Any 2025
Número 61
PVP: 5 euros

nodes

El butlletí de l'ACIA

La intel·ligència artificial generativa

Edita:
Associació Catalana d'Intel·ligència Artificial (ACIA)
www.acia.cat
@acia_cat
@AssocCatIA
@acia_cat_rep
aciacat

Consell editorial:
Vicent Costa (IIIA-CSIC), coord.
Josep Puyol Gruart (IIIA-CSIC)
Cecilio Angulo (IRI-UPC)
Ulises Cortés (UPC)
Emilia López-Iñesta (UV)
Lledó Museros (UJI)
Eloi Puertas (UB)
Marco Schorlemmer (IIIA-CSIC)
Marc Torrens (ESADE-URL)

Revisió ortogràfica:
Núria Altés

Dipòsit legal i ISSN:
GI 1598/2008
ISSN 2171-5602
ISSN 2014-5020 (internet)

Disseny original:
Toni Casals

Maquetació:
Jordi Pardinilla Vilaplana

Sumari

nodes

Editorial

Un any més, fem camí

Vicent Costa

Benvolgudes sòcies,
benvolguts socis,

L'any passat celebràvem, amb el número 60 de *NODES*, els 30 anys del naixement de la nostra associació, l'Associació Catalana d'Intel·ligència Artificial. Després de la festa, calia seguir construint; per això, aquest any hem continuat amb pas ferm, mirada al futur i la intenció de mantenir la mateixa il·lusió que tenia el col·lectiu de persones de diverses universitats catalanes que va fer néixer l'associació 31 anys enrere. En definitiva, trobem que no hi ha un altre camí per fer comunitat.

Aquest número de *NODES* continua amb la divisió habitual dels darrers anys. D'una banda, pel que fa a les seccions fixes del butlletí, hem comptat amb una bona collita de contribucions per als apartats de «Societat» i «Tendències i futur». Ens trobem amb l'article «Intel·ligència artificial i ús dual: entre la potència creadora i el dilema ètic», escrit conjuntament per Alger Sans Pinillos, Marc Teixidor Vilarrasa, Adrián Ortega Novillo, Adrià Quintanas Corominas i Joan Farnós. Els autors ens plantegen reptes estratègics, ètics i socials del desplegament de tecnologies amb potencial dual, de manera que, un any més, s'afiança la secció de «Societat» com una constant en aquesta nova etapa de *NODES*. Per tancar aquesta part del número, tenim el luxe de comptar amb Joan Genís Valverde i també de retrobar-nos amb Llorenç Valverde, qui, com ja sabreu la majoria de vosaltres, va contribuir incommensuradament al desen-

Editorial

volupament del nostre butlletí des dels seus inicis. En aquest cas, pare i fill ens rebaten la classificació clàssica de sistemes d'intel·ligència artificial i fan una proposta excel·lent amb l'article «Una nova classificació per a sistemes d'IA», el qual s'emmarca en la secció de «Tendències i futur».

D'altra banda, a més de les seccions habituals, hi ha dos articles que tracten la discussió sobre el tema monogràfic d'aquest número: la intel·ligència artificial generativa. Aquest és un assumpte en boga que desperta atenció, admiració i preocupació a parts iguals. Així, d'una banda, Camilo Chacón Sartori proposa una anàlisi de la intel·ligència artificial generativa a l'article «En defensa de la IA generativa», sense caure en visions pessimistes, ingènues o resignades. De l'altra, com és habitual, hem preguntat a la comunitat de l'ACIA sobre el tema del monogràfic. L'anàlisi de les respostes i comentaris que hem rebut es troba a l'article «Què en pensen?», signat per Lucia Roda i Vicent Costa. Clou aquest número 61 de *NODES* el resum de la tesi doctoral guardonada amb el Premi Marc Esteva Vivanco 2024, que ens presenta Thanasis Zoumpekas.

Finalment, en nom del consell editorial, vos done, en primer lloc, les gràcies a les autores i als autors de les contribucions que formen aquest número 61 de *NODES*. En segon lloc, vos comunique que, en el proper número del butlletí, tenim previst incloure totes les propostes que ens heu fet arribar durant els darrers mesos.

Intel·ligència artificial i ús dual: entre la potència creadora i el dilema ètic

Alger Sans Pinillos, Marc Teixidor Vilarrasa, Adrián Ortega Novillo, Adrià Quintanas Corominas, Joan Farnós

1. El Grup de Tecnologies d'Ús Dual al Barcelona Supercomputing Center

De vora quatre anys ençà, específicament des de gener de 2022, que el Grup de recerca de tecnologies d'ús dual (en anglès, Dual-Use Technologies Research Group) es va gestar dins del Computer Applications in Science and Engineering (CASE) Department del Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS). Avui, més consolidat que mai, l'equip lidera iniciatives implicant de manera transversal una cinquantena d'investigadors en projectes vinculats a tecnologies d'ús dual, abastant des de l'ús de la computació d'alt rendiment (*High Performance Computing*) per a la modelització de sistemes físics, fins a l'aplicació de la intel·ligència artificial (IA), passant per l'abordatge de l'ètica digital per fer front als dilemes i les problemàtiques que aquestes tecnologies plantegen. Perquè, sorprenentment, l'ús dual no ha arrelat del tot, ni tan sols entre els especialistes.

> **«Per exemple, components o sistemes concebuts en l'àmbit civil poden acabar incorporats al sector de la defensa (i viceversa). »**

Per això, a tall d'introducció, permeteu-nos una breu pinzellada sobre aquest concepte. 'L'ús dual' engloba aquelles tecnologies desenvolupades per a aplicacions civils i defensa, sempre amb un enfoc de ciència per la pau i la seguretat, així com aquelles que poden ser emprades tant amb finalitats *beneficioses* com les *no beneficioses* (Miller, 2018; *NATO* Parliamentary Assembly, 2024)[1]. A finals de 2024, Manuel Heitor, president del grup d'alt nivell de la Comissió Europea per a *Horizon Europe* i el proper *10th EU Framework Programme for Research and Innovation*, va afirmar que ja no té sentit identificar i classificar les tecnologies 'd'ús dual'.

