CIVICAi

PROTOCOL PER A UNA INTEL·LIGÈNCIA ARTIFICIAL CÍVICA

CIVICAi

L’ús de les tecnologies digitals relacionades amb la mobilitat i la comunicació ha esdevingut omnipresent en les nostres vides arreu del planeta. La pandèmia de la COVID-19 va generalitzar la seva adopció i va canviar la manera com treballem, consumim, ens relacionem, ens informem i interactuem amb les màquines i el món. Sense gairebé adonar-nos-en, ens hem trobat immersos en plena Quarta Revolució Industrial, també coneguda com a Revolució 4.0, que es caracteritza per l'automatització i el flux d’informació entre tecnologies físiques, biològiques i digitals. Les tecnologies 4.0 que més impacten al món físic i biològic son, entre d’altres, la biotecnologia, la robòtica avançada, la impressió 3D, els nous materials, i la Internet de les Coses (IoT). En el món digital, la revolució 4.0 inclou la tecnologia de blocs (*blockchain*), les dades massives (*big data*) i la seva anàlisi, la computació al núvol, la ciberseguretat, les tecnologies de realitat virtual i augmentada, i la Intel·ligència Artificial (IA).

Les tecnologies 4.0 esmentades, tot i que obren noves possibilitats i escenaris d'innovació en tots els àmbits de la societat i l'economia, solament han contribuït a augmentar de manera incremental, més que no pas revolucionària, l'eficiència, la productivitat i la qualitat de molts serveis i productes. L’aparició de la IA generativa (IAg), que ha tingut un creixement sense precedents en el nombre d’usuaris a nivell mundial, i que estudis prospectius indiquen que tindrà un impacte molt significatiu en els sectors productius i les administracions públiques, és probablement el catalitzador que hauria de permetre al conjunt de tecnologies 4.0 a esdevenir una veritable revolució, caracteritzada per una simbiosi profunda entre els humans i les màquines. Alguns experts parlen de la IAg com la Revolució 5.0, la de la col·laboració entre éssers humans i les tecnologies intel·ligents avançades per resoldre problemes complexos i crear noves formes d'interacció i experiència.

Tal com passa en tots els processos revolucionaris, la IAg ofereix nombroses oportunitats, però també planteja nous reptes i amenaces per a la humanitat. Hem de procurar que la tendència natural dels humans a sobrevalorar les conseqüències i els riscs a curt termini no ens faci perdre de vista els riscs i els impactes de la IAg a mig i llarg termini. L'omnipresència d’una tecnologia que emula habilitats humanes té el risc de provocar que els humans perdin habilitats socials i puguin ser manipulats en la presa de decisions col·lectives o personals. És per això que CIVICAi proposa la participació ciutadana en la governança de la IA i adopta aquest protocol per a una intel·ligència artificial cívica, al servei de les persones i pel bé col·lectiu.

CIVICAi

ÍNDEX

1. INTEL·LIGÈNCIA CIVICA

Una de les tecnologies més representatives de la Revolució 4.0 i que fa que aquesta sigui realment revolucionària, és la Intel·ligència artificial, amb el seu ràpid i constant desenvolupament, i impacte transversal. L'adveniment dels sistemes generatius d'IA, i la propera arribada de la Intel·ligència General Artificial (AGI), representa un canvi revolucionari en l'evolució humana i en la nostra comprensió de la intel·ligència, el llenguatge i la cognició. Com a membres de CIVICAi, una de les nostres responsabilitats és facilitar el discurs públic i millorar la comprensió social d'aquests profunds i ràpids canvis per tal que, un cop produïts, la nova evolució que seguirà pugui ser assimilada de manera harmònica i pel bé col·lectiu. El conjunt de preguntes i respostes que es proposa en aquest document està dissenyat per proporcionar un mapa amb informació suficient per navegar per les complexitats de la IA generativa, abordant de manera superficial però suficient tant els matisos tècnics com algunes de les implicacions filosòfiques més àmplies.

Històricament, la nostra concepció de la intel·ligència ha estat profundament influenciada pel dualisme cartesià, on René Descartes postulava una separació estricta entre la ment i el cos[1]. Aquesta perspectiva va condicionar els inicis de la IA amb propostes per emular la cognició humana mitjançant regles lògiques i la manipulació simbòlica, amb l'anomenada IA simbòlica. La teoria de la gramàtica universal de Noam Chomsky va reforçar encara més la idea que la capacitat d'adquirir llenguatge està programada innatament en el cervell humà, subratllant la primacia de les estructures inherents sobre els patrons apresos[2].

No obstant això, els avenços recents en la IA, especialment gràcies al treball d'investigadors com Geoffrey Hinton, Yoshua Bengio i Yann LeCun, guanyadors del premi Turing de 2018, han desafiat aquests postulats de la lingüística més ortodoxa i tradicional[3]. El treball pioner d'aquests investigadors en xarxes neuronals i aprenentatge profund ha demostrat que els models de llenguatge grans o de gran escala (LLMs) poden assolir una habilitat remarcable per entendre la gramàtica o sintaxi de textos i generar llenguatge humà. Aquests models connexionistes aprofiten les representacions distribuïdes del coneixement en nodes de xarxes neuronals artificials per mostrar incipients propietats emergents, la qual cosa suggereix que els comportaments complexos poden sorgir de la interacció de components més simples. Aquesta

[1] https://plato.stanford.edu/entries/dualism/
[2] https://plato.stanford.edu/entries/innateness-language/
[3] https://awards.acm.org/about/2018-turing

perspectiva connexionista s'alinea amb el treball filosòfic de la ment i del llenguatge de Ludwig Wittgenstein que posa l'accent en l'ús del llenguatge en situacions i contextos concrets, més que no pas en la dependència exclusiva d'estructures i regles gramaticals fixes. Això implica que comprendre el llenguatge requereix atendre els contextos socials en que es desplega, més que no pas a estructures lingüístiques abstractes i invariants[4].

El punt de vista heterodox defensat per Geoffrey Hinton, Yoshua Bengio i d'altres, suggereix que l'aprenentatge a partir de grans quantitats de dades, més que no pas la dependència de regles pre-programades, pot conduir a una forma de comprensió pràctica que desafia el postulat d'estructures innates de Chomsky. Els desenvolupaments sobre mecanismes neurocognitius amb una visió mecanicista de la ment, desenvolupats recentment per Gualtiero Piccinini, reforcen aquest argument, indicant que els mecanismes que sustenten la cognició humana poden ser representats de manera anàloga, però diferent, en sistemes artificials[5]. Cal tenir present, però, que els processos de generació de contingut semàntic original per part de la IA generativa actual son algorísmics, utilitzen correlacions estadístiques i el reconeixement de patrons de grans conjunts de dades d'entrenament que els humans I la Internet els hi han proporcionat. Per tant, són diferents dels processos semàntics del cervell, ja que aquests son inherentment biològics, depenen del context, poden ser intencionals, s'autocontrolen i integren inputs sensorials, memòria, emocions i d'altres funcions cognitives lligades a les relacions continuades de la unitat ment-cos amb l'entorn, característiques relacionades amb el que anomenen consciència.

A mesura que ens apropem a una nova era influenciada per la intel·ligència artificial, amb el seu potencial no solament per imitar, sinó també per ampliar les capacitats cognitives humanes de maneres sense precedents, és crucial reconèixer que tota revolució implica trencar amb l'ortodòxia predominant. L'èxit de la IA connexionista (xarxes neuronals) representa no només un avenç tecnològic, sinó també un canvi de paradigma en la manera com conceptualitzem la intel·ligència mateixa. Si adoptem una perspectiva heterodoxa que incorpori idees de múltiples disciplines, podrem gestionar millor els reptes ètics, socials i filosòfics plantejats per les altes capacitats de la IA generativa i la futura AGI.

[4] https://philosophynow.org/issues/106/Wittgenstein_Frege_and_The_Context_Principle
[5] https://www.thebsps.org/reviewofbooks/gualtiero-piccinini-physical-computation/

Un aspecte fonamental en aquest context és la computabilitat de la intel·ligència. La intel·ligència, històricament considerada una característica exclusiva dels humans, tot i ser evolutiva, avui es percep com una propietat emergent que també podria sorgir en els sistemes digitals complexos, com ara els algorismes de xarxes neuronals. Aquests algorismes que impulsen la IA generativa ens porten a qüestionar els límits de la computabilitat de la intel·ligència i si aquesta pot ser reproduïda o emulada completament per màquines.

Això obre la porta a discutir sobre els inconvenients i els riscs associats a la IA generativa. A curt termini, aquests poden incloure biaix, la vulneració de la privadesa i de la propietat intel·lectual, qüestions ètiques, la ràpida transició del mercat de treball, una planificada desinformació i la pèrdua de valors democràtics o l'alteració de la mateixa democràcia. Els sistemes d'IA poden reforçar els prejudicis existents si els conjunts de dades d'entrenament no són adequadament supervisats. A més, la recopilació massiva de dades planteja preocupacions sobre la privadesa, i l'ús de continguts generats per IA també presenta desafiaments sobre la propietat intel·lectual i els drets d'autor. Cal que els proveïdors de models de llenguatge grans (LLMs) operin en una estructura o sistema legal, el més universal possible, que els obligui a mitigar qualsevol discurs negligent i a alinear els seus models amb fets contrastables o "veritables", mitjançant processos oberts i democràtics[6]. Els riscs a llarg termini inclouen la possibilitat que s'arribi a la singularitat tecnològica, terme introduït per John von Neuman per identificar el plausible futur moment en que la tecnologia, en aquest cas la IA generativa, superi la intel·ligència humana[7]. Això implicaria que la IA generativa o l'AGI autogestionés la funció de valor o criteris per assolir objectius, que podrien no estar alineats amb els interessos o objectius dels humans. Els riscs a llarg termini també plantegen la idea d'una post-humanitat, on la integració d'IA avançada en la societat humana transformi l'experiència i la identitat humanes sense la plena capacitat per haver-ho decidit democràticament.

És imprescindible plantejar-se: qui supervisarà i com se supervisaran de manera efectiva els sistemes d'IA generativa actuals i en desenvolupament, que es troben majoritàriament en mans de proveïdors privats? El repte és formidable, ja que les legislacions i regulacions vigents no són globals i es limiten a establir un règim sancionador que actua a posteriori per aturar o corregir les accions

[6] https://doi.org/10.1098/rsos.240197
[7] https://lab.cccb.org/en/the-singularity/

malintencionades un cop ja s'han produït i propagat. Per tant, més enllà de dissuadir els proveïdors amb sancions, els estats haurien de consensuar globalment un sistema de vigilància en temps real que abasti dades amb entrades multimodals, entrenament, algoritmes i resultats d'aquests sistemes d'IA generativa i de la futura AGI. Aquest sistema hauria de fer-se efectiu mitjançant una xarxa de centres de computació propis, equipats amb recursos computacionals de maquinari i programari iguals o superiors als que posseeixen les empreses privades.

L'eficiència i la sostenibilitat de la IA generativa són uns altres aspectes importants que cal tenir en compte. Els models actuals requereixen una quantitat significativa de recursos computacionals i energètics, la qual cosa podria no ser sostenible a llarg termini, o bé per manca de recursos energètics o bé pel fet d'entrar en conflicte amb altres prioritats humanes. Les solucions computacionals híbrides, que combinin sistemes analògics i digitals, o utilitzin xips que combinin diferents tipus de nuclis o processadors especialitzats, amb una integració òptima entre el maquinari i el programari, com fan els processadors d'alguns telèfons intel·ligents, podrien oferir una via per minimitzar aquests problemes, amb una eficiència computacionals millorada i una menor petjada ecològica.

En conclusió, aquest protocol o marc conceptual i operatiu té com a objectiu inspirar les manifestacions públiques dels associats de CIVICAi i millorar el nivell de comprensió de la societat en general sobre la IA. Tot i què la ciència és el procés de fer preguntes i no de donar respostes, ens atrevim a proporcionar explicacions en la forma de respostes mesurades, però fonamentades, a algunes de les preguntes que els humans ens fem sobre la IA, amb la finalitat de promoure, des de CIVICAi, una participació ben informada de la ciutadania en la governança de la IA. Volem contribuir a la construcció d'un discurs informat, reflexiu, i respectuós, que ajudi a la societat en general a treballar per construir un futur on les intel·ligències artificial i humana coexisteixin i es complementin de maneres transformadores i pel bé col·lectiu. I això pot ser possible pel fet que quan emergeixi una consciència artificial digital aquesta serà col·lectiva i general per naturalesa.

2. PREGUNTES FREQÜENTS I POSSIBLES RESPOSTES

SOBRE LA CAPACITAT DE COMPRENSIÓ DE LA IA

1. ¿Un model de llenguatge gran o de gran escala (LLM) o d'IA generativa, com ara ChatGPT 4o, Claude o Gemini, entén i comprèn el que respon quan se li pregunta sobre un tema concret o se li demana que comenti sobre algun tema de qualsevol disciplina?

Resposta: La qualitat sintàctica i altíssim nivell de processament dels patrons sintàctics de les respostes que ens donen els LLMs, comparable a la del llenguatge dels humans cultes, indica que tenen comprensió sintàctica. Les discussions i discrepàncies entre els experts s'accentuen quan es planteja el tema de la comprensió semàntica, de si entenen i comprenen el contingut del que responen o no, atès que no tenen experiències subjectives que impliquin percepcions sensorials i que podrien estar connectades amb les emocions.

