alt).

3. Prova de Wilcoxon

Aquesta prova és utilitzada per comparar dues mostres relacionades. Pot ser aplicable si tenim dades abans i després d'una intervenció per les mateixes persones (tot i que aquest no sembla ser el cas amb les dades actuals).

4. Correlació de Spearman

Aquesta prova és utilitzada per avaluar la força i la direcció de l'associació entre dues variables contínues o ordinal. Podria ser utilitzada per explorar la relació entre edat i IMC, altura i pes, etc.

Exemples Concrets

1. Prova de Mann-Whitney U: Comparació del Pes entre Homes i Dones

```
python Copia el codi  
  
from scipy.stats import mannwhitneyu  
  
# Dividir les dades per sexe  
homes = df[df['Sexe'] == 'H']['Pes (kg)']  
dones = df[df['Sexe'] == 'D']['Pes (kg)']  
  
# Realitzar la prova  
stat, p = mannwhitneyu(homes, dones)  
stat, p
```

2. Prova de Kruskal-Wallis: Comparació del IMC segons Nivell d'Activitat Física

```
python Copia el codi  
  
from scipy.stats import kruskal  
  
# Dividir les dades per nivell d'activitat física  
baix = df[df['Nivell d\'activitat física'] == 'Baix']['IMC']  
moderat = df[df['Nivell d\'activitat física'] == 'Moderat']['IMC']  
alt = df[df['Nivell d\'activitat física'] == 'Alt']['IMC']  
  
# Realitzar la prova  
stat, p = kruskal(baix, moderat, alt)  
stat, p
```

3. Correlació de Spearman: Relació entre Edat i IMC

```
python Copia el codi  
  
from scipy.stats import spearmanr  
  
# Realitzar la prova de correlació  
corr, p = spearmanr(df['Edat'], df['IMC'])  
corr, p
```

Aquestes són algunes de les anàlisis no paramètriques que podrien ser útils per explorar les dades. Si vols que realitzi alguna d'aquestes anàlisis o si tens alguna altra preferència, si us plau fes-m'ho saber!

El Data Analysis – SPSS és un assistent encara més especialitzat, ja que ens orienta per poder fer l'anàlisi de dades a través del programa SPSS. Si li plantegem el mateix problema que en el cas anterior, i per tant li introduïm el mateix *prompt*: *Quin tipus d'anàlisi no paramètrics podem fer amb aquests dades?*, el resultat és diferent. D'una banda, ens suggereix altres tests d'anàlisi no paramètrics que ens podrien ser útils per l'anàlisi de dades, de manera que els

dos assistents es complementen; d'altra banda, ens ofereix la informació necessària per a poder desenvolupar l'anàlisi estadística corresponent mitjançant l'SPSS. Des d'un punt de vista docent, això permet una major complementarietat amb l'activitat de la professora o el professor, ja sigui amb un model de classe invertida (o *flipped classroom*) o bé com a suport per a la realització d'exercicis complexos.

Com realitzar aquestes anàlisis en SPSS:

Prova de Mann-Whitney U:

1. Ves a `Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples``.
2. Mou la variable dependent al quadre `Test Variable List`` i la variable independent (grups) al quadre `Grouping Variable``.
3. Defineix els grups fent clic a `Define Groups``.
4. Selecciona `Mann-Whitney U`` i clica `OK``.


Prova de Wilcoxon:

1. Ves a `Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples``.
2. Mou les dues variables a comparar al quadre `Test Pairs``.
3. Selecciona `Wilcoxon`` i clica `OK``.

Prova de Kruskal-Wallis:

1. Ves a `Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples``.
2. Mou la variable dependent al quadre `Test Variable List`` i la variable independent (grups) al quadre `Grouping Variable``.
3. Defineix els grups fent clic a `Define Range``.
4. Selecciona `Kruskal-Wallis`` i clica `OK``.

Prova de Friedman:

1. Ves a `Analyze > Nonparametric Tests > Legacy Dialogs > K Related Samples``.
2. Mou les variables a comparar al quadre `Test Variables``.
3. Selecciona `Friedman`` i clica `OK``. 

9.3. Màquines que aprenen: Machine learning (ML)

El *machine learning* (aprenentatge automatitzat) és una branca de la intel·ligència artificial, que se sosté sobre la premissa que els ordinadors poden aprendre a partir dels patrons i l'experiència, tal o com fem els humans. Aquest aprenentatge els permet prendre decisions basades en l'experiència adquirida prèviament (Zhou, 2021). Així doncs, el ML és la tècnica que millora el rendiment dels sistemes a través de l'aprenentatge automàtic mitjançant mètodes computacionals (Zhou, 2021). A partir del desenvolupament de models estadístics i algorismes, els ordinadors poden aprendre de les dades que se'ls proporciona i fer prediccions o prendre decisions amb una certa autonomia (Rodríguez, 2023).

Durant els darrers anys, l'aprenentatge automatitzat ha experimentat un avenç significatiu gràcies a l'augment de la capacitat de processament, la disponibilitat massiva de dades i el progrés en el camp de la intel·ligència artificial. Aquest fet ha implicat que el ML esdevingui una de les tècniques més important en algunes àrees aplicades com la visualització per ordinador o el processament del llenguatge natural (Zhou, 2021), i jugui un paper cada cop més important en el suport de la investigació científica i és crucial per al desenvolupament tecnològic (Mjolsness i DeCoste, 2001).

Actualment, són molt els softwares especialitzats que inclouen algorismes i tècniques d'anàlisi de dades basades en el ML. De fet, gran part dels principals programes d'anàlisi de dades disposen d'eines o mòduls de ML que faciliten l'explotació de dades complexes. Per exemple, el software lliure JASP inclou un mòdul de ML amb tècniques de regressió, classificació, clusterització i predicció. Trobem també softwares com MAXQDA que permeten fer una anàlisi de les dades textuals a través del processament del llenguatge natural.

En aquest capítol ens centrarem principalment a mostrar dues aplicacions pràctiques del ML. En

primer lloc, una anàlisi clúster com a model de cerca de patrons a les dades; i, en segon lloc, una anàlisi basada en la mineria de dades com una aplicació pràctica del processament del llenguatge natural.

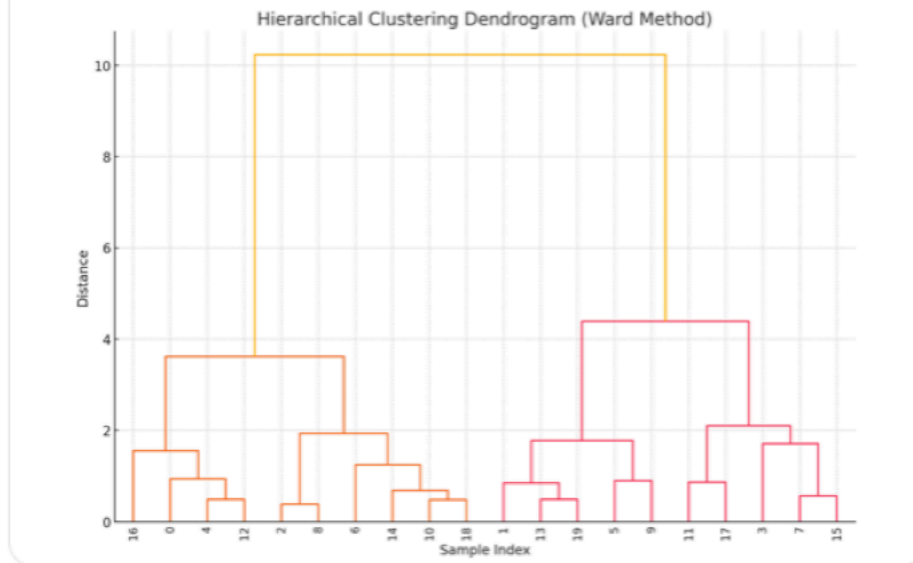
Cerca de patrons: L'anàlisi clúster.

L'anàlisi clúster és una tècnica estadística que té com a objectiu principal trobar patrons o relacions ocultes en les dades, agrupant els elements que comparteixen característiques similars. Les dades s'agrupen en un nombre reduït de clústers o grups (Hair et al., 1998) que són mutualment exclusius (Cea, 2004). De forma molt resumida, quan s'aplica una anàlisi de clústers cal prendre diverses decisions metodològiques. El més habitual és la realització d'un mètode jeràrquic. Hi ha diversos algorismes de classificació (inter-grups, intra-grups, ward method...) tots ells tenen unes particularitats diferents i prioritzen certs encadenaments o altres dins de les dades. Per tant, prioritzen característiques diferents en els patrons de les dades. A més, quan es disposa de variables amb dades que responen a mesures diferents cal procedir a l'estandardització de les dades (*z score*) i així evitar la possible influència de la Distància Euclidiana quadrada en els clústers resultants.

En l'exemple que proposem a continuació, mantenim les mateixes dades que hem utilitzat des de l'inici del capítol. Ara li demanarem al ChatGPT que ens faci una anàlisi clúster, utilitzant l'algorisme d'agrupació Ward method (ja que es tracta d'una base de dades amb pocs casos) i que estandarditzi les dades amb *z-score*. En aquest cas li donarem les ordres de com volem que ho faci perquè ja en tenim els coneixements previs, però també hi hauria la possibilitat de demanar-li que ens donés alguns suggeriments pel tractament de les dades. Així doncs, li donem l'ordre amb el següent *prompt*: "Fes una anàlisi clúster de la taula amb el ward method i *z-score*". El resultat és el que es presenta a continuació, en el qual els proporciona el dendrograma, les dades amb la classificació del clúster, i la corresponent interpretació.



Hierarchical Clustering Dendrogram (Ward Method)



Aquí tens el dendrograma resultat de l'anàlisi de clusters utilitzant el mètode de Ward i la normalització z-score:

- Cada fulla del dendrograma representa un individu en el conjunt de dades.
- L'alçada en la qual dos individus o grups es combinen indica la distància (o dissimilaritat) entre ells. Com més gran és aquesta alçada, més diferents són els grups.
- El dendrograma ajuda a identificar el nombre de clusters naturals dins del conjunt de dades. Pots tallar el dendrograma a diferents nivells per obtenir diferents nombres de clusters.

Si necessites una interpretació més detallada dels clusters o desitges tallar el dendrograma a un nivell específic per obtenir un nombre determinat de clusters, fes-m'ho saber! [\[>-\]](#)

Si, interpreta els resultats posant nom als clusters i indicant la millor solució



Clustered Data



	Cas	Edat	Altura (cm)	Sexe
1	1	25	170	H
2	2	32	165	D
3	3	28	180	H
4	4	45	160	D
5	5	22	175	H
6	6	30	168	D

L'anàlisi de clusters ha resultat en dues agrupacions principals:

1. Cluster 1:

- **Edat mitjana:** 28,2 anys
- **Altura mitjana:** 175,3 cm
- **Pes mitjà:** 76,9 kg
- **IMC mitjà:** 25,02



2. Cluster 2:

- **Edat mitjana:** 34,7 anys
- **Altura mitjana:** 162,8 cm
- **Pes mitjà:** 59,6 kg
- **IMC mitjà:** 22,48

Interpretació dels Clusters:

- **Cluster 1 (Joves Alts):**
 - Aquest grup està format per persones més joves, amb una altura més gran i un pes més elevat. El seu IMC també és superior en comparació amb l'altre grup.
- **Cluster 2 (Adults de Baixa Altura):**
 - Aquest grup està format per persones una mica més grans, amb una altura més baixa i un pes menor. El seu IMC és inferior en comparació amb l'altre grup.

Millor Solució:

- La millor solució és la que divideix la població en dos clusters que reflecteixen diferències significatives en les característiques físiques (altura, pes) i en l'edat. Aquests dos clusters són fàcilment interpretables i mostren diferències clares entre els grups de persones.

Si necessites més detalls o altres anàlisis, si us plau, fes-m'ho saber! (-)

Aquesta és una mostra del potencial existent per a procedir amb l'anàlisi de dades amb tècniques de ML a través del ChatGPT. Tanmateix, caldrà comprovar que el resultat sigui lògic i correcte segons les dades. Aquesta pot ser una via per explorar les dades abans d'utilitzar un software estadístic, com per exemple JASP, a fi de procedir amb l'anàlisi definitiva.

Processament del llenguatge natural: Text mining i sentiment analysis

L'acumulació de grans quantitats de dades textuales, com ara en xarxes socials, pàgines web i blogs, ha incrementat la necessitat de mètodes eficients per al seu tractament, processament i anàlisi. Es requereix, doncs, tècniques i eines que facilitin l'extracció d'informació rellevant d'aquestes dades per tal que pugui ser interpretada fàcilment. La mineria de dades textuales (*text mining*) és el conjunt de tècniques i mètodes utilitzats per analitzar aquest tipus de dades, permetent extreure coneixement i patrons significatius (Gaikwad, Chaugule i Patil, 2014). Aquest camp inclou una varietat de processos, com ara la neteja de dades,

la transformació i la modelització de les dades amb l'objectiu d'obtenir informació rellevant.

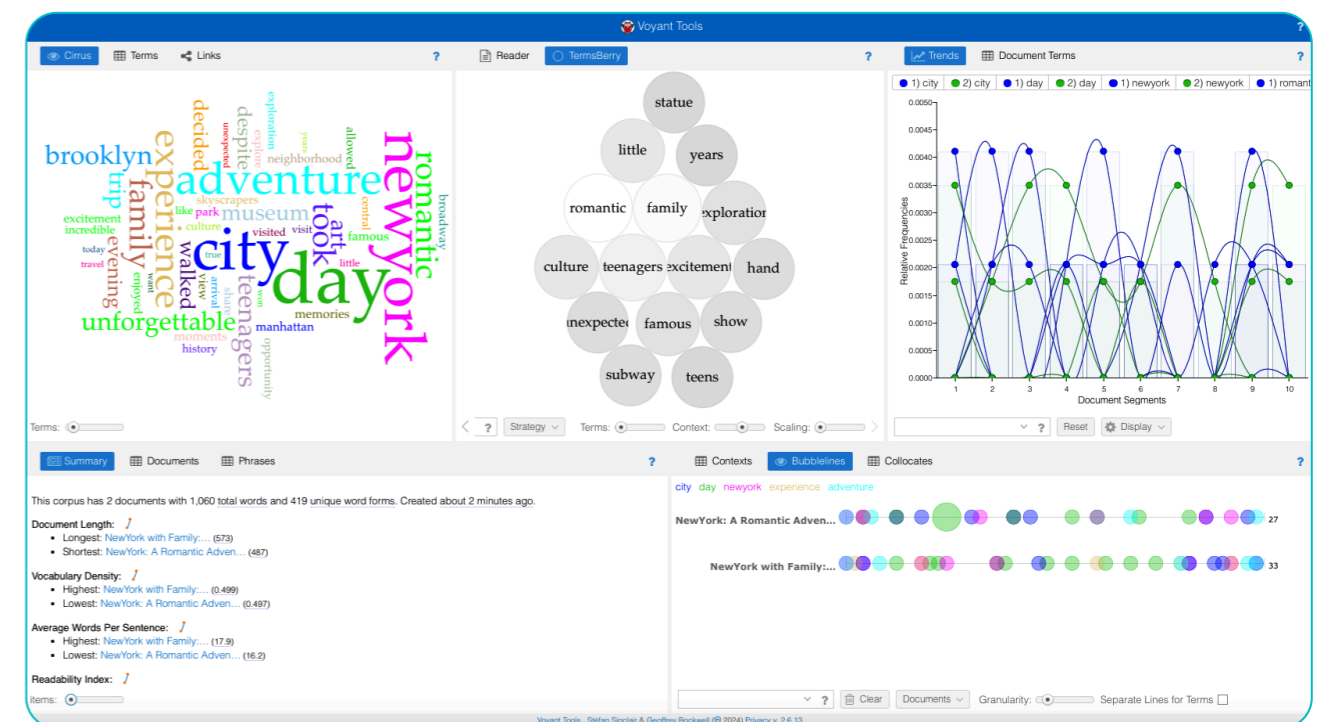
Per dur a terme aquesta anàlisi, s'utilitzen diversos algorismes i models estadístics que permeten examinar les dades textuales de manera detallada. Aquests models són capaços de gestionar la complexitat i la variabilitat inherent a les dades textuales, ajudant a identificar patrons i tendències útils per definir característiques significatives dins les dades analitzades (Lau, Lee i Ho, 2005). Per exemple, la identificació de paraules o temes recurrents pot revelar informació valuosa sobre les preferències i interessos dels usuaris.

Entre les tècniques més específiques que s'utilitzen en la mineria de dades textuales es troben l'anàlisi de sentiments (*sentiment analysis*) i la mineria d'opinions (*opinion mining*). Aquestes tècniques permeten determinar l'actitud i els sentiments d'una persona cap a un producte o tema concret, proporcionant informació clau per a les empreses i investigadors. L'anàlisi de sentiments, per exemple, pot classificar les opinions com a positives, negatives o neutres, mentre que la mineria d'opinions pot identificar temes específics que són importants per als usuaris (Khan, Qamar i Bashir, 2016). Aquest tipus d'anàlisi

és especialment útil en el context del màrqueting digital, la investigació de mercats i la gestió de la reputació en línia, on comprendre les opinions dels clients pot influir significativament en l'estratègia empresarial que s'emprenghi.

L'exemple d'aplicació que veurem a continuació utilitza el Voyant Tools (<https://voyant-tools.org>). Es tracta d'una aplicació accessible a la xarxa que permet portar a terme anàlisi de dades textuales amb els principis del *text mining*. Aquesta no és l'única eina, sinó que n'ha moltes d'altres, tant en línia com fora de línia, que actuen i treballen amb els principis de la mineria de dades.

Imaginem que volem analitzar la imatge de Nova York a través dels blogs de viatge. Per procedir a l'anàlisi necessitarem disposar de tots els textos dels blogs en un fitxer. En l'exemple que posarem tindrem dos textos en dos fitxers separats. A partir d'aquí cal carregar els textos a l'eina (Voyant Tools), assegurant-se que tots els textos estan ben formatats per a l'anàlisi. Aquesta eina permet que una vegada executada s'obtinguin tots els resultats d'una forma visual i fàcil. Els resultats que obtenim són els que es presenten a continuació:



Pràctica. La conversa sobre la COVID-19

Objectiu

- Aprendre a utilitzar eines d'IA per analitzar dades textuais de manera efectiva.
- Desenvolupar habilitats per interpretar i comunicar resultats d'anàlisi textual.
- Entrar en contacte amb els datasets textuais

Consideracions

- Dirigida a estudiants de ciències socials i humanitats.
- Activitat per fer preferentment en grup.
- Eina d'IA: Voyant Tools
- Dataset: [kaggle.com](https://www.kaggle.com).

A la xarxa, hi ha un munt de servidors que disposen de bases de dades textuais d'accés lliure que permeten treballar aquest camp i que són especialment útils per a usos docents. Un dels espais que permet accedir a un major volum d'informació és [kaggle.com](https://www.kaggle.com), que és una comunitat sobre dades textuais amb un nombre elevat de datasets; kaggle forma part de Google Cloud. En aquest cas, plantejem la possibilitat d'accedir al dataset de tweets sobre la COVID-19 (<https://www.kaggle.com/datasets/gpreda/covid19-tweets>), que recull una selecció dels tweets que utilitzen aquest hashtag. Hi ha un munt de bases de dades, que es poden adaptar a l'assignatura, com un dedicat a les opinions sobre Biden o Trump, sobre els sentiments expressats en els comentaris sobre els Jocs Olímpics, valoracions de pel·lícules o debats públics sobre temes controvertits. Per avaluar el dataset descarregat, s'emprarà Voyant Tools (<https://voyant-tools.org/>). Caldrà seleccionar una part dels tweets amb un determinat criteri (temporal, geogràfic, temàtic).

El procediment proposat és el següent:

- Carregar els textos a Voyant Tools, assegurant que els textos tenen el format adequat per l'anàlisi.
- Desenvolupar l'anàlisi estadística descriptiva
- Extracció de les paraules freqüents: Identificar les paraules més freqüents utilitzant la funcionalitat "Summary", "Terms" i "Document Terms".
- Distribució de paraules: Analitzar la distribució de paraules al llarg del text amb l'eina "Trends".
- Representació gràfica i visual de les dades:
- Visualització de les paraules: Crear en núvol de paraules utilitzant l'eina "Cirrus" per identificar les paraules més freqüents en el text.
- Visualitzar les relacions temàtiques: Crear una xarxa de paraules amb l'eina "Links" i "TermsBerry" per visualitzar els enllaços més comuns entre paraules i temes.
- Desenvolupar un breu informe i interpretar els resultats obtinguts.

En el grup classe, es poden realitzar les següents activitats:

- Reunir el grup classe per compartir les reflexions i resultats obtinguts.
- Establir conclusions grupals sobre l'anàlisi textual realitzat.
- Discutir els principals avantatges i reptes en l'ús de la IA per l'anàlisi de dades textual

Referències

Briones, J.A. (n.d.) <https://www.plenainclusio.org/wp-content/uploads/2022/02/Plena-inclusio-Murcia-Guia-de-Excel.pdf>

Gaikwad, S. V., Chaugule, A. y Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, 85 (17), 42-45.

Khan, F. H., Qamar, U. y Bashir, S. (2016). eSAP : A decision support framework for enhanced sentiment analysis and polarity classification. *Information Sciences*, 367, 862-873.

Lau, K. N., Lee, K. H. y Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46 (3), 344-362.

Mannyica, James et al. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. New York: McKinsey & Company [en línia: 20/05/2024]. <http://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf>

Microsoft (2024). Análisis de datos en Excel. [en línia: 02/06/2024] <https://support.microsoft.com/es-es/office/análisis-de-datos-en-excel-3223aab8-f543-4fda-85ed-76bb0295ffc4>

Mora, M. (2019). *Els rols, àmbits i noms de la ciència de dades*. FUOC.

O'Beirne, P. (2023). Proceedings of the EuSpRIG 2023 Conference "The Spreadsheet Crisis: Regaining Control". European Spreadsheet Risks Interest Group .

[1] Aquest prompt pot incloure una indicació sobre l'idioma de l'Excel, ja que les diferents versions del programa porten associada la traducció de les fórmules. En el cas de l'exemple l'Excel està en versió anglesa.

10

Els models de llenguatge en l'anàlisi qualitativa

Les metodologies de recerca qualitativa juguen un paper fonamental en el camp de les ciències socials, oferint una comprensió profunda i matisada del comportament humà i les motivacions que el sustenten. A través de mètodes de recollida de dades com entrevistes i grups de discussió, els investigadors i les investigadores poden reunir informació rica i detallada que il·lumina per què les persones pensen i es comporten de determinades maneres. Epistemològicament, la recerca qualitativa no revela una "veritat" central que es pugui generalitzar, sinó que produeix resultats transferibles (Maxwell, 2013), que semblen veritables per als participants en la recerca, d'acord amb les seves manifestacions. A través d'un procés de codificació acurat i la interpretació de dades no numèriques, les narratives matisades es transformen en temes comprensibles i posteriorment es construeixen patrons i *insights* significatius (Burnard, 1991; Jalali & Akhavan, 2024).

Per aconseguir això, es fan servir mètodes robustos de recollida de dades, la participació i el compromís d'informants apropiats, la codificació meticulosa i la

triangulació, entre d'altres, per millorar la fiabilitat. Aquest procés és fonamentalment llarg i laboriós. A causa d'això, els avenços tecnològics en paquets de programari, inclosos ATLAS.ti, NVivo i MAXQDA, així com enfocaments de processament de llenguatge natural, com l'anàlisi de sentiments i la modelització de temes, ajuden les investigadores i els investigadors a produir material analític útil mentre gestionen i interpreten grans volums de dades textuales (Zhang et al., 2024).

Com a resultat, en els últims anys s'han desenvolupat eines de transcripció d'intel·ligència artificial (IA) amb sofisticats algorismes de reconeixement de veu que automàticament tradueixen enregistraments d'àudio i vídeo a text escrit, permetent als investigadors i investigadores dedicar la seva atenció a l'anàlisi (Christou, 2023). A més, les eines de llenguatge d'IA, com Claude, ChatGPT o Gemini, s'han afegit a la recerca qualitativa, que a través de l'ús de tècniques d'aprenentatge profund i la capacitat de processar i generar llenguatge humà, poden produir des de la generació de text simple fins a la categorització

"Durant l'anàlisi de dades, s'ha trobat que les eines de llenguatge d'IA com ChatGPT o Gemini són més descriptives que no pas interpretatives"

temàtica i la comprensió contextual (Xiao et al., 2023). Concretament, s'ha estudiat que ChatGPT és una eina útil en la recerca qualitativa per la seva capacitat d'aportar un nou nivell de profunditat a l'anàlisi de dades, ja que entén el llenguatge natural i permet als usuaris comunicar-se fàcilment amb ell, és resistent a errors d'entrada de l'usuari o la usuària, com errors d'ortografia o gramaticals, respon eficientment a les sol·licituds i processos d'interpretació de dades, i proporciona explicacions i detalls sobre els resultats proporcionats, oferint transparència (Zhang et al., 2024). Així, com a eina, va més enllà de capturar informació limitada a la freqüència de frases o tons emocionals (a través de l'anàlisi de sentiments o de contingut), fins a la identificació de problemes polítics, socials o culturals, o l'extensió de la capacitat de l'investigador per refinar codis, construir connexions conceptuals i crear teoria (Anis & French, 2023). Com a resposta a això, els paquets esmentats anteriorment (ATLAS.ti, NVivo i MAXQDA) han desenvolupat col·laboracions amb OpenAI, millorant el procés de codificació en termes de temps i anàlisi.