[1] En anglès, la distinció habitual es fa entre "beneficial" i "harmful" (sovint traduït com "perjudicial"). Tanmateix, en contextos com el de la defensa, el contrapunt d'un ús "beneficiós" no implica necessàriament un perjudici actiu, sinó que pot correspondre simplement a la prevenció d'un dany sense cap benefici directe. Per això, optem per l'expressió "no beneficiós", reservant el terme "mals usos" per a aquells casos il·legítims (habitualment perjudicials), que queden fora de l'abast i les finalitats d'aquest escrit (cf. Sans Pinillos et al., 2025).

Segons ell, cal assumir-ne la ubiqüitat i inevitabilitat en el panorama tecnològic actual i futur (Greenacre i Zubascu, 16 d'octubre de 2024).

Aquesta afirmació resumeix amb precisió el gir històric que estem presenciant: no només per l'acceleració tècnica en el sector civil i en l'àmbit de la defensa, sinó també per la transformació profunda en la manera de concebre el desenvolupament tecnològic. Per exemple, components o sistemes concebuts en l'àmbit civil poden acabar incorporats al sector de la defensa (i viceversa). Això comporta que solucions tecnològiques desenvolupades per a un sector específic poden esdevenir funcionalment operatives en àmbits molt diferents mitjançant modificacions dirigides. De la mateixa manera, línies de recerca amb beneficis evidents poden adquirir una deriva maliciosa si es desenvolupen sense supervisió ni enteniment dels riscos previstos, ni de la preparació per facilitar una avaluació i minimització dels danys col·laterals.

«Si la IA actua com a catalitzador de projectes d'innovació a gran escala, també pot catalitzar, sense supervisió adequada, vectors d'amenaces que desplacin l'equilibri geopolític.»

D'aquí sorgeix l'anomenat dilema de l'ús dual: el conflicte ètic que es produeix quan una tecnologia o un coneixement creat amb finalitats legítimes pot esdevenir un mecanisme per causar perjudici (Miller & Selgelid, 2007). Com és sabut, un dilema es caracteritza per ser una situació en la qual un agent té raons morals per dur a terme dues accions, però no pot fer-les alhora, per la qual cosa fracassa moralment, ja que qualsevol acció serà incorrecta (McConnell, 2022: §2). Pel que ens ocupa, el dilema de l'ús dual no només posa de manifest els riscos associats a l'ús indegut, sinó que també planteja qüestions ètiques i pràctiques sobre com dissenyar, distribuir i regular aquestes tecnologies per minimitzar-ne el potencial destructiu.

En aquest punt, el lector ja entén que el dilema de l'ús dual no és un simple exercici acadèmic, sinó una qüestió de primer ordre per a qualsevol institució amb ambicions de lideratge tecnològic. El repte consisteix a conjugar la injec-

ció massiva de capacitat computacional amb un marc de controls ètics i operatius capaç de mitigar riscos estratègics. Si la IA actua com a catalitzador de projectes d'innovació a gran escala, també pot catalitzar, sense supervisió adequada, vectors d'amenaces que desplacin l'equilibri geopolític. Cal, doncs, desenvolupar protocols de governança i quadres de comandament de riscos que, amb la mateixa agilitat que un model d'entrenament, permetin monitoritzar i ajustar en temps real l'impacte de cada nova funcionalitat. Amb aquesta visió integral, podem abordar tant el dilema ètic com les oportunitats d'optimització i seguretat que la IA aporta a empreses, governs i centres d'investigació.

2. La IA com a tecnologia d'ús dual: oportunitats i reptes

Com és ben sabut, la IA està redefinint gairebé, si no tots, els aspectes de les nostres vides. Deixant de banda la recerca de la IA forta o el somni d'alguns —malson per altres— (Coeckelbergh 2020), de la singularitat, és indubtable que en els darrers anys s'ha experimentat un auge desmesurat en la implementació d'aquesta tecnologia, el qual sembla no tenir aturador.

«No és d'estranyar: la IA s'ha consolidat com una palanca tecnològica clau en l'enginyeria computacional.»

No és d'estranyar: la IA s'ha consolidat com una palanca tecnològica clau en l'enginyeria computacional. Accelera el desenvolupament de solucions, redueix riscos en processos d'R+D i optimitza la despesa en recursos computacionals i humans. Això es tradueix en un impacte tangible en àmbits com la simulació de materials avançats, la dinàmica de fluids computacional, l'optimització de sistemes complexos i, naturalment, reforcen línies de recerca intrínsecament d'ús dual.

Un dels vectors clau d'aquest impacte és la seva integració amb tècniques de simulació computacional avançada, com ara la simulació numèrica de sistemes físics complexos a gran escala. Per exemple, mitjançant modelatge computacional d'alta fidelitat, es poden avaluar components ae-

ronàutics (com el perfil d'una ala d'avió) en centenars de milers de condicions diferents: angle d'atac, material, velocitat del flux d'aire i factors ambientals (ràfegues de vent, temperatura, humitat, gel, etc.). Aquesta capacitat permet no només accelerar el disseny i la validació virtual, sinó també demostrar el compliment normatiu mitjançant simulació (el que es coneix com a Certification by Analysis o CbA)[2].

De manera anàloga, aquestes arquitectures de processament massiu s'apliquen al monitoratge intel·ligent del terreny, on xarxes de sensors i drons recullen dades multimodals (òptiques, infraroges, acústiques i químiques) en temps quasi real. Amb tècniques de visió per computador i detecció d'anomalies, s'identifiquen patrons de canvi (erosió, moviments de masses, presència d'objectes o persones) abans que siguin crítics. Aquest flux de dades s'integra en bessons digitals del territori, on es simulen escenaris meteorològics extrems, desplaçaments poblacionals o operacions coordinades, que permeten validar estratègies de resposta, optimitzar rutes logístiques i assignar recursos amb màxima eficiència, tant en contextos civils (prevenció de riscos naturals, gestió d'emergències) (Monterrubio-Velasco et al., 2024) com en missions de seguretat i defensa.