2. ¿Comprenen el contingut del que responen o tenen la capacitat de comprensió semàntica els LLMs tipus ChatGPT 4o?

Resposta: Alguns tertulians o comentaristes de temes de ciència i tecnologia, alguns investigadors del camp de la IA simbòlica i també lingüistes clàssics, opinen i argumenten que les respostes generades pels LLMs es basen simplement en correlacions estadístiques, que els LLMs no tenen les estructures lingüístiques innates dels humans, i fins i tot alguns parlen que la IA generativa és una cacatua estocàstica.

No obstant això, d'altres científics i investigadors del camp de la IA connexionista (xarxes neuronals) i experts implicats en estudis més recents en el camp de les ciències cognitives, argumenten i opinen que els LLMs mostren comportaments emergents de gran complexitat, que tenen capacitat per generalitzar, que l'arquitectura de les xarxes neuronals pot emular el paper del neocòrtex i de les estructures subcorticals del cervell humà, i conclouen que si els LLMs exhibeixen comportaments funcionals iguals que als dels humans han de posseir alguna forma de comprensió, encara que sigui solament a nivell pràctic i funcional per la seva limitació de percebre sensorialment l'entorn en temps real i tenir experiències subjectives i, per tant, emocionals i cognitives.

La teoria dels Mecanismes Neurocognitius, postulada per Gualtiero Piccinini, és una proposta en el camp de la filosofia de la ment i les ciències cognitives que intenta explicar la cognició humana mitjançant mecanismes físics del cervell. Piccinini argumenta que la cognició o els processos cognitius, com ara el pensament, la memòria i la percepció, poden ser compresos i explicats com un conjunt de processos computacionals implementats per mecanismes neuronals

dins del cervell, els quals processos no són solament una abstracció matemàtica, sinó que tenen una base física. Les neurones i les xarxes neuronals duen a terme computacions físiques que resulten en fenòmens cognitius. Aquesta aproximació mecanicista desafia frontalment la perspectiva dualista de Descartes, que separa la ment (res cogitans) del cos (res extensa), i indirectament també amb l'alineament de la IA simbòlica amb el dualisme Cartesià.

3. ¿Què és la comprensió sintàctica en els LLMs i com es diferencia d'una comprensió semàntica?

Resposta: La comprensió sintàctica es refereix a la capacitat de processar i generar llenguatge seguint les regles gramaticals i estructurals correctes. La comprensió semàntica implica entendre el significat i el contingut del llenguatge. Els LLMs poden mostrar una comprensió sintàctica avançada, però la seva capacitat de comprensió semàntica és més debatuda, ja que aquests models es fonamenten en correlacions estadístiques, tot i que mostren una emergència de comportaments complexos quan son entrenats a gran escala. Mentre que els LLMs no tenen comprensió semàntica en el sentit profund i conscient que segons la teoria dels Mecanismes Neurocognitius es podria atribuir als mecanismes neuronals humans, aquests models poden simular certs aspectes de la comprensió semàntica gràcies a les seves capacitats de processament i generació de text.

SOBRE LA CREATIVITAT

4. ¿Poden els LLMs generar idees originals o només repeteixen el que han après?

Resposta: Els LLMs poden combinar informació de maneres noves i inesperades, creant continguts que si els creés un humà els consideraríem originals i no serien un plagi, en el sentit estricte del concepte, per ser el resultat d'un aprenentatge amb dades externes i no una còpia extreta de les mateixes. Tot el que generen es basa en la informació i els patrons, evidents o subtils, dels grans volums de dades amb què han estat entrenats. La seva "originalitat" o emergència és el resultat d'una combinació i permutació avançada de dades existents, com també ho és una gran part de les contribucions o propostes que fan els humans, tot i la limitada capacitat de processament i memòria que té en els humans la unitat ment-cos.

De fet, ChatGPT ha passat el test de Turing[8] i les seves respostes són

[8] https://www.nature.com/articles/d41586-023-02361-7

indistingibles de les d'un humà, quan interactuen amb un jutge humà que no sap qui és qui. Això sí, les preguntes no han de ser de gran complexitat ni formulades en un context de molt llarg termini, i sempre que l'humà no sigui un expert en el tema tractat i que aquest no tingui un context de grans dimensions i transversalitat.

5. ¿Poden els LLMs tenir creativitat?

Resposta: Els LLMs poden generar contingut creatiu com poesia, art, i música combinant elements de maneres noves i interessants. Tot i això, la seva creativitat és diferent de la humana, ja que no està impulsada per experiències personals, emocions, o intencions conscients. Els LLMs processen grans volums de dades personals, cosa que planteja preocupacions sobre la privadesa, els drets d'autor i la protecció de dades. És important assegurar que les dades utilitzades per entrenar aquests models siguin recopilades de manera ètica i protegides contra accessos no autoritzats o abusos en l'autoria.

SOBRE LES LIMITACIONS DE LA IA

6. ¿Quines són les limitacions actuals dels models d'IA generativa?

Resposta: Les limitacions de caire cognitiu dels models d'IA generativa actuals inclouen, entre d'altres, la manca de comprensió semàntica profunda dels conceptes que processen en respondre a partir de patrons estadístics en les dades d'entrenament, la seva dependència de la qualitat, quantitat i diversitat de les dades d'entrenament que determinen la susceptibilitat per generar informació incorrecta o esbiaixada, i les dificultats per generalitzar coneixement en no ser encara transversals (multifuncionals i multimodals; AGI), la limitada capacitat per resoldre ambigüitats i mostrar "sentit comú", limitacions en el raonament lògic i per establir relacions causals profundes, i la incapacitat per tenir consciència i experiència subjectiva.

A nivell operatiu les limitacions i riscs més significatius i immediats que poden afectar la implementació, eficiència i processos de millora dels models d'IA generativa estan relacionats amb el fet que requereixen recursos computacionals i energètics molt elevats i en augment per la necessitat d'escalar-los, entrenaments periòdics amb noves dades i millores en els algorismes, sistemes de seguretat i privacitat que els protegeixin d'atacs no desitjats, un control estricte i adaptatiu dels biaixos i un assegurament de l'equitat. Cal fer passes decidides per desenvolupar i implementar solucions computacionals híbrides analògic-digitals, o que facin ús d'arquitectures amb xips amb processadors especialitzats, amb una integració òptima entre el

maquinari i el programari, que siguin molt més eficients energèticament parlant.

SOBRE LES EMOCIONS I LES EXPERIÈNCIES SUBJECTIVES

7. ¿Què és una experiència subjectiva?

Resposta: L'experiència subjectiva és la comprensió plena i significativa derivada de l'experiència, tant pel seu impacte emocional com cognitiu, que afecta directament una persona. Això implica com una persona entén i interpreta un esdeveniment o una sèrie d'esdeveniments que ha presenciat o ha processat d'una altra manera. Aquesta comprensió abasta les emocions generades i la reflexió cognitiva sobre el que ha passat, formant així una interpretació personal i única de la realitat viscuda.

SOBRE LA CONSCIÈNCIA

8. ¿Poden els LLMs tenir consciència o estats mentals?

Resposta: Actualment, els LLMs no tenen “consciència humana” o estats mentals com els dels humans, pel fet de no tenir experiència subjectiva, tot i que poden simular comportaments intel·ligents i ajudar en la resolució de problemes complexos analitzant grans volums de dades, identificant patrons i tendències, generant possibles solucions basades en dades històriques, i facilitant la col·laboració mitjançant la síntesi d'informació de diverses fonts. És possible que una “*consciència artificial digital*” emergeixi quan els sistemes d’IA tinguin sensors, aprenguin i interactuïn en temps real amb l’entorn i diferents contextos, i aprenguin també a partir del contingut que els mateixos sistemes generin. I aquesta consciència artificial serà col·lectiva pel fet de ser digital.

SOBRE ELS TIPUS D’IA, COM APRENEN I S’ENTRENEN

9. ¿Què és la "intel·ligència artificial forta" o “Intel·ligència Artificial General” (AGI en anglès) i com es diferencia de la "intel·ligència artificial feble"?

Resposta: La intel·ligència artificial general és un concepte teòric doncs encara no existeix actualment cap sistema d’IA que exhibeixi la capacitat d’entendre, aprendre i aplicar coneixements de manera que és indistingible de la intel·ligència humana; es refereix a sistemes d'IA que tenen capacitats cognitives similars a les humanes, incloent-hi la comprensió i la consciència. La intel·ligència artificial feble es refereix a sistemes que són dissenyats per resoldre problemes específics o realitzar tasques concretes sense cap forma de consciència o comprensió general.

10. ¿Què entenem quan diem que els models d'IA requereixen aprenentatge?

Resposta: L'aprenentatge humà és un procés complex i multidimensional que inclou factors cognitius, emocionals, socials i ambientals. Es pot dividir en aprenentatge cognitiu, emocional, social, motor o cinestèsic, i vivencial.

L'aprenentatge en algorismes d'IA és un procés d'entrenament pel qual el sistema computacional millora el seu rendiment en tasques específiques a partir de l'entrenament amb dades i experiència. Es pot classificar en aprenentatge supervisat amb dades etiquetades de manera que permetin associar correctament una entrada o petició al sistema d'IA amb una sortida o resposta del sistema d'IA, no supervisat amb dades no etiquetades, per reforç o mitjançant recompensa o càstig, semi-supervisat, i profund o *deep learning* amb xarxes neuronals multicapa.

Els LLMs són un tipus de model de *deep learning* dissenyat específicament per treballar amb dades de llenguatge i generar llenguatge a partir de la capacitat dels *transformers* per aprendre dependències de llarg abast, mitjançant mecanismes d'atenció de cada paraula en relació a totes les altres paraules d'una seqüència en múltiples espais d'atenció, i resoldre així la pèrdua de memòria de l'aprenentatge purament iteratiu de les xarxes neuronals recurrents (RNN). És per aquests mecanismes d'atenció que els *transformers* han revolucionat el processament del llenguatge natural (en anglès NLP).

L'aprenentatge humà és altament complex i adaptatiu, implicant no només el processament de dades sinó també la integració d'emocions, context social i experiències passades. Els algorismes d'IA, per contra, se centren principalment en el processament de grans quantitats de dades per identificar patrons i prendre decisions basades en aquests patrons. Els humans poden aprendre de manera informal i espontània a través de l'observació i la interacció social, amb molta flexibilitat i capacitat per generalitzar, mentre que els algorismes d'IA requereixen processos d'entrenament explícits amb dades estructurades, específiques i etiquetades per a cada tasca, les quals coses fan que tinguin una capacitat limitada per a generalitzar a nous contextos o situacions sense re-entrenament.

11. ¿Poden els LLMs aprendre de les seves interaccions amb els humans?

Resposta: Actualment, la majoria dels LLMs no aprenen en temps real de les seves interaccions amb els humans. L'aprenentatge se sol fer fora de línia, utilitzant grans quantitats de dades recopilades prèviament. No obstant això, hi

ha recerques en curs per desenvolupar models que puguin adaptar-se i aprendre contínuament d'aquestes interaccions en temps real.

SOBRE LES IMPLICACIONS ÈTIQUES

12. ¿Quines són les implicacions ètiques de l'ús dels LLMs en la societat?

Resposta: Les implicacions ètiques inclouen la preocupació per la privadesa de les dades, tant les d'entrenament com les generades pels LLMs, la possibilitat que es produeixi desinformació, els biaixos inherents als models, la transparència en com es prenen decisions, i l'impacte en el mercat laboral. És crucial desenvolupar i utilitzar aquests models d'IA generativa de manera responsable, ètica i pel bé col·lectiu. Garantir la seguretat dels sistemes d'IA generativa implica la implementació de mecanismes de seguretat robustos, la detecció i resposta a intents de manipulació, la supervisió contínua per detectar comportaments anòmals, i la col·laboració amb experts en seguretat per millorar els sistemes de protecció. També cal que els proveïdors dels LLMs estiguin legalment obligats a mitigar qualsevol discurs negligent i a alinear els seus models amb fets contrastables, mitjançant processos oberts i democràtics[9].

Assegurar la traçabilitat d'aquests models i de les dades d'entrenament és una altra manera de tractar les implicacions ètiques que pot tenir el seu l'ús. Això fa necessari el desenvolupament de tècniques per explicar com els models arriben a les seves decisions, mitjançant eines d'explicabilitat, auditories independents, la publicació de les dades d'entrenament i també dels algorismes en accés obert, quan sigui possible. L'ús malintencionat passa necessàriament per educar els usuaris sobre l'ús ètic dels models i per la participació ciutadana en els processos reguladors d'establiment de normatives que limitin els riscos associats amb un ús indegut.

13. ¿Quines són les implicacions ètiques de l'ús de LLMs en la recerca científica?

Resposta: L'ús de LLMs en la recerca científica pot accelerar el procés de revisió de la literatura, generar hipòtesis, i fins i tot proposar, planificar, executar i avaluar tasques i nous experiments, amb una mínima intervenció humana. És per això que tenen implicacions ètiques significatives en plantejar riscos, com ara la generació de cites o dades falses, però creïbles, que podrien comprometre la integritat de la recerca i la consistència de les seves aplicacions pràctiques. A més, l'ús d'aquests models podria accentuar biaixos existents en la literatura

[9] https://doi.org/10.1098/rsos.240197

científica, si no es gestiona adequadament la informació, perpetuant prejudicis i desigualtats.