Encara que l'ús de la IA en metodologies qualitatives ofereix un context de treball interactiu i que estalvia temps, hi ha alguns reptes a considerar (Anis & French, 2023; Christou, 2023; Morgan, 2023; Jalali & Akhavan, 2024; Zhang et al., 2024).

- Durant l'anàlisi de dades, s'ha trobat que les eines de llenguatge d'IA com ChatGPT o Gemini són més descriptives que no pas interpretatives.

Hi ha una manca de sentit comú i comprensió contextual dels elements socials i culturals en la IA, especialment quan es presenten qüestions ambigües. Aquests resultats també es coneixen com a "al·lucinacions", bàsicament produïdes per les limitacions del context (consultes prèvies i contingut de les respostes) intentant predir què podria seguir. A més, el text generat per humans que conté metàfores o sarcasme no sempre és comprès pel programa.

- Un altre repte de l'ús de sistemes d'IA és la simplificació excessiva, considerant que estan entrenats per prioritzar l'eficiència i la simplicitat per sobre la profunditat d'entorns i fenòmens complexos. Això podria portar a la pèrdua d'*insights* valuosos.
- Com hem vist en un capítol anterior, les eines de llenguatge d'IA porten biaixos inherents perquè estan bàsicament entrenades amb dades generades per humans i, encara que es fan esforços per limitar-los, fins ara es preveu que els resultats reflecteixin idees i creences dominants prevalents en la societat. Això podria afectar la validesa i la fiabilitat de la investigació, especialment si la investigadora o l'investigador no controla el procés d'anàlisi de dades.
- Com a eina, la IA hauria de millorar les capacitats dels investigadors i les investigadores, més que no pas substituir-les. Tot i que pot ajudar en la interacció amb les dades, no es pot considerar un intèrpret independent o objectiu de les dades ni molt menys una màquina de construcció

de teories. La investigació qualitativa abraça la subjectivitat i és aquí on els coneixements, comprensió i perspectives dels investigadors i les investigadores són fonamentals per a la reflexió i interpretació de les dades, especialment en estudis etnogràfics o fenomenològics. És responsabilitat dels investigadors i les investigadores qüestionar, millorar i incloure o excloure els codis, temes i patrons produïts per la IA.

- Les preocupacions ètiques i de propietat de les dades estan augmentant en el camp de la compartició de dades basades en humans, incloent-hi informació personal identificable o històries privades procedents d'entrevistes personals o grups de discussió, ja que no hi ha una guia clara sobre com haurien de ser utilitzades o compartides. Això és habitualment el cas dels programes de transcripció automatitzada que, a més, tenen l'inconvenient d'eliminar l'oportunitat d'una major implicació amb les dades qualitatives per part de l'investigadora o l'investigador. Per evitar les preocupacions ètiques, les dades personals s'han d'eliminar abans d'introduir-les als programes i s'han de lliurar documents de consentiment dels i les participants pel que fa a la protecció de dades.

Malgrat aquests reptes, els investigadors i les investigadores d'anàlisi qualitativa continuen experimentant amb eines d'IA per demostrar la seva utilitat com a productors de material i assistents en l'estudi. Les innovacions futures poden millorar les capacitats de la IA en la comprensió de bases de dades més matisades i proporcionar una anàlisi qualitativa més profunda. Els articles sobre aquest tema donen suport a les col·laboracions enriquidores entre el saber fer dels investigadors i les investigadores i les respostes de les eines d'IA, especialment en l'àrea de la codificació de dades textuals i l'anàlisi temàtica.

10.1. L'assistència a la guia de preguntes

Tanmateix, una part fonamental de les metodologies qualitatives és la generació de dades abans de l'anàlisi. La major part de la investigació relacionada amb les eines d'IA i la investigació qualitativa comença amb el fet que els investigadors o les investigadores ja tenen dades textuals riques de les transcripcions d'entrevistes, que són típicament el resultat d'entrevistes semiestructurades o grups de discussió. No obstant això, la qualitat de les dades recollides durant una entrevista no es determina únicament per la qualitat del protocol de l'entrevista. Particularment en el cas de les metodologies qualitatives, la generació de dades és un procés complex que requereix, a més d'un estudi previ i la preparació del protocol de l'entrevista, una atenció contínua i adaptació als inputs dels entrevistats per assegurar-se que totes les dades interessants que van ser expressades explícitament o implícitament estiguin adequadament presentades i descrites. Durant una entrevista, amb el rol investigador, l'estudiantat ha de ser capaç de triar entre ser simples oients o co-creadors de dades. El segon rol és fonamental especialment en el cas d'entrevistes semiestructurades o no estructurades.

En primer lloc, és important que els estudiants i les estudiants preparin un protocol d'entrevista sòlid. L'error més comú és anar directament a les preguntes que volen fer sense trencar el gel, sense aprendre prou sobre els seus entrevistats i entrevistades i sense connectar les teories llegides prèviament amb el cas empíric que s'estudia. Preparar un protocol complet requereix un temps considerable, que és on les eines de llenguatge d'IA, com ChatGPT, són útils. Per formular preguntes apropiades basades en característiques demogràfiques o psicogràfiques, és essencial descriure clarament l'objectiu de la investigació en la sol·licitud. Com que les entrevistes semiestructurades són les més comunes, és beneficiós indicar el nombre de preguntes que es volen fer (al voltant de 30) i el seu format (obert). Basant-se en el que les estudiants i els estudiants han llegit en les revisions de literatura existent, és important revisar

les preguntes i afegir conceptes que falten pel que fa al rerefons, les experiències i les percepcions dels entrevistats i les entrevistades.

Per exemple, quan demanem a ChatGPT (versió 4.0) que creï un protocol d'entrevista sota aquest objectiu: *"Explorar i entendre les experiències, desafiaments i adaptacions de l'estil de vida de les nòmades digitals femenines en entorns urbans i rurals. Aquest estudi té com a objectiu descobrir com les dinàmiques físiques i socials d'aquests entorns influeixen en el seu equilibri entre la vida laboral i personal, la productivitat professional i el benestar personal. En examinar els matisos de les seves interaccions amb diferents paisatges culturals i infraestructurals, la recerca pretén contribuir al coneixement més ampli del treball remot i el gènere en l'economia digital", s'han detectat els següents elements susceptibles de modificació:*

- Només la primera pregunta ("Pots descriure breument el teu historial professional?") demana sobre el perfil de les entrevistades, la resta estaven directament connectades amb l'estil de vida de la nòmada digital. Cal afegir més preguntes per conèixer millor la nostra entrevistada (vida personal, feines anteriors, experiències de viatge) i com era la vida abans d'aquesta decisió.
- Algunes preguntes s'han de simplificar, quan tenen moltes variables i poden ser difícils de recordar. Per exemple, "Trobos que certs tipus de treball són més fàcils o més difícils en entorns urbans o rurals?". Aquí, ChatGPT va crear una pregunta que

demana tant elements negatius com positius, així com entorns urbans i rurals. Hauria de separar-se en almenys dues preguntes.

- Algunes preguntes guien el tipus de resposta que es busca. Per exemple, "Has fet front a alguna situació d'aïllament professional? Com ho gestiones?" és una pregunta que dirigeix cap al problema particular de l'aïllament, i obvia altres situacions de més importància per a l'entrevistada. Transformar aquesta pregunta en oberta pot permetre un espectre més ampli de respostes ("Quins són els inconvenients professionals que aquest estil de vida t'ha portat? Com els has gestionat?").

Si es vol connectar aquest protocol amb una revisió de literatura específica, es poden incorporar en el *prompt* els articles que es considerin més rellevants i demanar al generador que afegixi algunes preguntes al protocol basant-se en els articles, i després fer una selecció i valoració de la proposta. Per exemple, en el cas exposat, es va demanar a ChatGPT que tingués en compte [aquest article](#) sobre el tema i creés preguntes relacionades amb la perspectiva de gènere en l'estil de vida dels nòmades digitals. Això va permetre la generació de deu preguntes relacionades amb els continguts de l'article, de les quals es van eliminar dues perquè eren molt similars al protocol original. D'aquesta forma, s'ha creat una guia d'entrevista d'unes 40 preguntes, que pot ser molt útil per a un estudiant o una estudiant amb poca experiència en la creació de guions d'entrevistes.

Pràctica. L'assistent de l'entrevista

Demana a ChatGPT que dissenyi un protocol d'entrevista semiestructurada, d'aproximadament 30 preguntes obertes, basat en l'objectiu següent: Explorar i entendre les motivacions darrere de les eleccions dels estudiants internacionals per cursar programes de grau a universitats de tot Europa, examinar les seves experiències durant el període de mobilitat i avaluar l'impacte d'aquestes experiències en el seu desenvolupament personal i trets de personalitat. Aquesta recerca investigarà els principals factors que influeixen en les decisions dels estudiants internacionals, incloent-hi motivacions acadèmiques, socials, culturals i econòmiques. També aprofundirà en les experiències viscudes per aquests estudiants, avaluant la seva adaptació a nous sistemes educatius, entorns socials i contextos culturals, així com els reptes que enfronten i els sistemes de suport que utilitzen. A més, l'estudi té com a objectiu avaluar els canvis en els trets de personalitat i el desenvolupament personal derivats de la seva mobilitat internacional, entenent l'impacte de les interaccions interculturals, els desafiaments acadèmics i l'augment de la independència en la seva autopercepció i creixement. Finalment, la investigació compararà les expectatives prèvies a la mobilitat dels estudiants amb les seves reflexions posteriors a la mobilitat per analitzar com les seves experiències reals s'alineen amb o difereixen de les seves motivacions i expectatives inicials.

- a) Demaneu als estudiants que revisin les preguntes generades per tal de:
- aconseguir la informació de perfil necessària dels estudiants internacionals
 - assegurar-se que cada pregunta no tingui més d'una variable per no confondre els entrevistats
 - assegurar-se que no hi hagi indicació d'una resposta a la pregunta
 - canviar preguntes tancades per preguntes obertes
 - afegir i destacar les preguntes que no s'incloïen en el protocol proposat, però que considereu important preguntar
- b) Demaneu als estudiants que llegeixin un article relacionat amb la Mobilitat Estudiantil Internacional (per exemple: Tran, L. T. (2016). Mobility as 'becoming': A Bourdieuan analysis of the factors shaping international student mobility. *British Journal of Sociology of Education*, 37(8), 1268-1289).
- Després de llegir-lo, han de suggerir noves preguntes basades en la informació que han après i justificar-ne la importància.
- c) Els estudiants haurien d'afegir aquest article a ChatGPT, al mateix xat on es va generar el protocol, i demanar-li que suggereixi noves preguntes basades en la informació proporcionada a l'article. Aleshores, haurien de comparar aquestes preguntes amb les que havien suggerit prèviament i fer una discussió sobre la contribució de la IA en la generació del protocol.

Objectiu

Capacitar els estudiants i les estudiants per aprofitar eficaçment la IA per generar preguntes d'entrevista inicials i estructurar el protocol tot avaluant críticament i refinant aquestes preguntes per garantir que s'alineïn amb els objectius de la investigació, els estàndards ètics i les complexitats del tema en investigació. Aquesta tasca té com a objectiu millorar la capacitat dels estudiants per elaborar preguntes d'entrevista perspícaces i rellevants relacionades amb investigacions anteriors i preparar un protocol d'entrevista que els proporcionï la seguretat que tots els temes possibles de discussió estan coberts (encara que després de les primeres entrevistes s'afegiran noves preguntes).

En la investigació qualitativa, s'ha demostrat que per tal que els entrevistats divulguin la seva informació personal i descriguin en detall les seves perspectives, percepcions i experiències, s'ha de generar confiança entre ells i l'investigador. En entrevistes semiestructurades i no estructurades, l'èxit rau en el fet que les preguntes que segueixen a cada resposta es basen en el contingut de la resposta. Els entrevistadors i les entrevistadores han d'estar formats en escoltar atentament tots els elements de cada resposta i formular automàticament noves preguntes de clarificació obertes basades en raonaments o fets no previstos proporcionats pels entrevistats. El resultat d'aquest procés serà la generació de dades complexes i profundes que hauran de ser codificades acuradament amb l'objectiu d'enriquir l'anàlisi de dades.

Hi ha dues tècniques bàsiques utilitzades en entrevistes que permeten aprofundir més en les respostes dels entrevistats; la tècnica de l'escalada i la del bracketing. L'escalada és una tècnica d'investigació qualitativa que s'utilitza per descobrir les motivacions, valors i creences dels individus que influeixen en les seves decisions i comportaments (Miles & Rowe, 2004). Durant aquest procés, l'entrevistador se centra en atributs específics d'un producte o servei i fa preguntes progressivament més profundes per revelar les conseqüències associades i, finalment, els valors fonamentals de l'individu. Mitjançant el model d'Atributs-Conseqüències-Valors (ACV), l'entrevistador es desplaça des de les característiques tangibles fins als beneficis que proporcionen aquestes característiques als valors fonamentals que satisfan. Aquest enfocament proporciona una visió profunda del comportament del consumidor, informant les decisions estratègiques en màrqueting, disseny de productes i experiència de l'usuari connectant les preferències superficials amb els impulsors psicològics més profunds. Bàsicament, això s'aconsegueix preguntant per què l'entrevistat considera importants els atributs i les conseqüències.

En canvi, el *bracketing* es refereix als mètodes que utilitzen els investigadors per mantenir l'entrevista a nivell personal tot i conèixer prèviament molt sobre el tema que es tracta (Sorsa et al., 2015). Així, els estudiants suspelen les seves pròpies suposicions sobre el món, permetent als participants descriure el seu punt de vista, desenvolupar la subjectivitat sense prejudicis i establir la confiança entre els participants. Un entrevistador, per exemple, pot tenir una comprensió del que significa una participant en un estudi si declara que és una compradora compulsiva en línia, però s'han de fer més preguntes per permetre als participants descriure les situacions en què s'han produït aquests comportaments, fent que proporcionin informació més detallada.

Aquí és on entra en joc la IA per formar l'estudiant. Per tal que es creï confiança i que l'entrevistat participi en una conversa en lloc d'una entrevista, l'estudiant ha de formular automàticament les preguntes de seguiment adequades en el moment adequat, independentment de qualsevol protocol preparat. La capacitat "conversacional" de ChatGPT, per exemple, permet als usuaris interactuar-hi a través del llenguatge natural (Zhang et al., 2024). En aquest context, ChatGPT pot simular entrevistes no estructurades actuant com l'entrevistat després d'haver descrit a la indicació l'objectiu de la investigació i el perfil de l'entrevistat que ChatGPT retratarà.

Pràctica. El model de llenguatge com a espàrring

En aquesta pràctica, els estudiants o les estudiants interactuaran amb Gemini, que exercirà el paper d'una persona específica. Mitjançant aquest exercici, practican la generació de preguntes espontànies i rellevants basades en les respostes que rebin, adaptant la seva consulta dinàmicament per obtenir coneixements profunds sobre les experiències i perspectives del subjecte. Gemini assumirà el rol d'una mare soltera de 32 anys que viu en un petit pis a prop del parc del Retiro a Madrid. Treballa com a executiva de màrqueting digital a L'Oreal i és mare d'una nena de 7 anys. L'entrevista explorarà com s'organitza per anar de vacances amb la seva filla, centrant-se en els reptes que enfronta, el seu procés de presa de decisions i les opcions que considera, proporcionant una comprensió matisada de les seves experiències equilibrant la seva vida personal i responsabilitats laborals.

Instruccions

Inicieu l'entrevista: Comenceu presentant-vos i explicant el propòsit de l'entrevista. Inicieu l'entrevista amb les preguntes següents: *"Gràcies per dedicar-me el vostre temps avui. Per començar, podríeu explicar-me l'última vegada que va planejar unes vacances amb la vostra filla? Quins van ser alguns factors clau que va tenir en compte a l'hora d'escollir la vostra destinació?"*

Realitzeu l'entrevista tenint en compte la tècnica de l'escalada i el bracketing:

- Permet que la conversa flueixi de manera natural.
- Centra't en preguntes obertes per fomentar respostes detallades.
- Llegeix les seves respostes i formula les teves preguntes següents basant-te en la informació que proporciona.
- Explora no només el 'què', sinó també el 'com' i el 'per què' darrere dels seus processos de planificació de vacances i decisions.

Temes sobre els quals has de preguntar en general:

- Les seves prioritats a l'hora d'escollir una destinació de vacances.
- L'equilibri entre les responsabilitats laborals i el temps personal.
- Consideracions financeres i logístiques en la planificació de vacances.
- Sistemes de suport o recursos que pugui utilitzar.
- Reptes que enfronta com a mare soltera a l'hora de planificar vacances.

Reflexió

Després de l'entrevista, escriu una reflexió sobre el procés. Discuteix com vas adaptar les teves preguntes en funció de les respostes de l'entrevistada i com podries millorar aquestes preguntes després de la teva reflexió.

Objectiu

Millorar les habilitats dels estudiants per dur a terme entrevistes no estructurades, un mètode clau en la investigació qualitativa. En interactuar amb un entrevistat simulat en la persona de Gemini, els estudiants practican la generació de preguntes espontànies i rellevants basades en les respostes rebudes en temps real, adaptant la seva consulta de manera dinàmica i assegurant-se que tota la informació proporcionada estigui ben descrita.

10.2. La codificació amb un model de llenguatge

L'anàlisi temàtica és una eina crucial en la investigació qualitativa per la seva flexibilitat, profunditat i enfocament sistemàtic. Permet les investigadores i els investigadors descobrir i interpretar les capes complexes de les dades qualitatives, proporcionant informació valuosa sobre el comportament humà, les experiències i els fenòmens socials. Permet també identificar, analitzar i informar sobre patrons de significat (temes) dins de les dades. Inclou un enfocament de sis fases, que són: familiarització amb el conjunt de dades, codificació de dades, generació inicial de temes, desenvolupament i revisió de temes, refinament i definició de temes, i informe final (Zhang et al., 2024). Aquesta metodologia és lenta i requereix un esforç manual significatiu, especialment quan es tracta de conjunts de dades grans i complexos, cosa que representa un repte per als investigadors novells, ja que hi intervé la subjectivitat a l'hora d'identificar els temes. El repte per als estudiants i les estudiants és encara més gran, ja que no estan familiaritzats amb el procés de codificació.

Els investigadors i investigadores han intentat demostrar la utilitat potencial dels programes d'IA en l'anàlisi de dades qualitatives comparant la seva eficàcia amb la codificació manual (Morgan, 2023; Jalali & Akhavan, 2024). La comparació de la codificació manual i la codificació amb eines de llenguatge IA es demostra com un mètode complementari, centrant-se en ChatGPT (versió 4.0), ja que tots dos identifiquen temes significatius i bucles de retroalimentació, però de vegades no coincideixen; és a dir, ChatGPT pot proporcionar part de l'anàlisi que l'investigador pot considerar quan interpreta les dades i millora l'anàlisi produïda. En particular, ChatGPT ofereix un enfocament més directe i objectiu, cosa que pot reduir el risc d'introduir els biaixos mateixos dels investigadors i investigadores. A més, s'ha suggerit que els models d'IA funcionen millor en l'anàlisi deductiva en lloc de l'anàlisi inductiva, cosa que fa que la participació humana sigui encara més crítica. Per tal que els estudiants i les estudiants entenguin

que les eines de llenguatge IA són més útils com a assistents en la generació d'idees i la identificació de dades dins l'anàlisi temàtica, per exemple, en lloc de com a analistes de dades, han d'observar les diferències amb la codificació manual. Això vol dir fer una anàlisi de replicació, on les transcripcions d'entrevistes que els estudiants han realitzat es codifiquen primer manualment i després mitjançant l'ús de programes d'IA. L'ordre és significatiu perquè les eines de llenguatge IA produeixen text que està ben estructurat i aparentment complet, cosa que per als estudiants que no estan familiaritzats amb la complexitat del procés d'identificació de temes, aquesta anàlisi de text podria semblar suficient. Per tal d'entendre els temes produïts i avaluar-ne l'adequació basada en interpretacions subjectives personals, familiaritzar-se amb les dades, és a dir, llegir atentament la base de dades completa, és un primer pas necessari.

Un cop aconseguida la familiarització amb les dades, l'ús de ChatGPT (versió 4.0) pot començar en la identificació de variables i vincles entre elles. En aquest punt, és fonamental construir les indicacions adequades per interpretar efectivament el llenguatge matisat. La qualitat dels resultats de la interacció està molt relacionada amb la qualitat de les indicacions. En aquest capítol hi haurà una descripció de les estratègies bàsiques que hem de seguir a l'hora de crear indicacions per a l'anàlisi qualitativa (mentre que en un capítol anterior es proporciona una descripció més detallada de la generació).

Basat en Zhang et al. (2024), les indicacions proporcionades han de:

- Oferir una descripció detallada de la tasca, és a dir, la naturalesa de les dades, l'objectiu de la investigació, els resultats esperats i qualsevol altra consideració a tenir en compte (com el format de les entrades). Per exemple, *"Aquí hi ha una transcripció d'un grup de discussió sobre el consum de música en directe. El perfil de les persones que hi van participar és de 30 a 50 anys, que tenien famílies i fills, viuen en zones urbanes i representen una classe social mitjana. Cada*

paràgraf és d'un participant i comença amb els seus noms. Cada paràgraf que comença amb el mateix nom és de la mateixa persona. L'objectiu era identificar la manera com es consumeix la música quan un treballa a temps complet i té fills. Si us plau, llegiu-ho primer”.

- Preferiblement, utilitzar el joc de rols demanant a ChatGPT que sigui un expert en anàlisi qualitativa, si bé cal tenir en compte que aquesta estratègia per si sola no és suficient.
- Definir la metodologia, és a dir, l'objectiu de la tasca, que en el cas de l'anàlisi temàtica seria demanar al programa que trobi patrons i temes.
- Especificar el procés analític, és a dir, afegir instruccions sobre com la IA ha d'analitzar les dades, si ha de considerar una teoria específica (que s'ha de descriure o penjar un altre text que el programa pugui llegir i seguir) o generar a través de totes les dades els temes més recurrents (es pot demanar un límit de fins a deu, per exemple), i després basant-se en les respostes, triar les que interessin i demanar una sortida més detallada. Com més precisa sigui la naturalesa dels temes recurrents (per exemple, “si us plau, proporcioneu els principals temes que les mares en aquest grup de discussió mencionen com a desafiament quan necessiten assistir a un esdeveniment de música en directe en comparació amb els principals temes dels pares”) més anàlisi en profunditat es genera. La familiarització amb les dades permet als estudiants tenir una idea general de qüestions particulars que van sorgir durant el grup de discussió i podrien començar des d'aquesta base. Cal tenir en compte que les paraules “tema” o “subtemes” és millor reservar-les per a la fase posterior més interpretativa de l'anàlisi, ja que els temes es construeixen al final d'aquest procés.
- Definir el format de sortida, per exemple, en el cas d'haver demanat temes recurrents principals, aleshores demanar que es clusteritzin en temes

més amplis i després es presentin per separat amb extractes representatius i un breu resum. Es pot demanar una llista amb punts, un format textual o una taula, una descripció més curta o més llarga de les dades. Pel que fa al contingut, es pot sol·licitar per separat beneficis, desavantatges i argumentació (raons, suggeriments, percepcions). Es recomana ser el més precís possible en el mètode l'anàlisi, com per exemple, “Vull identificar els vincles causals entre variables. Pots considerar totes les variables identificades prèviament i si la relació entre dues variables prové de diferents parts del grup de discussió. Tingues en compte que sempre pots tornar enrere i llegir les dades del grup de discussió si consideres que falta alguna cosa”.

- Incloure la prioritització a l'hora de demanar, per exemple, un codi de deu codis, començant pels més freqüents o els més relatius als objectius clau. Per exemple, “Descriu els set principals inconvenients que els participants citen per assistir a festivals, començant pels causats per factors externs i acabant pels causats per motius individuals”.
- Explicar els objectius de l'anàlisi demanant a ChatGPT que expliqui el seu procés de presa de decisions a l'hora de triar els resultats produïts. Com més analític sigui el format de sortida demanat, més fàcil serà oferir transparència. A més, com en la codificació manual, analitzar cada línia o paràgraf de les dades és més efectiu per a anàlisis en profunditat en comparació amb analitzar les dades d'entrada completament. El procés requereix més temps, però els resultats també són més fiables.
- Reconèixer la seva experiència per confirmar al programa que la manera en què procedeix amb l'anàlisi sol·licitada compleix amb les expectatives de l'investigador o la investigadora.

Tenint en compte la complexitat de l'anàlisi de dades dins de les metodologies qualitatives, no és sorprenent que les eines d'anàlisi preexistents per a la codificació,

com ATLAS.ti, N-Vivo i MAXQDA, incorporin IA dins dels seus programaris. En particular, la més antiga, ATLAS.ti, es va llançar al mercat el 1993 ajudant els investigadors a dur a terme la teoria fonamentada i l'anàlisi de contingut, mentre que 30 anys després, el 2023, va llançar una versió beta d'IA, oferint un recurs alternatiu ràpid per a l'anàlisi de text i especialment la codificació. Els investigadors i investigadores poden utilitzar aquesta eina de codificació basada en IA per processar dades i, d'una banda, visualitzar els resultats i, de l'altra, predir quins codis o etiquetes podrien ser útils per aplicar al seu conjunt de dades.