3. Navegant entre dades i biaixos

Per tal d'entrenar models d'IA es necessiten grans quantitats de dades. Fins ara, aquestes s'han obtingut de manera experimental. Tot i així, en molts casos, el gran desenvolupament actual d'aquests models d'IA fa necessàries més dades de les que es poden obtenir de manera experimental, ja sigui pel seu elevat cost, per l'escassedat de casos, per limitacions tècniques, per restriccions de privadesa o simplement perquè la quantitat necessària és tan elevada que és inviable. Davant la impossibilitat d'accedir a dades reals suficients, la generació de dades sintètiques es presenta com una solució per entrenar models d'aprenentatge automàtic. Això només és possible si es té un coneixement profund del domini per tal de que siguin útils i segures.

[2] El CbA substitueix parcialment els assaigs físics per models computacionals verificats i validats permetent, d'aquesta manera, quantificar incerteses i riscos associats (NASA report 2021). És aquí on conflueixen la IA, els models basats en dades i les tècniques de uncertainty quantification, essencials per garantir la fiabilitat dels resultats. Projectes com DIDEAROT (s. d.) o CAELESTIS (s. d.), on el BSC col·labora, en són exemples concrets.

Un exemple específic on les dades sintètiques resulten especialment apropiades és en el desenvolupament de models subrogats per a simulacions numèriques computacionalment costoses. Tot i que no impliquen la manipulació directa del món físic, les simulacions permeten reproduir condicions experimentals sota models controlats, oferint un entorn virtual per a la generació sistemàtica de dades.

El fet que les dades es basin en fenòmens físics mesurables no implica necessàriament l'absència de biaixos cognitius i epistèmics.[3] No obstant això, hi ha contexts, com el de l'enginyeria, on certs paràmetres poden operar sota una definició matemàtica ben establerta i limitada, on la interpretació del seu significat o l'aplicació de les seves regles es troba restringida per models consensuats i entorns formals, fet que pot conferir una major robustesa i estabilitat inferencial al seu ús. Un exemple el trobem en la caracterització de propietats de materials, obtingudes a través d'assaigs estandarditzats o models físics validats.

«El fet que les dades es basin en fenòmens físics mesurables no implica necessàriament l'absència de biaixos cognitius i epistèmics. »

Aquest tipus de paràmetres, matemàticament ben definits i empíricament acotats, resulten especialment adequats per a generar dades sintètiques. Un cas il·lustratiu és la predicció de distorsions en materials compostos, els quals pateixen deformacions induïdes durant el procés de fabricació a causa de la seva naturalesa heterogènia, que fa que la geometria final dels components no sigui la desitjada. Mitjançant la generació de dades sintètiques que incorporen variacions de les propietats del material, geometria, propietats dels panells, del procés del fabricació, etc., aconseguim suficients dades per entrenar models subrogats i així fer prediccions de les deformacions de manera ràpida i efectiva (Teixidor-Vilarrasa, et al., 2025).

Com és sabut, això vol dir que, tot i basar-se en models formulats amb un alt grau de precisió matemàtica i fidelitat física, aquests sistemes incorporen simplificacions necessàries

[3] L'objectivitat de les dades en contextos d'enginyeria pot resultar aparent si no es consideren els biaixos introduïts durant la formulació del problema, la selecció de variables, l'establiment de condicions de simulació, o fins i tot en els processos col·lectius de decisió i disseny. Diversos estudis en enginyeria del programari i enginyeria de sistemes han assenyalat que els biaixos cognitius (com ara el confirmation bias, anchoring o availability heuristic) poden afectar de manera sistèmica tant l'adquisició com l'ús de dades en processos tècnics que es presumeixen racionals. A més, els processos de generació de dades sintètiques poden incorporar errors sistèmics si el mostreig no reflecteix adequadament l'espai de variabilitat rellevant per al fenomen modelat (cf., Weber, 2019; Mohanani, et al., 2020.

i supòsits sobre l'anatomia, la fisiologia o les condicions dels escenaris modelitzats. La raó d'això es desprèn del que s'ha exposat anteriorment sobre la complexitat i el volum de les dades amb què es treballa, fet que fa impossible un seguiment i una gestió humana exhaustiva si no es recorre a processos de simplificació, abstracció i selecció paramètrica. Tanmateix, aquestes decisions impliquen la incorporació de supòsits, abstraccions i reduccions metodològiques que, tot i ser sovint imprescindibles per fer viable el càlcul computacional, poden esdevenir fonts d'incertesa i afectar la validesa dels resultats obtinguts. Aquest fenomen és especialment rellevant en el context de la IA, on aquests resultats s'integren sovint en arquitectures opaques de tipus caixa negra, que dificulten la traçabilitat de l'origen dels biaixos i n'amplien potencialment els efectes en aplicacions crítiques o d'ús dual.

«Tot i basar-se en models formulats amb un alt grau de precisió matemàtica i fidelitat física, aquests sistemes incorporen simplificacions necessàries i supòsits.»

Ara bé, sovint es dona per suposat que aquestes incerteses (les derivades de la simplificació dels models i l'abstracció dels paràmetres) són controlades de manera sistemàtica o, com a mínim, quantificades dins de marges acceptables. I en molts casos és així; atès que existeixen metodologies ben establertes per a l'anàlisi de sensibilitat, la validació creuada o l'ajust fi de paràmetres. Malgrat això, aquest enfocament tendeix a circumscriure el problema als límits del model formalitzat, i obviar una font igualment crítica d'indeterminació, a saber, les decisions prèvies de l'especialista. L'elecció de variables, la definició dels límits del domini, la selecció dels escenaris simulats o les mètriques de rendiment no són reduïbles a meres decisions i operacions tècniques: són també accions dutes a terme per agents cognitius i, per això, per molt entrenades que estiguin per donar resultats basats en factors epistèmics, estan inevitablement carregats de supòsits implícits i, sovint, d'asimetries de valoració.[4]