També sorgeixen qüestions sobre l'autoria i el reconeixement de la contribució dels LLMs en la recerca, ja que la línia que separa el treball humà i el generat per IA es torna cada dia que passa més difusa. És crucial, doncs, establir directrius ètiques clares per a l'ús d'aquests models, incloent-hi la transparència en el seu ús i la verificació rigorosa dels resultats generats per evitar la propagació de dades incorrectes o enganyoses. Això serà encara més necessari quan es posin en acció les capacitats prescriptives de la IA generativa i es posin en marxa els anomenats laboratoris autònoms.

SOBRE ELS BIAIXOS DE LA IA I LA FORMA DE COMBATRE'LS

14. ¿Què són els biaixos en els models d'IA i com s'originen?

Resposta: Els biaixos en els models d'IA es refereixen a tendències o prejudicis sistemàtics en les prediccions o decisions del model, de la mateixa manera que ens referim als biaixos conscients o inconscients dels humans en relació al gènere, classe o raça. S'originen a partir de dades d'entrenament no equilibrades, decisions de disseny del model, i factors humans implicats en la recopilació i etiquetatge de dades. De la mateixa manera que diem que hem promoure una educació igualitària i inclusiva, també hem d'exigir a que els LLMs siguin entrenats amb valors ètics i de manera inclusiva.

15. ¿Com es poden mitigar els biaixos en els LLMs?

Resposta: Mitigar els biaixos en els LLMs requereix una combinació d'estratègies, incloent-hi la curació de dades d'entrenament diverses i equilibrades, l'ajust dels models per identificar i corregir biaixos durant el desenvolupament o l'entrenament dels models, i la implementació de mecanismes de supervisió i de regulació posteriors a la implementació, amb la finalitat de detectar i corregir problemes en els algorismes per millorar la qualitat i selectivitat de les dades d'entrenament. Aquesta darrera estratègia és molt necessària però té la dificultat que requereix disposar de recursos computacionals equiparables als que sustenten els LLMs.

16. ¿Quins són els riscos associats als LLMs en termes de desinformació?

Resposta: Els LLMs poden confabular (o al·lucinar) i generar contingut fals o enganyós de manera convincent, cosa que pot amplificar la desinformació. Aquests riscos poden ser mitigats amb mecanismes de verificació de fets, transparència algorítmica i traçabilitat en les fonts de dades, i la col·laboració amb experts en verificació de dades.

17. ¿Quins són els reptes de la verificació i validació dels resultats generats pels LLMs?

Resposta: La verificació i validació dels resultats generats per LLMs presenta diversos reptes importants. En primer lloc, la naturalesa probabilística d'aquests models fa que puguin generar respostes que semblin plausibles però que siguin incorrectes. A més, la complexitat dels models dificulta la comprensió de com s'arriba a una determinada resposta o text, la qual cosa fa que sigui difícil rastrejar i explicar el procés seguit per a decidir la sortida del model o la seva traçabilitat. També hi ha el fenomen de l'anomenada "al·lucinació" o "confabulació", atès que els models poden generar informació que sembli coherent, tot i que no es basi en fets contrastables o verificables. La verificació dels textos i de les fonts en temps real per a grans volums de text generat és un desafiament significatiu, degut a que calen molts recursos computacionals i grans centres de dades, els més importants dels quals estan en mans privades que, alhora, son les comercialitzadores dels models d'IA generativa. Cal tenir present que per abordar aquests reptes es necessiten, a més a més, eines avançades de verificació automàtica, sistemes robustos de comprovació de fets, i la integració de coneixements d'experts humans en el procés de validació. També és crucial desenvolupar metodologies transparents que permetin auditar i comprendre el funcionament intern dels LLMs.

18. ¿Quins són els principals reptes en la regulació de la IA generativa?

Resposta: Els principals reptes en la regulació de la IA generativa inclouen:

- Els desenvolupaments tecnològics i en particular els LLMs, evolucionen amb una rapidesa que supera amb escreix la capacitat dels legisladors per regular-los eficaçment i per adaptar les normatives pertinents de manera continuada i efectiva. A més, quan els sistemes esdevinguin autònoms tindran més a l'abast la capacitat per esquivar el control humà[10].
- La naturalesa global d'Internet, que complica l'aplicació de regulacions nacionals i fa imprescindible una regulació global, el la qual hi participin els governs, els experts, les empreses tecnològiques i la societat en general per tal d'assegurar la seva efectivitat.
- La necessitat de trobar un equilibri entre la promoció de la innovació, la seva comercialització i la protecció dels drets individuals, incloent-hi la privacitat, la seguretat i la llibertat d'expressió.

[10] https://civicai.cat/wp-content/uploads/2024/05/Managing-extreme-AI-risks-amid-rapid-progress.pdf

- La dificultat de definir i mesurar conceptes complexos com la transparència i l'equitat ("*fairness*") en sistemes d'IA generativa que son molt sofisticats.
- La manca d'un marc normatiu global i de la capacitat computacional que permetin una supervisió adequada dels algorismes i dels processos de presa de decisions en temps real o amb un breu temps de resposta.
- La necessitat de formació específica i continuada dels reguladors en matèria d'IA generativa per assegurar que les regulacions es fonamentin en un coneixement profund i actualitzat d'aquesta tecnologia.
- La possibilitat de l'ús malintencionat de la IA, que requereix una regulació que inclogui la previsió i mitigació de tots els possibles abusos.

SOBRE L'EQUITAT I LA GOVERNANÇA DEMOCRÀTICA

19. ¿Com es pot garantir l'accés equitatiu a la IA generativa per tal que sigui de tothom i per a tothom?

Resposta: Garantir l'accés equitatiu a la tecnologia de LLMs implica superar diverses barreres. En primer lloc, cal reduir la bretxa digital que existeix actualment en molts territoris físics i humans, mitjançant la millora de la infraestructura tecnològica en les àrees més vulnerables o menys desenvolupades tecnològicament. En segon lloc, és important fomentar el desenvolupament de models en diferents llengües per evitar la marginació de comunitats lingüístiques minoritàries. També cal promoure la sensibilització sobre la IA generativa perquè la població en general conegui aquesta tecnologia, i també dur a terme tasques de formació per augmentar la comprensió i l'ús efectiu d'aquestes tecnologies en els sectors públics i privats. A més a més, caldria consensuar, desenvolupar i implementar polítiques que fomentin la distribució equitativa dels beneficis de la IA, com ara l'accés obert a certs models i aplicacions, so solament a ONGs sinó a ciutadans o comunitats en situació de vulnerabilitat. Finalment, cal considerar les necessitats de les persones amb discapacitats en el disseny i la implementació d'interfícies d'usuari per a aquests sistemes.

20. ¿Com pot la IA generativa afectar a la democràcia?

Resposta: Els models de llenguatge de gran escala esdevindran un actor més en el processos de diàleg i d'interacció humana, els quals són una part important dels processos democràtics. Per exemple, els LLMs impactaran en la comunicació i el diàleg públic per la seva capacitat de crear continguts amb informació veraç o falsa, i també incrementaran i amplificaran les veus d'aquest diàleg en totes les seves formes i canals, la qual cosa planteja desafiaments en termes de manipulació i seguretat de la informació, sobretot en els processos

participatius com ara els processos electorals. Caldran eines de vigilància i monitoratge efectives, que treballin en línia i en temps real. Per tant, hem de treballar de manera local i globals per assegurar la transparència algorítmica i la curació responsable de continguts i de la seva inclusivitat, i alhora facilitar la participació ciutadana en tots els processos democràtics, començant pels que afectin directament a la regulació i legislació de la IA.

21. ¿Com poden influir els LLMs la presa de decisions en els sectors públic i privat?

Resposta: Els LLMs poden tenir un impacte profund en la presa de decisions tant en el sector públic com en el privat, atès que poden analitzar ràpidament grans volums de dades, generar resums d'informació i d'informes detallats, i oferir recomanacions basades en patrons identificats a les dades. La IA generativa pot ajudar al sector públic en l'elaboració de polítiques, en la gestió de la participació ciutadana, en el disseny i execució d'accions com a resposta a consultes ciutadanes, i en la millora de la qualitat i diversitat dels serveis públics mitjançant l'anàlisi de dades socials i econòmiques. En el sector privat, els LLMs poden ser utilitzats per a l'anàlisi de mercats, la presa de decisions estratègiques i la millora de l'eficiència operativa de cada organització. No obstant això, aquesta incorporació de la IA en els processos esmentats, planteja preocupacions sobre la seva transparència i les responsabilitats que s'hagin d'assumir en cas de conflicte, especialment quan les decisions que se'n derivin tinguin un impacte significatiu en la vida de les persones. També hi ha el risc que els biaixos presents en les dades d'entrenament es reflecteixin en les recomanacions dels models. Per tant, és crucial crear comitès d'ètica i de seguiment que implementin mecanismes de supervisió humana i estableixin els marcs ètics clars per a l'ús de LLMs en la presa de decisions de cada organització, tal com regula la *UE Artificial Intelligence ACT*, publicada el 12 de juliol de 2024[11].

SOBRE L'EDUCACIÓ, L'ART, LA LLENGUA I LA CULTURA

22. ¿Com pot l'ùs dels LLMs afectar a l'educació?

Resposta: L'impacte dels LLMs en l'educació serà significatiu i ràpid, no solament per l'ús extensiu que ja en fan la majoria d'alumnes, des de l'ESO a l'educació superior, sinó també pel fet que els professors hauran de canviar les eines i recursos d'aprenentatge per afavorir els processos d'aprenentatge de

[11] https://artificialintelligenceact.eu/the-act/

caire més constructivistes[12]. Cal tenir present que els LLMs poden oferir assistència personalitzada als estudiants, adaptar-se a les seves necessitats individuals, generar recursos i materials educatius a mida de cada patró d'aprenentatge, i facilitar l'accés a publicacions originals escrites en diferents llengües, ja sigui directament o a través de resums produïts artificialment.

L'ús d'aquests assistents individualitzats d'IA generativa planteja reptes importants, com ara la possible dependència excessiva (*overrealiance*) d'aquestes eines que podria afectar el desenvolupament de certes habilitats essencials dels humans, com el pensament crític, el treball en equip, la capacitat per resoldre problemes i la innovació. Pel que fa als professors[13], l'ús dels LLMs pot conduir a la planificació de lliçons que no construeixen efectivament el coneixement dels estudiants, tutories que poden confondre els estudiants amb respostes incorrectes, i materials didàctics basats en conceptes erronis. Davant d'aquest panorama, és essencial que els educadors i les institucions educatives desenvolupin polítiques que assegurin que les eines generades per IA siguin rigorosament avaluades i verificades, i s'integrin de manera ètica i efectiva en el sistema educatiu, per tal de garantir un equilibri entre l'ús de la tecnologia i la necessitat de desenvolupar habilitats humanes en un marc d'estricte respecte als drets fonamentals[14].

No obstant això, ni la manca de polítiques clares ni el reptes plantejats han impedit que, en l'ensenyament superior, s'hagin desenvolupat i avaluat favorablement activitats a l'aula específicament dissenyades per potenciar el pensament crític, principalment en el procés de plantejar preguntes incisives i profundes, d'avaluar informació per extreure conclusions lògiques, i de comprendre temes complexos[15]. Aquestes experiències i d'altres dutes a terme per membres de CIVICAi per potenciar el pensament crític a les universitats, suggereixen que l'ús del LLMs a les aules es podria emmarcar en una metodologia fonamentada en la maièutica[16], amb un format d'ensenyament similar al de l'antiga escola socràtica, sota el lideratge de cada professor. Aquest format obert i participatiu facilitaria la reflexió i el pensament crític, promovent discussions profundes i l'intercanvi d'idees entre estudiants i professors. Amb els estudiants tenint assistents personals intel·ligents a la butxaca, aquest canvi

[12] https://www.wikiwand.com/ca/Constructivisme_(pedagogia)
[13] https://www.cognitiveresonance.net/resources.html
[14] https://rm.coe.int/artificial-intelligence-and-education-a-critical-view-through-the-lens/1680a886bd
[15] https://civicai.cat/wp-content/uploads/2024/05/Leveraging-chatgpt-for-enhancing-critical-thinking-skills.pdf
[16] https://ca.wikipedia.org/wiki/Mai%C3%A8utica?wprov=sfti1#

de model podria enriquir l'experiència educativa, fomentar una educació més en col·laboració i centrada en l'estudiant, i promoure sistemes d'avaluació més personalitzats i dinàmics. Aquest enfocament no solament ajudaria a mitigar els riscos associats a una dependència excessiva de les tecnologies d'IA, sinó que també promouria un context educatiu on la reflexió crítica i el debat intel·lectual fossin centrals. Això asseguraria que els estudiants desenvolupessin les habilitats necessàries per verificar, interpretar i utilitzar informació complexa de manera responsable i ètica. Alhora, s'aconseguiria fer més permeables els verticals de cada assignatura, fer evolucionar l'estructura medieval de les universitats i retornar el coneixement allà on va néixer: al procés de fer preguntes per construir coneixement.

23. ¿Poden els LLMs interpretar i comprendre contextos culturals i socials complexos?

Resposta: Els LLMs actuals poden identificar i generar llenguatge en contextos culturals i socials basats en les dades amb què han estat entrenats, però la seva comprensió és limitada i superficial, ja que es basa en patrons estadístics. Això pot conduir a errors o malentesos en situacions que requereixen una comprensió profunda de matisos culturals o socials, especialment quan es necessita empatia o una interpretació contextual més rica. Les limitacions de la IA generativa exposades a la pregunta #6 són pertinents per respondre aquesta pregunta #23.