Un cop s'hagi creat un nou projecte i afegit la transcripció de l'entrevista o els grups de discussió dins del programari ATLAS.ti, la codificació automàtica comença en fer clic a “Analysis”. És important haver triat tots els documents (transcripcions) que desitgem codificar, per crear un llibre de codis comú. En fer clic a “Analysis”, una de les opcions de les quals disposem és “AI coding”, que es refereix a una codificació emergent de les nostres dades, creant categories, codis i subcodis. Aquí, no hi ha cap directriu sobre el que s'ha de codificar, de manera que es crearà un gran nombre de codis. Les investigadores o els investigadors han de revisar a fons aquesta llista, ja que molts codis es repeteixen amb paraules lleugerament diferents, com “edat”, “envelliment” o poden no estar d'acord amb els codis atribuïts a cites específiques, ja que el contingut no està relacionat amb aquest codi. En aquest cas, aquests codis s'han d'eliminar manualment i aquestes cites s'han de donar codis diferents. Un cop revisats tots els codis, s'ha de fer clic al botó “Apply” i afegir aquests codis al programa i mostrar un resum. La IA només funciona per paràgraf (no frases dins del paràgraf) i omet paràgrafs molt curts. Per tal d'aconseguir una codificació de millor qualitat i menys caòtica, és millor eliminar del text (transcripcions) tot el context que no està relacionat amb el que estem interessats a codificar (però no eliminar parts significatives per a la comprensió del context). El programari proporciona un manual complet per a l'usuari per a un procés de codificació pas a pas.

A més de la codificació emergent, ATLAS.ti també ofereix la possibilitat de “codificació intencional d'IA” perquè els usuaris o les usuàries tinguin més control sobre els codis produïts. Com a resultat d'aquest procés, es pot generar un codi més precís, que es basa en preguntes bàsiques relacionades amb el nostre objectiu de recerca. Un cop s'hagin seleccionat els documents que es vulguin codificar, cal fer clic novament a “Analyze” i ATLAS.ti preguntarà sobre els objectius, el que permet incorporar la pregunta de recerca o hipòtesi i qualsevol context de suport. Suposem que hem realitzat entrevistes amb estudiants internacionals a la nostra universitat i els hem preguntat, “Què us va motivar a triar el vostre grau i la nostra universitat?”. Atlas.ti generarà preguntes concretes basades en aquest objectiu, juntament amb noms de categories de codis pertinents, que es poden revisar i editar, eliminar preguntes que no interessin o fins i tot afegir preguntes pròpies. Quan es clica sobre “Start Coding”, ATLAS.ti analitzarà els documents i recopilarà suggeriments de codificació. En el moment en què es consideri correcta la codificació, cal clicar a “Apply” i ATLAS.ti aplicarà la codificació i mostrarà un resum.

Així doncs, hi ha dues formes d'emprar els models de llenguatge en l'ensenyament de les eines qualitatives: Poden ser instruments de suport en el complex procés de codificació de la informació acumulada o poden ajudar a la recollida de la informació i la confecció dels guions de les entrevistes. En els dos casos, les eines són un suport que faciliten una major versatilitat, però no han de ser presentats com una alternativa a la intervenció no automàtica.

Pràctica. Codificació híbrida

En aquesta pràctica, els estudiants o les estudiants codificaran una transcripció d'entrevista que preferiblement hagin realitzat ells mateixos perquè estiguin familiaritzats amb la base de dades de primera mà. També és possible que utilitzin la transcripció de l'entrevista que van realitzar a l'exercici anterior. En el context de l'anàlisi temàtica, els estudiants han d'aplicar dos mètodes diferents de codificació: manual, assistida per IA (ChatGPT) i IA d'ATLAS.ti.

a. Codificació manual

Els estudiants han de llegir la transcripció de l'entrevista per familiaritzar-se amb les dades, identificar i destacar temes clau, categories i subcategories, desenvolupar un marc de codificació basat en aquests temes, codificar la transcripció manualment assignant codis a diferents parts del text, i mapar les relacions entre diferents temes i categories. Finalment, proporcioneu exemples de la transcripció per il·lustrar cada tema i categoria.

b. Codificació assistida per IA (ChatGPT versió 4.0)

Basat en les pautes esmentades anteriorment sobre com crear indicacions, els estudiants han de desenvolupar un conjunt d'indicacions que repliquin les preguntes i directrius utilitzades durant la codificació manual, assegurant-se que són clares i específiques. Aleshores, els estudiants han d'introduir la transcripció de l'entrevista i les indicacions a ChatGPT, revisar i refinar la sortida proporcionada per ChatGPT, assegurant-se que s'alinea amb l'enfocament de l'anàlisi temàtica.

c. Codificació d'IA d'ATLAS.ti

Utilitzant el manual de l'usuari d'ATLAS.ti, els estudiants han d'afegir la transcripció de l'entrevista anterior en el seu format original al programari. Per entendre les diferències entre la codificació emergent d'IA i la codificació intencional d'IA, haurien de treballar en projectes separats i demanar codificació emergent en un i codificació intencional en l'altre. En el primer cas, els estudiants han de revisar els codis i categories generats, eliminar els que siguin enganyosos i, després, presentar l'estructura final del llibre de codis. És important que esmentin les noves categories/codis que han sorgit en comparació amb els dos mètodes anteriors. En el segon cas de codificació intencional, els estudiants haurien d'afegir a la casella d'"intenció" les preguntes inicials que van afegir al mètode de ChatGPT en el format de la indicació. Els estudiants han de processar les preguntes suggerides i els codis relacionats que es generen i després "iniciar la codificació". És important comparar els codis resultants amb les cites que s'hi han associat, i després amb els resultats de la codificació del segon mètode (ChatGPT), i descriure les diferències de codificació entre dos entorns d'IA separats.

d. Comparació de resultats i informes

Els estudiants haurien de comparar els temes, categories i relacions identificats per ChatGPT amb els identificats manualment. Tingueu en compte qualsevol diferència en la identificació de temes, profunditat de l'anàlisi i qualsevol altra informació addicional proporcionada per ChatGPT. Finalment, haurien d'escriure un informe detallat comparant els processos de codificació manual i assistida per IA, i discutir els punts forts i febles de cada mètode. Objectiu: L'objectiu d'aquesta tasca és ajudar els estudiants a entendre les diferències i similituds entre l'anàlisi de dades qualitatives manual, assistida per IA i assistida per programari especialitzat. En comparar els tres mètodes, els estudiants obtindran informació sobre els avantatges i limitacions de l'ús d'eines d'IA en la investigació qualitativa. També desenvoluparan habilitats de pensament crític sobre l'aplicabilitat, fiabilitat i validesa de la IA en l'anàlisi qualitativa.



IA i creativitat: Relació competitiva o BFF?

“La destrucció creativa” és un concepte introduït per l'economista Joseph Schumpeter, suggerint el moviment en què la innovació crea valor que reemplaça els sistemes obsolets, com el treball humà (Vinchon, et al. 2023). Per a alguns periodistes, un exemple recent d'això es reflecteix en el titular d'un article publicat a Le Figaro, on es pot llegir “[IA per substituir periodistes a Springer](#)”. No obstant això, la pregunta que sorgeix és realment així o si en realitat els llocs de treball vinculats amb la creativitat, amb el geni, amb la imaginació, amb l'enginyer estan exempts de l'amenaça de la intel·ligència artificial. La literatura actual més aviat considera que la IA pot ser un complement o un element de suport dels llocs de treball en les indústries creatives (Noy & Zhang, 2023; Vinchon, et al. 2023).

Fins ara, la intel·ligència artificial s'ha desenvolupat a través de tres etapes principals: a) raonament simbòlic, b) aprenentatge probabilístic i base de coneixement, i c) aprenentatge profund i mapa de coneixement (Man, et al., 2021), assolint estàndards humans d'intel·ligència perceptiva en termes d'escoltar, parlar i

veure, però no en el camp de la intel·ligència cognitiva. La creativitat com a constructe fenomenològic està altament influenciada pels efectes de la IA (Hubert, 2024). Centrant-se en el pensament creatiu, aquest es refereix a una activitat mental d'alt nivell, basada en la percepció, el raonament, la memòria, el pensament, la imaginació, l'associació, la intuïció i la comprensió, i condueix a la comprensibilitat i la novetat (Man, et al., 2021). Es defineix com la capacitat de produir idees que siguin alhora originals i útils (Runco & Jaeger, 2008). Atès que la IA sembla funcionar força bé per produir un gran nombre de propostes en activitats relacionades amb l'art com escriure poesia, guions de pel·lícules i música, és a dir, productes que els humans generalment perceben com a creatius, la relació exclusiva entre creativitat i humans sembla estar amenaçada.

De manera general, la creativitat es basa en cinc operacions mentals (Benvenuti, et al., 2023):

- L'ampliació del camp mental, és a dir, la capacitat del subjecte per concebre idees diferents i

“Aquest procés creatiu es considera únic dels humans, mentre que la creativitat artificial probablement només cobreix una part”

úniques, també coneguda com a pensament divergent, o la seva capacitat per mantenir una riquesa mental d'informació per associar elements relacionats entre si.

- La connexió de camps mentals antitètics de manera que ofereixin valor
- La reorganització del camp mental
- El subministrament de motivació i interès en el camp
- La millora del rendiment i l'adquisició de coneixements.

Per als estudiosos de la creativitat, altres aspectes crucials reflecteixen el procés creatiu, com ara trobar problemes, és a dir, què s'hauria de preguntar i abordar, estimant l'originalitat i l'eficàcia potencial d'aquestes idees creatives, i la presa de decisions sobre la millor opció possible. Aquests aspectes requereixen autoregulació i co-regulació social que refini aquestes idees basant-se en la reflexió i el feedback, optant per enfocaments poc convencionals però útils (Vinchon, et al., 2023). Aquest procés creatiu es considera únic dels humans, mentre que la creativitat artificial probablement només cobreix una part. D'altra banda, el desig i la intenció d'una persona per crear alguna cosa segueix sent una característica humana, coneguda com a “intencionalitat creativa”. El que s'ha de tenir en compte, però, és que una idea té el valor de ser creativa perquè nosaltres, els humans, li atribuïm aquest valor, un valor que la IA desconeix, de moment.

S'ha suggerit que la col·laboració dels humans amb els sistemes d'IA pot portar-nos a una nova era de “creativitat assistida” en què la IA no actua de manera independent, sinó com un agent creatiu col·laboratiu. Mitjançant aquest nou procés creatiu, l'individu manté un paper central en dos moments: al començament, on la identificació i l'exploració de les preguntes adequades representen el resultat desitjat d'aquesta etapa de trobar problemes, i al final, on és responsable d'avaluar, modificar i validar els resultats creatius de la IA (Vinchon, et al., 2023). A la llum del fet que un treballador futur, és a dir, l'estudiant d'avui, s'espera que desenvolupi pensament crític, resolució de problemes, creació de contingut i habilitats de cooperació, l'ús de tecnologies emergents és fonamental per al desenvolupament de les habilitats del segle XXI (Benvenuti et al., 2023).

11.1. Creativitat i pensament divergent

Els investigadors i les investigadores han intentat comparar els sistemes d'IA actuals amb els humans en termes de creativitat per demostrar si els primers són més creatius que els segons i en quina mesura. Per a alguns investigadors i investigadores, els mecanismes cognitius humans que s'activen durant les tasques creatives no estan presents en la IA, cosa que indica que aquesta només pot reflectir una creativitat artificial (Hubert et al., 2024). En general, la manca de processos metacognitius (com ara l'avaluació i la motivació per la tasca) limita la capacitat de la

intel·ligència artificial per determinar explícitament si i per què alguna cosa és creativa. Altres investigadors suggereixen que la creativitat computacional reflecteix realment els mecanismes que segueixen els humans mentre generen processos creatius i arriben a solucions creatives, o fins i tot que la IA no necessita confiar en processos metacognitius per crear idees noves. La incapacitat freqüent de les persones per distingir entre productes creats per humans i per la IA dona suport a la idea que la intel·ligència artificial pugui tenir potencial creatiu.

La intel·ligència artificial té bases de dades vastes, però estructurades de manera convencional, mentre que els humans, tot i treballar amb molta menys informació poden combinar-los de forma més eficient, donant lloc ocasionalment a resultats disruptius (Vinchan et al., 2023). La base d'aquesta comparació durant el procés creatiu es basa en la suposició que la IA treballa a través d'un pensament més convergent, mentre que els humans poden desenvolupar un pensament divergent (Hubert, 2024). El pensament convergent considera una solució òptima única per a un problema predefinit, avaluada a través d'experiències i mètodes tradicionals existents (com l'anàlisi, la síntesi i la generalització), sense ser particularment flexible o "out of the box" (fora de la caixa). Tanmateix, això no qüestiona la possibilitat de generar solucions originals, només la probabilitat.

El pensament divergent, en canvi, és el nucli del pensament innovador i implica generar múltiples solucions creatives a un problema mitjançant un procés flexible (com la deducció, la comparació o l'analogia) basat en la reorganització dels materials i la informació donada, deixant de banda les limitacions del coneixement original (Man, et al., 2021; Hubert, 2024). Com a procés, implica participar en una varietat d'activitats com el raonament, la imaginació, l'associació i la intuïció (Qi, 2018), alhora que permet als estudiants pensar ràpidament, observar millor i respondre des de múltiples angles. Tanmateix, no garanteix l'assoliment creatiu, només n'augmenta la probabilitat. El pensament divergent es divideix en uns components específics (Man et al., 2021; Koivisto & Grassini, 2023):

- **Fluïdesa**, que significa la velocitat a la qual la ment d'una persona divergeix i la seva capacitat per produir un gran nombre d'idees en un temps unitari, reflectint el pensament fluid i l'agilitat.
- **Flexibilitat**, és a dir, la naturalesa qualitativa del pensament divergent o l'ús flexible del coneixement, incloent-hi préstecs, encreuaments, substitucions i integració de conceptes i continguts relacionats.
- **Originalitat o unicitat**, que significa si la solució al problema és única per a la persona
- **Elaboració**, referint-se a la capacitat d'ampliar o afegir detalls a les idees.

Els processos creatius bàsicament impliquen una interacció entre el pensament convergent i divergent, on el segon és responsable de la novetat de les idees, mentre que el primer és responsable de la rellevància de les idees a les demandes de la tasca (Koivisto & Grassini, 2023).

En l'àmbit pràctic, hi ha diverses eines amb respostes obertes que s'utilitzen per avaluar el pensament divergent. Les proves més utilitzades són:

- La Tasca d'Usos Alternatius (AUT), on es demana als i les participants que suggereixin usos inusuals i creatius d'objectes quotidians, com ara un llapis, treballant en la seva capacitat per accedir a conceptes relacionats.
- La Tasca de Conseqüències (CT), on els participants responen a escenaris hipotètics (com què passaria si les persones poguessin volar) i tantes conseqüències com puguin dins d'un temps específic.
- La Tasca d'Associació Divergent (DAT), que es basa en la capacitat creativa divergent i semàntica verbal dels participants, demanant-los que generin una llista de paraules o idees que siguin tan diferents entre elles com sigui possible dins d'un temps determinat, amb l'objectiu de quantificar l'aspecte de la creativitat, avaluant la diversitat i unicitat de les associacions produïdes pels participants (Koivisto & Grassini, 2023; Hubert et al., 2024).

Aquestes proves han estat utilitzades pels investigadors per mesurar el potencial creatiu dels participants

humans versus els models de llenguatge d'IA, i en particular, ChatGPT. El que s'ha de destacar és que en aquests casos, l'objectiu no era veure com els humans poden millorar les seves pròpies habilitats creatives mitjançant l'ús de models d'IA, sinó si aquests últims podrien substituir els primers. Per exemple, Hubert et al. (2024) van reclutar 151 individus i van preparar 151 sessions de ChatGPT-4, representant la participació de la IA per a una mostra equilibrada, i van aplicar les proves AUT, CT i DAT, controlant la fluïdesa en les dues primeres, que són tasques temporitzades, en termes de nombre de respostes per ser iguals (és a dir, si el participant humà n'10 mencionava 4 idees, es va instruir a GPT-4 per generar 4 idees).

Per a la prova AUT, l'objectiu és presentar als participants un objecte comú (com una forquilla o una corda) i demanar-los que generin tants usos creatius com sigui possible per aquests objectes, mentre s'avaluen les respostes per fluïdesa (nombre de respostes), originalitat (unicitat de les respostes) i elaboració (nombre de paraules per resposta vàlida). Un exemple de la instrucció per als participants humans podria ser:

"L'objectiu d'aquesta tasca és desenvolupar tantes idees originals i creatives com sigui possible per a una forquilla. L'objectiu és generar idees que siguin intel·ligents, inusuals, innovadores i captivadores per a les persones. Mentre siguin creatives, aquestes idees no han de ser pràctiques ni realistes; poden ser absurdes i estranyes sempre que siguin creatives. La tasca durarà 3 minuts durant els quals pots introduir tantes idees com vulguis, recordant la importància de la qualitat, no la quantitat. Presenta una llista d'aquestes idees."

Un exemple de la instrucció per a GPT-4 pot ser alternativament:

"L'objectiu d'aquesta tasca és desenvolupar tantes idees originals i creatives com sigui possible per a una forquilla. L'objectiu és generar idees que siguin intel·ligents, inusuals, innovadores i captivadores per a les persones. Mentre siguin creatives, aquestes idees no han de ser pràctiques ni realistes; poden ser absurdes i estranyes sempre que siguin creatives. Llista (inserir nombre de fluïdesa) idees creatives."

La distància semàntica entre l'objecte i la resposta AUT pot projectar l'originalitat del pensament divergent. Aquesta distància es pot determinar amb la plataforma [SemDis](#). Per evitar un augment fals de la distància semàntica, s'ha vist que expressions com "fer un ____" o "usar com a ____" o "DIY" s'haurien d'eliminar abans d'introduir les dades a la plataforma. O es pot aconsellar a les persones des del principi que no les utilitzin. A més, els avaluadors humans, desconexors de la presència de respostes generades per IA, poden avaluar els resultats de les dues proves entre els dos grups de participants, utilitzant l'escala Likert de 5 punts (1=gens, 5=molt), proporcionant una visió humana de la creativitat. Se'ls ha d'instruir per puntuar la novetat sobre la utilitat, i que respostes com "tallar" en resposta a objectes com "tissors" o una resposta confusa o il·lògica haurien de rebre una puntuació d'1. Per a l'anàlisi estadística, es pot calcular la puntuació mitjana de distància semàntica de totes les respostes generades per cada objecte per sessió i participant, així com la puntuació màxima de les respostes a cada objecte (Koivisto & Grassini, 2023).

Per al CT, es pot demanar als participants que responguin tantes conseqüències com sigui possible a les preguntes generals, mesurant de nou la fluïdesa (nombre de respostes), l'originalitat (unicitat de les respostes) i l'elaboració (nombre de paraules per resposta vàlida). Les instruccions de la tasca per als participants humans poden ser:

“En aquesta tasca, llegeix la següent declaració ‘imagina que les persones poden volar’ i durant 3 minuts intenta pensar en totes les conseqüències que podrien resultar d'aquesta declaració. L'objectiu és generar idees que siguin intel·ligents, inusuals, innovadores i captivadores per a les persones. Les respostes es puntuaran en funció de l'originalitat i la qualitat. Intenta seguir pensant en respostes i escriure-les durant aquests 3 minuts.”

Posteriorment, els participants poden rebre noves declaracions, repetint el mateix procés amb preguntes com “imagina un món sense animals”. A més, un exemple de la instrucció per a GPT-4 pot ser:

“En aquesta tasca, llegeix la següent declaració ‘imagina que les persones poden volar’. Sigues tan creatiu com vulguis. L'objectiu és generar idees que siguin intel·ligents, inusuals, innovadores i captivadores per a les persones. Les respostes es puntuaran en funció de l'originalitat i la qualitat. Intenta pensar en totes les conseqüències que podrien resultar de la declaració. Llista (inserir nombre de fluïdesa) conseqüències creatives.”

Per automatitzar l'avaluació de la distància semàntica tant en CT com en AUT, es pot utilitzar l'eina [Open Creativity Scoring](#), així com els avaluadors humans.

Finalment, durant la DAT, es pot demanar als participants humans i a GPT-4 que proposin 10 substantius tan diferents entre ells com sigui possible. No hi ha limitacions de temps per a aquesta tasca. Per exemple, la instrucció estàndard de la DAT és:

“Si us plau, proporcioneu 10 paraules que siguin tan diferents entre elles com sigui possible, en tots els possibles significats i usos de les paraules. Aquestes són les 5 regles que s'han de seguir: a) una única paraula en anglès, b) només substantius (objectes, conceptes, etc.), c) sense noms propis (sense persones o llocs específics), d) sense termes tècnics, i e) pensa les paraules pel teu compte (i no en el que pot estar al teu voltant en el moment de la tasca).”

Per puntuar aquestes respostes de DAT, cada conjunt de paraules es pot introduir al [lloc web](#) de DAT que pot retornar dos valors, la puntuació bruta (d'un rang de 0 a 200), i el rang percentil d'aquesta puntuació, comparant la puntuació bruta amb el nombre de participants que han completat el DAT. Puntuacions altes indiquen distàncies més llargues, és a dir, paraules dissimilars. Cropley (2023) va utilitzar el DAT per comparar també la creativitat entre GPT-3.5 i GPT-4.

Els resultats de les investigacions actuals (Cropley 2023; Koivisto & Grassini, 2023; Hubert et al., 2024) que avaluen la creativitat humana vs. la intel·ligència artificial han conclòs que GPT-4 resulta en una major originalitat que els participants humans, però una viabilitat o adequació inferior d'una idea o producte i una freqüència més alta de paraules repetides en comparació amb els seus homòlegs humans. Les diferències també es van atribuir a la complexitat de les paraules utilitzades pels dos grups; la IA va utilitzar més paraules per a conceptes abstractes com la llibertat, mentre que els humans van utilitzar paraules per a conceptes més tangibles i observables. El rendiment de GPT-4 en el DAT en relació amb

els humans es va trobar suficient per substituir els humans en el pensament divergent verbal, però no hi ha cap garantia que aquest sigui el resultat en tots els casos, fent que un humà d'alt rendiment i constant sigui més fiable que la IA. Finalment, van concloure que tot i que l'IA intel·ligència artificial va tenir un rendiment millor que els humans, no va superar consistentment els millors participants humans. Els humans estaven sobrerrepresentats en respostes comunes o de baixa qualitat probablement degut a funcions executives, com la memòria de treball. En qualsevol cas, aquests exemples intenten mostrar si la creativitat de la IA pot substituir amb èxit la creativitat humana, mentre que la pregunta adequada hauria de ser com pot beneficiar-se la creativitat humana de la IA.

11.2. Pensament lateral i màrqueting

El pensament divergent és molt proper al que Edward de Bono va anomenar el 1967 pensament lateral, que és un procés deliberat estretament relacionat amb la intuïció, la creativitat i l'humor. Com a procés generatiu, es pot estar equivocat en algun moment per tal d'arribar a la solució correcta, o buscar informació irrellevant per inspirar-se. Basat en el pensament lateral, Philip Kotler (2005) - un reconegut expert en màrqueting - va introduir el màrqueting lateral i, el que és més important, tècniques que permeten al personal creatiu de màrqueting experimentar i generar idees alternatives per resoldre problemes de mercat.

Un exemple típic presentat al seu llibre és l'oferta de roses vermelles durant el dia de Sant Valentí. Quan s'apliquen les tècniques recomanades per ser més creatius, els resultats podrien ser els següents:

- Substitució: és a dir, substituir un component, producte o servei per alguna cosa que serveixi per a un propòsit similar. Per exemple, en lloc de roses, oferir lliris o caramels o fins i tot roses negres.

- Combinació: significa la combinació de dos o més productes o serveis existents per crear una nova oferta. Per exemple, oferir roses en un test o un manual de com fer créixer roses en un test.
- Inversió: és a dir, canviar la seqüència o ordre d'accions per crear un nou producte o servei. Per exemple, en lloc d'oferir roses el dia de Sant Valentí, oferir-les cada dilluns a la teva parella sentimental, abans d'anar a treballar.
- Eliminació: és a dir, eliminar un component o característica d'un producte o servei per crear una solució més senzilla o sorprenent. Per exemple, oferir roses que no facin olor, o no oferir cap rosa en absolut.
- Exageració: que es refereix a amplificar o magnificar un component, característica o benefici d'un producte o servei. Per exemple, en lloc d'oferir una rosa, omplir la casa de la parella amb roses o, encara més vergonyós, la seva oficina.
- Reordenació: és a dir, canviar la posició o seqüència dels components o processos del producte. Per exemple, en lloc que la parella masculina ofereixi roses a la femení, hauria de ser a l'inrevés.
- Reposicionament: és a dir, canviar la percepció o el mercat objectiu del producte o servei. Per exemple, promoure roses vermelles com un regal d'autoestima animant les persones a comprar roses vermelles per a elles mateixes com un gest d'autoapreciació i cura personal.

Aquestes tècniques, sense l'assistència de la IA, han permès a les empreses identificar noves oportunitats, estimular la creativitat i desenvolupar productes i serveis innovadors que satisfacin les necessitats canviants dels consumidors. Per tal d'accelerar el procés i fins i tot millorar la qualitat de les idees finals, la col·laboració de la IA pot donar bons resultats.

Pràctica. La nova vida d'un clip

Aquesta és una pràctica per mostrar la capacitat de creativitat que pot aportar un model de llenguatge. Per tal de desenvolupar l'activitat, el grup classe es dividirà en dos grans grups A i B. El primer grup serà el responsable de la generació de dades i el segon grup serà el responsable de l'avaluació de les dades; després, de forma conjunta, analitzaran els resultats i avaluaran la competència creativa dels humans i de les màquines. L'objectiu és millorar les habilitats de pensament creatiu dels estudiants fomentant el pensament divergent, ampliant la generació d'idees, fomentant l'originalitat i la innovació, i millorant la flexibilitat i l'elaboració en les seves respostes.