[4] És sempre pertinent recordar la doble cara dels biaixos i de les inferències aparentment fal·laces. Atès que són inevitables en tot procés cognitiu no trivial, cal entendre'ls com formes d'optimització inferencial, sovint vinculades a l'economia de recursos en la presa de decisions. És el cas, per exemple, de l'ús d'heurístiques com la de disponibilitat o la de representativitat (p.e., la availability heuristic, cf., Tversky & Kahneman, 1973), pròpies de la racionalitat acotada (en anglès, bounded rationality, cf., Simon, 1957). Així mateix, allò que des de la lògica proposicional clàssica es categoritza com una fal·làcia (com l'afirmació del conseqüent) pot operar, des del punt de vista no-monotònic com a inferència abductiva, és a dir, com un recurs per arribar a hipòtesis plausibles en condicions d'incertesa, àdhuc d'ignorància (Sans Pinillos & Magnani, 2022) i formar part, en darrer terme, del que podem entendre grosso modo com a heurística de la creativitat (Vallverdú & Sans Pinillos, 2023).

Aquestes accions no només impliquen coneixement proposicional (*know-that*), com ara principis físics, fórmules o criteris de validació, sinó també coneixement tàcit, és a dir, procedimental i situat (*know-how*) (Norström, 2015). Aquesta distinció, àmpliament reconeguda en l'epistemologia de la tècnica, resulta especialment rellevant en el cas de la modelització computacional aplicada a tecnologies d'ús dual. Mentre que el *know-that* pot ser formalitzat, documentat i compartit, el *know-how* roman sovint tàcit, arrelat en l'experiència, en la capacitat d'interpretació contextual i en la intuïció operativa dels especialistes. Això implica que una part substancial de la presa de decisions (fins i tot en entorns altament automatitzats o regulats) és irreductiblement humana i, per tant, carregada de valoracions implícites.

Així doncs, podem dir que les decisions de l'especialista no són només passos previs a la formalització computacional: actuen com a filtres que delimiten l'espai de possibilitats del model. En definir què s'inclou, què s'omet i com s'ordena la rellevància dels paràmetres, aquestes decisions estableixen, de facto, les condicions epistèmiques del sistema. Una de les manifestacions més significatives d'aquest procés és el mostreig, el qual condiciona la generació de dades sintètiques, determina quins punts de l'espai de dades seran generats i amb quina distribució. Per aquest motiu, l'estratègia emprada ha de ser rigorosa. Independentment de quines variables es considerin, la distribució dels punts de mostreig dins l'espai determina la qualitat i representativitat del conjunt de dades generat.

Els mètodes de mostreig aleatori simple poden generar agrupaments i buits que comprometin la cobertura uniforme de l'espai paramètric i introdueixin biaixos en les dades. Així doncs, una estratègia inadequada pot conduir a models su-

brogats amb prediccions errònies en regions poc explorades o sobreajustament en zones densament mostrejades. Aquesta problemàtica es veu agreujada pel caràcter de doble ús de la tecnologia dual, on errors de mostreig en aplicacions civils poden propagar-se a contextos de defensa i seguretat, i viceversa.

Això interpel·la també el paper de l'especialista, i la consideració que ha de tenir en l'exercici de la seva tasca, a l'hora de minimitzar i, si escau, eliminar els possibles biaixos intrínsecs en les dades, o bé introduïts durant el procés de generació o d'entrenament dels models. Certament, en alguns casos, aquests biaixos no tenen implicacions que comprometin greument la validesa o l'aplicació pràctica del model. Per exemple, en el cas dels materials compostos en aplicacions civils, com ara bicicletes, avions comercials, trens o raquetes de pàdel, entre molts altres, malgrat la possible presència de biaixos, aquests no solen traduir-se en impactes significatius sobre el funcionament o l'ús esperat dels productes resultants.[5]

«La distribució dels punts de mostreig dins l'espai determina la qualitat i representativitat del conjunt de dades generat. »

Tot i així, el punt clau aquí és que la mateixa tecnologia i metodologia emprades en el sector civil són també, de facto, aplicables en entorns de seguretat i defensa. És el cas, per exemple, de l'optimització estructural d'elements aeronàutics o aeroespacials, així com del desenvolupament de tecnologies i protocols per a la presa de decisions en situacions de risc elevat. En aquests contextos, els possibles biaixos induïts en les dades o en els models poden tenir implicacions operatives crítiques; per exemple, una infrarepresentació sistemàtica de morfologies corporals diverses, biaixos de gènere o d'ètnia en les dades antropomètriques, o l'omissió de condicions relacionades amb certes discapacitats funcionals.

[5] Ens referim al fet que, tot i que fins i tot el disseny d'una raqueta de pàdel pot tenir efectes tangibles, com ara afavorir l'ús amb la mà dreta en detriment dels usuaris esquerrans, l'impacte d'aquesta decisió és, en la majoria dels casos, acotat i amb conseqüències limitades. No és que no hi hagi afectació, sinó que el seu abast difícilment assoleix dimensions estratègiques, estructurals o irreversibles, com sí pot ocórrer en àmbits com el de la ciència per a la pau i la seguretat. A tot això, cal afegir que, si aquest tipus de decisions no acostumen a ser unívocament determinants, és també perquè el mercat sovint ofereix un ventall ampli de productes alternatius que permeten mitigar o compensar aquestes mancances funcionals.

En escenaris d'alt risc, com els vinculats a operacions de defensa i seguretat, aquestes omissions poden traduir-se en febleses individuals que, acumulades, esdevenen falles estructurals en el desplegament operatiu o en la missió en conjunt, i que posen en risc no només l'eficiència sinó també la integritat del sistema. Aquest fet fa palès que el rigor tècnic no pot separar-se d'una vigilància crítica sobre els criteris de selecció i exclusió que operen en la fase de modelització, i reclama l'establiment de protocols transversals que integrin criteris de validesa tècnica, representativitat epistèmica i responsabilitat ètica en l'ús de dades per a sistemes d'ús dual.