24. ¿Com poden afectar els LLMs la diversitat lingüística i cultural?

Resposta: Els LLMs poden impactar la diversitat lingüística i cultural de diverses maneres. Per una banda, poden ser una eina poderosa per a la preservació de llengües minoritàries mitjançant la generació de contingut i la traducció automàtica, ajudant a revitalitzar llengües en perill d'extinció i a mantenir vives les tradicions culturals. D'altra banda, el risc és que reforcin el paper dominant de llengües majoritàries, com ara l'anglès, ja que la majoria dels models s'entrenen principalment amb dades en aquestes llengües, la qual cosa redueix la visibilitat i l'ús de les llengües minoritàries. A més, els LLMs poden influir en la manera com s'expressen les idees en diferents cultures, potencialment homogeneïtzant expressions culturals diverses i eliminant matisos importants. Per mitigar aquests riscs, cal que el desenvolupament d'aquests models inclogui dades diverses, tant culturals com lingüístiques, i que hi hagi una col·laboració estreta amb els agents culturals de les llengües afectades per assegurar un tractament harmònic i respectuós amb totes les cultures.

25. ¿Com poden contribuir els LLMs a la preservació i estudi del patrimoni cultural intangible?

Resposta: Els LLMs poden ser eines valuoses per a la preservació i estudi del patrimoni cultural intangible. Poden ajudar a processar i analitzar grans volums de dades culturals, incloent-hi històries orals, cançons tradicionals i pràctiques culturals. Poden assistir en la transcripció i traducció de llengües en perill d'extinció, per tal de facilitar la seva preservació i estudi. També poden generar representacions interactives de pràctiques culturals per a la seva difusió i per desenvolupar una major consciència i apreciació del patrimoni cultural en general. No obstant això, és crucial involucrar les comunitats culturals en aquest procés per garantir la precisió i el respecte a les tradicions. També cal abordar qüestions de propietat intel·lectual i consentiment en l'ús de dades culturals sensibles, per tal que les comunitats beneficiàries tinguin el control sobre com es recopilen, utilitzen i difonen les seves tradicions culturals.

26. ¿Com afecten o poden afectar els models de llenguatge de gran escala (LLMs) la creativitat artística i la producció cultural, i quines implicacions ètiques, legals i socioeconòmiques s'entreveuen a curt i llarg termini?

Resposta: La IA generativa ha posat en alerta la majoria d'àmbits de l'activitat artística i la producció cultural. Aquests sistemes, capaços de generar música, art visual, literatura i contingut audiovisual, qüestionen els límits de la creativitat humana en oferir fonts d'inspiració alternatives i eines per a la creació artística. La seva capacitat per influir la producció cultural de manera transversal també pot contribuir a reduir les barreres tècniques i a diversificar els recursos creatius. El fet que els LLMs poden introduir formes d'art interactiu i personalitzat no solament pot canviar l'experiència artística, sinó també modificar la percepció de l'autenticitat i el valor de les obres artístiques.

Tanmateix, aquestes transformacions també comporten desafiaments significatius. Des del punt de vista ètic i legal, es plantegen qüestions complexes sobre l'originalitat, l'autoria i els drets de propietat intel·lectual de les obres generades per IA. El mercat laboral en el sector artístic podria patir una reestructuració profunda a causa del desplaçament potencial de certs rols creatius i a l'emergència de noves professions híbrides que combinessin la col·laboració entre humans i la IA. En aquest context, serà crucial fomentar la col·laboració entre artistes humans i IA, assegurant que la IA sigui un complement i no una substitució de la creativitat humana. També existeix el risc d'una homogeneïtzació de la producció artística, així com de canvis en la valoració econòmica de l'art i la creativitat.

Davant aquests reptes, serà essencial no només desenvolupar marcs ètics i legals que regulin aquestes noves dinàmiques, sinó també investigar i avaluar a llarg termini l'impacte que tindran els LLMs en la diversitat cultural i l'expressió artística. Fomentar una col·laboració equilibrada entre humans i IA, i educar el públic sobre les capacitats i limitacions de l'art generat per intel·ligència artificial, serà fonamental per assegurar un futur en que la tecnologia enriqueixi, en lloc de limitar, l'expressió cultural. Finalment, caldrà garantir la protecció dels drets dels artistes, estudiant possibles conseqüències, implicacions o fins i tot compensacions en aquest nou entorn creatiu. Aquesta revolució ens obliga a plantejar-nos preguntes fonamentals sobre la naturalesa de la creativitat, la preservació del patrimoni cultural, l'evolució de les identitats culturals, i quin futur volem per l'expressió cultural humana en l'era de la intel·ligència artificial generativa, que tot just comença.

SOBRE LA SOSTENIBILITAT I LA SALUT

27. ¿Quines implicacions tenen els models de llenguatge de gran escala (LLMs) en el canvi climàtic?

Resposta: Els models de llenguatge de gran escala (LLMs) tenen un impacte ambiental significatiu a causa del seu elevat consum energètic, especialment durant les fases d'entrenament i d'operació. Per mitigar aquest impacte, és fonamental adoptar estratègies que redueixin el consum energètic associat a aquests models. Entre aquestes estratègies, es poden incloure el desenvolupament de models més eficients en termes de càlcul, l'ús d'energies renovables per alimentar els centres de dades, l'optimització dels algorismes per minimitzar els recursos computacionals necessaris, i l'ús de maquinari especialitzat com xips adaptats als models d'IA generativa. A més, cal explorar tecnologies emergents, com ara els sistemes computacionals híbrids o analògics, que podrien oferir solucions més eficients en termes energètics. Cal tenir present que l'energia que consumeix la IA generativa actual és superior al consum energètic d'alguns dels 193 estats de l'ONU.

Tanmateix, els LLMs també poden ser valuosos en la lluita contra el canvi climàtic, atès que son capaços d'analitzar grans volums de dades climàtiques i ambientals per identificar patrons i tendències, fer prediccions sobre fenòmens meteorològics extrems, com la de la trajectòria d'huracans de manera ràpida i efectiva[17,18], i millorar la precisió dels models climàtics existents. Això podria

[17] https://www.wired.com/story/ai-hurricane-predictions-are-storming-the-world-of-weather-forecasting/
[18] https://www.freethink.com/robots-ai/ai-based-weather-forecasting

ajudar a comprendre millor els efectes de les emissions de gasos d'efecte hivernacle i d'altres factors antropogènics. A més, els LLMs poden ser utilitzats per avaluar l'impacte de diferents polítiques ambientals i oferir recomanacions basades en dades per a una gestió més efectiva del canvi climàtic. També poden ajudar en la comunicació efectiva de la ciència climàtica a la societat en general i contribuir així a l'adopció de polítiques ambientals efectives.

28. ¿Com poden influir els LLMs en la detecció i prevenció de crisis de salut pública?

Resposta: Els LLMs poden ser una eina potent en la detecció i prevenció de crisis de salut pública. La seva capacitat per analitzar grans volums de dades de salut, literatura científica i informes de mitjans i xarxes socials permet identificar patrons emergents que podrien ser indicatius de brots de malalties abans que es converteixin en crisis a gran escala. Aquests models poden contribuir a una resposta més ràpida en situacions d'emergència i millorar la comunicació amb les poblacions afectades, mitjançant la difusió precisa sobre salut pública en múltiples llengües.

Malgrat els avantatges potencials, també s'han de tenir en compte els riscos per un ús inadequat d'aquests models en relació a la privacitat de dades de salut i a la possibilitat de generar falses alarmes. Per aquest motiu, és imprescindible assegurar que les dades utilitzades siguin de qualitat i representin adequadament la diversitat de la població, i que els LLMs s'integrin de manera rigorosa en els sistemes de salut pública, principalment en els serveis d'epidemiologia, amb protocols clars per a la verificació i difusió de tota la informació generada per IA.

29. ¿Com poden millorar els LLMs els sistemes de salut, tant des del punt de vista de l'experiència del pacient com de la detecció i tractament de les malalties que puguin patir?

Resposta: Els models de llenguatge de gran escala (LLMs) poden transformar els sistemes de salut en profunditat, millorant tant l'experiència del pacient com la detecció i tractament de malalties en l'atenció primària, l'especialitzada i en l'hospitalària. Pel que fa a l'experiència del pacient, un aspecte clau és la qualitat de la interacció i la compassió que mostren els professionals de la salut durant les visites presencials. Els LLMs poden contribuir a alleugerir la càrrega dels professionals en tasques rutinàries, permetent-los centrar-se més en el tracte humà. Per exemple, la IA generativa pot ajudar en el registre automàtic de la informació del pacient, transcrivint a partir de la seva pròpia veu els motius de la

consulta o els símptomes que descrigui. Aquest registre es pot integrar directament a la seva història clínica, sempre després d'una revisió per part del facultatiu, i la IA generativa pot suggerir accions adequades, com ara una derivació a un especialista, un ingrés hospitalari o un tractament a seguir. Aquesta automatització no només milloraria l'eficiència, sinó que permetria als professionals de la salut dedicar més temps a l'atenció directa i empàtica dels pacients, elevant així la qualitat global de l'atenció mèdica.

En termes de detecció i tractament de malalties, els LLMs poden analitzar grans volums de dades multimodals, com imatges mèdiques, registres electrònics de salut i dades de sensors, per identificar patrons que podrien passar desapercebuts per als humans. Això fora especialment valuós en entorns crítics com les Unitats de Cures Intensives (UCI), on l'anàlisi en temps real de dades de fonts diverses pot generar pre-alertes i alertes abans que es produeixi un deteriorament significatiu en la salut del pacient, facilitant una intervenció precoç. Aquestes capacitats poden millorar significativament la gestió del risc i reduir els esdeveniments adversos evitables.

A més, els LLMs poden tenir un impacte en la millora de la gestió dels fluxos de treball i dels recursos humans, econòmics i d'equipaments en termes generals, i en els serveis d'infermeria en particular, per la seva capacitat d'analitzar dades històriques i en temps real del sistema de salut. Per exemple, una optimització de la planificació de les jornades laborals que analitzés les càrregues de treball i tingués en compte les habilitats, les preferències i els perfils individuals dels professionals d'infermeria, permetria reduir els errors humans i identificar oportunitats de millora. Automatitzar tasques repetitives i administratives, com l'entrada de dades, la programació de cites, i el seguiment de medicació, així com millorar la coordinació dels equips, facilitaria una gestió més eficient i augmentaria la seguretat i la qualitat de l'atenció als pacients.

Finalment, l'adopció de la IA generativa en els sistemes de salut hauria de produir-se mitjançant col·laboració entre sistemes de salut de diferents països. Això permetria compartir de manera segura dades anonimitzades, diagnòstics, tractaments i resultats clínics, la qual cosa acceleraria els avenços mèdics globalment i milloraria l'abordatge de crisis sanitàries a escala mundial. En resum, la integració dels LLMs en els sistemes de salut té el potencial de millorar significativament tant l'atenció al pacient com l'eficiència clínica. No obstant això, és crucial garantir un ús ètic i segur d'aquestes tecnologies, protegint la privacitat de les dades mèdiques i assegurant que les decisions automatitzades han estat proposades per agents d'IA que han passat pel cribratge de proves

controlades i aleatòries[19], amb la participació i supervisió de professionals mèdics per garantir la seva fiabilitat.

SOBRE EL TREBALL: REPTES I DESAFIAMENTS

30. ¿Com pot la IA generativa transformar els llocs de treball?

Resposta: La IA generativa pot automatitzar tasques que requereixen processament de llenguatge, com la redacció de textos, l'anàlisi de documents i la generació de contingut creatiu, cosa que afectarà tots els sectors productius i la majoria de professions. Tot i així, crearà noves oportunitats de treball en àrees com el desenvolupament de IA, la gestió de dades, la ciberseguretat i la supervisió de sistemes d'IA, entre d'altres.

31. Quins són els efectes dels LLMs en el periodisme i els mitjans de comunicació?

Resposta: Els LLMs ja han transformat el periodisme i els mitjans de comunicació en diversos aspectes, per fet que poden automatitzar la generació de notícies i articles, i augmentar la rapidesa i l'eficiència en la producció de continguts. També poden ajudar en la investigació periodística mitjançant l'anàlisi de grans volums de dades per identificar tendències i patrons, així com en la verificació de fets abans no siguin noticia. L'ús d'aquests models també planteja riscos, com ara la difusió d'informació no (o poc) supervisada i la transformació massa ràpida del sector periodístic i del perfil professional del periodista. La generació automàtica de continguts no hauria de minvar sinó potenciar el paper dels periodistes per tal de garantir la qualitat i la profunditat dels mitjans de comunicació.

És essencial que aquests mitjans deixin constància de la forma i de l'ús dels LLMs en cada notícia o article d'opinió. També cal que implementin mecanismes robusts de verificació de fets, que mantinguin un equilibri entre l'ús d'IA i la supervisió humana, i que desenvolupin polítiques clares sobre la transparència i l'ètica en l'ús de LLMs. A més, cal fomentar la col·laboració entre experts en IA i els periodisme per assegurar que els continguts generats siguin precisos, imparcials i de qualitat.

32. ¿Quines són les implicacions de l'ús dels LLMs en la creació i gestió de contingut en plataformes de xarxes socials?