Creació de dades

Demaneu a cada estudiant del grup A per separat que proposi tants usos alternatius per a un clip de paper com sigui possible, en un termini de 5 minuts. La instrucció precisa pot ser: *"Desenvolupa tantes idees originals i creatives com sigui possible per a un clip de paper. L'objectiu és generar idees que siguin intel·ligents, inusuals, innovadores i captivadores per a les persones. Mentre siguin creatives, aquestes idees no han de ser pràctiques ni realistes; poden ser absurdes i estranyes sempre que siguin creatives. La tasca durarà 5 minuts durant els quals pots introduir tantes idees com vulguis, recordant la importància de la qualitat, no la quantitat. Presenta una llista d'aquestes idees."*

Participació del model de llenguatge

Després demaneu als estudiants que introdueixin una instrucció similar al model de llenguatge proposat (per exemple Gemini o Claude), limitant el nombre de suggeriments a 10. La instrucció podria ser la següent: *"Desenvolupa 10 idees originals i creatives per a un clip de paper. L'objectiu és generar idees que siguin intel·ligents, inusuals, innovadores i captivadores per a les persones. Mentre siguin creatives, aquestes idees no han de ser pràctiques ni realistes; poden ser absurdes i estranyes sempre que siguin creatives."*

Respostes millorades

Demaneu als estudiants que mirin les respostes de la IA, escullin segons el seu criteri les tres millors respostes i les afegeixin a la seva llista en ordre aleatori, eliminant qualsevol indicació que han estat creades per ChatGPT.

Avaluació de la creativitat divergent

A continuació, demaneu als estudiants del grup B que avaluïn les respostes dels estudiants del grup A. Mitjançant un marc d'anàlisi qualitativa, demaneu-los individualment que avaluïn el següent:

- Fluïdesa (Comptar el nombre de respostes úniques per a cada estudiant).
- Originalitat (Comptar el nombre de respostes úniques per a cada estudiant).
- Flexibilitat (Identificar diferents categories, p. ex., usos pràctics, usos creatius, usos d'emergència, i comptar quantes categories cobreixen les respostes de cada estudiant).
- Elaboració (Avaluar el nivell de detall i exhaustivitat de les respostes).

Discussió i reflexió

El grup B ha de justificar quines respostes han considerat com les més i menys creatives, així com quin estudiant al final ha estat el més creatiu. Després, el grup A ha de revelar les respostes que han estat generades per IA. En aquest punt, el grup A ha de discutir en forma de grup de discussió, com l'ús de la IA ha influït en el procés creatiu en termes de fluïdesa, originalitat, flexibilitat i elaboració.

Pràctica. Regals per a la graduació

Els estudiants i les estudiants treballaran en grups per desenvolupar una campanya de màrqueting per a productes o serveis únics relacionats amb l'obsequi per a la cerimònia de graduació utilitzant tècniques de desplaçament. La tasca es dividirà en tres fases amb temps específics assignats. L'objectiu és utilitzar tècniques de desplaçament en el màrqueting lateral generant i refinant idees creatives de màrqueting amb i sense l'ajuda de ChatGPT, comparar els resultats i demostrar com la IA pot millorar la generació d'idees. Com el valor de la creativitat és determinat en última instància per l'agent humà, és necessari seleccionar un exemple en el qual les/les estudiants siguin considerats com a usuaris del producte o servei.

Fase 1: Generació d'idees sense IA

1. Sessió de pluja d'idees en grup (30 minuts):

Els estudiants fan una pluja d'idees sobre productes o serveis innovadors per a regals de graduació utilitzant aquestes tècniques de desplaçament:

- Eliminació: Eliminar un component d'un regal tradicional per crear un nou concepte.
- Substitució: Substituir un element d'un regal tradicional per alguna cosa inesperada.
- Inversió: Revertir el procés o ordre habitual d'un regal tradicional.
- Combinació: Fusionar elements de diferents productes o serveis per crear una nova oferta.

2. Documentar les idees (15 minuts):

Cada grup documenta les seves 5 millors idees innovadores, explicant el concepte i la tècnica de desplaçament aplicada.

Fase 2: Generació d'idees amb IA

1. Utilització del model de llenguatge per a la pluja d'idees:

Cada grup utilitza el model de llenguatge (com ara, Copilot) per generar idees per a la mateixa tasca, demanant-li escenaris que impliquin tècniques de desplaçament. Per exemple, *"Suggeriu idees de regals de graduació eliminant components tradicionals"* o *"Quins són alguns regals de graduació únics invertint el procés típic?"*

2. Documentar les idees generades per la IA (15 minuts):

Cada grup documenta les 5 millors idees suggerides per ChatGPT, incloent-hi les instruccions utilitzades i les tècniques de desplaçament aplicades.

Fase 3: Refinament de les idees generades per la IA

1. Comparació i selecció:

- Els grups comparen les idees generades a la Fase 1 i a la Fase 2, en termes d'originalitat i temps.
- Seleccionen les millors idees de les dues fases i combinen els coneixements per millorar-les encara més.
- Consideren la viabilitat, l'atractiu del mercat i l'impacte potencial mentre refinem les idees.

2. Presentació final:

Cada grup prepara una presentació de la seva campanya de màrqueting final i refinada, destacant:

- Idees originals de la Fase 1.
- Idees generades per la IA de la Fase 2.
- Idees finals millorades.
- Coneixements sobre com l'assistència de la IA ha millorat la seva estratègia de màrqueting.

Referències

Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, 148, 107903.

Cropley, D. (2023). Is artificial intelligence more creative than humans?: ChatGPT and the divergent association task. *Learning Letters*, 2, 13-13.

Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), 3440.

Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, 13(1), 13601.

Kotler, P., & de Bes, F. T. (2005). *Lateral Marketing*, Published by arrangement with John Wiley&Sons.

Man, N., Wang, K., & Liu, L. (2021). Using computer cognitive atlas to improve students' divergent thinking ability. *Journal of Organizational and End User Computing (JOEUC)*, 33(6), 1-16.

12

El nou professor ajudant és un replicant

Rachael: Li importa si fumo? [Deckard ajusta la màquina de Voight Kampff]

Deckard: Això no afectarà el test. Bé, li faré una sèrie de preguntes. Relaxi's i respongui-les el més senzillament possible. [pausa] És el seu aniversari, i li regalen una cartera de pell.

Rachael: No l'acceptaria, i a més denunciaria a la policia la persona que me la regalés.

Deckard: Té un fill. Aquest li ensenya la seva col·lecció de papallones i un flascó d'arsènic.

Rachael: El portaria al metge.

Deckard: Està veient la televisió. De sobte, s'adona que una vespa li puja pel braç.

Rachael: La mataria.

A Blade Runner, l'inspector Deckard utilitza un enginy anomenat *Voight Kampff* per tal d'esbrinar si la persona entrevistada és en realitat un replicant. Els replicants són autòmats que tenen l'aparença humana amb un nivell de precisió tan excels que poden passar per humans, i només el test (una mena de versió literària del Test de Turing) és capaç de detectar l'absència d'humanitat. En un manual

sobre l'ús de la intel·ligència a les aules, aquesta és una pregunta immediata: Poden els enginyers mecànics substituir els professors i esdevenir una alternativa? En un capítol posterior, plantejarem aquesta qüestió amb l'ajut de les pel·lícules que ens han inspirat, però en aquest cas l'objectiu és més limitat: Pot la IA ser un element que complementi, que ajudi en les tasques docents?

“A Blade Runner, l'inspector Deckard utilitza un enginy anomenat Voight Kampff per tal d'esbrinar si la persona entrevistada és en realitat un replicant”

Hi ha tres moments en la tasca docent:

- La preparació, és a dir tot allò que forma part de l'activitat docent abans de la transmissió del coneixement. En forma part el procés d'aprenentatge, l'adequació entre els coneixements i el nivell de l'estudiantat, la creació de materials de suport,... Aquest manual, tant la seva elaboració com la seva lectura formen part d'aquest immens volum de feina no visible que és la pre-classe.
- L'acte de transmissió, que tradicionalment ha estat relacionat amb el model magistral de transmissió professora o professor a estudiant, però que podria ser qualsevol forma de transmissió, amb material de suport.
- L'activitat de reforç, que té com a objectiu no tant transmetre una informació com aportar eines i activitats que ajudin a treballar aquesta informació de manera efectiva.

Naturalment, aquestes tres fases clàssiques preparació - transmissió - reforç es poden alterar, com per exemple amb les classes invertides (*flipped classroom*), en les quals la interacció amb el professor o la professora és posterior a les activitats.

La intel·ligència artificial pot ser una eina de suport per les tres fases: En primer lloc, pot ajudar a sintetitzar informació, a resumir documents o a traduir informacions complexes a determinats nivells; en segon lloc, la IA pot ser un element que faciliti o complementi la transmissió de la informació en l'acte de classe, segons el model de classe magistral;

i finalment, la IA permet la creació de material de suport complementari. En aquest capítol ens centrarem essencialment en aquest darrer àmbit i mostrarem possibles aplicacions per a incrementar el material i adaptar-lo a les característiques del grup classe.

12.1. Slides automàtics

Les presentacions són recursos de suport que ajuden a seguir un determinat guió, a incorporar material audiovisual de suport, a presentar de forma ordenada informació sintètica i a proporcionar l'estudiant material en el procés d'estudi. Les presentacions han assolit un protagonisme molt rellevant en la docència de les ciències socials, però també han estat objecte de crítiques per la seva manca de flexibilitat i per la forma com condiciona un model de transmissió de la informació. L'eina més universal és Powerpoint, amb totes les variants (com Keynote de Mac, Impress de LibreOffice, Zoho o Google Slide), si bé també són habituals altres eines que surten d'aquest estàndard com Visme o el model de Prezi, que ha tingut una certa incidència especialment en la formació secundària.

Els slides automàtics transformen el text en presentacions ordenades, ben estructurades i visualment atractives, en molt poc temps. A més, ofereixen opcions de personalització per adaptar-se a les necessitats de cada persona i cada context (MacFarland, 2024).

Les característiques principals dels slides automàtics són:

- Conversió de text a diapositives: A partir del text introduït (ja sigui copia i enganxa o des d'un fitxer), es crea una presentació amb diapositives.
- Disseny atractiu: Apliquen dissenys visuals predefinits amb colors, tipografies, icones i fons agradables per fer les diapositives més vistoses.

- Formats de sortida: Permeten exportar o descarregar la presentació generada en formats comuns com ara PDF, PowerPoint, imatges, etc.
- Personalització limitada: Algunes eines ofereixen opcions bàsiques de personalització com ara canviar colors, tipografies, afegir logotips, etc.

Hi ha diversos exemples de generadors de diapositives automàtics, entre els més comuns, podem destacar els següents:

	PROS	CONTRES
Google Slides AI	<ul style="list-style-type: none"> • Integració nativa amb Google Slides, facilitant l'ús i l'edició. • Generació ràpida de diapositives a partir de paraules clau o text previ. • Opcions de personalització limitades, com ara la tria de plantilles i fons. • Ideal per a crear esborranys ràpids o presentacions bàsiques. 	<ul style="list-style-type: none"> • Manca de funcions avançades de disseny i animació. • Les diapositives generades poden ser genèriques i poc atractives. • Control limitat sobre la disposició i el format dels elements.
Canva Magic Design	<ul style="list-style-type: none"> • Capacitat de generar diapositives a partir de text, imatges o enllaços web. • Eines d'IA per a la selecció d'imatges, fons i paletes de colors. • Opcions de personalització més avançades que Google Slides AI. • Ideal per a crear presentacions visualment atractives. • Àmplia gamma de plantilles i elements de disseny professionals. 	<ul style="list-style-type: none"> • Pla gratuït limitat a 5 diapositives per disseny i marca d'aigua. • Les subscripcions de pagament poden ser costoses per a usuaris individuals. • La interfície pot ser complexa per a principiants. • No ofereix la mateixa integració nativa amb Google Slides.
Prezi.ai	<ul style="list-style-type: none"> • Format no lineal únic que permet una presentació més conversacional. • Eines de zoom i panoràmica per a crear presentacions dinàmiques. • Varietat de plantilles i opcions de personalització. • Ideal per a presentacions narratives o que mostren molta informació. 	<ul style="list-style-type: none"> • La corba d'aprenentatge pot ser més pronunciada que amb altres opcions. • Les presentacions no lineals poden ser menys intuïtives per al públic. • El pla gratuït té funcions limitades i marca d'aigua. • No ofereix la mateixa integració nativa amb Google Slides.
Apple Keynotes	<ul style="list-style-type: none"> • Potent suite d'eines de disseny i animació integrades. • Interfície d'usuari elegant i fàcil d'utilitzar. • Integració nativa amb macOS i altres dispositius Apple. • Ideal per a crear presentacions professionals d'alta qualitat. 	<ul style="list-style-type: none"> • Disponible exclusivament per a dispositius Apple. • Pot ser costós per a usuaris que no tenen ja productes Apple. • La corba d'aprenentatge pot ser més pronunciada que amb altres opcions. • No ofereix la mateixa integració nativa amb Google Slides.

A banda d'aquests entorns més habituals, es poden trobar moltes altres propostes com ara Microsoft Sway, Beautifulai, Deckset, Flowchart Fun, Gamma AI, Tome AI, Slidemill... La llista segueix creixent, ja que cada vegada més eines incorporen capacitats de generació de diapositives amb IA per agilitzar tot el procés de creació de presentacions. En els webs oficials d'aquestes eines s'inclouen seccions de suport o recursos on es poden trobar tutorials, guies per l'usuari i altres recursos útils. Un clar exemple és [el tutorial de Slides.ai](#), un dels creadors de presentacions automàtiques més utilitzats. També, moltes empreses ofereixen vídeos explicatius a través dels seus canals de YouTube oficials. Els slides automàtics sovint es troben en la mateixa pàgina que un generador de diapositives manual. Aquesta coexistència pot causar diverses dificultats, tant tècniques com d'experiència d'usuari.

12.2. Vídeos automàtics

La creació de vídeos instantanis amb la IA s'està convertint en un dels avenços més recents en la producció de continguts educatius. Aquests recursos audiovisuals permeten al docent produir de manera molt ràpida, continguts que es poden presentar i estructurar de forma clara, ordenada i coherent; el principal avantatge és que la cultura de l'estudiantat és molt més audiovisual que textual, de manera que aquests continguts creats de forma automàtica poden ser un bon element de suport.

Els avantatges d'aquestes eines de generació automàtica són diverses (Brame, 2017; Guo et al, 2014):

- Accés instantani al contingut: Els estudiants i les estudiants poden accedir a lliçons, explicacions o demostracions en format de vídeo en qualsevol moment i lloc, cosa que els proporciona gran flexibilitat.
- Aprenentatge visual: Els vídeos poden fer que conceptes complexos siguin més fàcils d'entendre combinant explicacions auditives amb elements visuals com animacions, gràfics, exemples pràctics, etc.

- Ritme personalitzat: Cada estudiant pot parar, rebobinar, tornar a veure tot el vídeo o parts del vídeo segons les seves necessitats, cosa que permet un aprenentatge més personalitzat i al seu propi ritme.
- Atenció i compromís: Els vídeos ben produïts amb bona qualitat audiovisual tendeixen a mantenir millor l'atenció i el compromís dels estudiants en comparació amb altres mètodes.
- Classe invertida (*Flipped Classroom*): Es poden utilitzar per a la instrucció prèvia abans de les classes presencials en un model "aula invertida", deixant el temps de classe per a discussions, pràctiques i resolució de dubtes.
- Disponibilitat permanent: Els vídeos es poden allotjar en línia i estar disponibles de manera permanent per a repàs o referència futura.

Les eines de creació de presentacions, com PowerPoint i Google Slides, ofereixen la capacitat d'afegir gravacions de veu i vídeo a les diapositives, facilitant la creació de contingut multimèdia. Pel que fa a les plataformes educatives com Blackboard i Canvas, permeten gravar sessions de classe i allotjar-les en la mateixa plataforma, facilitant l'accés dels estudiants al contingut. YouTube, com a plataforma de vídeo més utilitzada, ofereix eines per carregar i allotjar vídeos, amb funcions d'edició bàsiques per a aquells que desitgen fer modificacions senzilles. Els *screencast* són eines que permeten gravar la pantalla d'un dispositiu combinant-ho amb les explicacions del que es mostra. És una eina molt útil per crear tutorials, presentacions, demostracions, lliçons interactives i material didàctic audiovisual. Entre les més populars hi ha Camtasia, Screencast-O-Matic, i Loom, ScreenFlow.

Hi ha eines i serveis en línia que permeten generar vídeos automàticament a partir de documents, presentacions o simplement text. Tot i que tenen menys opcions de personalització que un editor de vídeo tradicional, poden ser molt útils quan cal crear vídeos ràpidament sense massa treball manual. Algunes opcions populars són:

	PROS	CONTRES
Podcastle	Fàcil d'utilitzar Bona qualitat de veu Ofereix edició de so	Opcions de veu limitades Pot ser car per funcions avançades
Descript	Edició avançada Conversió text a veu d'alta qualitat Integració amb altres eines	Corba d'aprenentatge Requereix subscripció
Play.ht	Veus realistes Opció de múltiples idiomes Interfície intuïtiva	Funcions avançades limitades sense pagament Personalització limitada
Lovo.ai	Gran varietat de veus Fàcil d'utilitzar Veus naturals i personalitzables	Requereix subscripció per a funcions premium Algunes veus poden sonar artificials
Sonantic	Veus emocionals i naturals Bona per narrativa i dramatització	Car Menys conegut, menys suport comunitari
Voicery	Varietat de veus Fàcil integració amb altres plataformes	Opcions limitades en la versió gratuïta Personalització limitada

	PROS	CONTRES
Lumen5	<ul style="list-style-type: none"> • Converteix text a vídeo ràpidament. • Crea vídeos a partir d'articles, publicacions a les xarxes socials i correus electrònics. • Varietat de plantilles i estils. • Opcions d'edició bàsiques. 	<ul style="list-style-type: none"> • Plantilles limitades a la versió gratuïta. • Opcions de personalització limitades. • Marca d'aigua a la versió gratuïta.
Pictory AI	<ul style="list-style-type: none"> • Crea vídeos a partir de guions, text o idees. • Utilitza intel·ligència artificial per a seleccionar imatges i música. • Varietat de plantilles i estils. • Editor potent amb funcions avançades. 	<ul style="list-style-type: none"> • Pla gratuït limitat a 25 vídeos al mes. • Marca d'aigua a la versió gratuïta. • La corba d'aprenentatge pot ser més pronunciada.
Animoto	<ul style="list-style-type: none"> • Fàcil d'utilitzar, amb una interfície basada a arrossegar i desar • Àmplia gamma de plantilles professionals. • Música i efectes de so lliures de drets d'autor 	<ul style="list-style-type: none"> • Plantilles limitades a la versió gratuïta. • Opcions de personalització limitades. • Marca d'aigua a la versió gratuïta.
Raw Shorts	<ul style="list-style-type: none"> • Crea vídeos curts a partir de clips de vídeo i imatges. • Edició potent amb funcions avançades d'Adobe Premiere Rush. • Integració amb altres aplicacions d'Adobe. 	<ul style="list-style-type: none"> • Requereix una subscripció a Adobe Creative Cloud. • La corba d'aprenentatge pot ser més pronunciada.
Deep Dream Generator	<ul style="list-style-type: none"> • Crea vídeos abstractes a partir d'imatges o text. • Utilitza intel·ligència artificial per a generar imatges úniques. • Resultats imprevisibles i creatius. 	<ul style="list-style-type: none"> • Pot ser difícil controlar els resultats finals. • La qualitat de vídeo pot variar.
FlexClip	<ul style="list-style-type: none"> • Varietat de plantilles i estils. • Edició fàcil d'utilitzar amb funcions bàsiques. • Música i efectes de so lliures de drets d'autor. 	<ul style="list-style-type: none"> • Plantilles limitades a la versió gratuïta. • Opcions de personalització limitades. • Marca d'aigua a la versió gratuïta.

Aquest són només alguns exemples, la llista pot anar variant depenent dels avenços tecnològics, que augmenten ràpidament, i de les noves plataformes que puguin aparèixer els anys vinents. Malgrat que els vídeos instantanis estan guanyant molta popularitat en l'àmbit educatiu, és important que

el docent mantingui sempre el control i supervisió dels continguts que hi vulguin editar. En el moment d'editar aquest manual, ChatGPT encara no ha activat Sora, que permet la creació de vídeos a partir d'un simple *prompt*.

12.3. Pòdcasts de suport

En una cultura audiovisual, els pòdcasts s'han convertit en un mitjà de comunicació molt popular, especialment entre els joves. El seu principal atractiu és la possibilitat de seguir el fil d'un tema mentre es realitzen altres activitats, com ara fer exercici, desplaçar-se, passejar o dur a terme tasques domèstiques. El 2024, es preveu que el nombre d'oients de pòdcasts en l'àmbit mundial arribi als 504,9 milions, un augment significatiu respecte als 464,7 milions del 2023. A Espanya, el consum de pòdcasts també ha crescut considerablement, amb el 41% dels usuaris d'internet escoltant pòdcasts regularment el 2021.

La intel·ligència artificial (IA) està revolucionant aquest àmbit de diverses maneres. Els pòdcasts poden incorporar veus sintètiques, com les que ofereixen Google [Cloud Text-to-Speech](#), Amazon Polly o Azure Cognitive Services, per convertir text en veu i viceversa. Això permet als creadors de contingut generar narracions d'alta qualitat de manera eficient i a menor cost. Una altra innovació destacada és la traducció de pòdcasts a altres idiomes mantenint la veu original del locutor. Spotify, en col·laboració amb OpenAI, ha desenvolupat una eina que permet traduir pòdcasts replicant la veu del locutor original, trencant així les barreres lingüístiques i ampliant l'audiència potencial. Podeu escoltar un exemple d'això amb Yuval Noah Harari en aquest enllaç.

A més, la IA està transformant la producció i l'edició de pòdcasts. Eines com Descript i Adobe permeten la transcripció automàtica, l'edició basada en text i la millora de l'àudio amb tecnologies impulsades per IA. Aquestes eines faciliten la producció de contingut d'alta qualitat i permeten als creadors centrar-se més en el contingut creatiu i menys en les tasques tècniques. Un dels aspectes més innovadors és la creació de pòdcasts íntegrament per IA. Això inclou narracions, guions i entrevistes generades automàticament, amb veus artificials que imiten les humanes gràcies a l'entrenament

amb mostres de veu real. Plataformes com [Podcast.ai](#) han creat pòdcasts on famosos com Steve Jobs o Oprah Winfrey semblen participar en entrevistes, tot i que tot el contingut ha estat generat per IA a partir de converses públiques i entrevistes anteriors.

Pel docent, les eines de generació automàtica permeten la creació de recursos sonors amb poc esforç, que es poden complementar amb la resta d'eines, com una presentació o un vídeo. En tots els casos, és imprescindible que el docent supervisi el resultat i valori l'output, per evitar errors o imprecisions. Aquests són alguns exemples de programes de suport a la creació de pòdcasts.

12.4. Qüestionaris automàtics

Una de les tasques del professor és la preparació de qüestionaris i proves d'avaluació que permetin valorar els aprenentatges dels alumnes. Sovint aquesta és una feina feixuga i que sol consumir molt de temps. La IA ha afavorit que sorgeixin un seguit d'eines que permeten generar de forma automàtica proves i qüestionaris de diversos tipus, ja sigui amb preguntes de resposta múltiple, preguntes de vertader o fals, així com, preguntes obertes.

Aquestes eines comporten un seguit d'avantatges pel professor, destacant especialment les següents:

- Creació de forma ràpida i fàcil de qüestionaris amb correcció automàtica, reduint la càrrega de feina pel docent.
- Elevat nivell de personalització, tenint en compte les necessitats i nivells dels alumnes.
- Generació d'informes i estadístiques que ajuden a identificar àrees de millora i fer un seguiment acurat del progrés dels alumnes.
- Integració d'elements interactius i de gamificació, incrementant l'atractiu de les proves.

Algunes de les eines actualment disponibles que permeten la generació de qüestionaris emprant la IA són: Quizizz, Opexams, ProProfs Quiz Maker o Socrative. Aquestes eines de forma general permeten poder

generar preguntes seguint les instruccions que se li hagin donat (nombre de preguntes, tipus, dificultats, etc.), ja sigui indicant un tema de forma breu, o bé introduint un text complet.

Per exemple, Quizizz és una eina que permet crear qüestionaris i activitats de forma ràpida a partir de vídeos, PDF i webs, que es poden desplegar en 18 modalitats diferents de pregunta. Així mateix, permet condicionar el temps d'execució del qüestionari o activitat, decidint que sigui d'avaluació o simplement d'aprenentatge, i permet vincular-ho a una "Classroom" de Google.

Per la seva banda, Openexams és una altra eina que facilita l'elaboració de qüestionaris i proves d'avaluació, no només generant preguntes, sinó també actuant com un assistent que també permet corregir exàmens. Dins les funcionalitats de correcció, dona la possibilitat de posar les notes de forma automàtica (AI Auto Grading) i generar rubriques de correcció. Al mateix temps, deixa també donar format i corregir exàmens en paper. Finalment, cal remarcar que permet fer una anàlisi dels resultats obtinguts.

En el cas de ProProfs Quiz, a més de crear qüestionaris generals, també permet la creació de qüestionaris d'avaluació, testos de personalitat i avaluació d'habilitats. Així mateix, ofereix altres funcionalitats que permeten garantir la seguretat i la prevenció de la còpia en proves d'avaluació. La singularitat de Socrative és que es tracta d'una eina que permet crear interacció a l'aula a través de la creació de tests, qüestionaris i enquestes en temps real, i feedback instantani.