Tot plegat posa de manifest que la naturalesa de les dades, per si sola, no garanteix ni neutralitat ni fiabilitat. És l'ús concret que se'n fa, ja sigui correcte o incorrecte, pertinent o desajustat, juntament amb el context d'aplicació, així com la finalitat operativa i les condicions materials de desplegament, allò que determina quines consideracions han de prevaler. En el marc de les tecnologies d'ús dual que ens ocupa, aquesta dimensió contextual no és ni accessòria ni secundària, sinó estructural. Ignorar-la pot comprometre tant la validesa inferencial dels models com la seva legitimitat estratègica, ètica i política.

«La IA no pot ser abordada únicament des d'una perspectiva tècnica o funcionalista.»

4. Governança adaptativa de tecnologies d'ús dual

La complexitat que emergeix d'aquest escenari evidencia que la IA no pot ser abordada únicament des d'una perspectiva tècnica o funcionalista, una limitació que s'agreuja encara més quan actua simultàniament com a tecnologia d'ús dual i com a catalitzador dins d'infraestructures ja concebudes amb aquesta doble vessant. La naturalesa de les dades, els biaixos epistèmics que poden arrossegar, la cen-

tralitat del coneixement tàcit en la presa de decisions i les condicions materials de desplegament exigeixen una anàlisi contextual i transversal que situï l'expert, o sia, l'agent cognitiu, al centre del procés.

En paral·lel, la responsabilitat no recau exclusivament en els algorismes, sinó també en les institucions i en els marcs geopolítics en què aquestes tecnologies es desenvolupen i s'apliquen. Això resulta especialment rellevant en centres de recerca com el BSC, on la participació en projectes de recerca i desenvolupament obre reptes estratègics i normatius: el desplegament de tecnologies amb potencial dual en entorns civils comporta implicacions polítiques, institucionals, tecnològiques, ambientals i socials de gran abast.

Aquesta situació posa en qüestió la suficiència dels marcs normatius vigents, com ara la *Llei d'intel·ligència artificial (EU AI Act*, Regulation N 2024/1689. Artificial Intelligence Act), o el llibre blanc *Trustworthiness for AI in Defence* (TAID) (European Defence Agency, 2025), que han evidenciat certes limitacions a l'hora d'abordar les especificitats de les tecnologies emergents i disruptives, precisament pel seu potencial dual. Això fa palesa la necessitat d'impulsar formes de governança més adaptatives, tant a escala nacional com internacional, que vagin més enllà dels enfocaments merament reguladors. Aquesta demanda es fa encara més evident amb l'aparició de noves iniciatives com el projecte europeu ReArm Europe (cf.,White paper for European defence - Readiness 2030 a European Comission, 2025), que reclama explícitament una coordinació estratègica en matèria de seguretat i tecnologia amb un fort enfocament d'ús dual.

En aquest sentit, les institucions de recerca, com el BSC, tenen una responsabilitat activa i ineludible: disposen d'informació de primera mà, d'experiència directa en el desplega-

ment de tecnologies emergents i disruptives, així com d'una visió transversal del seu impacte. Per aquest motiu, el Grup de recerca en tecnologies d'ús dual hi exerceix un paper central, no només a través de la seva participació en comitès d'ètica vinculats a programes de recerca per a la pau i la seguretat, sinó també mitjançant assessorament normatiu i la creació d'espais deliberatius. Això garanteix que les mesures adoptades no només siguin legítimes, sinó també factibles, tècnica i èticament viables.

Finançament

Alger Sans Pinillos i Adrià Quintanas Corominas han estat finançats per la iniciativa "Generación D", Red.es, del Ministerio para la Transformación Digital y de la Función Pública, per a l'atracció de talent (C005/24-ED CV1), amb fons del programa NextGenerationEU de la Unió Europea a través del PRTR.

Declaracions ètiques

Finançat per la Unió Europea - Next Generation EU. No obstant això, les opinions i els punts de vista expressats són responsabilitat exclusiva dels autors i no reflecteixen necessàriament els de la Unió Europea ni els de la Comissió Europea. Ni la Unió Europea ni la Comissió Europea poden ser considerades responsables del seu contingut.

Referències

CAELESTIS. (s.d.). CAELESTIS. Recuperat de https://caelestis-project.eu/

Coeckelbergh, M. (2020). AI Ethics. MIT Press.

DIDEAROT. (s.d.). DIDEAROT. Recuperat de https://www.didearot-project.eu/

European Commission. (2024). Introducing the White Paper on European Defence and the ReArm Europe Plan for Readiness 2030. Directorate-General for Defence, Industry and Space. Recuperat de https://defence-industry-space.ec.europa.eu/eu-defence-industry/introducing-white-paper-european-defence-and-rearm-europe-plan-readiness-2030_en

European Defence Agency. (2025). White paper: Trustworthiness for artificial intelligence in defence – Towards ethical and trustworthy AI systems for European defence. Recuperat de https://eda.europa.eu/publications-and-data/thematic-policy-reports/whitepaper-trustworthiness-for-artificial-intelligence-in-defence

Greenacre, M., & Zubascu, F. (2024, October 16). Time for a major shakeup of how the EU funds research, expert group says. Science|Business. Recuperat de https://sciencebusiness.net/news/fp10/time-major-shakeup-how-eu-funds-research-expert-group-says

McConnell, T. (2022, July 25). Moral dilemmas. En E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2020 Edition). Stanford University. (Originalment publicat el 15 d'abril de 2002). Recuperat de https://plato.stanford.edu/entries/moral-dilemmas/#ConMorDil

Miller, S., & Selgelid, M. J. (2007). Ethical and philosophical consideration of the dual-use dilemma in the biological sciences. Science and Engineering Ethics, 13(4), 523–580. https://doi.org/10.1007/s11948-007-9043-4

Miller, S. (2018). Dual-use technology in the twenty-first century: The GPS case. Technology and Security Journal, 5(2), 61–76.