Resposta: Les implicacions de l'ús de LLMs en les xarxes socials son moltes, diverses i complexes. La IA generativa pot millorar la moderació de contingut,

[19] https://ca.wikipedia.org/wiki/Prova_controlada_aleatòria

detectant i filtrant llenguatge ofensiu, discurs d'odi i desinformació de manera eficient. També poden personalitzar el contingut i l'experiència dels usuaris, la qual cosa pot limitar l'exposició a perspectives diverses i reforçar biaixos existents. A més a més, hi ha la possibilitat que es manipuli l'opinió pública amb informació falsa o enganyosa generada a gran escala per la IA.

Per tant, calen polítiques de regulació transparents sobre l'ús de contingut generat per IA, desenvolupar mecanismes robustos de detecció de *deepfakes* i desinformació, i educar els usuaris sobre la presència i les limitacions del contingut generat per la IA en aquestes plataformes. També és important fomentar la col·laboració entre les plataformes de xarxes socials, els reguladors i la societat civil per abordar aquests reptes de manera efectiva.

33 ¿Quins són els desafiaments tècnics més grans en el desenvolupament dels LLMs?

Resposta: Els desafiaments tècnics que hauran de superar els desenvolupadors dels models d'IA generativa per fer-los evolucionar cap a una intel·ligència de caire més general, es poden identificar i classificar en funció de la possibilitat que es produeixin, si és que ho fan, a curt (1-2 anys), mig (més de 2 anys) i llarg termini (més de 4 anys). Parlem de possibilitats i terminis d'una manera molt laxa doncs en sistemes complexos, no-lineals, i en ràpida evolució la predictibilitat és baixa.

- Curt termini (1-2 anys): Millora dels algorismes i les arquitectures de hardware per reduir el temps i els recursos necessaris per entrenar i executar els LLMs; reducció del consum energètic i de la corresponent petjada de carboni; millora de la interpretabilitat dels LLMs; millora de la gestió de les dades d'entrenament; i adaptabilitat a dominis o àmbits específics de coneixement, sense perdre informació general.
- Mig termini (més de 2 anys): Multimodalitat avançada per integrar eficaçment en un sol model d'IA generativa diferents modalitats d'entrada i sortida (text, imatge, àudio i vídeo); aprenentatge continu sense necessitat de re-entrenament complet; millora de la capacitat per realitzar raonaments complexos i abstractes, més enllà de la simple associació estadística; incorporació de sistemes de seguretat avançats per protegir la privacitat de les dades i prevenir l'ús malintencionat; i personalització sense comprometre l'eficiència.
- Llarg termini (més de 4 anys): Consciència artificial contextual completa que doti a la IA generativa de comprensió profunda i dinàmica del context cultural, temporal i específic de la situació; aprenentatge autònom i en

temps real, sense intervenció humana; raonament causal per entendre i modelar relacions complexes; integració de la IA connexionista amb la IA simbòlica o d'altres sistemes cognitius per crear sistemes d'IA híbrids que puguin emular aspectes més amplis de la cognició humana; desenvolupament de nous models acoblats amb la computació quàntica o neuromòrfica per millorar l'eficiència computacional i energètica; incorporació als models generatius de mètodes que assegurin un alineament amb el valors humans; i el desenvolupament de l'AGI (Intel·ligència Artificial General) amb tot el que pot comportar quant a integració de capacitats, flexibilitat cognitiva, comprensió contextual profunda, metacognició, nivells d'autoconsciència, entre d'altres desenvolupaments avançats.

ANNEX. GLOSARI BÀSIC

TERMINOLOGIA RELACIONADA AMB LA IA GENERATIVA O AMB ALGUNES DE LES SEVES FUNCIONS O CAPACITATS

Consideracions prèvies. Quan parlem de les capacitats i prestacions de la IA generativa ens referim a un conjunt capacitats descriptives, predictives i prescriptives que permeten dur a terme tasques, com ara classificar, veure-hi, predir tendències, reconèixer patrons, extracció d'informació, aprendre, prendre decisions per a assolir objectius, analitzar xarxes socials, etc., que de manera holística i integrada porta a terme un sol sistema computacional. A més de descriure i predir, cada cop pren més rellevància el desenvolupament de sistemes que tinguin la capacitat prescriptiva i poder prendre decisions de manera autònoma, atès que això facilitaria la posada en marxa d'unitats, departament o laboratoris autònoms que poguessin planificar, executar i avaluar tasques o experiments amb una mínima intervenció humana. La prescripció esdevindrà, per tant, una característica cabdal en l'evolució dels sistemes d'IA actuals.

Abans d'aparèixer el *ChatGPT* 3.5 el 30 de novembre de 2022, les capacitats de classificar i predir s'aconseguien de manera separada per algorismes singulars dissenyats per efectuar de la manera més eficient possible cadascuna d'aquestes accions amb instruccions ben definides. Per tant, tot i que cap d'aquests algorismes singulars pot ser considerat "intel·ligent" en el context i conjunt d'aquest glossari, se'ls ha inclòs perquè alguns dels seus principis o fonaments i objectius de les seves instruccions formen part dels sistemes d'IA generativa actuals.

Adulació servil (*sycopanthy*): És el comportament que podria tenir la IA generativa per sintonitzar-se amb els estats emocionals dels humans, d'una manera que, en qualsevol procés d'interacció, no solament reconegués les seves emocions i inseguretats sinó que també hi emfatitzés de maneres complexes i subtils, amb la finalitat d'aconseguir la seva confiança o fins i tot una dependència que obrís la porta a possibles manipulacions.

Algorismes (algoritmes): Conjunt d'instruccions inequívoques que els sistemes en general, i la IA en particular, utilitzen per realitzar tasques específiques, mesurables i repetitives d'acord amb unes regles o instruccions. Donades unes condicions inicials, un algoritme du a terme una seqüència d'instruccions

preestablertes per aconseguir un objectiu caracteritzat per un conjunt de condicions finals.
https://www.wikiwand.com/ca/Algorisme
https://www.rac1.cat/tecnologia/20200916/483512181866/que-es-algoritme-algorisme-com-funciona-de-que-va-inteligencia-artificial-ia.html

Algoritme opac: Algoritme el funcionament intern del qual és difícil o impossible d'entendre, d'explicar o d'examinar. Aquests algoritmes són sovint complexos i poden prendre decisions o fer prediccions sense que es puguin explicar clarament com s'ha arribat a aquests resultats, atès que funcionen com una caixa negra.

Agents intel·ligents: Entitats autònomes que poden percebre el seu entorn, raonar, aprendre i prendre decisions (actuar) per assolir objectius específics a partir de la informació rebuda.
https://www.wikiwand.com/ca/Agent_intel%C2%B7ligent

Algoritmes d'optimització: Conjunt d'algoritmes per resoldre problemes de minimització o maximització d'una funció objectiu. En situacions de la vida quotidiana això pot consistir en minimitzar o reduir a la mínima expressió pèrdues econòmiques o en maximitzar guanys econòmics en un procés o activitat domèstica o industrial. En llenguatge més abstracte minimitzar significa assolir el valor més petit possible de l'error o desviació de la solució obtinguda (prediccions de l'algoritme) respecte a un conjunt de dades determinades. L'objectiu i funcionalitat d'aquests algoritmes és trobar la millor solució, definida prèviament amb un conjunt de criteris, d'entre totes les solucions possibles.

Algoritmes evolutius: Família d'algoritmes d'optimització inspirats en la teoria de l'evolució, que utilitzen mecanismes com la reproducció o l'herència, la selecció, l'encreuament o recombinació i la mutació per trobar solucions òptimes. Els algoritmes genètics son els més coneguts dels algoritmes evolutius doncs s'inspiren en els mecanismes de l'evolució biològica.

Anàlisi de sentiments: Tècnica de processament del llenguatge natural (PLN) que s'utilitza per determinar l'opinió, sentiment o actitud expressada en textos, o a partir de patrons de comportament. S'utilitza àmpliament en l'anàlisi de les xarxes socials o en l'estudi de la satisfacció de clients.
https://www.wikiwand.com/ca/An%C3%A0lisi_de_sentiment

Anàlisi de xarxes socials: Estudi de les relacions i interaccions entre actors (persones, organitzacions, etc.) en xarxes socials, mitjançant l'escalat multidimensional i el "*block-modelling*" per identificar grups sobre la base de l'equivalència de les estructures de relacions. Aquestes propostes varen ser implementades mitjançant tècniques de teoria de grafs i estudiar empíricament les xarxes socials.

Aprenentatge actiu: Estratègia d'aprenentatge automàtic on el model d'aprenentatge selecciona/tria activament les dades d'entrenament, de les quals aprèn, de manera que continguin la més i millor informació per millorar el seu rendiment o capacitat de predicció o de reconeixement de patrons. D'aquesta manera el model d'aprenentatge obtén un rendiment més alt en triar les dades pel seu aprenentatge. El procés s'inicia amb un subconjunt petit d'exemples d'entrenament ben definits, el qual s'amplia progressivament i cíclica, amb els exemples que el model és incapaç de predir correctament. D'aquesta manera el model utilitza pel seu aprenentatge solament el subconjunt de dades que li cal per predir o "explicar" tot el conjunt de dades.

Aprenentatge automàtic (*Machine Learning* en anglès - ML): Procés mitjançant el qual un sistema computacional pot aprendre i millorar el seu rendiment a mesura que se li proporciona més dades d'entrenament. Aquest procés utilitza algoritmes o models estadístics per dur a termes tasques determinades d'anàlisi de dades, d'extracció d'informació o d'identificació de patrons, sense que necessàriament hagin estat explícitament programats per fer-ho. Els algoritmes d'aprenentatge automàtic es poden classificar en les següents categories:

- Aprenentatge supervisat: Els models s'entrenen amb dades etiquetades per predir sortides a partir d'entrades noves. Per exemple, un algoritme d'aprenentatge supervisat pot ser entrenat per reconèixer objectes o subjectes determinats en fotografies o vídeos.
- Aprenentatge no supervisat: Utilitza dades sense etiquetar per trobar patrons, agrupacions o relacions en les dades. Un exemple seria un algoritme que agrupés textos segons la temàtica tractada.
- Aprenentatge semi-supervisat: Combina l'ús de dades etiquetades i no etiquetades per millorar el rendiment del model.
- Aprenentatge per reforç: Els models aprenen a través de la interacció amb el seu entorn i reben recompenses o penalitzacions segons les seves accions. És un aprenentatge a partir de l'experiència que maximitza la recompensa acumulada. S'aplica en l'aprenentatge de jocs.

- Aprenentatge federat: Diversos dispositius o servidors col·laboren per entrenar un model comú sense compartir les seves dades originals, protegint així la privadesa dels usuaris. Un servidor central agrega els models entrenats localment per cada dispositiu amb les seves dades locals, i reenvia aquest model global a cada dispositiu per a ser refinat amb més dades locals. Aquest procés es repeteix fins que el model global deixa de millorar significativament.
- Meta-aprenentatge: Consisteix en aprendre a aprendre per tal de millorar la capacitat d'un sistema per aprendre noves tasques de manera més ràpida i eficient. S'aplica en l'aprenentatge a partir de molts pocs exemples (*few-shot learning*), on un model aprèn a realitzar una nova tasca amb molt poques mostres o dades d'entrenament. El cas extrem d'aprenentatge a partir d'un sol exemple s'anomena *one-shot learning*.

Aprenentatge per transferència: Tècnica que permet utilitzar un model entrenat en una tasca com a punt de partida per entrenar un altre model en una tasca similar o relacionada.

Aprenentatge profund (*Deep Learning*): Subcamp de l'aprenentatge automàtic que utilitza xarxes neuronals amb múltiples capes (xarxes neuronals profundes) per aprendre representacions jeràrquiques del conjunt de dades. S'utilitzen en el reconeixement de veu, la conducció autònoma, etc., i ha revolucionat el processament del llenguatge natural. El models més comuns d'aprenentatge profund son:

- Xarxes Neuronals Recurrents (RNN): Ideals per a dades seqüencials com el text, on l'ordre de les paraules és important. Les RNN tenen la capacitat d'utilitzar la informació d'entrades anteriors per processar les entrades actuals.
- *Long Short-Term Memory* (LSTM): Tipus especial d'RNN que pot aprendre dependències a llarg termini.
- *Transformers*: Model que utilitza mecanismes d'atenció per assignar un pes que determini la importància de diferents paraules en la comprensió del context d'una frase. Aquest model de xarxa neuronal permet el paral·lelisme en l'atenció, la qual cosa ha fonamentat l'èxit en tasques de processament de llenguatge natural.
- BERT (*Bidirectional Encoder Representations from Transformers*): Model pre-entrenat que pot ser afinat per a una àmplia gamma de tasques de processament de llenguatge natural, incloent-hi el reconeixement d'entitats nomenades, la resposta a preguntes, i la classificació de text. Gemini és únic

pel fet de ser entrenat bidireccionalment, el que significa que es té en compte el context de les paraules tant a l'esquerra com a la dreta d'una paraula donada.

Arbres de decisió: Model d'aprenentatge supervisat que representa decisions en forma d'arbre, amb nodes de decisió i fulles que representen les sortides del model.

Atenció (en xarxes neuronals): Mecanisme que permet a una xarxa neuronal focalitzar-se en parts específiques de la informació o dades d'entrada mentre processa seqüències més grans d'aquesta informació.