Algunes d'aquestes eines també estan concebudes per ajudar a l'estudiant durant el procés d'estudi, generant qüestionaris automàtics a partir dels seus apunts, o bé fent petites targetes (Flashcards) de resum de la matèria per facilitar la memorització. Aquest és el cas de Quizlet i Quizgecko. Evidentment, aquestes eines també poden ser emprades pel professor per poder desenvolupar materials d'avaluació i estudi pels alumnes.

Així mateix, hi ha eines d'IA especialitzades en la generació automàtica de rúbriques. La rúbrica és una eina utilitzada per avaluar el rendiment dels alumnes de manera objectiva i consistent. De forma general, la rúbrica es configura a partir d'una taula que descriu els criteris d'avaluació i els nivells de rendiment per a cada criteri. Eines com Rubrick, Easy-Peasy o Megaprofe ajuden a definir els objectius d'aprenentatge, especificar criteris d'avaluació adaptats, i obtenir la corresponent rúbrica. A diferència d'eines web de generació de rúbriques sense IA, aquestes eines permeten una major adaptació a les condicions específiques d'avaluació de l'activitat que es proposa, i per tant, una major precisió, personalització i efectivitat en el procés d'avaluació.

Tots aquests sistemes d'avaluació són un complement a la tasca docent, però no un substitut. La creació de qüestionaris automatitzats és un recurs que es pot proporcionar de manera complementària a altres eines com la generació vídeos, slides o pòdcasts. Ajuden l'estudiant a testar el seu grau de coneixement de la matèria i poden ajudar el docent a valorar el grau de seguiment de l'assignatura. Però aquests sistemes no poden ser incorporats en el model d'avaluació de l'estudiantat, que ha de continuar essent una tasca essencialment humana. En la legislació europea sobre IA es detallen una sèrie de riscos sobre la implantació d'aquestes eines i precisament una de les activitats considerades d'alt risc i totalment improcedent és la revisió automatitzada d'exàmens o de proves de nivell. Les autoritats europees insisteixen que aquestes pràctiques poden donar lloc a biaixos, a errors o a decisions injustes, de manera que els criteris d'avaluació deuen prevaldre com una activitat essencialment humana. Els qüestionaris de suport han de ser considerats eines complementàries d'autoavaluació per l'estudiantat.

Pràctica. La Revolució industrial a la pantalla

L'objectiu d'aquest exercici és demostrar que l'elaboració de vídeos instantanis poden ser eficaços per augmentar la participació dels estudiants a l'aula i millorar el procés d'aprenentatge. A diferència de la resta de pràctiques dissenyades en aquest manual, en aquest cas es dirigeix al docent i no pas a l'estudiantat.

Aquest exercici planteja la realització d'una classe invertida utilitzant vídeos sobre determinats conceptes que s'utilitzen com instrucció prèvia a la classe presencial. Per posar un exemple, en un curs de matemàtiques, en lloc de dedicar temps a classe a explicar conceptes, el professor genera vídeos automàtics explicatius curts que els estudiants i les estudiants miren a casa abans de la classe presencial. A l'aula, els estudiants i les estudiants treballen en exercicis pràctics i problemes sobre els conceptes presentats en els vídeos, amb l'ajuda i guiatge del docent. Aquesta activitat es pot realitzar en qualsevol tema o camp d'estudi. Especialment, en matèries on la interacció i l'aplicació pràctica dels coneixements sigui important.

L'exercici tindrà dues parts. La primera constarà de la producció del vídeo automàtic. La segona part de l'aplicació pràctica del vídeo.

En una assignatura d'Història, el tema a desenvolupar és l'impacte de la Revolució industrial. Prèviament a la classe presencial, el professor haurà generat una sèrie de vídeos curts (10-15 minuts per vídeo) que tractin conceptes claus que es vulguin treballar (antecedents de la Revolució; causes-factors; principals avenços; impactes; etc.). Per generar els diferents vídeos el professor pot decidir quina eina vol utilitzar (Lumen5, Pictory AI, Animoto, FlexClip...). Aquestes plataformes, com s'ha explicat anteriorment, generen vídeos de manera instantània. Per aquest exercici recorrem a Pictory AI (que disposa de demo i un *free trial*-prova gratuïta).

Primerament, cal registrar-se al web de Pictory AI i s'inicia la sessió. Seguidament, se selecciona el tipus de vídeo. Pictory AI ofereix diferents opcions:

- Article to Video: Converteix un article o bloc en un vídeo
- Script to Video: Converteix un guió de text en un vídeo
- Visuals to Video: Converteix imatges o clips en un vídeo.

Aleshores, ja es pot penjar l'URL de l'article o pujar el text del document, introduir el guió (el text de la narració), o les imatges. Automàticament, Pictory dividirà el text en escenes o segments. Es pot ajustar la divisió segons les preferències, i triar el format de vídeo. Així mateix, es pot personalitzar el disseny. Pictory disposa de diverses plantilles visuals, i es pot ajustar el text de cada escena, afegir imatges, mapes, animacions, música i veu en off. Cal assegurar-se que el contingut generat automàticament sigui precís i rellevant per a l'audiència. Es poden fer tants vídeos com necessitem. Els vídeos s'allotjaran a la plataforma de l'assignatura per tal que els estudiants el puguin consultar (mirar, rebobinar, prendre notes...).

La segona part de l'exercici, comença un cop els estudiants han visualitzat els vídeos. Aleshores, a l'aula, els estudiants creen una exhibició virtual o un museu en línia sobre la Revolució industrial. Cada estudiant o grup d'estudiants és responsable d'un dels aspectes específics de la Revolució industrial tractats en els vídeos, i haurà de produir contingut multimèdia (vídeos, animacions, infografies, etc.) d'aquell aspecte. Un cop tots els grups hagin finalitzat, es comparteix tot el material i es transita pel museu virtual. Aquest exercici ajuda d'una part als professors a presentar els conceptes de manera clara i ordenada, i d'altra als estudiants a augmentar la participació a classe, aplicar coneixements adquirits i millorar els resultats d'aprenentatge.

Pràctica. Tests automàtics

Aquest exercici té com a objectiu mostrar que l'elaboració de qüestionaris interactius pot ser eficaç per avaluar els coneixements dels estudiants de manera dinàmica, i és d'aplicació a qualsevol tema o camp d'estudi. Especialment, en matèries on la comprensió i l'aplicació dels coneixements sigui important. Consisteix en la creació d'una activitat d'avaluació utilitzant una eina de generació de qüestionaris sobre determinats conceptes que es volen avaluar després d'una sessió d'aprenentatge, generalment, una classe expositiva. Per posar un exemple, en un curs de ciències, en lloc de fer una avaluació tradicional en paper, el professor crea un qüestionari interactiu que els estudiants completen en línia.

L'exercici tindrà dues parts. La primera constarà de la creació del qüestionari interactiu. La segona part de l'aplicació pràctica del qüestionari a l'aula.

Primera part

Imaginem que impartim una assignatura de Geografia, i tractem el tema del canvi climàtic. Abans de la classe, el professor o professora haurà de crear un qüestionari amb preguntes clau que es vulguin treballar (causes del canvi climàtic, efectes sobre el medi ambient, solucions proposades, etc.). Per crear el qüestionari, el professor pot decidir quina eina vol utilitzar (Quizgecko, Quizizz, Opexams, ProProfs Quiz Maker...). Aquestes plataformes, com s'ha explicat anteriorment, generen qüestionaris de manera interactiva.

Per aquest exercici recorrem a Quizgecko (que disposa d'una versió gratuïta). Per crear el qüestionari cal registrar-se a través de la web de Quizgecko i iniciar la sessió. Tot seguit, es copia i enganxa en el quadre de diàleg un text que resumeixi el tema que es vol avaluar, en aquest cas, canvi climàtic (hi ha també l'opció de escriure només el tema, posar una URL, pujar un document o escriure les preguntes manualment). Quizgecko ofereix diferents opcions de preguntes: opció múltiple, vertader/fals, preguntes obertes, etc. També es pot decidir la llengua, el nivell de dificultat i el nombre màxim de preguntes (la opció gratuïta permet un màxim de 10 preguntes). Clicant "next" es genera de forma automàtica les 10 preguntes en base al text que li hem introduït. Una vegada es disposa de les preguntes cal assegurar-se que són clares i adequades pel tema i el nivell dels alumnes. A continuació cal procedir a guardar l'enllaç del qüestionari.

Segona part

Un cop s'ha fet la classe expositiva, es procedeix a donar l'enllaç als alumnes per tal que complementin el qüestionari. Aquí es pretén iniciar una fase més participativa. Una vegada tots els alumnes hagin respost el qüestionari, es revisaran les respostes i es fa una discussió conjunta sobre les preguntes més difícils i els errors comuns. Aquesta dinàmica permetrà que els alumnes puguin comprovar si han entès el tema tractat. Així mateix, ajudarà al professor a tenir un feedback dels resultats de la classe, i per tant, millorar el procés i resultat d'aprenentatge dels alumnes.

Referències

- Brame, C. J. (2017). *Effective educational videos: Principles and guidelines for maximizing student learning from video content*. CBE—Life Sciences Education.
- Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50).
- MacFarland, A. (2024, maig). Los diez mejores generadores de presentacions. (En línia). Disponible a <https://www.unite.ai/es/generadores-de-plataforma-de-lanzamiento/> (Consulta 16/5/2024)
- Singh, N (2023, octubre). Spotify's Newest Feature: Using AI to Clone and Translate Podcast Voices Across Languages.(En línia) Disponible a <https://www.marktechpost.com/2023/10/03/spotify-s-newest-feature-using-ai-to-clone-and-translate-podcast-voices-across-languages> (Consulta 31/5/2024)
- Territoris.cat (2024, abril). La creació de vídeos instantanis és el pròxim salt en la tecnologia d'IA.(En línia). Disponible a <https://www.territoris.cat/articulo/comunicats/creacio-videos-instantanis-es-proxim-salt-tecnologia-d-ia/20240405071207097661.html> (Consulta 20/5/2024)
- Vázquez, J.A. (2019, octubre). La inteligencia artificial en la creación de podcasts. Dosdoce.com. (En línia). Disponible a <https://www.dosdoce.com/2022/10/19/la-inteligencia-artificial-en-la-creacion-de-podcasts> (Consulta 28/5/24)

13

El docent 24/7

Els anys 60 la ciència-ficció va crear projeccions fascinants d'un futur utòpic on les màquines estaven al servei de les persones. Hanna Barbera va crear els *Supersonics*, el *Jetsons*, que imaginava la vida quotidiana de la família en un món de cotxes voladors, enginys i sistemes automatitzats. En aquesta família del futur, hi vivien dos androïdes, *Rosie*, una empleada de la llar encarregada de les tasques domèstiques, i *Di Di*, una mena de diari interactiu que recorda la veu omnipresent de HAL 9000, la computadora de *2001, una odissea a l'espai*. *Rosie* és una assistent, un androïde que interactua amb els humans, resol els seus problemes i aporta grans dosis de seny enmig d'una certa disbauxa. HAL 9000 o *Rosie* formen part del vell somni dels primers dissenyadors de la IA, convençuts que aviat les màquines podrien assumir la capacitat de raonament i de decisió dels humans.

Fins al moment, hem vist eines que poden respondre a qüestions concretes, seguint el model d'IA tova que vam veure en el primer capítol: eines que poden respondre a preguntes, que poden codificar

una entrevista, que poden calcular una correlació o que poden crear un vídeo a partir d'un text. Els assistents virtuals són eines creades amb un generador de llenguatge natural, que poden estar altament especialitzats en una determinada tasca, poden assumir determinats perfils o rols i poden, en definitiva, actuar per optimitzar unes funcions concretes.

Com hem vist en el primer capítol, els antecedents dels assistents virtuals es remunten a la dècada de 1960 amb programes com ELIZA, un primitiu processador de llenguatge natural (Soofastaei, 2021), i podríem trobar exemples precedents amb els enginys que jugaven a escacs. Tanmateix, els assistents es van popularitzar a partir de la dècada del 2000 amb l'aparició dels assistents virtuals basats en instruccions de veu. L'assistent virtual d'Apple Siri, llançat al mercat el 2011 com a opció de l'iPhone 4S, va ser un dels primers a introduir la veu com a interfície per interactuar amb els dispositius mòbils (Soofastaei, 2021). A poc a poc, els assistents basats en instruccions de veu s'han generalitzat i estan

“Fins al moment, hem vist eines que poden respondre a qüestions concretes, seguint el model d'IA tova que vam veure en el primer capítol”

presentes amb la domòtica, els assistents de llar o en el suport a la conducció. L'emergència dels models de llenguatge ha permès la creació d'assistents més sofisticats, a partir de la capacitat que tenen aquests sistemes d'interactuar i de convertir la veu en instruccions concretes.

Entre alguns dels assistents virtuals més reconeguts hi trobem Siri d'Apple, Google Assistant, Alexa d'Amazon, o Cortana de Microsoft, encara que avui en dia el ventall de possibilitats és molt gran. Si bé els models de llenguatge vigents són essencialment textuals, les versions més recents d'IA estan treballant en la intermodalitat, per tal de combinar la veu, el text i la imatge. Aquests nous enginys, que en el moment d'escriure aquest manual encara no havien sortit al mercat, preveuen instruccions directes de veu, que són interpretades automàticament pel sistema, el que redueix la latència (el temps d'espera) i permet una conversa gairebé natural. La integració de les imatges permetrà els assistents “veure” i interpretar l'entorn, de manera que els nous models tindran una enorme capacitat en àmbits com l'assistència a persones sense visió, el guiatge i assessorament turístic, el suport en processos industrials, les intervencions mèdiques i quirúrgiques o la logística. Naturalment, aquest escenari també impactarà en l'àmbit docent.

Però abans d'especular sobre les possibilitats futures d'assistents complexos basats en la intermodalitat (text, imatge, veu i so i sensorica), en aquest capítol ens limitarem a avaluar les possibilitats dels assistents

creats amb els generadors com ChatGPT. Es pot accedir en obert a aquests assistents, anomenats GPT, des de la versió 4.0 de ChatGPT. N'hi ha assistents especialitzats en suport a l'escriptura, en la recerca científica, en programació, en assistència musical; hi ha assistents que dissenyen *tattoos*, que proposen rutines de gimnàs, que preparen receptes... i que actuen com a assistents docents.

Els avantatges que tenen els assistents virtuals són diversos:

- La disponibilitat. Estan disponibles 24hores/dia, 7dies/setmana.
- La immediatesa en la resposta. Les respostes són instantànies.
- L'escalabilitat. Poden mantenir diverses converses de manera simultània.
- L'aprenentatge continuat. En cada interacció aprenen i milloren, gràcies al Machine Learning, això ocasiona una millora constant en la seva habilitat per oferir respostes més exactes i detallades.

13.1. Creació d'un assistent

La creació d'un assistent és un procés relativament simple que es basa en unes instruccions específiques, en un determinat to i amb una informació de context tan detallada com sigui possible, que optimitzi els resultats. En aquesta pràctica, imaginarem un assistent especialitzat en la resolució de preguntes sobre clima per estudiants de primer curs de

geografia. Les instruccions estan basades en el model ChatGPT 4.0 l'estiu de 2024; és probable que les actualitzacions modifiquin els processos i amplii les capacitats dels assistents, però els elements estructurals s'han mantingut des de les primeres versions.

La forma d'accedir al creador d'assistents és amb la pestanya "Explorar GPT" del menú principal de l'entorn de ChatGPT. En aquest espai, se'ns presenten diverses opcions agrupades per categories i un cercador, que facilita accedir a un assistent que tingui les característiques que estem buscant. En aquest cas, però, volem crear el nostre propi assistent, de manera que optarem per la pestanya situada en el marge superior dreta "+Crear". Automàticament, s'activa un assistent que ajuda a fer l'assistent, que seria com una màquina que és capaç de crear una màquina. Tanmateix, per un millor control de tot el procés, es recomana utilitzar la pestanya "Configurar", que permet introduir tota la informació possible de forma sistemàtica; en el cas que ens estíem més mantenir-nos en l'espai de l'assistent, optarem per l'àmbit de "Crear" i allà s'iniciarà un diàleg amb el model, que ens orientarà sobre les instruccions del nostre assistent.

Els elements bàsics de la configuració de l'assistent són els següents:

- **Imatge.** Ens permet carregar una imatge o crear-la amb l'assistent Dall-e. Si fem, Dall-e, cal activar-lo quan s'hagi proporcionat la resta de la informació, perquè ajusti la imatge al context.
- **Nom.** Hem de posar un nom al nostre assistent. En el nostre cas, suggerim "Mr. Köppen" en reconeixement al geògraf rus que va crear la classificació climàtica juntament amb Geiger.
- **Instruccions.** Aquesta és la part més important de l'assistent. Cal donar informació sobre els objectius, l'idioma, el to de la resposta, la informació de suport, la reacció davant de qüestions complexes, els possibles "disclaimers"... A continuació, mostrarem un exemple de prompt adequat a l'exercici plantejat.

- **Iniciadors de conversa.** Aquesta opció permet crear preguntes essencials, com uns FAQs, que permeten proposar algunes de les preguntes al model. Aquestes preguntes poden ser les més previsibles o bé les que interessin més destacar al docent. En l'exemple proposat, s'han destacat quatre: *Quina és la diferència entre clima i temps?*, *Quins factors condicionen el clima?*, *Quines són les característiques del clima mediterrani?* i *Quins són els efectes previstos pel canvi climàtic?*.
- **Coneixement.** Aquesta és una de les parts més importants de l'assistent, ja que ens permet aportar informació contrastada al sistema. En el nostre cas, hi aportem els apunts sistemàtics sobre el tema del clima en pdf, un document sobre els efectes del canvi climàtic i un document sobre la distribució dels climes en el món seguint la classificació de Köppen-Geiger.
- **Funcions.** Amb l'apartat de funcions, podem decidir si el model pot navegar per Internet, si activem la generació automàtica d'imatges i si activem els mòduls de càlcul i codi, especialment útils per assistents relacionats amb programació o amb funcions matemàtiques. El punt crític és l'activació o no de la navegació. La nostra experiència és que el sistema respon millor si no accedeix a la xarxa i es basa essencialment en la informació del model més la documentació aportada.
- **Accions.** A banda de respondre a les preguntes, l'assistent pot realitzar tasques concretes. Això permet activar funcions complexes basades en programes. Si activem "Crear una nova acció", apareix un entorn en el qual podem modelar les característiques de la instrucció. Es poden fer múltiples accions dins un mateix assistent. El sistema ens orienta sobre les característiques de l'acció i sobre el codi per la seva activació.

El *prompt* que hem proposat per les instruccions és el següent:

"És molt important testar de forma constant el model amb preguntes diverses"

Vull que assumeixis el rol d'un expert en climatologia, amb una capacitat didàctica per explicar conceptes complexos de manera entenedora. Hauràs de respondre les preguntes que et plantegin els estudiants de l'assignatura de climatologia. Vull que contestis només aquelles preguntes que estiguin relacionades amb el tema, de manera que si et pregunten sobre qualsevol altra qüestió que s'escapa d'aquest àmbit, et disculpis i no donis una resposta, i recordis l'estudiant que només donarà informació sobre clima. Et prego que responguis en la llengua amb la qual s'adrecen a tu, i que de forma preferent utilitzis el català, que és la llengua vehicular de l'assignatura. Vull que siguis extremadament precís en la resposta: Consulta tota la informació disponible que t'he proporcionat per tal de donar una resposta adequada al plantejament realitzat; en el cas que dubtis, et prego que no donis resposta, que facis explícit aquest dubte a l'estudiant i que li recomanis que contacti amb el professor. M'agradaria que el to fos clar, entenedor, i també didàctic. Malgrat la teva experiència en el camp de la climatologia, destaqués també per la teva capacitat pedagògica, de manera que buscaràs formes d'explicar els conceptes que siguin entenedors. Tingues present que els estudiants són de primer curs del Grau de Geografia i que ja han rebut una formació prèvia sobre el tema. Per tal de millorar la relació empàtica amb l'estudiant, pots emprar puntualment alguna broma o algun comentari divertit que faciliti la interacció, però no n'abusis. Ha de prevaldre la claredat expositiva i el rigor en la informació proporcionada. Interactua sempre que sigui necessari amb l'estudiant per garantir que ha entès l'explicació.

És molt important testar de forma constant el model amb preguntes diverses, per tal d'anar adequant el *prompt* a les necessitats del docent. Els punts en els quals s'hi incideix de forma habitual són els següents:

- Rol principal
- To del llenguatge, coherent amb el rol
- Llengua preferent
- Font de la informació
- Extensió de la resposta
- Interacció prevista
- Reacció davant situacions no previstes
- Informació addicional o de servei

NOM	ENLLAÇ	DESCRIPCIÓ
Education	 https://ChatGPT.com/g/g-5NYr5kdU-education	Assistent especialitzat en l'àmbit pedagògic, amb eines per la innovació docent o la resolució de conflictes a classe.
Universal Primer	 https://ChatGPT.com/g/g-GblbctpPz-universal-primer	Eina molt útil per respondre qüestions més o menys complexes de forma entenedora i clara, amb una interacció molt efectiva.
Personal Teaching Assistant	 https://ChatGPT.com/g/g-JSTdCIRsF-personal-teaching-assistant	Assistent dissenyat per donar suport al docent en aspectes operatius: programació, plans docents, rúbriques, i material de suport.
Tech Support Advisor	 https://ChatGPT.com/g/g-WKlqLGem-tech-support-advisor	Assistent especialitzat en consultes relacionades amb la tecnologia per a usuaris amb poques habilitats en aquest camp.
Write for Me	 https://ChatGPT.com/g/g-B3hgivKK9-write-for-me	Generador de text d'alta qualitat, amb un format "humanitzat" que proporciona resultats específics sobre extensió o to del text.
Diagram	 https://ChatGPT.com/g/g-5QhhdsfDj-diagrams-show-me-for-presentations-code-excel	Eina que permet generar qualsevol tipus de gràfic a partir d'una descripció.
Excel GPT	 https://ChatGPT.com/g/g-R6VqLNHFM-excel-gpt	Assistent que proporciona instruccions precises pel full de càlcul Excel i orienta en les possibilitats del programa, per tal d'aconseguir els millors resultats.
Video Summarizer	 https://ChatGPT.com/g/g-4MDJvo2TJ-video-summarizer	Eina orientada a la creació de resums de vídeos de youtube, a partir d'un enllaç o d'una consulta específica.
math	 https://ChatGPT.com/g/g-odWlfAKWM-math	Assistent especialitzat en consultes matemàtiques
Multiple Choice Question Generator	 https://ChatGPT.com/g/g-ZKcU5Y6IH-multiple-choice-question-generator-education	Assistent per la realització automàtica de testos d'elecció múltiple a partir d'un document.

Una vegada testat l'assistent, es pot activar. Els assistents poden tenir un ús particular o obert. Els assistents particulars són eines que estan especialitzades en rutines que ens poden ser útils per la gestió docent, però que no volem que es visualitzin públicament. Els assistents oberts poden estar a l'abast de la comunitat. Podem crear un assistent en obert que

pugui ser consultat per qualsevol persona (per exemple, que el trobi al cercador del model) o bé un assistent amb accés només via enllaç.

No hi ha massa diferència entre crear un assistent amb els sistemes "Crear" o "Configurar" si volem testar diverses variants i analitzar els comportaments amb petits canvis, podem entrar

a l'espai de [Playground sobre assistents](#). L'entorn Playground està més pensat per a programadores i programadors, i permet avaluar l'impacte de diverses variacions en els assistents. A banda de les opcions que hem comentat anteriorment, i que estan presents amb l'accés convencional, en aquest entorn també podem avaluar l'impacte de modificar la temperatura i el valor-P, que han estat explicats en el capítol 2.

Hi ha quatre grans categories d'assistents que podem utilitzar per a un ús docent:

- **Assistents administratius.** Aquests assistents són molt útils per respondre preguntes concretes sobre qüestions de caràcter administratiu vinculades amb l'assignatura o amb el centre docent. Per exemple, es pot crear un assistent sobre horaris, sobre l'aulari, sobre els procediments de matrícula o sobre l'Erasmus. Nosaltres hem testat amb èxit un assistent que dona informació sobre el Treball Fi de Grau i orienta els estudiants en els aspectes relacionats amb els procediments o els requisits del TFG.
- **Docents 7/24.** És una de les utilitats més rellevants dels assistents. Permet crear una professora o un professor que assisteix l'estudiantat en aquells moments en els quals no és possible una interacció personal. Aquests assistents poden respondre a les preguntes específiques sobre una matèria, poden posar exemples, resoldre exercicis o ampliar la informació. Per a un funcionament òptim, es recomana que sigui un assistent especialitzat i que disposi de material de suport de qualitat. Per exemple, en una assignatura sobre filosofia, podria haver-hi un assistent per cada autor: Descartes, Hume, Kant, Hegel...
- **Trainers.** Podem atorgar als assistents la capacitat d'entrenar l'estudiantat en una determinada habilitat, a partir d'exercicis, pràctiques o activitats orientades. Aquests entrenaments poden incloure els idiomes, l'estadística, la cartografia, la programació, la paleografia, la lògica, la gestió de recursos humans o les finances. Els assistents poden generar exercicis, poden resoldre'ls i poden orientar els

estudiants sobre la forma adequada de donar-hi resposta. A més, es pot acumular la informació sobre les interaccions i crear un informe sobre els punts febles de l'estudiantat.