Mohanani, R., Salman, I., Turhan, B., Rodríguez, P., & Ralph, P. (2020). Cognitive biases in software engineering: A systematic mapping study. IEEE Transactions on Software Engineering, 46(12), 1318–1339. https://doi.org/10.1109/TSE.2018.287775

Monterrubio-Velasco, M., Callaghan, S., Modesto, D., et al. (2024). A machine learning estimator trained on synthetic data for real-time earthquake ground-shaking predictions in Southern California. Communications Earth & Environment, 5, 258. https://doi.org/10.1038/s43247-024-01436-1

NATO Parliamentary Assembly. (2024). Dual-use technologies: Enhancing military capabilities through civilian innovation. NATO Parliamentary Assembly. Recuperat de https://www.nato-pa.int/document/2024-dual-use-technologies-report-baldwin-051-esc

Norström, P. (2015). Knowing how, knowing that, knowing technology. Philosophy & Technology, 28(4), 553–565. https://doi.org/10.1007/s13347-014-0178-3

Sans Pinillos, A., Vallverdú, J., & Farnós, J. (2025). Dual-use technologies in a VUCA world: Ethical abduction and wargames as responses to radical ignorance in scientific development. Lato Sensu: Revue de la Société de philosophie des sciences. [Properament publicat]

Simon, H. A. (1957). Models of man. John Wiley.

Teixidor-Vilarrasa, M., Quintanas-Corominas, A., Ortega, A., Zárate, I., Marquinez, E., Otero, I., & Guillamet, G. (2025). A high-fidelity HPC workflow for predicting process-induced distortions in composites using surrogate models. Materiales Compuestos. Advance online publication. Recuperat de https://www.scipedia.com/public/Vilarrasa_et_al_2025a

Timothy, M., Alonso, J., Andrew, C., Lee, V., Malecki, R., Mavriplis, D., Gorazd, M., Schaefer, J., & Slotnick, J. (2021, May). A guide for aircraft certification by analysis (Certification by Analysis Technical Requirements Report, NASA/CR–20210015404). National Aeronautics and Space Administration. Recuperat de https://ntrs.nasa.gov/citations/20210015404

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Weber, C. (2019). Engineering bias in AI. IEEE Pulse, 10(1), 15–17. https://doi.org/10.1109/MPULS.2018.2885857

Una nova classificació per a sistemes d'IA

Joan Genís Valverde,, Llorenç Valverde

En general, es classifica els sistemes d'IA segons les seves capacitats com a Intel·ligència Artificial Estreta (ANI, per les seves sigles en anglès), Intel·ligència Artificial General (AGI, per les seves sigles en anglès) o Superintel·ligència Artificial (ASI, per les seves sigles en anglès). Aquesta famosa classificació, que es troba en milers d'articles, va ser proposada inicialment per Nick Bostrom a "Superintelligence: Paths, Dangers, Strategies" (Bostrom, 2014). Tanmateix, malgrat el seu ús generalitzat, el fet és que aquesta classificació ha quedat obsoleta com demostren els profunds debats i reptes que ha generat. Per consegüent, sembla necessària una revisió ateses les evidents limitacions de la classificació existent, ja que s'ha tornat insuficient per descriure l'acolorida gamma de capacitats i complexitats de la IA moderna. Calen més classes per classificar els diversos sistemes d'IA disponibles avui dia d'una manera ben definida i universalment aplicable. Per abordar-ho, proposem un nou sistema de classificació que inclou classes addicionals per categoritzar millor els diversos sistemes d'IA disponibles avui dia, i anar més enllà de les limitacions del marc ANI, AGI i ASI.

Atesa la naturalesa complexa i evolutiva del panorama de l'AGI, no només és beneficiós sinó també necessari continuar un debat global i sòlid entre els experts en IA. Aquest debat, amb les seves diverses perspectives sobre la proximitat de l'AGI, serveix com a testimoni de la naturalesa dinàmica del nostre camp i de la necessitat d'una discussió i reavaluació contínues de la nostra comprensió de la IA.

«Calen més classes per classificar els diversos sistemes d'IA disponibles avui dia.»

El missatge del bàndol liderat i representat per OpenAI i Sam Altman sembla més motivat per la seva necessitat d'inversió que

per declaracions sostingudes. En aquest cas, es pot percebre més com una estratègia de seducció, a través de somnis salvatges i promeses emocionants, que com a proves produïdes científicament. Per tant, podrien estar alimentant l'enrenou de l'AGI de manera poc realista o 'rebaixant' el concepte d'AGI a alguna cosa inferior al que la majoria de la gent entén com a AGI. Un altre enfocament del grup *optimista* és la predicció de Ray Kurzweil (Kurzweil, 2005) que la potència computacional del 2030 farà possible l'AGI en termes de quantitat. Les nostres objeccions a la hipòtesi de Kurzweil són dues:

«La quantitat (potència informàtica) per si sola no és suficient. »

·D'una banda, aquesta hipòtesi no té una 'forta' arrel científica (és més aviat una observació i deducció empíriques).

·D'altra banda, la quantitat (potència informàtica) per si sola no és suficient. Hi ha factors addicionals a tenir en compte, com la qualitat o la complexitat. Com dèiem als primers temps de la consultoria, duplicar la capacitat (pel que fa als caps humans o a la potència de càlcul) no garanteix duplicar el resultat ni reduir a la meitat el temps necessari. Hi ha molts aspectes addicionals a considerar. Sembla descabellat i poc realista dir que, quan assolim la potència de càlcul per emular el poder de pensament de tota la humanitat, tots els altres aspectes addicionals necessaris per aconseguir-ho també estaran disponibles.

Mentrestant, l'altra part, representada per líders en IA com Yann LeCun (Lead AI de Meta) (Varanasi, 2025) o Carlos Escapa d'Amazon Web Services (AWS) (Niño, 2024), entre d'altres, sembla més interessada a gestionar les expectatives de l'increïble enrenou que els experts en IA estem experimentant amb les IA generatives, com ChatGPT, Gemini i similars. Defensen una opinió més realista i basada en la ciència, que és crucial perquè puguem avançar amb confiança.

«I si el problema és que, de fet, estem parlant de diferents tipus d'IA amb el mateix nom »

I si el problema és que, de fet, estem parlant de diferents tipus d'IA amb el mateix nom? I si l'ambigüitat d'un sistema de mesura obsolet és la veritable causa d'aquestes diferents visions del futur proper de la humanitat pel que fa a la Intel·ligència Artificial i les seves capacitats previsibles?