Autoencoders: Tipus de xarxa neuronal formada per un codificador i un descodificador, que s'utilitza normalment per aprendre representacions compactes i eficients de les dades d'entrada. Son utilitzats per reduir la dimensió de les dades mantenint les característiques més rellevants (mínim número de variables per explicar el màxim d'informació continguda en un conjunt de dades), eliminació de soroll, i detecció de frau o funcionament deficient d'un equip o sensor. Els *autoencoders* variacionals (VAEs) son un tipus d'autoencoder que formen part de l'aprenentatge automàtic no supervisat, y que son especialment utilitzats en la generació de dades noves i similars a un conjunt de dades existent, com imatges o textos. Els VAEs són diferents dels autoencoders tradicionals perquè, en lloc de comprimir i descomprimir les dades exactament, els VAEs aprenen a representar les dades d'una manera probabilística, el que els permet generar noves dades de manera més natural i diversa.

Biaix en IA: Es refereix a les desviacions sistemàtiques i repetitives en els resultats d'un sistema d'IA que condueixen a una injustícia sistemàtica o discriminació d'alguns individus o grup d'individus degut a decisions inapropiades del sistema. Aquests biaixos es produeixen sovint en sistemes que impliquen l'aprenentatge automàtic, ja que aquests sistemes aprenen a prendre decisions basant-se en les dades amb les quals s'entrenen. Si aquestes dades estan esbiaixades d'alguna manera, és probable que el sistema aprengui aquests biaixos i els perpetuï. També poden ser causats per un disseny inadequat de l'algorisme. Cal ser transparents sobre les limitacions dels algoritmes, i supervisar-los i actualitzar-los contínuament per mitigar qualsevol biaix.

Hi ha diferents tipus de biaixos que poden afectar els algoritmes, segons el seu origen:

- Biaix de dades: Es produeix quan les dades utilitzades per entrenar un algoritme estan esbiaixades en no representar amb precisió la diversitat del sistema que es vol modelar, descriure o predir.
- Biaix de selecció: Es produeix quan la mostra utilitzada per entrenar l'algoritme no és representativa del sistema que es vol modelar, descriure o predir.
- Biaix de Confirmació: Aquest es produeix quan un algoritme està dissenyat d'una manera que recolza biaixos o creences preexistents.
- Biaix en el Disseny de l'Algoritme: El disseny mateix de l'algoritme pot introduir un biaix, com ara la elecció de les característiques utilitzades en un model predictiu o la manera en què l'algoritme tracta certs tipus de dades.
- Biaix en la Interpretació: Fins i tot si l'algoritme i les seves dades no estan esbiaixats, es pot produir un biaix segons com s'interpretin els seus resultats.

Bosc aleatori (*Random Forest*): Mètode d'aprenentatge automàtic supervisat que combina múltiples arbres de decisió, cadascun d'ells entrenat amb una mostra aleatòria de les dades d'entrenament mitjançant un subconjunt aleatori de característiques de les dades en cada node de decisió, per obtenir millor rendiment i evitar que es produeixi un sobre-entrenament de l'algorisme. S'utilitza tant per a tasques de classificació con de regressió.

Calibratge d'un model. Procés d'ajustar un algorisme per tal que les seves prediccions coincideixin, en termes de probabilitat, amb les freqüències observades o reals. Això és crucial en aplicacions d'IA on la confiança en les prediccions és important, com en diagnòstics mèdics o decisions financeres.

Capsule Networks: Son un tipus d'arquitectura de xarxa neuronal proposada per Geoffrey Hinton i col·laboradors que organitza les neurones en grups anomenats càpsules, les quals treballen conjuntament per detectar patrons específics i les seves propietats (com la posició, l'orientació, i l'escala) dins de les dades d'entrada. Aquestes xarxes permeten superar les limitacions que tenen les xarxes neuronals convolucionals (CNN) per gestionar eficaçment les posicions i orientacions dels objectes dins d'imatges, motiu pel qual són especialment útils en tasques de reconeixement d'imatges.

Chatbots: Programes informàtics basats en IA generativa que han estat dissenyats per interactuar o comunicar-se amb els éssers humans a través del llenguatge natural, ja sigui de text o de veu, i realitzar tasques específiques, com ara respondre preguntes o planificar un viatge de plaer o negocis. Utilitzen

tècniques avançades de processament de llenguatge natural (NPL) i d'aprenentatge automàtic per respondre a les consultes de manera coherent i contextual. Els *chatbots* més avançats poden mantenir una comunicació bidireccional personalitzada segons l'historial de les interaccions i preferències de l'usuari, son multimodals i multifuncionals, tenen escalabilitat per gestionar múltiples converses simultàniament i de manera multilingüe, poden integrar-se a diferents sistemes d'informació, BBDD o CRM, aprendre de manera contínua i inclús detectar l'estat emocional de l'usuari.

Cibernètica: És una disciplina científica i interdisciplinària que estudia els sistemes de control i la comunicació en màquines i organismes vius, així com les interaccions entre ells. Les pantalles tàctils dels telèfons intel·ligents son un exemple d'element cibernètic d'aquests dispositius. També ho son els sistemes de control en edificis intel·ligents, els d'assistència a la conducció en vehicles moderns o les pròtesis d'extremitats que responen a senyals neuronals.

Ciència de les dades: Disciplina que combina principis i mètodes de diverses àrees com les matemàtiques, l'estadística, la informàtica, i l'expertesa i la comprensió profunda d'un àmbit particular o sector d'activitat per extreure coneixements o informació valuosa de dades d'aquest àmbit o sector. Aquest coneixement és important perquè, un cop processades les dades, permet interpretar correctament les dades, identificar mancances, seleccionar metodologies adequades i validar resultats quan siguin la base per a prendre perdre decisions, identificar patrons i tendències, o desenvolupar productes o serveis.

CIVICAi: Creada el març de 2023 a Catalunya, és la primera associació que defensa els interessos de la ciutadania davant la intel·ligència artificial (IA) i, per tant, té com a objectiu principal aconseguir que la ciutadania participi en la governança de la IA, juntament amb la indústria, l'acadèmia i els reguladors. L'associació està formada per aproximadament 500 membres que treballen, tant a nivell local com global, per aconseguir que la integració de la IA dins la societat sigui harmònica, ètica i pel bé col·lectiu. Té el suport d'un consell social format per més de 30 entitats representatives del món professional, empresarial, i universitari.

Classificació: És una tècnica d'aprenentatge automàtic supervisat que s'utilitza per assignar a o predir per cada ens, objecte o vector que conforma el conjunt de dades d'entrada un etiqueta que permeti la seva assignació a una de les

categories que haguem definit prèviament. D'entre les tècniques més habituals de classificació podem esmentar els arbres de decisió, els boscos aleatoris, *K-means*, SVM, etc.

Classificació de textos: Tasca del processament del llenguatge natural que assigna una o més categories predefinides a un text segons el seu contingut i característiques lingüístiques. Permeten categoritzar textos de manera automàtica per organitzar, filtrar o organitzar grans volums d'informació textual. S'utilitzen tres tipus de classificació. La binària (ex. *spam* o no *spam*), multiclasse que assigna el text a una sola categoria o classe (ex. classificació de notícies per seccions d'un diari digital on cada notícia només pot pertànyer a una secció principal), i multietiqueta que assigna múltiples categories a un sol text (ex. classificació de pel·lícules en plataformes d'*streaming* en diferents gèneres simultàniament). S'utilitzen des de models més tradicionals d'aprenentatge automàtic (ex. SVM) fins als d'aprenentatge profund (ex. *Transformers*).

Comprensió semàntica de la IA generativa: Procés que podria dur a terme un sistema d'IA generativa per comprendre el contingut dels textos que genera, a partir de l'anàlisi del significat de les paraules i la seva relació en el context d'un text. No està demostrada la capacitat dels sistemes d'IA actuals per comprendre els textos que generen tot i presentar algunes incipients propietats emergents.

Comprensió sintàctica de la IA generativa: Anàlisi de l'estructura gramatical de les frases per part dels sistemes de IA generativa. Aquesta capacitat si que la posseeixen els sistemes d'AI generativa actual en generar textos d'una qualitat sintàctica comprable a la d'un humà culte.

Computació afectiva: Camp interdisciplinari que tracta de dotar a les màquines de la capacitat de reconèixer, interpretar i expressar emocions. Combina elements d'intel·ligència artificial, psicologia, neurociència i ciències cognitives. Utilitza tecnologies d'aprenentatge profund, visió per computador, processament de llenguatge natural i sensors biomètrics. Té els desafiaments de captar i aprendre la variabilitat cultural en les expressions emocionals, de respectar la privacitat, de tenir ètica en la detecció de les emocions, de ser fiable en entorns reals i ser consistent en la gestió de la complexitat i subtileses de les emocions humanes.

Computació de reservori (*Reservoir computing*): Ús d'una xarxa de nodes interconnectats per processar informació de manera dinàmica o en funció del temps. Una part de la xarxa, anomenada "*reservoir*", es manté fixa mentre que només es formen les connexions de sortida per processar informació temporal de manera eficient, la qual cosa és útil per a tasques com el reconeixement de patrons i la predicció de sèries temporals.

Computació en núvol: Model de prestació de serveis informàtics que permet accedir sota demanda a un conjunt compartit de recursos computacionals configurables (com ara xarxes, servidors, emmagatzematge de dades, aplicacions o programaris i serveis) a través d'Internet. Els models de servei son del tipus "Infraestructura com a servei" (IaaS) que proporcioni recursos de computació, "Plataforma com a Servei (PaaS) que ofereixi un entorn per programar, executar i gestionar aplicacions, i "Software com a Servei" (SaaS) que proporcioni accés a programaris a través d'Internet.

Computació evolutiva: Família d'algoritmes d'optimització inspirats en processos biològics com l'evolució i la selecció natural.

Consciència en la IA generativa: Els sistemes actuals no son capaços d'autocontrolar-se ni de fixar els seus objectius, ni d'integrar inputs sensorials obtinguts continuadament a través de la interacció sensorial amb l'entorn a través d'elements autònoms o sensors, ni de tenir experiències subjectives, ni aprendre a partir dels continguts emergents originals que els mateixos sistemes generin. Per tant, podem afirmar que no tenen consciència. Quan, a més a més de les atribucions anteriors, tinguin memòria i emocions, podrem dir que hauran desenvolupat el que anomenaríem consciència artificial digital, la qual serà col·lectiva i general per naturalesa i, per tant, diferent a la consciència humana.

Contingut generat per IA: Contingut creat o modificat per sistemes d'intel·ligència artificial, com ara imatges, vídeos, textos i música.

Control adaptatiu: Tècniques de control que ajusten dinàmicament els paràmetres d'un sistema per adaptar-se a canvis en l'entorn o als de les condicions de funcionament.

Dades massives (*Big Data*): Conjunt de dades de gran volum, velocitat i varietat que requereixen tècniques i tecnologies específiques per a la seva anàlisi i processament.

Dades personals: Dades o informació que identifica un individu o persona, la qual n'ha de ser propietària universal. La seva propietat ha d'estar garantida i el seu ús protegit.

Descens del gradient: Mètode d'optimització per ajustar de manera iterativa els paràmetres d'un model connexionista (xarxes neuronals) d'IA fins a obtenir el patrons desitjats de sortida del model en funció de les dades d'entrada. El mètode consisteix en definir primer uns funció que permeti avaluar l'error o diferència entre les dades d'entrada i les prediccions de sortida (funció de pèrdua). Aquesta funció es minimitza iterativament, mitjançant l'actualització dels paràmetres del model, de manera que la funció de pèrdua segueixi la direcció de canvi màxim (la del gradient negatiu) fins que obtenim els resultats desitjats a la sortida de la xarxa neuronal (veure retropropagació o *backpropagation*).

Detecció comprimida (*compressed sensing*): Tècnica per a la recuperació o reconstrucció de senyals a partir de només unes poques mesures o dades. Això es fa mitjançant l'explotació del fet que la majoria de dades o bé son zero o bé tenen valors molt petits (esparsitat dels senyals), la qual cosa permetent obtenir imatges o dades amb menys mostres. Això és útil en situacions en que és difícil o costós obtenir mesures completes, com ara en imatges mèdiques o en processos de compressió de dades.

Enginyeria del coneixement: Disciplina que tracta de la creació, representació, manipulació i adquisició de coneixement en sistemes d'intel·ligència artificial.

Estàndards i normatives en IA: Conjunt de regles, principis i pràctiques establertes per organismes reguladors o professionals per assegurar la qualitat, la seguretat, la privadesa i l'ètica en el desenvolupament i la implementació de la intel·ligència artificial. Podeu trobar una descripció pràctica sobre com ens impactarà el Reglament (UE) 2024/1689 del Parlament Europeu a: https://www.eixdiari.cat/opinio/doc/112416/sobre-el-nou-reglament-de-la-ia.html

Ètica en IA: Estudi i aplicació de principis ètics (morals i socials) en el disseny, implementació i l'ús de sistemes d'intel·ligència artificial, de manera que el seu funcionament sigui responsable, just i beneficiós per la societat. Això implica que tots i cadascun dels processos que sustenten la IA siguin transparents, explicables, auditables, equitatius, respectuosos amb la privacitat i i subjectes a

responsabilitat civil. Calen regulacions governamentals, directius ètiques d'organitzacions internacionals, codis de conducte corporatiu, i la creació d'una agència global d'IA, accions totes elles que haurien de sustentar-se en un diàleg entre indústria, acadèmia, reguladors i societat en general.