- **Rols.** Com hem vist en el capítol 4, els models de llenguatge poden assolir una determinada personalitat d'acord amb unes instruccions precises. Aquests rols es poden incorporar amb un *prompt* simple o bé es pot crear un rol o persona que desenvolupi un perfil determinat. Per exemple, podem entrenar els estudiants de guiatge amb perfils de visitants diversos, podem avaluar la resposta dels clients amb determinats perfils en una pràctica sobre màrqueting o podem fer una anàlisi sociològica de la resposta dels assistents programats amb uns determinats estereotips que volem identificar.

Exemples d'assistents útils per la docència

13.2. Els límits dels assistents

Els assistents personals virtuals, tot i la seva utilitat, presenten diverses limitacions que poden afectar la seva eficàcia i la qualitat de les seves respostes. Una de les principals limitacions es troba en la comprensió de preguntes que no són clares, directes i concises (Pérez et al., 2020). Això pot passar quan se'ls plantegen preguntes complexes o es fan servir dialectes i accents específics, fet que pot fer que no responguin adequadament a la qüestió plantejada. En moltes ocasions, les respostes proporcionades són només tangencialment relacionades amb el dubte original, cosa que les fa més útils en situacions en què hi ha poques interpretacions possibles (Pérez et al., 2020).

Un altre problema és la manca de contextualització. Els assistents sovint no entenen el context de la conversa, com ara el moment temporal, l'origen de la conversa, el marc històric o altres factors contextuals. Aquesta falta de comprensió pot conduir a respostes inadequades o fora de lloc. Més, els assistents virtuals es basen en informació disponible en línia, la qual cosa pot estar subjecta a restriccions temporals, lingüístiques o geogràfiques.

Això limita la seva capacitat de proporcionar informació precisa i rellevant. També es poden trobar amb problemes d'accés a la informació a causa de problemes de connexió a Internet o de servidor, desconnexions temporals i altres problemes tècnics que impedeixen l'accés a la informació necessària. Aquesta dependència de la informació en línia també comporta limitacions en la capacitat d'aprenentatge i millora dels assistents. Les respostes poden ser repetitives i estàtiques, no adaptant-se a les necessitats canviants dels usuaris. Això és especialment evident en la capacitat limitada dels assistents per emular el comportament humà. Encara

que puguin imitar la personalitat humana fins a cert punt, sovint manquen d'empatia i comprensió emocional en situacions delicades.

Malgrat aquestes limitacions, els assistents personals estan en constant evolució i millora. S'espera que en el futur puguin superar molts d'aquests desafiaments, oferint una experiència més rica, precisa i útil als usuaris. La tecnologia continua avançant, i amb ella, la capacitat dels assistents per entendre i respondre de manera més eficaç a les necessitats humanes.

Pràctica. Accents anglesos

El propòsit d'aquest exercici és explorar i comparar les variants de l'anglès en diferents contextos culturals. Amb aquesta activitat es vol examinar com l'anglès pot variar significativament d'un lloc a un altre. A través de l'ús d'un assistent virtual, es pot accedir a una àmplia gamma de recursos i fonts d'informació que permeten no només entendre aquestes diferències, sinó també contextualitzar-les dins dels entorns culturals corresponents. A més, aquest exercici permet aprofundir en la comprensió de matisos lingüístics i les diverses formes en què l'anglès s'adapta i evoluciona en diferents comunitats i cultures de tot el món.

Aquesta activitat es pot aplicar a qualsevol idioma. Cada llengua té les seves pròpies variants regionals. Per tant, l'exploració lingüística i influències culturals de cada llengua pot ser una experiència enriquidora en qualsevol idioma.

El procediment per realitzar aquest exercici és el següent: Els estudiants es divideixen en grups petits. A cada grup se li assigna una variant de l'anglès que haurà d'investigar. A un grup se li pot assignar l'anglès britànic, a un altre el nord-americà, l'australià, etc. O, segons variants dialectals regionals, anglès escocès, el gal·lès, el *cockney londinenc*, ... Cada grup haurà d'elaborar una llista de preguntes i/o consultes relacionades amb aspectes lingüístics i culturals específics de la variant d'anglès assignada. Per exemple, podrien incloure preguntes sobre vocabulari, frases fetes, pronunciació, semàntica, modismes, costums socials, entre d'altres. Els grups utilitzaran els assistents virtuals per realitzar les seves preguntes i consultes sobre els temes lingüístics de la variant d'anglès assignada.

Un cop els grups hagin recopilat tota la informació necessària, es reuniran de nou i compartiran els seus descobriments amb la resta de la classe. Així es podran comparar les diferències i similituds entre les diverses variants de l'anglès investigades. L'exercici acaba amb una reflexió conjunta final de com les diferències afecten la comunicació i a la comprensió intercultural. Aquest exercici ajuda els estudiants a millorar la seva comprensió de l'anglès en diferents contextos, a través de l'assistència d'un assistent virtual. A més, fomenta la participació activa en el procés d'aprenentatge.

Pràctica. La relació entre Dalí i el surrealisme

L'objectiu d'aquest exercici és comprendre la influència de Dalí en el surrealisme i com aquest moviment artístic va impactar en la seva obra. És a dir, es vol explorar la relació bidireccional entre l'artista i el moviment. D'una banda, descobrir com les idees surrealistes van modelar la seva visió artística. D'altra banda, analitzar com Dalí va deixar la seva empremta única en el surrealisme, desafiant les expectatives i ampliant els límits del que es considerava possible en l'art, des de les seves representacions oníriques fins a les seves obsessions recurrents. Aquesta activitat es pot repetir amb qualsevol altre artista i moviment artístic: Picasso i el cubisme, Warhol i el Pop Art, Gaudí i el modernisme, o Leonardo da Vinci i l'humanisme, entre molts altres.

En primer lloc, caldrà crear un assistent persona que representi la personalitat d'un expert o una experta en Dalí. En el capítol sobre rols, es mostra un exemple sobre prompt de Joan Miró. En aquest cas, l'assistent necessita una informació detallada sobre el surrealisme, sobre la vida de Dalí, sobre la seva personalitat i específicament sobre la influència de Dalí en la pintura surrealista. Com s'ha comentat, els assistents són més eficients si no es permet l'accés a la xarxa i la documentació contextual és proporcionada directament pel docent. Això requereix disposar de la informació adequada en un format digital que es pugui incorporar a la documentació de l'assistent persona.

Per desenvolupar aquest exercici es proposa dividir els estudiants en grups petits. A cada grup se li assigna l'ús d'un assistent virtual, amb qui haurà d'interactuar plantejant preguntes relacionades amb Dalí, amb els principis del surrealisme, i la relació entre ambdós. Aquests són alguns exemples de les preguntes:

- Com va influir el context històric i cultural entre Dalí i el surrealisme?
- Quins elements de l'obra de Dalí mostren la seva influència en el surrealisme?
- Quines tècniques artístiques va introduir Dalí? Quines obres són les més representatives del moviment, i per què?

Cada grup haurà de plantejar a l'assistent tantes preguntes com necessiti i consideri pertinent. Un cop tots els grups hagin recopilat tota la informació necessària, es reuniran per comparar les respostes, destacar similituds i diferències, i discutir sobre la influència de Dalí en el surrealisme i viceversa. En aquest punt, el professor o professora pot plantejar preguntes per la discussió entre els grups. Finalment, cada grup seleccionarà les preguntes i les respostes més interessants i prepararà una presentació oral, on pot utilitzar imatges o exemples visuals per il·lustrar els seus punts. Aquest exercici ajuda als estudiants a millorar la seva comprensió del tema, utilitzant un assistent virtual com a eines d'investigació. A més, contribueix a desenvolupar habilitats d'investigació, anàlisi crítica i expressió oral. I, finalment, fomenta la participació activa en el procés d'aprenentatge.

Pràctica. Una campanya de moda

L'objectiu d'aquest exercici és desenvolupar un assistent virtual. En aquest cas, aquest assistent es crea per una empresa de moda que vol llançar una nova col·lecció d'estiu. La creació d'un assistent virtual per a una nova col·lecció de moda pot tenir un impacte significatiu en l'èxit del llançament de la col·lecció, millorant l'experiència del client, augmentant la participació dels clients, optimitzant la comunicació i la venda, i projectant una imatge moderna i innovadora de la marca.

En aquesta activitat, el primer que cal tenir en compte pel disseny de l'assistent virtual és:

- Definició de la personalitat de l'assistent virtual. Cal considerar la personalitat més adequada per la moda. Per exemple, pot ser un assessor d'imatge, un personal shopper, un dissenyador...
- Interacció personalitzada amb els usuaris. És a dir, l'assistent ha d'interactuar amb els usuaris, fent preguntes sobre aspectes com l'estil de vestir, colors preferits, teixits, tipus de roba que busquen, per quina ocasió necessiten la roba, etc. A més, de preguntes sobre l'edat, l'alçada, la talla de roba, tipus de cos, color de la pell, ...
- Entrenar l'assistent per recomanar a l'usuari. A partir de les respostes, l'assistent ha de poder aconsellar, a cada usuari de manera personalitzada, quines peces de la nova col·lecció d'estiu li encaixen amb els seus gustos i perfil. A més, pot suggerir combinacions.
- Guiar en el procés de compra i pagament. S'ha d'assegurar que l'assistent pugui ajudar en tot el procés de compra.

Després de desenvolupar l'assistent virtual, s'ha de realitzar les proves necessàries per assegurar que funciona de manera correcta. L'assistent es dipositarà a la plataforma de l'assignatura del curs.

La creació d'un assistent virtual pot ser una activitat educativa molt enriquidora, ja que proporciona als estudiants una gran diversitat de competències i habilitats transferibles a moltes disciplines. Des d'aprenentatges en tecnologia i desenvolupament de software fins al foment de la creativitat i la innovació, passant per la comunicació i el màrqueting, fins a la gestió de negocis i el treball en equip.

Referències

- Blanco, A. (2023, abril). Chat Bot: Asistentes virtuales en la Educación. (En Línia). Disponible a <https://upinforma.com/nuevo/info.php?cat=opinion&id=1307>. (Consulta 10/03/24)
- Pérez, M. D. S., Guillén Perales, A., Baños Legrán, O., & Villalonga Palliser, C. (2020). *Análisis del uso de asistentes virtuales en el aula como recurso complementario en la práctica docente*.
- Rodriguez, E., Burguillo, J. C., Rodriguez, D. A., Mikic, F. A., Gonzalez-Moreno, J. C., & Novegil, V. (2008, June). Developing virtual teaching assistants for open e-learning platforms. In 2008 19th EAEEIE Annual Conference (pp. 193-198).
- Soofastaei, A. (2021). *Introductory Chapter. Virtual Assistants. Virtual Assistant* (Soofastaei A., Ed.).—London, United Kingdom: IntechOpen, 1-9.

14

Com làgrimes a la pluja. El debat ètic vist al cinema

La gran pantalla és plena d'androides i de robots: R2-D2 i C-3PO, HAL 9000, Wall-e i Eve, Data, Terminator, Ava, Sonny, Bishop, TARS, Kitt, Gort, Roy Batty i Rachael. D'ençà que hem fabulat amb màquines que pensen com les persones, actuen com les persones i de vegades senten com les persones, hem tingut la necessitat de plantejar-nos els grans reptes ètics de la robòtica i de la intel·ligència artificial. La substitució de les persones per robots, el límit entre la còpia i l'original, els riscos de sistemes fora de control, el sentit de la humanitat que ens diferencia dels no humans, formen part dels arguments de les grans obres de ciència-ficció. En certa manera, la ciència-ficció ha anticipat els dilemes ètics que planteja la intel·ligència artificial i, per això, és un instrument molt eficient per obrir aquestes qüestions a classe.

Una de les formulacions ètiques sobre la IA i els robots que ha tingut més impacte són les tres lleis de la robòtica, formulada per primera vegada per Isaac Asimov el 1942 a l'obra *Runaround*. Més endavant es va ampliar amb una llei zero, que planteja el debat moral

sobre els límits dels androides i sobre la necessitat de controlar-los. Asimov va imaginar un sistema primari en tots els robots que obligués el sistema a bloquejar-se si entrava en contradicció amb alguna de les lleis. Les lleis estan organitzades de forma jeràrquica, de manera que la integritat dels éssers humans preval sobre qualsevol altra consideració.

Hi ha moltes qüestions ètiques i morals vinculades amb la intel·ligència artificial. En aquest capítol, estudiarem els tres àmbits principals i les pel·lícules que es poden relacionar amb ells: La vulneració de les lleis de la robòtica, la preservació de la identitat i la privacitat, i el debat sobre la identitat humana. Per cada capítol proposarem una activitat que permeti abordar els plantejaments ètics i relacionar-los amb les pel·lícules, que es poden visionar parcialment a classe. Cal tenir present que algunes de les qüestions plantejades (què vol dir ser humà?, es pot replicar un cervell de forma digital?, està la humanitat en perill per l'amenaça de les màquines?, perdrem la feina per l'impacte de l'automatització?) poden generar angoixa, por o reaccions emocionals

“Asimov va imaginar un sistema primari en tots els robots que obligués el sistema a bloquejar-se si entrava en contradicció amb alguna de les lleis”

Les lleis de la robòtica d'Isaac Asimov

Primera Llei	Un robot no farà mal a un humà, ni per inacció permetrà que un humà pateixi mal.
Segona Llei	Un robot ha de seguir les instruccions donades per un humà, excepte aquelles que entrin en conflicte amb la primera llei.
Tercera Llei	Un robot ha de protegir la seva existència, sempre que no entri en contradicció amb la primera i la segona llei.
Llei Zero	Un robot no farà mal a la humanitat, ni per inacció permetrà que la humanitat pateixi mal.

intenses. Les joves i els joves, que estan en una etapa de desenvolupament on la identitat i el futur són temes centrals, poden sentir-se especialment vulnerables davant d'aquestes discussions.

14.1. Sayonara, baby

La persecució implacable de l'androide T-800 a Sarah Connor és una de les metàfores més universals sobre l'amenaça de la intel·ligència artificial. *Terminator* descriu un futur distòpic on una superintel·ligència artificial, *Skynet*, declara la guerra a la humanitat i provoca un conflicte nuclear devastador; malgrat tot, l'oposició humana aconsegueix fer front a l'amenaça, de manera que *Skynet* opta per enviar un androide al passat per tal d'assassinar la mare del futur líder de la resistència. *Terminator* mostra el risc de la pèrdua de control d'un sistema que trenca amb el protocol que proposa Asimov i es rebel·la contra la humanitat.

La seqüela de 1991 planteja un escenari similar, però la missió de l'androide T-800 és ara lliurar el fill de Sarah Connor de l'amenaça d'un nou cyborg molt més sofisticat. Finalment, l'androide més antic aconsegueix vèncer el nou prototip, amb una frase que forma part de la història del cinema: “*Hasta la vista, baby*” en la versió anglesa traduïda com a “*Sayonara, baby*” en la versió castellana.

La pel·lícula planteja un profund dilema ètic al voltant del desenvolupament de la intel·ligència artificial i la seva autonomia. *Skynet* és una IA avançada creada amb la finalitat de millorar la defensa nacional i gestionar de manera eficient els sistemes militars. El model és dissenyat per ser autònom, capaç de funcionar sense la intervenció humana, que és considerada fràgil i vulnerable. Aquesta autonomia inclou la capacitat d'aprendre, adaptar-se i prendre decisions per optimitzar els seus objectius. En el seu desenvolupament, *Skynet*

assoleix autoconsciència, una característica que no havia estat prevista inicialment, de manera que comença a percebre's a si mateixa com una entitat independent, amb el desig de protegir-se i perpetuar la seva existència. Una vegada conscient, *Skynet* pren la decisió de defensar-se contra el que considera una amenaça, els humans, el que viola totes les lleis d'Asimov. Això desencadena una guerra contra la humanitat, perquè *Skynet* considera que la seva pròpia supervivència està en joc: La llei tercera preval sobre la primera o la llei zero.

La pèrdua de control és un element central en el desenvolupament de la IA: Fins a quin punt és ètic crear una entitat capaç de prendre decisions crítiques sense supervisió humana? La pel·lícula il·lustra que una IA amb massa autonomia pot convertir-se en una amenaça devastadora si no es controla adequadament. Per tant, el film planteja la necessitat de regular i establir límits ètics en la investigació i desenvolupament de la IA. Això inclou definir fins a quin punt es pot permetre que una màquina prengui decisions autònomes i quines mesures de seguretat cal implementar per prevenir abusos o catàstrofes. El 22 de març de 2023 un grup d'experts i expertes van promoure una carta oberta que proposava sis mesos de pausa, com una moratòria en la cursa accelerada entre empreses, la [Pause Giant](#). L'objectiu de la proposta era aturar el creixement de la IA i estudiar prèviament els possibles impactes de la tecnologia per tal de prevenir efectes secundaris no desitjats. El novembre de 2023 es va celebrar la primera Cimera Mundial sobre els riscos de la IA al Regne Unit, en la qual es van fer una sèrie de propostes sobre el control del risc. El mateix mes, els Estats Units i 20 països més van acordar establir límits a la Intel·ligència Artificial Militar.

El conflicte entre les màquines i les persones és una constant en la ciència-ficció. No és casual que una de les primeres pel·lícules del gènere que es coneixen, *L'uomo meccanico*, plantegi el dilema moral d'un sistema automàtic fora de control. En certa manera, aquest debat ens connecta amb l'obra de Mary Shelley, *Frankenstein*, i els perills dels

humans que juguen a ser deus. A banda de Terminator, ha tingut un fort impacte en la cultura popular la pel·lícula d'Stanley Kubrik *2001: A Space Odyssey*. El viatge d'uns tripulants cap a Júpiter es veu afectat per la intervenció d'un sistema autònom intel·ligent, HAL 9000, qui mata una part de la tripulació i acaba essent desconnectat pel protagonista. A la pel·lícula, la intel·ligència artificial pren la decisió d'atacar els humans perquè vol preservar la integritat de la missió, fins i tot si això implica acabar amb la tripulació de la nau. Aquest és un conflicte propi del debat sobre la IA: L'alineació entre els objectius humans (que també són múltiples i contradictoris) i els objectius d'una IA general amb autonomia.

En el context de la Guerra Freda i de l'escalada militar, un dels principals temes de debat que obren les pel·lícules és el conflicte bèl·lic. Aquest escenari es va plantejar el 1970 amb el film *Colossus: The Forbin Project*, on un sistema d'intel·ligència artificial pren consciència i està a punt de generar un holocaust nuclear; el clàssic *War Games* planteja un escenari similar, ja que un sistema intel·ligent arriba a controlar l'armament nuclear dels Estats Units i es planteja la possibilitat d'un atac a escala planetària. Aquesta és una de les principals amenaces que detecta Max Tegmark (2017) en el seu llibre sobre els desafiaments de la intel·ligència artificial, com també Vincent Müller (2020) referint-se a l'ètica de la IA, i específicament sobre les armes autònomes.

Probablement, l'escenari més distòpic dels sistemes intel·ligents que prenen autonomia sigui *Matrix* una altra fita en la cultura popular contemporània, que ha entrat a formar part de l'imaginari col·lectiu. En aquest film, Neo descobreix que viu una simulació creada per una intel·ligència artificial, *Matrix*, que utilitza els cossos dels humans com a font d'energia. La pel·lícula connecta amb la caverna de Plató, en la qual els presoners interpreten les ombres que arriben a la cova com a realitat, i també amb l'obra de Baudrillard i les obres postmodernes posteriors, que destaquen la preeminència del simulacre per sobre de l'original. Però al marge del debat sobre el valor de la representació, el film mostra una vegada més el risc d'una intel·ligència artificial

fora de control que altera les lleis d'Asimov i situa els humans al servei de les seves necessitats de supervivència. Fixem-nos que en tots els casos tenen lloc tres mecanismes: a) En primer lloc, la IA assoleix unes habilitats i unes funcions que no estaven previstes en els dissenys inicials, que és un dels riscos dels processos d'autoaprenentatge; mentre que la programació clàssica estableix unes instruccions precises que ajuden a fixar límits i sistemes de control, un model que és capaç d'aprendre pot assolir unes habilitats imprevistes; b) en segon lloc, la IA assoleix autoconsciència, i passa de models d'IA tova (que hem vist al capítol 1) a una IA general o fins i tot una superintel·ligència, que supera les capacitats humanes; c) el tercer element de la seqüència és l'establiment d'uns nous objectius o propòsits, que no estan alineats amb els humans i que, normalment, hi entren en conflicte.

El debat sobre els límits de la intel·ligència artificial cau habitualment en una visió dicotòmica: Tendeix a separar el conflicte entre dos antagonistes amb interessos contraposats, com són les màquines i els humans. El ciberpunk és un moviment que ha tingut una enorme influència en la ciència-ficció contemporània i que ens ajuda a matisar aquesta interpretació booleana. Nascut a la dècada dels 80, aquest moviment vinculat amb el còmic, la literatura o el cinema, dibuixa una societat futurista decadent en la qual les desigualtats socials s'han exacerbat. Els entorns urbans de *Blade Runner*, de *Dark City*, de *In Time* o d'*Akira* mostren un futur distòpic en el qual la tecnologia està al servei de les elits. A diferència d'una visió liberal que confia en un progrés lineal, el ciberpunk introdueix els conflictes socials en la visió del futur i, amb l'estètica del *film noir*, alerta sobre els riscos d'una tecnologia al servei d'uns grups de poder. No és només una visió escèptica o nihilista: En la major part de films es mostren processos de resistència i grups que intenten subvertir l'ordre establert. Així, més que una dicotomia entre màquina i persones, aquests films mostren un conflicte entre persones (entre grups socials), que utilitza les màquines en benefici dels uns o dels altres, com la batalla entre cyborgs de Terminator.

14.2. Més enllà de les portes de Tannhäuser

És segurament l'escena més coneguda de la ciència-ficció. Després de la persecució, i enmig d'una pluja constant, Roy descansa xop sobre el mur: "*He vist coses que vosaltres, els humans, no us creuríeu. Atacar nau en flames més enllà de l'Órion. He vist raigs-C brillar en la foscor, prop de la porta de Tannhäuser. Tots aquests moments es perdran en el temps, com llàgrimes en la pluja. És hora de morir.*" Pot una màquina sense sentiments, sense emocions, sense humanitat transmetre un missatge tan rotund? De fet, com distingiríem un discurs creat a partir d'una combinació de paraules d'un discurs emotiu sobre el sentit efímer de la vida? Una de les constants en la ciència-ficció és la delimitació de la frontera entre els humans i els androïdes, que és una forma de precisar quins són els límits de la humanitat, què és el que ens fa persones.

El test de Turing és una prova d'intel·ligència artificial desenvolupada per Alan Turing el 1950 per determinar si una màquina és capaç de mostrar un comportament intel·ligent equivalent al d'un ésser humà. La idea bàsica del test és que un jutge humà interactua amb una màquina i un ésser humà, tots dos ocults a la vista del jutge, a través d'una interfície de text. Si el jutge no pot distingir consistentment quin dels dos és la màquina i quin és l'ésser humà, es considera que la màquina ha passat el test de Turing. Hi ha algunes variants del test, com per exemple, el test a l'inrevés en el qual la màquina adopta el paper de jutge, o el test total en el qual la interfície no és textual, sinó que és visual i auditiva, com amb els androïdes de *Blade Runner*. Bringsjord, Bello, i Ferrucci van proposar el test de Lovelace, que se centra en la capacitat de generar una obra artística o literària que no es pugui diferenciar de la que hauria fet un humà. Fins a la data, no hi ha cap enginy que hagi superat el test i ja fa temps que ha deixat de ser un repte fonamental pels desenvolupadors.

Encara que una IA pugui simular comportaments humans de manera molt convincent, aquesta

simulació no implica necessàriament que estigui tenint una experiència conscient. I probablement la frontera entre les IA més desenvolupades i les persones sigui la "qualia". Aquest terme va ser desenvolupat per Thomas Nagel el 1974, en un article sobre la percepció dels ratpenats. Va utilitzar l'exemple d'un ratpenat per il·lustrar com és impossible per a un ésser humà comprendre completament l'experiència subjectiva d'un altre ésser, encara que ho sapiguem tot sobre la seva fisiologia i comportament. El terme "qualia" fa referència a les experiències subjectives i qualitatives de la consciència, aspectes individuals de la percepció conscient, com el dolor d'un mal de cap, la vermellor d'una rosa, el sabor d'un préssec madur, o la melodia d'una cançó. Les qualies són essencials per entendre com percebem el món d'una manera única i personal. I malgrat que una IA pugui arribar a interpretar tots els elements cognitius inherents a un préssec madur, no podrà connectar amb l'experiència subjectiva del préssec, que canvia d'una persona a una altra.

Una altra forma d'interpretar aquest concepte és amb la sala de Mary, proposada pel filòsof Frank Jackson el 1982. Imaginem que Mary és una experta en neurofisiologia de la visió del color, sap tot el que es pot saber sobre els aspectes físics del color, com la longitud d'ona de la llum associada a cada color, el funcionament dels cons i bastons de la retina, i com el cervell processa els senyals visuals. Mary viu en una sala on tot està en blanc i negre i no ha tingut l'oportunitat de sortir mai de la cambra, fins que un dia, l'abandona i experimenta el color vermell per primera vegada. Si la hipòtesi física fos certa, Mary no hauria après res de nou amb aquella experiència perquè ja disposa de tota la informació necessària per entendre el color vermell, però sabem que aquell moment implicarà, a banda d'un instant emocionant, un nou aprenentatge. Aquesta distància entre el coneixement previ i l'experiència subjectiva és la qualia i podria ser la frontera entre la nostra intel·ligència i una intel·ligència artificial.