El sistema de classificació clàssic obsolet

Actualment, la classificació que la majoria de la gent utilitza i amb la qual està familiaritzada és la següent (Colomer, 2023) (Joshi, 2019):

·ANI: Intel·ligència Artificial Estreta. Aquesta categoria fa referència a usos molt particulars i restringits de la IA: aplicacions molt específiques de casos d'ús amb una capacitat totalment limitada. Aquí podem incloure totes les IA utilitzades abans del 2010 i la majoria de models i algoritmes utilitzats per a una tasca altament emmarcada. L'ANI fa bé una tasca específica, millor que un humà.

·AGI: Intel·ligència Artificial General (abans també anomenada Àmplia) fa referència a un tipus d'IA que és igual a la intel·ligència humana. Això significa que actua d'una forma que és indistingible d'un ésser humà.

·ASI: Superintel·ligència Artificial: es refereix a les IA distòpiques que normalment veiem a les pel·lícules apocalíptiques. Aquestes intel·ligències es tornen autoconscients: assumeixen la càrrega de considerar-se vives. S'adonen que existeixen, i poden igualar o superar la intel·ligència humana.

Les super IA, immerses en la noció d'existència, creen noves pors: gelosia de les màquines, supremacisme artificial, rivalitat o màquines que odien els humans. Els humans estan terroritzats per la idea que les IA vegin els humans com una amenaça que necessita extermini o com una raça inferior que requereix su-

«El principal problema que sorgeix de la definició d'AGI és arribar a un acord sobre 'Què és la intel·ligència?' Són tots els humans intel·ligents? Tota la intel·ligència és humana?

pervisió i guia autoritàries. Aquestes reaccions s'alimenten per la possibilitat que les IA mostrin capacitats que no només són iguals a les humanes individualment, sinó superiors a les capacitats de tota la humanitat junta. És per això que la paraula *súper* s'utilitza àmpliament. Però no us regireu encara: la super IA ja està confinada a la ciència-ficció. Primer hem d'arribar a la intel·ligència artificial general (AGI) abans d'afrontar l'apocalipsi de l'ASI.

El principal problema que sorgeix de la definició d'AGI és arribar a un acord sobre 'Què és la intel·ligència?' Són tots els humans intel·ligents? Tota la intel·ligència és humana? Sense un consens, construir tot el coneixement lliure d'ambigüitats (i, per tant, de discussions) és impossible. Potser sigui impossible posar-se d'acord sobre una definició inequívoca d'intel·ligència i de la intel·ligència humana.

La intel·ligència humana com a patró per a l'AGI

El nostre objectiu no és definir en profunditat què són la intel·ligència i la intel·ligència humana d'una manera acceptada mundialment. Seria un esforç tan agradable i emocionant com improductiu i quimèric. En canvi, proposem un enfocament alternatiu més específic, mesurable, assolible, realista i amb un termini definit.

Tenint en compte la definició àmpliament acceptada d'AGI (Karthik, 2024), necessitem alguns criteris per avaluar si una IA mostra una profunditat de pensament indistinguible del pensament humà. Per 'profunditat de pensament', ens referim a la capacitat d'una IA per comprendre i raonar sobre conceptes complexos, aprendre de l'experiència i aplicar coneixements en situacions noves. Per a aquest propòsit, proposem sis capacitats de la intel·ligència humana que haurien de ser evidenciades per una IA i que són fonamentals per considerar aquest 'procés de pen-

sament indistinguible del d'un humà'. Aquestes capacitats permetran a tothom avaluar si una IA ha assolit un nivell proper a la intel·ligència humana i, per tant, per ser considerada general segons la definició àmpliament acceptada d'AGI. Abans d'aprofundir en les capacitats humanes necessàries per considerar un Sistema d'Intel·ligència Artificial com a General (IAG), primer justificarem una capacitat que no hauria de ser (i no serà) necessària.

Sentiments

No totes les capacitats humanes haurien de ser necessàries de manera sensata en un entorn d'IA. Preveiem, per exemple, que cap intel·ligència artificial mai no podrà mostrar sentiments. Per una raó purament pragmàtica, qui invertiria temps i diners per desenvolupar alguna cosa que sovint es percep com una de les majors debilitats dels éssers humans? Per què necessitaríem la IA per enfadar-nos, molestar-nos o alegrar-nos? Com saben els antropòlegs, aquests trets sorgeixen de la selecció natural per a la nostra lluita per la supervivència. Ens van ajudar a saber quan necessitàvem augments d'adrenalina i altres hormones per millorar les nostres reaccions a diverses amenaces. Seria sorprenent veure inversions en el desenvolupament de funcions tan inútils i poc pragmàtiques en l'actualitat. Considerar aquesta capacitat com una *obligació* en la cursa per la IAG no seria racional. De totes maneres, semblava raonable esmentar-ho i justificar la nostra posició.

Dit això, ara podem aprofundir en les capacitats que les IA haurien d'assolir per poder considerar-les IA generals:

1. Abstracció de patrons i pensament lateral (creativitat) (Genesistrainingsolutions, 2023)

Un requisit àmpliament acceptat per a les IA generals és la capacitat d'abstracció real: extreure patrons subjacents en un context i aplicar aquestes idees a contextos completament diferents.

Aquesta és una de les característiques definidores de la intel·ligència humana. Els processos de pensament *out of the box*, com el pensament lateral, són contranaturals als enfocaments lògics i estadístics actuals de la IA, molt més emmarcats, limitats i enunciats amb precisió.

«La IA actualment simula alguns tipus d'intel·ligència humana sense entendre el context subjacent i les implicacions de la resposta donada»

2. Comprensió

La crítica de la Sala Xinesa a la IA plantejada per John Searle (Searle, 1980) i (Kharbari, 2019) ens ensenya que la IA actualment simula alguns tipus d'intel·ligència humana sense entendre el context subjacent i les implicacions de la resposta donada. No té comprensió. Aquest problema és més greu que el problema de la caixa negra i l'explicabilitat, ja que les IA actuals no entenen què està passant en l'àmbit conscient.