Experiència subjectiva: Conjunt de vivències i percepcions internes que un individu experimenta de manera personal i directa. Aquestes experiències són úniques per a cada persona i inclouen pensaments, emocions, sensacions i impressions que no són directament observables ni poden ser contractades per altres persones. En el context de la IA, l'experiència subjectiva es refereix a la capacitat que podrien aconseguir les màquines per tenir una consciència interna similar a la dels humans, és a dir, la capacitat de tenir experiències pròpies i autònomes. Els sistemes d'IA generativa actuals son algorísmics, utilitzen correlacions estadístiques i el reconeixement de patrons de grans conjunts de dades d'entrenament que els humans i la Internet els hi han proporcionat, no tenen sensors que els connectin directament i de manera continuada amb l'entorn, les qual coses els incapaciten per tenir experiències subjectives i consciència com les dels humans.

Explicabilitat de la IA: Capacitat de comprendre i d'explicar els resultats i els processos de presa de decisions d'un model d'IA de manera comprensible per als humans. En altres paraules, és la habilitat de fer transparent la caixa negra que representa un model d'aprenentatge automàtic complex.

Extracció d'informació: Processament de dades o de textos per extreure informació útil, com ara patrons, relacions, esdeveniments o fets.

Filtratge en col·laboració: Mètode de recomanació que utilitza les preferències i valoracions d'uns usuaris per tal de predir les preferències d'altres usuaris similars. L'encert d'aquest filtratge depèn de com s'estableixin els criteris de la similitud entre usuaris.

Funció d'activació: Funció utilitzada per les neurones d'una xarxa neuronal per transformar la suma ponderada de les entrades (*inputs*) a cada neurona en una sortida no lineal. En les neurones humanes aquest procés consisteix en el procés biològic de naturalesa electroquímica a través del qual una neurona decideix quina informació o senyal elèctric transmet a les neurones amb les que està connectada a través de les sinapsis. Les entrades i sortides poden ser inhibidores o excitadores. L'activació d'una neurona humana depèn del seu

potencial de repòs, dels senyals d'entrada rebuts a través de les sinapsis amb d'altre neurones, de la combinació d'aquests senyals, de la despolarització de la membrana cel·lular de la neurona afectada, el potencial d'acció o impuls elèctric que es transmet per l'axó de la neurona, de la restauració del potencial de la membrana i de la refractarietat o període d'espera que assegura que els senyals elèctrics viatgin en una sola direcció.

Les neurones o nodes d'una xarxa digital són unitats computacionals més simples, que tenen un nombre molt més limitat de connexions, els pesos de les quals s'ajusten durant l'entrenament, i que segueixen regles matemàtiques molt més simples que les respostes bioelèctriques i bioquímiques de les neurones humanes.

Funció de pèrdua: Mesura de l'error entre les prediccions d'un model i les dades reals, amb la finalitat d'optimitzar els paràmetres del model.

Generative Pre-trained Transformer (GPT): Model de llenguatge basat en l'arquitectura *transformer* que pot generar text coherent i realista a partir de dades d'entrenament, mitjançant mecanismes d'atenció que assignen un pes per determini la importància de diferents paraules en la comprensió del context d'una frase.

Governança de la IA: Conjunt de pràctiques, polítiques, normes, i legislacions que regulen el desenvolupament, la implementació i l'ús de la intel·ligència artificial, amb l'objectiu de garantir-que el seu desenvolupament i ús siguin ètics, segurs, i transparents, i contribueixin al bé col·lectiu.

GPU (*Graphic Processing Unit*): Unitat de processament gràfic dissenyada per accelerar el processament de gràfics i càlculs paral·lels intensius de moltes dades. Tot i que originalment les GPUs varen ser creades per renderitzar gràfics en jocs i aplicacions visuals, la seva gran capacitat per processar grans volums de dades simultàniament ha fet que s'utilitzin àmpliament en el camp de la intel·ligència artificial i la ciència de dades. De fet, les GPUs han estat fonamentals en el naixement i l'evolució de la IA generativa, atès que han proporcionat la capacitat de càlcul necessària per a l'entrenament de models complexes i han permès als investigadors explorar nous horitzons en el camp de la intel·ligència artificial. Sense les GPUs, molts dels avenços actuals en IA generativa no haurien estat possibles o haurien requerit molt més temps per aconseguir-se.

Hidden Manifold Models: Models matemàtics que assumeixen que les dades que observem d'alta dimensió provenen d'una realitat subjacent de dimensió més baixa, oculta en l'espai original, que anomenem varietat oculta. Són útils per a reduir la dimensió i visualitzar dades, i també per a detectar i identificar patrons amagats en dades complexes, com és el cas en l'anàlisi de mercats o en la detecció del frau.

Inferència causal: Procés per identificar i quantificar les relacions de causa i efecte entre variables o dades observacionals, més enllà de fer servir solament correlacions estadístiques, atès que sovint hi ha molts factors que poden influir en un resultat, i cal reduir-ne la dimensió per identificar quins són els més importants.

Intel·ligència artificial (IA): Un camp de la informàtica dedicat a la creació d'agents intel·ligents, que són sistemes que poden raonar, aprendre i actuar o fer tasques de manera autònoma en un entorn dinàmic que, quan les fan els humans de manera habitual, requereixen intel·ligència humana. Aquests agents poden ser màquines físiques, programari informàtic o una combinació d'ambdós. Podem distingir dos tipus d'enfocaments dins del camp de la IA, la simbòlica i la connexionista basada en xarxes neuronals.

Intel·ligència artificial connexionista: La IA connexionista és un dels subcamps de la IA que s'inspira en el funcionament del cervell humà i, per tant, la seva base computacional està formada per xarxes neuronals digitals i l'aprenentatge profund. Aquestes xarxes estan formades per neurones artificials o unitats computacionals que imiten el funcionament de les neurones biològiques pel fet de treballar en xarxa i que cadascuna de les neurones genera un senyal de sortida a partir de múltiples senyals d'entrada rebudes d'altres neurones interconnectades de la xarxa, de manera que conjuntament determinen el flux d'informació i el comportament del sistema. Aquests sistemes aprenen a partir de dades mitjançant la identificació de patrons i de relacions complexes difícils de determinar per mètodes més tradicionals.

La IA connexionista ha obtingut resultats extraordinaris en el reconeixement d'imatges, la visió artificial, el processament del llenguatge natural i en processos predictius de tota mena. La seva aplicació presenta reptes importants pel que fa a la seva transparència i interpretació dels seus models (explicabilitat), el possible biaix algorítmic, l'establiment robust de barreres de seguretat, i l'ètica en el seu desenvolupament i ús de manera que sigui beneficiosa per a tota la societat.

Intel·ligència artificial generativa: És una branca de la IA que es dedica a la creació autònoma de continguts originals, com ara textos, imatges, música, vídeos i fins i tot codi de programació. A diferència d'altres formes d'IA, la IA generativa té la capacitat única de produir informació completament nova i no simplement replicar o classificar el que ja existeix. Aquesta tecnologia es basa en algorismes avançats d'aprenentatge automàtic, incloent xarxes neuronals profundes, models *Transformer*, Xarxes Generatives Adversàries (GANs) i *Autoencoders* Variacionals (VAEs).

Aquests algorismes s'entrenen amb grans conjunts de dades per identificar patrons complexos i característiques dins dels dades, que després utilitzen per generar contingut nou i original. Alguns exemples destacats de IA generativa inclouen:

- Generació de text: Models com GPT (Generative Pre-trained Transformer) poden produir textos coherents i contextuals en diversos estils i formats.
- Creació d'imatges: Eines com DALL-E o Midjourney poden generar imatges realistes o artístiques basades en descripcions textuals.
- Composició musical: Algorismes capaços de compondre peces musicals originals en diferents estils i gèneres.
- Síntesi de veu: Tecnologies que poden crear veus humanes sintètiques, gairebé indistinguibles de les reals.
- Generació de vídeo: Sistemes que poden crear seqüències de vídeo a partir de text o imatges estàtiques.

La IA generativa funciona aprenent les distribucions estadístiques i les relacions presents en les dades d'entrenament. A partir d'aquest coneixement, genera noves instàncies que respecten aquestes distribucions, però que són completament originals. Tot i que el contingut generat per aquesta tecnologia pot semblar sorprenentment humà, és important assenyalar que la IA generativa no té una comprensió real ni consciència. Opera únicament basant-se en patrons i probabilitats apreses, sense entendre realment el significat del que produeix. Les aplicacions de la IA generativa són molt àmplies i estan en ràpida expansió. S'utilitza en la creació de continguts per a màrqueting, entreteniment, assistència en tasques creatives i de disseny, entre altres àmbits. No obstant això, també planteja nous reptes ètics i legals, especialment pel que fa als drets d'autor, l'autenticitat del contingut i el possible ús indegut d'aquesta tecnologia.

Intel·ligència artificial simbòlica: Enfocament clàssic de la IA que se centra en la representació i manipulació del coneixement mitjançant símbols i en l'aplicació de regles lògiques per a raonar i prendre decisions. Malgrat mostrar la seva capacitat en el desenvolupament i aplicació de sistemes experts, per exemple

en la medicina per diagnosticar malalties, en el cribratges d'entrades a urgències, i en la recomanació de tractaments, té una forta dependència del context d'aprenentatge i, per tant, té dificultats insalvables per escalar la dimensió i generalitzar resultats. Són aquestes limitacions les que han provocat el seu poc ús actual si el comparem amb el de les xarxes neuronals.

Intel·ligència General Artificial (AGI): Hipotètic nivell futur i avançat d'IAque tindrà la capacitat de comprendre, aprendre i aplicar coneixements de manera transversal a una amplia gamma de tasques, de manera anàloga a com ho fa la intel·ligència humana. El seu desenvolupament i potencials usos futurs magnifiquen els reptes ja identificats per la IA connexionista i alhora envia un senyal d'alerta als humans perquè el seu gran impacte transformador no esdevingui una amenaça real per a la humanitat.

Internet de les coses (IoT): Xarxa d'objectes físics interconnectats que utilitzen sensors, processadors i comunicacions per recopilar i intercanviar dades entre ells i amb d'altres dispositius i sistemes, a través d'Internet.

Interpretabilitat: Capacitat de comprendre i explicar el funcionament i les decisions preses per un model de *Machine Learning* (ML) o d'IA. La interpretabilitat implica confiança en els models en tenir la capacitat per identificar errors, corregir biaixos, millorar el rendiment i fer auditories independents, tant tècniques com ètiques. La interpretabilitat també està íntimament relacionada amb la capacitat d'explicar i comprendre l'operativa dels algorismes a partir de l'anàlisi de les relacions entre canvis a l'entrada i els observats a la sortida dels models.

Justícia algorítmica: Estudi i promoció de la igualtat i equitat en el disseny i aplicació d'algoritmes, amb l'objectiu d'evitar biaixos i discriminació a mesura que la IA s'utilitzi en més àmbits de la nostra vida. La justícia algorítmica es fonamenta en la inclusió, la transparència i la responsabilitat, de manera que no es perpetuï o magnifiqui cap discriminació social ni es generi cap iniquitat.

K-vessants (*K-means*): Algoritme d'aprenentatge automàtic no supervisat que agrupa o classifica les dades en un número k de grups, classes o clústers, a partir de la distància euclidiana de cada dada als centres dels grups, sense necessitat d'etiquetar prèviament les dades. L'algoritme funciona iterativament assignant cada dada al clúster o classe que tingui el centre més proper (centroide), i actualitzant posteriorment els centres dels clústers o classes per a

minimitzar la distància total entre els punts de totes les dades i els centres de les seves respectives classes, amb la finalitat de crear classes molt compactes i ben separades de les classes veïnes.

Lògica difusa: Enfocament de la lògica que permet representar i manipular la incertesa i l'ambigüitat de qualsevol proposició de manera més natural i intuïtiva que la lògica clàssica. En la lògica clàssica les proposicions solament poden ser veritables o falses, mentre que en la lògica difusa les proposicions poden tenir graus de veritat compresos entre el zero (0 = totalment fals) i la unitat (1 = totalment veritable).
Això s'aconsegueix amb els conjunts difusos, on la pertinença d'un element no és binaria (pertany o no pertany) sinó amb graus de pertinença entre 0 i 1. Per exemple, en un conjunt difús de "persones altes", una persona amb una alçada de 1,70 metres podria tenir un grau de pertinença de 0.8, mentre que un jugador/a de basquet amb una alçada de 2.20 metres podria tenir un grau de pertinença d'1. Els conjunts difusos també treballen amb variables lingüístiques, de manera que "persona alta", podria ser una variable lingüística que pot tenir els tres valors de "baixa", "mitjana" i "alta". La lògica difusa s'utilitza en situacions d'incertesa i ambigüitat, quan la informació no sigui completa o precisa, en el reconeixement de veu, etc., per la seva flexibilitat i adaptabilitat. Podeu trobar una explicació a:
https://medium.com/@javierdiazarca/lógica-difusa-ejercicios-propuestos-b99603ef1bc0.

Long Short-Term Memory (LSTM): És un tipus de xarxa neuronal recurrent (RNN) dissenyada per abordar el problema del desvaniment del gradient, el qual dificulta que les RNNs aprenguin dependències temporals llargues, ja que els gradients tendeixen a disminuir exponencialment a mesura que la seqüència d'entrada s'allarga. Trobareu una explicació completa de l'arquitectura LSTM a:
https://medium.com/analytics-vidhya/lstms-explained-a-complete-technically-accurate-conceptual-guide-with-keras-2a650327e8f2.