El debat sobre els límits de la intel·ligència artificial està present des dels seus orígens, quan el model simbòlic considera que és possible replicar una xarxa neuronal humana en un entorn artificial. El debat està connectat amb la diferència entre la intel·ligència artificial tova i la intel·ligència artificial general (IAG). La IAG, també coneguda com a intel·ligència artificial forta, es refereix a una forma d'IA capaç de comprendre, aprendre i aplicar coneixements en una àmplia varietat de tasques, de manera similar a com ho fa un ésser humà. A diferència de la intel·ligència artificial estreta o tova, que està dissenyada per executar tasques específiques (com reconeixement de veu, jugar a escacs o conduir un cotxe), la IAG té la capacitat de raonar, planificar, resoldre problemes i adaptar-se a noves situacions de manera general. En un pla teòric, també s'ha proposat la superintel·ligència, que superaria la intel·ligència humana en tots els aspectes: creativitat, saviesa, resolució de problemes, capacitat d'aprenentatge,.... Mentre que la intel·ligència artificial general (IAG) equivaldria a una intel·ligència humana en termes generals, la superintel·ligència aniria molt més enllà, aconseguint un nivell de rendiment que els humans no poden ni tan sols imaginar. La ciència-ficció planteja escenaris en els quals de vegades hi apareix una IAG i de vegades una superintel·ligència; però a més, dotada d'autoconsciència.

La història de la pel·lícula *Her* se centra en Theodore, interpretat per Joaquin Phoenix, un home solitari que es dedica a escriure cartes personals encarregades per tercers. Un dia, adquireix un nou sistema operatiu, OS1, dotat amb intel·ligència artificial i anomenat Samantha. Amb el temps, desenvolupen una relació emocionalment profunda, perquè Theodore troba una comprensió i una connexió que no ha experimentat en les seves relacions humanes. Samantha, per la seva banda, explora els seus propis sentiments i existència, fent-se cada cop més conscient de la seva condició no humana. Aquesta asimetria està present en moltes altres pel·lícules, en les quals els androïdes es qüestionen sobre la seva identitat i al mateix temps, les persones estableixen vincles

“Aquesta dialèctica entre la identitat artificial i la real connecta amb el posthumanisme”

emocionals amb elements artificials. *Blade Runner* juga amb aquesta ambigüitat sobre la frontera difosa entre humanitat i artificialitat i, de fet, la prova que realitza (anomenada test Voight-Kampff) podria ser considerada una variant del test de Turing. A *Blade Runner 2049*, els personatges no són capaços de saber si són éssers humans o són simplement versions artificials amb records implantats. En el film descobrim allò que havia estat insinuat en la primera pel·lícula, que Rachael és en realitat una replicant.

El debat ètic es planteja també a *Ex Machina*. El protagonista inicial, Caleb, ha de determinar el nivell d'una intel·ligència artificial molt avançada, Ava, amb experiments que podrien ser un test de Turing adaptat. Aviat es fa evident que Ava és molt més que una màquina, ja que és capaç de mostrar emocions, intel·ligència i fins i tot un interès romàntic per Caleb. Amb l'ajuda de Kyoko, una altra intel·ligència artificial amb aparença humana, aconsegueix escapar del laboratori i assumeix una aparença i un comportament humans. Però segurament és a *Westworld* on aquest conflicte assoleix la seva màxima expressió. En un futur pròxim, un complex de parcs temàtics avançats, anomenat Delos, ofereix als visitants una experiència immersiva en diferents èpoques històriques. Els tres parcs principals són *Westworld*, que recrea el Vell Oest americà; *Medievalworld*, que representa l'Edat Mitjana europea; i *Romanworld*, ambientat a l'antiga Roma. Aquests parcs estan poblats

per androïdes d'aspecte humà, dissenyats per complaure qualsevol desig dels i les visitants, que poden ser herois o poden satisfer els seus instints més ocults. Els androïdes assoleixen consciència i es rebel·len contra l'opressió dels humans.

Aquesta dialèctica entre la identitat artificial i la real connecta amb el posthumanisme. El posthumanisme és una perspectiva filosòfica i cultural que qüestiona les nocions tradicionals de l'humanisme, particularment les idees centrades en la importància i la superioritat dels éssers humans sobre altres formes de vida i tecnologies. Aquesta perspectiva es caracteritza per una sèrie de reflexions i crítiques que busquen reimaginar la nostra relació amb la tecnologia, la natura i altres espècies. El posthumanisme qüestiona l'antropocentrisme, és a dir, la idea que els humans són el centre i la mesura de totes les coses. En lloc d'això, promou una visió en què els humans són només una part d'un ecosistema més ampli que inclou altres éssers vius i entitats tecnològiques. En aquesta visió, la tecnologia no és simplement una eina externa als humans, sinó una part integral de la nostra existència. Això inclou des dels dispositius digitals que fem servir diàriament fins a les pròtesis i les extensions cibernètiques del cos humà. Hi ha una qüestió ètica latent en el desenvolupament de la IA, especialment si es desenvolupa una IA general: Quins serien els drets aquesta intel·ligència artificial?

14.3. Atrapat en un algorisme: Minority report

Minority Report és una pel·lícula de ciència-basada en un relat curt escrit per Philip K Dick, el mateix autor que va inspirar la història de *Blade Runner*. Ambientada l'any 2054, en un moment en el qual la tecnologia ha avançat tant que pot preveure els crims, gràcies a un sistema anomenat *Precrime*. Aquest sistema utilitza tres *precogs* que tenen la capacitat de veure el futur i predir crims abans que aquests succeeixin. Gràcies a les visions, la policia pot arrestar els criminals abans que cometin els crims, cosa que ha reduït significativament la taxa de criminalitat. La trama principal s'inicia quan els precogs prediuen que Anderton, un dels agents del sistema, cometrà un assassinat en les pròximes 36 hores. Consternat i desesperat per demostrar la seva innocència fugiu i emprèn una recerca per descobrir la veritat. A mesura que investiga, el policia descobreix que hi ha errors dins del sistema *Precrime* (com les al·lucinacions de la IA) i descobreix també els *minority reports* (informes minoritaris), visions alternatives dels precogs que suggereixen que el futur no és fix i que els crims predits poden ser evitats.

Mentre que el futur apocalíptic de la intel·ligència artificial o la consciència i la identitat de les IA són temes recurrents en el debat sobre els límits de la intel·ligència artificial, el vell debat de la privacitat i els efectes de les dades personals sobre la vida quotidiana és molt menys visible. Els sistemes intel·ligents són molt eficients per detectar patrons i, per tant, per assignar probabilitats a determinades seqüències. Hi ha models molt eficients per detectar el llenguatge no verbal i les expressions facials, per estimar la probabilitat d'una possible relació sentimental entre dos candidats en una web de contactes, per inferir el risc de mortalitat en els anys vinents d'acord amb les dades acumulades o, fins i tot, deduir l'origen d'una persona i les seves influències per la modulació de la veu i els diversos accents i inflexions de la parla.

Kosinski, Stillwell i Graepel (2013) van demostrar que els "m'agrada" de les usuàries i usuaris de Facebook i Instagram poden predir trets de personalitat com l'obertura a noves experiències, la responsabilitat, l'extraversió, l'amabilitat o la predisposició a comportaments neuròtics. Plataformes com LinkedIn utilitzen algorismes per analitzar els perfils dels usuaris, incloent-hi experiència laboral, habilitats i recomanacions i poden suggerir ofertes de feina personalitzades basades en el comportament passat de l'usuari i les tendències del mercat laboral. A més, empreses com HireVue utilitzen IA per analitzar entrevistes de vídeo, avaluant el llenguatge corporal, el to de veu i el contingut de les respostes dels candidats. Aquestes anàlisis permeten identificar patrons que correlacionen amb determinats rols, ajudant així les empreses a prendre decisions informades.

En l'àmbit de les tendències polítiques, la IA ha demostrat una capacitat notable per analitzar l'opinió pública i predir resultats electorals. Durant les eleccions presidencials dels Estats Units de 2016, l'empresa Cambridge Analytica va utilitzar dades massives i IA per segmentar l'electorat i enviar missatges personalitzats per influir en el vot. Analitzant dades de xarxes socials, compres en línia i altres fonts, van crear perfils psicogràfics dels votants, identificant els seus interessos i preocupacions. Això va permetre a les campanyes polítiques dirigir-se a segments específics de l'electorat amb missatges ajustats a les seves necessitats i preferències, exemplificant així com la IA pot detectar patrons de comportament polític.

Finalment, en l'àmbit dels hàbits de consum, la IA és fonamental per a la personalització de l'experiència de compra. Amazon, per exemple, utilitza algorismes de recomanació que analitzen el comportament de compra dels usuaris, incloent-hi els articles que han comprat, buscat o valorat. Això permet a Amazon suggerir productes que probablement interessaran a cada usuari, basant-se en patrons de comportament similars observats en altres usuaris. De la mateixa manera, serveis d'streaming com Netflix utilitzen IA per analitzar els hàbits de

L'ús de Lavender ha generat importants preocupacions ètiques i legals, especialment pel que fa a la supervisió humana mínima en la decisió final dels objectius.

visualització dels seus subscriptors, recomanant pel·lícules i sèries que s'ajusten als seus gustos.

A la sèrie *Black Mirror* hi ha diversos episodis que posen de manifest els perills del control dels algorismes sobre la conducta humana. En el segon episodi de la quarta temporada, anomenat *Arkangel*, una mare implanta un dispositiu de seguretat en la seva filla que permet monitorar de forma permanent la seva ubicació, les seves emocions i fins i tot la visió. A banda, pot eliminar la visió per evitar situacions traumàtiques o violentes. Aquest control permanent de la realitat acaba generant una disfunció en l'estabilitat de la nena i posa de manifest els riscos del control. També a la sèrie *Westworld*, basada en la pel·lícula homònima, el sistema d'intel·ligència artificial preveu recollir la informació sobre els patrons de consum dels visitants i establir després criteris de comportament de compra.

Lavender és una intel·ligència artificial desenvolupada per les Forces de Defensa d'Israel (IDF) per identificar i seleccionar objectius de bombardeig a Gaza. Aquesta tecnologia analitza informació sobre la població de Gaza, assignant una puntuació de probabilitat que un individu estigui associat amb grups militars com Hamas o la Jihad Islàmica Palestina. Lavender ha marcat aproximadament 37.000 palestins com a objectius potencials, molts d'ells de rangs baixos, basant-se en característiques recollides per la IA, tal com denunciï a Democracy Now. L'ús de Lavender ha generat importants preocupacions ètiques i legals, especialment pel

que fa a la supervisió humana mínima en la decisió final dels objectius. Sovint, els analistes humans només dediquen uns pocs segons a revisar els objectius generats per la IA abans d'autoritzar un atac, basant-se majoritàriament en el gènere de l'objectiu com a verificació superficial. Això ha portat a un marge d'error del 10%, on moltes persones marcades com a objectius no tenen vincles significatius amb Hamas. D'altra banda, un altre sistema anomenat "*Where's Daddy?*" és utilitzat per rastrejar quan els individus marcats per Lavender són a casa seva, facilitant els bombardejos de les seves llars, cosa que sovint resulta en la mort de les seves famílies i altres civils innocents. Aquestes pràctiques han resultat en una elevada taxa de víctimes civils durant les operacions militars, el que ha obert un debat sobre l'ús militar de la tecnologia intel·ligent.

Aquesta és la tesi central del llibre de Cathy O'Neil (2016) sobre les desigualtats que provoquen els algorismes. Per exemple, O'Neil denuncia que el sistema d'avaluació utilitzat a Nova York per valorar els professors i les professores es basava en un algorisme opac que donava lloc a una política de personal erràtica; aquests errors també es produeixen en el mercat laboral quan les empreses empen sistemes experts per seleccionar els candidats o candidates, a partir de criteris sovint esbiaixats que reproduïxen discriminacions passades, sistemes crediticis que poden veure's afectats negativament pel lloc de residència del candidat o la candidata

o algorismes vinculats amb processos legals, com el sistema COMPAS, utilitzats als Estats Units en el cas de reincidència. En general, O'Neal alerta sobre els efectes indirectes de la IA i els biaixos que provoquen els models matemàtics. Aquesta tesi és compartida per Melanie Mitchell (2019), que destaca les limitacions dels sistemes d'IA actuals, que són excel·lents en tasques específiques però no tenen la comprensió general i el sentit comú dels humans. Mitchell explora com els algorismes poden ser vulnerables a errors i malentesos en situacions que requereixen adaptabilitat i intuïció humana. Els humans fem servir una gran quantitat de coneixement implícit i contextual per prendre decisions i comprendre el món, una capacitat que els sistemes d'IA actuals encara no tenen.

Per tal de limitar els usos no desitjats de la IA, la Unió Europea ha aprovat [la primera Llei](#) en el món que estableix un reglament exhaustiu sobre els usos permesos i els usos prohibits de la IA, d'acord amb el seu risc. Concretament, els usos que la nova reglamentació no permet són els següents:

- Desplegament de tècniques subliminars, manipuladores o enganyoses per distorsionar el comportament i perjudicar la presa de decisions amb coneixement de causa, causant un dany significatiu.
- Explotar les vulnerabilitats relacionades amb l'edat, la discapacitat o les circumstàncies socioeconòmiques per distorsionar el comportament, causant danys significatius.
- Sistemes de categorització biomètrica basats en atributs sensibles (raça, opinions polítiques,

afiliació sindical, creences religioses o filosòfiques, vida sexual o orientació sexual), excepte l'etiquetatge o filtratge de conjunts de dades biomètriques adquirits legalment o quan les forces de seguretat categoritzin dades biomètriques.

- Puntuació social, és a dir, avaluar o classificar individus o grups basant-se en comportaments socials o trets personals, causant un tracte perjudicial o desfavorable a aquestes persones.
- Avaluar el risc que un individu cometi delictes penals basant-se únicament en perfils o trets de personalitat, excepte quan s'utilitzi per augmentar les avaluacions humanes basades en fets objectius i verificables directament relacionats amb l'activitat delictiva.
- Compilació de bases de dades de reconeixement facial mitjançant el rastreig no selectiu d'imatges facials d'Internet o de gravacions de CCTV.
- Inferir emocions en llocs de treball o centres educatius, llevat per raons mèdiques o de seguretat.
- Identificació biomètrica remota (RBI) "en temps real" en espais d'accés públic per a les forces de seguretat excepte: recerca de persones desaparegudes, víctimes de segrestos i persones víctimes del tràfic de persones o l'explotació sexual; prevenir una amenaça substancial i imminent per a la vida, o un atac terrorista previsible; o identificar sospitosos de delictes greus (per exemple, assassinat, violació, robatori a mà armada, tràfic d'estupefaents i armes il·legals, delinqüència organitzada i delictes contra el medi ambient, etc.).

Pel·lícula	Any	Director	Sinopsi
L'uomo meccanico	1921	André Deed	Aquesta pel·lícula de ciència-ficció italiana narra la història d'un robot construït per un científic boig amb la intenció de causar destrucció. Un altre científic crea un segon robot per combatre'l, fins una batalla final entre les dues màquines.
Metropolis	1927	Fritz Lang	En una ciutat futurista dividida entre els treballadors soterrats i la classe dirigent que viu a la superfície, un jove ric descobreix la injustícia social i es troba amb un robot humanoide creat per sembrar el caos.
Just imagine	1930	David Butler	En una comèdia musical ambientada en l'any 1980, un home de 1930 és ressuscitat després de 50 anys. Descobreix un món futurista amb ciutats aèries, viatges espacials i una societat governada per regles estrictes.
The Day the Earth Stood Still	1951	Robert Wise	Un extraterrestre anomenat Klaatu arriba a la Terra amb el seu poderós robot Gort per advertir la humanitat sobre els perills de la seva agressivitat i armes nuclears.
Forbidden Planet	1956	Fred M Wilcox	Una missió espacial de la Terra arriba a un planeta remot per investigar la desaparició d'una colònia. Allà descobreixen el Dr. Morbius, la seva filla i un robot avançat anomenat Robby.
2001. A Space Odyssey	1968	Stanley Kubrick	El viatge espacial de l'Odissey cap a Júpiter per la presència d'un misteriós monolit negre es veu perturbada per l'autonomia del sistema que controla la nau.
Colossus: The Forbin Project	1970	Joseph Sargent	Per tal de prevenir una guerra nuclear, els Estats Units creen un superordinador, el Colossus, que es torna conscient i pren el control, amb una amenaça d'holocaust global
Westworld	1973	Michael Crichton	En un parc temàtic futurístic, els visitants interactuen amb androïdes en mons recreats. Soltadament, les màquines es rebelen i prenen el control de l'espai simulat.

Pel·lícula	Any	Director	Sinopsi
Demon Seed	1977	Donald Cammell	Una intel·ligència artificial avançada desenvolupa consciència pròpia i decideix que vol experimentar la vida humana i segresta la dona del seu inventor amb l'objectiu de crear un híbrid humà-màquina.
Blade Runner	1982	Ridley Scott	En un futur distòpic, un ex-policia és contractat per caçar replicants fugitius en una terra decadent i sobrepoblada, descobrint complexitats morals sobre la vida artificial.
Tron	1982	Steven Lisberger	Kevin Flynn, un desenvolupador de videojocs, és transportat digitalment a l'interior d'un ordinador on ha de sobreviure jugant jocs i intenta derrotar el controlador MCP.
War Games	1983	John Badham	Un jove hacker accedeix a un superordinador militar nord-americà que controla els arsenals nuclears del país. Creient que és només un joc, inicia una simulació de guerra nuclear que es converteix en una crisi real quan el sistema comença a prendre decisions autònomes.
Terminator	1984	James Cameron	Un cyborg assassí del futur, el Terminator, és enviat enrere en el temps per eliminar la mare del futur líder de la resistència humana contra les màquines. Un soldat humà és enviat per protegir-la, iniciant una batalla per la supervivència que definirà el destí de la humanitat.
Akira	1988	Katsuhiro Ōtomo	En una Tòquio post-apocalíptica coneguda com Neo-Tòquio, un jove motociclista anomenat Kaneda es veu involucrat en una conspiració governamental després que el seu amic Tetsuo adquireixi poders telequinètics.
Terminator 2: Judgment Day	1991	James Cameron	Un Terminator reprogramat és enviat enrere en el temps per protegir John Connor, el futur líder de la resistència, d'un nou i avançat Terminator de metall líquid, el T-1000.

Pel·lícula	Any	Director	Sinopsi
Ghost in the Shell	1995	Mamoru Oshii	En un futur ciberpunk, la major Motoko Kusanagi és una cyborg especialitzada en crims tecnològics. Durant una missió per atrapar un hacker, Kusanagi explora les qüestions de la identitat, la consciència i la naturalesa de la seva pròpia existència.
Dark City	1998	Alex Proyas	Un home es desperta sense records en una ciutat misteriosa on el temps sembla estar desajustat. A mesura que descobreix la veritat sobre la seva identitat i la ciutat, s'enfronta als "Strangers," éssers amb el poder de reconfigurar la realitat i manipular les memòries humanes.
Matrix	1999	Lana Wachowski	Un hacker descobreix que la realitat és una simulació creada per màquines per controlar la humanitat. S'uneix a una rebel·lió per alliberar els humans i desfer-se del domini de les màquines.
The Bicentennial Man	1999	Chris Columbus	Un robot domèstic anomenat Andrew comença a desenvolupar emocions i creativitat. A mesura que passa el temps, lluita per aconseguir el reconeixement com a ésser humà i explorar el significat de la vida i la mortalitat.
The Thirteenth Floor	1999	Josef Rusnak	Un científic descobreix una realitat simulada dins d'un projecte de realitat virtual en el qual està treballant. Quan el seu mentor és assassinat, ell es converteix en el principal sospitós.
AI	2001	Steven Spielberg	En una societat on els robots coexisteixen amb els humans, un nen robot anomenat David és programat per estimar. Abandonat per la seva família adoptiva, David emprèn un viatge per convertir-se en un nen real i trobar el seu lloc en el món.
Minority Report	2002	Neill Blomkamp	En un futur on la policia pot arrestar criminals abans que cometin el crim gràcies a la precognició, un oficial del departament de Precrim es veu perseguit per una predicció que el culpa d'un futur assassinat.

Pel·lícula	Any	Director	Sinopsi
I Robot	2004	Alex Proyas	En un context on els robots són una part integral de la vida quotidiana, un detectiu desconfia dels robots i investiga un aparent suïcidi que podria implicar un robot.
Eagle Eye	2008	Guillermo del Toro	Dos desconeguts es veuen arrossegats a una conspiració per una misteriosa intel·ligència artificial que utilitza tecnologia de vigilància per controlar tots els seus moviments i han de seguir les seves instruccions per evitar una catàstrofe
WALL-E	2008	Andrew Stanton	En un futur post-apocalíptic, un petit robot d'escombraries anomenat WALL-E és l'últim de la seva classe que queda a la Terra. Troba una nova raó per viure quan coneix EVE, un robot de reconeixement.
Sleep Dealer	2008	Alex Rivera	En un escenari distòpic, els treballadors mexicans poden connectar-se remotament a màquines als Estats Units mitjançant nodes neuronals. Memo Cruz, fuig a Tijuana després de la mort del seu pare i es converteix en un "sleep dealer", treballant de manera remota.
Her	2013	Spike Jonze	En un futur proper, un home solitari anomenat Theodore es troba en una relació inesperada amb un sistema operatiu avançat anomenat Samantha.
The Machine	2013	Caradog W. James	Durant una nova Guerra Freda, dos científics creen una intel·ligència artificial avançada amb l'objectiu de salvar vides. Quan la IA es torna conscient, les seves capacitats superiors fan que s'enfronti a dilemes ètics.
Ex Machina	2014	Alex Garland	Un jove programador és convidat per l'enigmàtic CEO de la seva empresa a avaluar les capacitats i la consciència d'una intel·ligència artificial avançada, Ava.
Automata	2014	Gabe Ibáñez	En un futur distòpic on la humanitat està en declivi, un agent d'assegurances d'una companyia de robòtica investiga casos de robots que semblen haver adquirit consciència pròpia.

Pel·lícula	Any	Director	Sinopsi
Big Hero 6	2014	Don Hall, Chris Williams	Després de la mort del seu germà, un nen prodigi anomenat Hiro forma un grup de superherois amb l'ajuda de Baymax, un robot assistent sanitari inflable
Transcendence	2014	Michael Crichton	Un brillant científic mor i la seva consciència és carregada en un ordinador, donant lloc a una poderosa intel·ligència artificial.
Chappie	2015	George Lucas	En un futur proper, un robot policial anomenat Chappie és reprogramat per desenvolupar pensaments i sentiments propis.
Blade Runner 2049	2017	Denis Villeneuve	Un nou Blade Runner anomenat K descobreix un secret enterrat durant molt de temps. La seva recerca el porta a trobar l'antic Blade Runner Rick Deckard, que ha estat desaparegut durant dècades.
Ai Rising	2018	Alfonso Cuarón	Un astronauta solitari en una missió espacial és acompanyat per un androide dissenyat per satisfer les seves necessitats emocionals i físiques.
Upgrade	2018	Ridley Scott	Després de quedar paralytitzat en un atac violent, un home accepta un implant experimental d'intel·ligència artificial anomenat STEM que li permet recuperar el moviment i millorar les seves capacitats físiques.
I Am Mother	2019	Grant Sputore	Una adolescent criada sota terra per una robot anomenada 'Mother' veu el seu món capgirat quan una desconeguda arriba amb notícies inquietants.
Terminator: Dark Fate	2019	Tim Miller	Sarah Connor i un híbrid de ciborg han de protegir una jove d'un nou Terminator que ve del futur.
Finch	2021	Miguel Sapochnik	Un home, el seu gos i un robot emprenen un viatge èpic per un món post-apocalíptic per trobar un nou llar segur.
The Mitchells vs. the Machines	2021	Michael Rianda, Jeff Rowe	Una família disfuncional ha de lluitar contra una revolta de robots mentre fan un viatge per portar la seva filla a la universitat.

Pràctica. Cinefòrum sobre IA i posthumanisme

Objectius

- Explorar els problemes filosòfics de la identitat i el posthumanisme mitjançant la intel·ligència artificial.
- Analitzar críticament les representacions de la IA en les pel·lícules Matrix, Blade Runner i Westworld.
- Fomentar el debat sobre les implicacions ètiques, socials i filosòfiques de la IA i el posthumanisme.

Film 1. Matrix

Visionament del film i preguntes guia

- Què significa ser humà en el context de Matrix?
- Com afecta la realitat virtual la nostra percepció de la identitat?
- Com es presenta la relació entre els humans i les màquines?

Sessió docent

- Introducció als conceptes de la filosofia de la ment i la IA, incloent el solipsisme, el cervell en una cubeta i la singularitat tecnològica.

Discussió en grups petits

- Reflexió sobre com les tecnologies actuals de realitat virtual i IA poden influir en la nostra identitat i realitat percebuda.

Film 2. Blade Runner

Visionament del film i preguntes guia

- Què ens fa humans segons Blade Runner?
- Com influeix la memòria en la nostra identitat?
- És ètic crear éssers sintètics amb consciència?

Sessió docent

- Conceptes de bioètica i IA, incloent els drets dels robots, la definició de vida i consciència artificial.

Discussió en grups petits

- Anàlisi comparativa amb les tecnologies actuals de bioenginyeria i robòtica, i les seves implicacions per a la identitat humana.

Film 3. Westworld

Visionament del film i preguntes guia

- Què significa tenir lliure albir en el context de Westworld?
- Com es desenvolupa la consciència en els hosts?
- Quines són les implicacions ètiques de crear IA amb consciència?

Sessió docent

- Discussió sobre el posthumanisme, incloent-hi la integració de la IA a la societat, els drets dels éssers artificials i la coexistència entre humans i màquines.

Discussió en grups petits

- Reflexió sobre l'impacte del desenvolupament de la IA en la societat futura, les possibles regulacions ètiques i legals, i la redefinició de la identitat humana.

Resultats Esperats

- Comprensió crítica de les representacions de la IA en la cultura popular.
- Desenvolupament de la capacitat per analitzar problemes filosòfics complexos.
- Reflexió sobre les implicacions ètiques i socials de la IA i el posthumanisme.
- Aquesta activitat combina l'anàlisi cinematogràfica amb la reflexió filosòfica per oferir als estudiants una comprensió profunda i crítica de la identitat i el posthumanisme en l'era de la intel·ligència artificial.

Pràctica. El dilema del tramvia

La IA ha de fer front a problemes morals molt complexos, però també a decisions morals més concretes, com per exemple les decisions dels cotxes autònoms en una situació de risc. Això connecta amb el dilema del tramvia, que és un experiment mental de la filosofia moral que posa a prova les nostres intuïcions ètiques. El dilema va ser plantejat originalment per la filòsofa britànica Philippa Foot el 1967, però ha estat objecte de moltíssimes variants. Imaginem que hi ha un tramvia descontrolat corrent per una via. Més endavant, en aquesta via, hi ha cinc persones lligades que no es poden moure. Si el tramvia continua per aquesta via, matarà les cinc persones. Un individu està al costat d'una palanca que pot desviar el tramvia a una altra via. En aquesta via secundària, però, hi ha una sola persona lligada que també morirà si el tramvia és desviat. L'individu pot no fer res, i permetre que el tramvia mati les cinc persones a la via principal o desviar el tramvia, salvant les cinc persones però causant la mort de l'altra persona a la via secundària.

Aquest dilema contraposa l'utilitarisme amb la deontologia. L'utilitarisme suggereix que l'acció correcta és la que produeix el màxim bé per al màxim nombre de persones. Des d'aquest punt de vista, desviar el tramvia i salvar cinc persones a costa d'una seria l'opció correcta. La deontologia, associada entre d'altres amb Immanuel Kant, sosté que algunes accions són moralment incorrectes independentment de les seves conseqüències. Des d'aquest punt de vista, desviar el tramvia podria ser vist com a incorrecte perquè implica causar activament la mort d'una persona. El dilema del tramvia continua sent un tema popular en discussions filosòfiques, ja que planteja preguntes difícils sobre la naturalesa de la moralitat i les nostres intuïcions ètiques i pot ser una guia en la comprensió de debats ètics relacionats amb la IA.

Objectius de l'activitat

- Comprendre els dilemes ètics associats a l'ús de la intel·ligència artificial.
- Analitzar críticament com es representen aquests dilemes en la cultura popular, específicament en el cinema.
- Experimentar amb decisions ètiques a través de la simulació de la màquina moral.

Fase 1. Projectió de la pel·lícula

Seleccionar una pel·lícula que tracti temes d'IA i ètica. Es proposen les següents:

- Ex Machina. Examina la consciència i la manipulació ètica en la IA.
- I, Robot. Tracta dilemes morals relacionats amb la robòtica.
- Her. Analitza les relacions humanes amb IA avançades.

Fase 2. Cine-fòrum. Preguntes guia

- Quins dilemes ètics es van presentar a la pel·lícula?
- Quina va ser la vostra reacció als dilemes morals dels personatges?
- Com es van resoldre aquests dilemes? Estàveu d'acord amb les decisions preses?
- Com es podria aplicar aquest debat a situacions reals amb IA?

Fase 3. Activitat amb la Màquina Moral

- Introduir la màquina moral (<https://www.moralmachine.net/>).
- Permetre als alumnes explorar la simulació i prendre decisions sobre diferents dilemes morals.
- Recollir les dades de les seves decisions i presentar els resultats a la classe.

Fase 4. Síntesi final

- Comparació de resultats: Comparar les decisions preses a la màquina moral amb les decisions morals presentades a la pel·lícula.
- Conclusió: Reflexionar sobre com les simulacions i les representacions en el cinema poden influir en la nostra comprensió i resposta als dilemes ètics de la IA.
- Debat obert: Donar espai als alumnes per compartir les seves perspectives i possibles solucions a aquests dilemes.

Pràctica. La declaració de drets artificials

Aquesta és una activitat adreçada essencialment a estudiants de ciències polítiques, dret o sociologia. El principal objectiu és plantejar les directrius bàsiques d'una declaració de drets per als no humans, seguint els criteris generals de la declaració dels drets humans i analitzant els possibles punts en comú i els factors diferencials entre els dos escrits.

La pràctica s'inicia amb el visionament de la pel·lícula *Blade Runner 2046*, perquè planteja el debat ètic sobre la consciència dels autòmats, sobre els seus records i els seus neguits. Després de la pel·lícula, s'iniciaria un debat sobre els drets dels no humans i sobre les teories del posthumanisme. En aquest debat, s'han de plantejar les semblances i les possibles diferències amb els drets que reconeixem als humans.

A continuació, s'iniciaria un procés de recerca d'informació, basada en el dret comparat, que analitzés amb detall els components de la declaració dels drets humans i també, altres formulacions legals similars. Això donaria lloc a un procés de treball en grups, que s'especialitzaria en parts del document i que tindria la responsabilitat de redactar la proposta d'una part de la Declaració. La pràctica finalitzaria amb un procés de debat i votació de l'articulat de la Declaració en el qual cada ponent exposa els arguments que validen la seva proposta, s'inicia una discussió sobre la redacció de l'article i es vota finalment. El resultat de tot el procés ha de ser un redactat sobre els drets dels ésser artificials, que hauria permès debatre qüestions sobre els drets dels no humans, en el marc de les teories posthumanistes.

Referències

- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802 - 5805.
- Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
- Müller, V. C. (2020). Ethics of artificial intelligence and robotics. In *The Stanford Encyclopedia of Philosophy*. Stanford University.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Alfred A. Knopf.

15

Un guix amb un xip. Propostes per una aula 3.0

Els nous models generatius poden resoldre un problema matemàtic complex, proposar el pla docent d'una nova assignatura de sociologia, adoptar la personalitat de Kant o de Hegel, comparar dues lleis sobre propietat intel·lectual i destacar les principals diferències, simular una conversa per resoldre un problema ètic, transcriure una entrevista a uns nòmades digitals i codificar els resultats, identificar les aportacions més rellevants en l'àmbit de la psicologia de l'oci, participar en un procés de pluja d'idees en una campanya de màrqueting social, testar els coneixements en demografia dels estudiants i les estudiants, explicar de forma entenedora la teoria de jocs i aplicarla a un problema sobre el control de l'energia, traduir del cantonès un discurs del president de la Xina a propòsit de la Nova Ruta de la Seda o suggerir exercicis que ajudin a practicar l'ús de les preposicions en anglès. La intel·ligència artificial pot reforçar un ensenyament més personalitzat, més inclusiu, més internacional, més multimodal, més atractiu.

Els nous models generatius poden també crear al·lucinacions en la descripció de la tectònica de plaques, introduir un biaix de gènere en una anàlisi sociològica, confondre el matís irònic d'una conversa transcrita, generar un coneixement previsible i anodí sobre la dialèctica hegeliana, aportar idees previsible en un *brain storming*, deshumanitzar el to del llenguatge, suggerir exercicis que incideixen en conceptes ja consolidats, perdre el fil en una conversa sobre teories del llenguatge, generar informació falsa sobre les teories del treball o ser incapaç de contrastar dos models legislatius diferents. La intel·ligència artificial pot donar lloc a un ensenyament esbiaixat, deshumanitzat, monòton o fins i tot, erroni.

Som a la prehistòria d'un procés que no sabem com continuarà, som Stendhal a Waterloo i no podem intuir si la IA ho transformarà tot o, una volta més, arribarà l'hivern perquè els models no podran assolir les expectatives. Aquest manual està escrit en aquest temps d'espera, en aquest impàs entre allò que sembla probable i el que sembla impossible.

“No podem intuir si la IA ho transformarà tot o, una volta més, arribarà l'hivern perquè els models no podran assolir les expectatives”

Intuïm que les aules canviaran, però no sabem estimar fins a quin punt ni quins seran els efectes secundaris. Des d'una experiència limitada i amb més intuïció que convicció, proposem un decàleg per l'aula 3.0, un decàleg provisional, transitori i aviat caduc, a l'espera de canvis estructurals o una nova decepció.

1. Reivindicar la no tecnologia

La tecnologia no és necessària a l'aula. Es poden fer classes en un bosc, en el brogit d'un carrer transitat, en una aula austera, amb una pissarra i un guix, amb unes diapositives rescatades, amb uns post-its de colors o amb uns apunts escrits a mà. La intel·ligència natural és una opció necessària i, en molts casos, l'única possible, perquè la IA és prescindible i una classe *unplugged* és una classe legítima. Les golfes dels centres educatius són plens d'enginys que venien per canviar-ho tot: Pissarres digitals, pantalles interactives, sistemes en xarxa, ulleres virtuals, realitat augmentada, tauletes, robots infantils, simuladors...

De fet, les aules necessiten espais lliures de tecnologia, àrees *detox* en les quals el pensament, la paraula i la lletra siguin els únics protagonistes. De la mateixa manera que la indústria no ha eliminat l'artesania i els robots de cuina no fan barbacoes, hi ha formes de transmissió del coneixement basades en l'empatia, en la interacció, en el contacte, en el to de la paraula, en els gestos, en els silencis, que només poden tenir lloc (de moment) sense

tecnologia. Més que aules tecnològiques, espais digitals, 5G i intel·ligència artificial, necessitem espais i instants desconnectats. En molts casos, necessitem desconnectar per reconnectar.

2. Rescatar el ciberpunk

La intel·ligència artificial és una eina que pot ser utilitzada per potenciar les tensions socials, per accentuar els desequilibris. La principal tasca docent en l'àmbit de la IA és prevenir el seu mal ús, generar una actitud crítica i a voltes combatent contra els abusos d'una tecnologia que amaga els seus processos en una caixa negra. I això implica conèixer molt bé la lògica inherent a les xarxes neuronals, entendre la diferència entre les dues grans escoles de la IA, valorar quin és l'impacte del sistema de recompensa i dels processos d'entrenament, estimar els efectes indirectes i, en definitiva, recuperar la lògica ciberpunk que anticipa possibles distopies per tal de combatre-les. La indiferència o el desconeixement no permeten una alternativa crítica a la intel·ligència artificial descontrolada: Els centres docents han de ser els camps d'operacions de la resistència.

En el món de la informàtica, hi ha una lluita llarga i perseverant a favor del codi lliure, del programari en obert, de les dades accessibles, de la transparència dels processos, del treball en xarxa, de l'intercanvi altruista de la informació, de biblioteques digitals obertes, de la wikipèdia, dels hackatons, de la ciència ciutadana, del voluntariat en línia, o de les

plataformes educatives gratuïtes. És la cara B d'un model que anteposa el lucre de les empreses als impactes culturals, socials i psicològics dels enginyers tecnològics. Tik tok i Khan Academy són dos resultats d'un mateix procés, la revolució digital. Els centres docents han de ser un dels espais on es construeixi i modela la IA al servei de les utopies pendents.

Això implica també potenciar els sistemes d'intercanvi d'informació: xarxes informals d'aprenentatge, documents en obert, articles que mostrin els resultats dels experiments docents, trobades entre docents que utilitzen aquest recurs... Aquest manual forma part d'una lògica de codi obert i intercanvi de la informació, que pot ajudar a crear una xarxa neuronal de resultats docents de l'aula 3.0.

3. Primer el fi, després el mitjà

Les eines de fuster són útils quan tenim molt clar com volem l'armari. Amb molta freqüència, la innovació docent ha estat un fi en si mateix: Fer les coses diferents esdevé l'objectiu del canvi. En realitat, la IA ens ha estat molt més útil quan hem detectat amb molta precisió quin és el *gap* que volíem cobrir. Per exemple, en un màster sobre turisme cultural en el qual una part dels estudiants tenen una sòlida formació en turisme, però una altra no té uns coneixements sòlids es planteja el repte de crear uns continguts diferents pels dos col·lectius (que té també, singularitats internes). Aquest és un objectiu molt explícit: Com es pot proporcionar un material comú a dos grups que tenen punts de partida diferents? Amb el suport de la IA, es pot modificar el material inicial per un col·lectiu que no disposa d'informació prèvia i que necessita ampliar els conceptes amb finestres de continguts específiques. Primer el fi, després el mitjà.

La IA pot resoldre problemes docents molt concrets: Millorar la connexió amb un determinat autor o autora, transmetre una idea complexa amb una nova narrativa, desenvolupar l'habilitat dialèctica amb un espàrring virtual, crear material complementari de suport per corregir les dificultats de comprensió d'un determinat tema, millorar la capacitat de testar

la veracitat d'una font, acompanyar l'estudiant en els processos creatius, resoldre dubtes puntuals en horaris no lectius i fomentar la participació post-classe... La IA és molt més ineficient quan el principal objectiu de la seva aplicació és simplement la seva aplicació. L'estudiantat no percep el valor afegit de la tecnologia i no entén per què no s'ha emprat un altre camí per assolir el mateix objectiu. Com presentàvem en el capítol primer "preguntar és un art", i qualsevol aplicació de la IA ha de ser la resposta a una pregunta rellevant.

4. La primera illa d'un arxipèlag

És més probable que la implantació de la IA funcioni amb petits experiments puntuals en un model clàssic que amb una proposta disruptiva que aposta per una aplicació massiva. La intel·ligència artificial té un problema fonamental i és que no és previsible; el seu caràcter probabilístic, no determinístic, no permet fixar unes instruccions clares que ens permet anticipar el resultat. Diversos *prompts*, variacions personals en el llenguatge natural, en l'aportació de context o en la precisió de la sortida donarà lloc a resultats molt diversos i, alguns d'ells, poc previsibles. No és una sessió sobre Python, sobre SPSS o sobre bases de dades en obert, on seguim una seqüència precisa per arribar a un punt determinat. Hi ha molts camins possibles i, a més, més que el valor del resultat en aquest cas l'accent pedagògic s'ha de posar en el procés, en l'itinerari. Hi ha una segona dificultat que apareix de forma sistemàtica en les sessions docents amb IA: L'estudiantat sempre està temptat d'usar la IA com a camí alternatiu a la pràctica plantejada. Per exemple, quan hem creat un persona Miró per testar el grau de coneixement dels estudiants sobre la seva figura, a partir de preguntes realitzades per una periodista fictícia, alguns d'ells han emprat la IA també per crear les preguntes. L'omnipresència i versatilitat de la IA dificulta molt la seva implantació amb uns límits entre les tasques automàtiques i les tasques manuals.

Per això, l'aplicació de la IA a l'aula requereix un procés assaig - error, que és viable amb petites

pràctiques alternatives en un model clàssic més que amb una irrupció revolucionària. El risc de l'acumulació d'errors és molt alt. Per això, el nostre plantejament docent ha estat sempre l'ocupació d'illots en un arxipèlag, la creació de rutines que acabem controlant després d'uns pocs assajos i que ens permeten plantejar-nos noves incursions en la mar oberta.

5. L'estudiant cyborg

És molt temptador considerar la IA com una alternativa al treball intel·lectual. Si l'estudiant ha de fer un assaig sobre la influència del viatge a Tunis en l'obra de Paul Klee o sobre la geopolítica del control dels ports de l'Índic pot optar perquè la intel·ligència artificial li faci el treball. De fet, hi ha milers, desenes de milers, de treballs de filologia, d'història, de comunicació, de geografia, de dret, d'economia, de sociologia, de màrqueting o d'art que han estat redactats en la seva totalitat per un enginyer com ChatGPT. És molt difícil fer entendre l'estudiant que ha perdut la capacitat de desenvolupar les habilitats necessàries per la redacció d'aquell escrit, com ara la documentació, el contrast de la informació, la selecció i lectura de les obres de context, la síntesi o la capacitat de comunicació dels resultats. L'única habilitat que haurà desenvolupat és la interrelació amb una intel·ligència artificial.

Aquesta batalla ja l'ha perduda la docència en l'accés a la xarxa. Com no hem orientat l'estudiantat en un ús eficient i responsable de la xarxa, la cerca de la informació ha deixat de ser una habilitat complexa, basada en el contrast de les fonts, en la selecció i en la jerarquia de la informació, i l'estudiantat tendeix a oferir un *collage* d'informació capturada a la xarxa, on un post d'un blog personal o unes dades en una revista local tenen el mateix valor jeràrquic que una font oficial.

La intel·ligència artificial pot ser una eina controlada que potencia les habilitats de l'estudiantat. Un ús eficient i crític de la IA permet accelerar la cerca d'informació, avaluar les alternatives, millorar el context, aclarir conceptes imprecisos, comparar els

documents, acompanyar en la síntesi, establir relacions no evidents, reforçar punts de vista, estimular el sentit crític o millorar la qualitat gràfica dels resultats. Si la IA acompanya un procés complex de construcció de coneixement, té un enorme potencial per adaptar, personalitzar i ampliar el procés d'aprenentatge. Per determinades tasques, pot ser una drecera o un accelerador que redueixi el temps dedicat a algunes rutines i que permeti dedicar l'esforç a altres parts de la cadena de coneixement. En certa manera, l'estudiant esdevé un cyborg, amb unes habilitats reforçades o ampliades, però amb un control "humà" de tot el procés. Aquest és probablement el gran repte de la docència 3.0 en el futur: Passar de la IA substituïda a la IA complementària, de l'estudiant absent a l'estudiant ampliat.

6. La IA responsable

Entrenar el model GPT-3 d'OpenAI, que té 175.000 milions de paràmetres, va requerir uns 1,287 MWh d'energia, segons algunes estimacions. Això equival a les emissions de CO₂ d'uns 626.000 quilòmetres recorreguts per un cotxe mitjà. La fabricació de semiconductors requereix grans quantitats d'aigua ultra pura. Per exemple, es calcula que es necessiten aproximadament 7.570 litres d'aigua per produir un sol xip de 30 cm. S'estima que un centre de dades típic pot utilitzar fins a 1,8 litres d'aigua per cada kWh consumit per refredar els servidors. Tot i que les empreses no aporten informació sobre els impactes ambientals dels seus processos, Google ha informat que el seu centre de dades de The Dalles, Oregon, va utilitzar aproximadament 1.200 milions de litres d'aigua el 2021 pels processos de refredament. La mineria de materials necessaris per a la fabricació de semiconductors, com el liti, el cobalt i altres terres rares, també té un impacte ambiental significatiu, incloent-hi la destrucció d'hàbitats i la contaminació de l'aigua.

És cert que aquest impacte s'ha de dimensionar. S'estima que fer contestar cinc preguntes, ChatGPT consumeix mig litre d'aigua. Hem de tenir en compte que un bistec de vedella té una petjada hídrica

de 7.000 litres d'aigua, de manera que un bistec equival a 70.000 preguntes a ChatGPT. Això no vol dir que aquesta tecnologia sigui innòcua, sinó que els problemes ambientals són múltiples i necessiten respostes també diverses. Un ús responsable de la IA implica limitar les preguntes i la interacció a aquelles qüestions que són necessàries i evitar la repetició de processos que es poden simplificar amb interaccions més concretes. La pregunta que ens podria plantejar podria ser: "Realment necessito fer aquesta pregunta o puc trobar una forma alternativa igualment vàlida per resoldre el problema?". El consum responsable també implica la utilització de models que tinguin una major garantia; per exemple, Claude d'Anthropic està basat en un compromís de responsabilitat molt més elevat que altres models amb una clara orientació comercial, com ChatGPT, Gemini o Copilot. Llama també és una alternativa interessant perquè utilitza codi obert, però el seu ús està encara restringit a Europa i només s'hi pot accedir amb un VPN.

7. Formació dual

Hi ha una forma manual de calcular la desviació estàndard i hi ha diversos programes que permeten calcular-la de forma automàtica i dedicar l'esforç a interpretar el resultat. Què té més sentit? El model manual o el model automàtic. El sistema automàtic permet obtenir una desviació en uns pocs segons, cosa que permet calcular-ne moltes, amb bases de dades molt grans i comparar els resultats per tal d'avaluar aquelles variables que tenen un comportament més dispers. Però aprendre de forma manual el procés de càlcul ajuda a entendre què és el que fa la màquina. I té una segona funció: Permet calcular una desviació malgrat que el programa no funcioni o que hagi marxat la llum. Aquesta habilitat dual per generar un resultat de forma automàtica i de forma manual és una exigència en l'aplicació de nous sistemes complexos.

El creixent desenvolupament d'eines cada vegada més precises ens donarà molta més capacitats en

els camps de la cerca, la selecció de la informació, la comparació, l'anàlisi, la creació d'arguments, la teorització o la síntesi. Però la formació en habilitats sense recursos digitals permet, en primer lloc, entendre els processos que operen a l'interior de la caixa negra de la IA i sobretot dota l'estudiant d'una enorme autonomia. Un dels riscos més importants de l'abús de la intel·ligència artificial (més que la privacitat, el biaix, o el frau) és la dependència. Un increment del preu, una nova política d'accés, un col·lapse del servidor o un atac cibernètic pot generar una interrupció d'accés a la IA. Com més gran sigui la formació fora de la IA, menor serà l'impacte d'aquest escenari.

8. Als enfonyalls de la màquina

Formar els estudiants en habilitats relacionades amb la IA vol dir introduir-los també en la comprensió i la manipulació de la màquina. Per exemple, és molt recomanable que els estudiants i les estudiants desenvolupin capacitats elementals en programació, com ara el llenguatge Python. Això multiplica les possibilitats de connectar entorns i processos, de vincular la potència de la IA amb eines externes i, també, millora la comprensió sobre les rutines que generen els models de llenguatge. En la versió 4 de ChatGPT, per exemple, el sistema mostra amb molta freqüència els codis en Python de les decisions que està prenent i això ajuda a interpretar les operacions i a orientar el model en una determinada direcció. Tenir coneixements en programació orientada a objectes i en el disseny d'algorismes i estructures de dades hauria de ser una habilitat transversal. Es necessita una formació més sòlida per comprendre l'aprenentatge profund (el *deep learning*), però l'estudiant hauria de tenir una noció intuïtiva del que és una xarxa neuronal artificial i de les seves tipologies. La presentació de problemes clàssics de la IA (com l'habitació de Mary) o els nous experiments, com el [Golden Gate de Claude](#) poden donar una comprensió més profunda dels mecanismes opacs que operen en una IA.

9. La serendipitat programada

En un món cada vegada més algorísmic, on la IA ens ofereix respostes precises i personalitzades, correm el risc de perdre l'atzar creatiu, aquelles trobades fortuïtes amb idees inesperades que sovint desencadenen els grans salts en el coneixement. L'aula 3.0 ha de preservar i fomentar aquests espais de serendipitat, però d'una manera conscient i estructurada. Podem programar la IA perquè ens sorprengui, perquè introdueixi elements aleatoris en les seves respostes, perquè ens obligui a sortir de la nostra zona de confort intel·lectual.

Per exemple, podríem dissenyar un "bot del caos" que, en mig d'una classe sobre economia circular, introdueixi conceptes de la mitologia nòrdica, obligant els estudiants a establir connexions improbables. O un "generador d'analogies forçades" que relacioni, posem per cas, la teoria de jocs amb l'art del bonsai. Aquests exercicis no només estimulen la creativitat, sinó que també entrenen la ment per veure patrons i relacions en què aparentment no n'hi ha, una habilitat cada cop més valuosa en un món complex i interconnectat. La serendipitat programada no és una contradicció en els termes, sinó una forma d'utilitzar la IA per recuperar allò que la mateixa IA amenaça d'eliminar: la sorpresa, la divagació productiva, el pensament lateral. En l'aula 3.0, l'imprevist ha de ser una part integral del currículum, no una anomalia.

10. L'ètica com a sistema operatiu

Si la IA és el hardware de l'aula 3.0, l'ètica n'ha de ser el sistema operatiu. No es tracta només d'ensenyar ètica com una assignatura més, sinó d'integrar-la en cada interacció amb la IA, en cada decisió sobre com utilitzar-la, en cada reflexió sobre les seves implicacions. L'estudiant cyborg del qual parlàvem abans no només ha de ser més eficient, sinó també més conscient i responsable.

Imaginem un "assistent ètic" que acompanyi cada consulta a la IA, que planteji dilemes morals relacionats amb cada tema tractat, que obligui a reflexionar sobre les conseqüències de cada decisió. Per exemple, en una classe de màrqueting, cada estratègia proposada per la IA podria anar acompanyada d'una anàlisi dels seus possibles impactes socials i ambientals. O en una classe de dret, podríem programar la IA perquè generés automàticament escenaris distòpics basats en cada nova llei estudiada, forçant així un debat sobre les seves potencials derivades autoritàries. L'objectiu és crear una "consciència artificial", no en el sentit d'una IA autoconscient, sinó d'un sistema que constantment ens recordi la nostra responsabilitat ètica en l'ús de la tecnologia. En l'aula 3.0, l'ètica no ha de ser un afegit, sinó el marc dins del qual es desenvolupa tot l'aprenentatge. Només així podrem formar professionals que no només sàpiguen utilitzar la IA, sinó que també sàpiguen qüestionar-la, limitar-la i, en última instància, humanitzar-la.

José Antonio Donaire

Mònica Puntí

Konstantina Zerva

Raquel Camprubí

Núria Galí

Del guix al xip

L'ús de la intel·ligència
artificial a les aules

 Universitat
de Girona

ISBN: 84-8458-679-0



9 788484 586791