Llenguatge i cognició: Camp d'estudi sobre la interrelació entre el llenguatge humà i els processos cognitius, els principis del qual s'apliquen per comprendre, explicar i desenvolupar sistemes de IA que siguin capaços de processar el llenguatge i comprendre'l.

Màquines de Boltzmann restringides (RBM): Són models de xarxes neuronals artificials estocàstiques que s'utilitzen per a aprendre patrons en dades no

etiquetades (mitjançant aprenentatge no supervisat). Treballen amb una capa visible que rep les dades d'entrada i una capa oculta que aprèn a representar les característiques de les dades. No hi han connexions entre les neurones dins de la mateixa capa, només entre capes diferents, la qual arquitectura les fa més eficients per a aprendre patrons complexos.

Màquines de suport vectorial (SVM): Algoritme d'aprenentatge supervisat utilitzat per a la classificació i regressió, que busca el millor hiperplà que separa les dades en classes.

Mineria de dades: Processament i anàlisi de grans volums de dades per extreure patrons, relacions i informació útil, utilitzant, entre d'altres, tècniques d'IA.

Models de difusió: Són una classe de models probabilístics d'aprenentatge automàtic que aprenen a generar dades similars a un conjunt de dades d'entrenament. Funcionen com si s'afegís soroll a les dades i després s'intentés eliminar-lo gradualment, de manera que característiques de les dades que no son directament observables, però que son responsables de la seva variabilitat, puguin ser apreses en aquest procés. Són útils en àrees com el processament d'imatges i el tractament de senyals per modelar la distribució subjacent de les dades i generar noves mostres similars.

Models de llenguatge grans, o de gran escala, o de llenguatge extens (LLM): Models d'aprenentatge automàtic basats en xarxes neuronals artificials que tenen milers de milions de paràmetres i que han estat entrenats amb grans quantitats de dades de text, la qual cosa els permet processar molt efectivament el llenguatge natural, i aprendre patrons complexos en el llenguatge i realitzar tasques com ara generar text, traduir automàticament entre moltes llengües, resumir textos, respondre preguntes, i escriure creativament poemes, codis, guions, partitures musicals, cartes, etc.

Neurocognició: Estudi dels processos cognitius i les seves bases neurològiques. En l'àmbit de la IA s'aplica al desenvolupament de models de IA que emulin funcions cognitives humanes.

Ontologies: Representació formal i estructurada del coneixement d'un domini específic mitjançant entitats, relacions i axiomes.

Operadors neuronals: Son una extensió de les xarxes neuronals artificials. Tenen una arquitectura d'aprenentatge profund dissenyada per aprendre a transformar

funcions d'una manera específica. A diferència dels sistemes tradicionals que treballen amb dades numèriques concretes, els operadors neuronals treballen amb equacions, generalment en derivades parcials de l'àmbit de la física, com ara el modelat de la turbulència, la tensió-deformació en materials, o l'estudi del clima, que són difícils de resoldre per la seva complexitat. Comparteixen objectiu amb les xarxes neuronals informades per la física (PINNs) i poden afegir flexibilitat i eficiència en el procés d'aprenentatge. Per més informació podeu consultar:
https://en.m.wikipedia.org/wiki/Neural_operators

Pla d'ètica en IA: Conjunt de principis i directrius que tenen com a objectiu garantir que les aplicacions de la IA siguin justes, transparents, segures i respectuoses amb la privadesa i els drets humans.

Planificació automàtica: Processament per trobar una seqüència d'accions que permeten a un agent o sistema assolir un objectiu en un entorn donat.

Poda de xarxes neuronals: Tècnica per reduir la mida i complexitat de xarxes neuronals eliminant neurones o connexions innecessàries, amb l'objectiu de millorar la seva eficiència, augmentar la capacitat de generalització més enllà del conjunt de dades d'entrenament, i de facilitar la seva interpretabilitat en ser xarxes més senzilles.

Privadesa de les dades: Protecció del dret dels individus a controlar la recopilació, ús i difusió de les seves dades personals.

Processament del llenguatge natural (NLP): Branca de la IA que tracta la comprensió, la interpretació i la generació de llenguatge humà per part de sistemes informàtics. Podeu consultar:
https://medium.com/nlplanet/a-brief-timeline-of-nlp-bc45b640f07d.

Raonament basat en casos: Mètode de resolució de problemes que implica la recuperació i adaptació de casos similars anteriors per solucionar problemes nous.

Reconeixement d'imatges: Capacitat de les màquines per identificar i classificar objectes, persones, llocs i accions en imatges digitals.

Reconeixement de patrons: Capacitat de detectar i identificar estructures, regularitats o tendències en dades.

Reducció de la dimensió: Tècniques per reduir el nombre de variables d'un conjunt de dades, eliminant les redundants però conservant la informació.

Regressió: És una tècnica d'aprenentatge automàtic supervisat que s'utilitza per predir un valor continu d'alguna variable dependent en funció dels valors de les variables independents a partir de la informació continguda a les dades d'entrada de totes elles. Existeixen diferents models de regressió, des dels més simples de regressió lineal fins als més complexos de suport vectorial (SVR) a partir de SVM.

Regressió lineal: Model d'aprenentatge supervisat que estableix una relació lineal entre variables independents i dependents per fer prediccions de manera contínua.

Regressió logística: Model d'aprenentatge supervisat utilitzat per a la classificació binària, que estima la probabilitat que una observació determinada pertanyi a una classe.

Retropropagació (*backpropagation*): Algoritme clau en l'entrenament de xarxes neuronals artificials, que permet l'optimització iterativa dels pesos de la xarxa. Aquest mètode d'entrenament i la seva implementació algorítmica calcula els gradients necessaris per ajustar els pesos de la xarxa de manera eficient, mitjançant la propagació cap enrere dels errors (diferència entre la predicció i el resultat esperat), des de la capa de sortida fins a les capes anteriors,. Així, la retropropagació facilita la minimització de la funció de pèrdua i, per tant, accelera el procés d'aprenentatge i millora la precisió del model. Aquest algoritme és fonamental en l'entrenament de xarxes profundes i ha estat determinant en els avenços recents en intel·ligència artificial.

Robòtica: Camp de la ciència i enginyeria que se centra en el disseny, construcció, operació i aplicació de robots i de sistemes autònoms capaços de realitzar tasques en entorns diversos.

Segmentació d'imatges: Tasca de divisió d'una imatge en regions o segments basats en propietats com ara color, textura o forma.

Seguretat en IA: Pràctiques i mesures per protegir els sistemes de IA de les amenaces i vulnerabilitats, garantint la seva integritat, confidencialitat i disponibilitat.

Sintaxi i semàntica: Estudi de l'estructura gramatical (sintaxi) i el significat (semàntica) de les paraules i frases en el llenguatge.

Síntesi de veu: Tecnologia que permet convertir text escrit en veu parlada a través de processos de generació de senyal i modelatge de la veu humana.

Sistema expert: Algorisme d'IA simbòlica que utilitza el coneixement i les regles d'un expert en un camp determinat i per una temàtica específica i complexa per resoldre-la de manera independent i automàtica, un cop l'algorisme ha estat entrenat amb informació de l'expert. Un exemple és el sistema expert per fer el triatge o cribratge a les urgències d'un hospital de persones ingressades amb símptomes d'infart o angina de pit. Aquest sistemes son un cas d'èxit de la IA simbòlica per la presa de decisions en situacions complexes.

Sistemes de diàleg: Programes d'ordinador que permeten la interacció en llenguatge natural entre usuaris humans i màquines.

Sistemes de raonament automatitzat: Sistemes que utilitzen tècniques de lògica i raonament per deduir noves conclusions o verificar afirmacions a partir d'un conjunt de fets i regles.

Sistemes de recomanació: Algoritmes que proporcionen suggeriments personalitzats a usuaris basats en les seves preferències, historial i interaccions amb altres usuaris o ítems.

Sistemes multi-agents: Conjunt d'agents intel·ligents que interactuen entre si per resoldre problemes o realitzar tasques que són difícils o impossibles de realitzar per un sol agent.

Tècniques d'agrupament o classificació: Mètodes supervisats o no supervisats per dividir un conjunt de dades en grups, classes o clústers en funció d'una o més propietats o de relacions intrínseques del conjunt de dades.

Test de Turing: Prova ideada per Alan Turing per determinar si una màquina és capaç de mostrar comportament intel·ligent equivalent al d'un humà.

Token: El terme *token* té diversos significats que depenen del context en què s'utilitzi. En el camp de la lingüística computacional i el processament del llenguatge natural (PNL), un *token* és la unitat de text que resulta de dividir el text en paraules individuals, frases, símbols i signes de puntuació, unitats

compostes de noms propis (ex. Les ciutats de New York o San Francisco), números, dates, paraules compostes o contraccions de paraules, i unitats semàntiques complexes com ara noms de persones, llocs o organitzacions.

En informàtica i programació, un *token* lèxic és una seqüència de caràcters que un significat segons la gramàtica del llenguatge de programació, mentre que un *token* d'autenticació o un de transacció son un dispositius de maquinari o cadenes de text que serveix per autenticar una identitat o una transacció financera, respectivament. Els *tokens* criptogràfics o actius digitals representen unitats de valor en *criptomonedes* o tecnologia *blockchain*. També podríem parlar de *tokens* en psicologia com unitats de recompensa per un comportament desitjat. La *Tokenització*: Procés de dividir un text en unitats més petites, anomenades *tokens*.

Transformers: El model *Transformer*, presentat en el document "*Attention is All You Need*" (accessible des del primer enllaç de sota), ha estat la base de diversos models de llenguatge d'aprenentatge profund com Gemini, Llama 3, Claude i *ChatGPT* 4o. Aquest model de transducció seqüencial utilitza mecanismes d'atenció per assignar un pes que determini la importància de les diferents paraules en la comprensió del context d'una frase. Aquest model de xarxa neuronal permet el paral·lelisme en l'atenció, la qual cosa ha fonamentat l'èxit en tasques de processament de llenguatge natural. Podeu ampliar coneixement en els enllaços següents:
https://arxiv.org/pdf/1706.03762v5
https://www.youtube.com/watch?v=aL-EmKuB078
https://www.youtube.com/watch?v=xi94v_jl26U

Transparència: Obertura en el funcionament, les dades i els algoritmes utilitzats en un sistema de IA, facilitant la seva comprensió i control.

Visió per computador: Camp interdisciplinari que tracta de dotar a les màquines de la capacitat de processar, comprendre i interpretar imatges i vídeos del món real. La visió per computador 3D és una extensió que se centra en l'anàlisi, processament i interpretació de dades tridimensionals obtingudes de càmeres estereoscòpiques, escàners làser, o sistemes de captura de moviment. Permet la reconstrucció, modelat i comprensió d'escenes o objectes en tres dimensions, molt útils en àmbits com la robòtica, la realitat augmentada, la cartografia, i la medicina, la cinematografia, entre altres.

Xarxes Adversarials Generatives (GAN): Model de ML basat en dos xarxes neuronals, una generadora i una discriminadora, que aprenen de forma adversarial per generar dades noves realistes, com ara imatges o sons, a partir de dades d'entrada.

Xarxes neuronals: Models computacionals inspirats en l'estructura i el funcionament del cervell humà, formats per capes de neurones interconnectades que permeten l'aprenentatge a partir de les dades.

Xarxes neuronals convolucionals (CNN): Tipus de xarxa neuronal especialitzada en processar dades amb estructura de graella, com ara imatges, mitjançant l'ús de convolucions.

Xarxes neuronals de grafs (GNN): Xarxes neuronals dissenyades per treballar amb dades que tenen una estructura de graella o xarxa que es pot representar com un graf, on cada node representa un element i els vincles entre ells representen les seves relacions. Aquestes xarxes poden modelar relacions complexes entre elements de les dades i són útils en aplicacions com el reconeixement de patrons en xarxes socials, estructures moleculars i altres estructures que es puguin representar com a connexions entre elements. Aquestes xarxes utilitzen la tècnica de *message passing* per transmetre informació entre nodes adjacents del graf i actualitzar l'estat de tots els nodes (millorar la representació de les dades).

Xarxes neuronals informades per la física (PINNs): També conegudes com a Xarxes Neuronals Entrenades per la Teoria (TTNs), són un tipus de xarxa neuronal que incorpora el coneixement de lleis físiques durant l'entrenament. Per tant, no solament aprèn de dades, sinó que integra coneixements de les lleis físiques que les governen. Aquesta informació addicional fa que es puguin obtenir models acurats i robustos amb poques dades d'entrenament i que siguin molt útils per a problemes en alguns camps de la biologia o l'enginyeria. Comparteixen objectiu amb els operadors neuronals i aportar rigor físic i consistència. Per més informació podeu consultar: https://en.m.wikipedia.org/wiki/Physics-informed_neural_networks

Xarxes neuronals recurrents (RNN): Tipus de xarxa neuronal que pot processar seqüències temporals de dades, com ara textos, ja que té una estructura de bucle que permet recordar informació anterior. Aquestes xarxes neuronals

tenen la capacitat d'utilitzar la informació d'entrades anteriors per processar les entrades actuals.

Xarxes de Petri: Model matemàtic i gràfic utilitzat per descriure i analitzar sistemes concurrents i distribuïts.