Per exemple la falta de comprensió por portar a un sistema a respondre afirmativament la pregunta: 'Hauríem de matar aquests nens?', atenent a criteris pràctics i objectius (són subversius contra l'ordre establert) però sense entendre les implicacions a altres (Niño, 2024) nivells. Això ens porta als dos punts següents.

«La IA no pot justificar ni explicar la raó que hi ha darrere del seu punt de vista ètic/moral. Sembla crucial per al desenvolupament d'una IA general (AGI) garantir que l'ètica sigui un component del raonament de la IA per disseny. »

3. Ètica i moral

En el context anterior, no comprendre el que l'entitat està fent, decidint o responent realment implica que és impossible aplicar dimensions ètiques o morals. Com en l'exemple anterior, matar nens mai no és èticament acceptable per a molts humans, per molt subversius que siguin. No comprendre les implicacions impedeix l'acció (o la inacció) ètica o moral. La intel·ligència real sap que no sempre és poc ètic construir una bomba o un explosiu (no és el mateix si estic treballant en demolicions o àrees similars que si planejo fer-ho per raons poc ètiques). Tanmateix, els sistemes d'IA actuals no poden establir una línia vermella adaptativa i han d'optar per tallar aquesta línia de discussió.

Els sistemes d'IA actuals no semblen tenir una posició ètica. Podria passar que un sistema fes servir indicacions ocultes per bloquejar elements poc ètics i immorals, però aquestes són sim-

plement regles codificades, inflexibles i inhumanes, més similars a la programació imbricada que a la intel·ligència real. A més, la IA no pot justificar ni explicar la raó que hi ha darrere del seu punt de vista ètic/moral.

Sembla crucial per al desenvolupament d'una IA general (AGI) garantir que l'ètica sigui un component del raonament de la IA per disseny d'una manera similar (aplicant el pensament lateral) a la que va proposar l'escriptor de ciència-ficció i doctor Isaac Asimov: IA (cervells positrònics) amb principis ètics (les lleis de la robòtica) tan profunds en la seva estructura que intentar manipular-los de qualsevol manera resultaria la ruptura catastròfica del sistema intel·ligent.

4. Empatia genuïna i intel·ligència emocional

Com s'ha comentat inicialment, no preveiem que els ordinadors tinguin mai sentiments, però ja són força bons identificant (per separat) les nostres emocions en contextos específics, ja sigui en text, parla, expressió facial o llenguatge corporal. Les IA també són potencialment bones per determinar quines respostes (aparentment) empàtiques són les més adequades per gestionar aquestes situacions, però hi ha un risc significatiu. Si no identifiqueu l'emoció correcta, podeu causar situacions dramàticament incòmodes.

«Si no identifiqueu l'emoció correcta, podeu causar situacions dramàticament incòmodes.»

Els humans sondegen els seus interlocutors en situacions ambigües mitjançant preguntes indirectes, hipòtesis i avaluació del context. 'Va ser alguna cosa que vaig dir?', 'Et trobes bé?', 'Hi ha algun problema?', etc. Aquest instint, que podem anomenar Empatia Real o l'etiqueta que preferiu, deriva de les habilitats esmentades anteriorment. Hauríem d'esperar-ho d'una IA real, sobretot si imaginem un món on se suposa que els humans i les màquines coexisteixen i col·laboren.

5. Simbolisme (Banafa, 2024)

La IA simbòlica tenia profundes limitacions pel que fa a l'autonomia, l'autoentrenament i l'escalabilitat. Això va ser un gran in-

convenient a suportar, i va permetre la presa de control de la IA per part de sistemes subsimbòlics, com ara els basats en l'aprenentatge automàtic, inclòs l'aprenentatge profund. Tanmateix, els sistemes d'IA simbòlica almenys mantenen el valor de la veritat inherentment assignat a cada element processat.

Per tant, les capacitats, els beneficis i les limitacions dels mètodes simbòlics i subsimbòlics semblen tan oposats com el yin i el yang, però també són naturals i inherents al pensament humà. Així doncs, és inevitable cobrir la bretxa actual entre ells si es vol aconseguir finalment la IA.

A més, fins i tot la IA simbòlica està lluny de la intel·ligència humana: les IA simbòliques que es basen en la lògica (aristotèlica, probabilística/heurística i difusa) se centren en el valor de veritat que se'ls assigna. Als elements amb diferents nivells de simbolisme se'ls assigna un valor (de veritat). A partir d'aquest punt, s'ignora la naturalesa i les característiques reals del component. Només la sortida es transforma de nou en un element probable.

«És raonable esperar una millora en les capacitats simbòliques de les IA abans d'arribar a la categoria d'AGI. »

Les IA subsimbòliques ignoren les característiques i identitats de la font i les transformen en figures i vectors. Aquesta manca de consideració de les veritats elementals inicials sobre l'element que s'està processant sovint condueix a resultats paradoxals i sense sentit. En els Grans Models de Llenguatge (LLM) les anomenem al·lucinacions. Aquests models utilitzen el que es podria anomenar lògica probabilística per generar el resultat més probable atesa una entrada complexa. Aquestes situacions, en alguns contextos, poden ser divertides, però, en la majoria de situacions, són simplement perilloses.

En conclusió, cal tancar la bretxa entre les IA simbòliques i subsimbòliques i superar les seves limitacions a l'hora de tractar amb valors de veritat en lloc de amb les veritats mateixes. És raonable

esperar una millora en les capacitats simbòliques de les IA abans d'arribar a la categoria d'AGI.

6. Curiositat

Finalment, si hi ha alguna cosa que fa que la intel·ligència humana sigui única, és la curiositat humana. De petits, els humans aprenen a través de la curiositat com interactuar amb el món i quins comportaments són acceptables i inacceptables. Descobreixen quines coses són molestes i quines ens fan sentir bé a nosaltres i als altres. I, el que és més important, l'avanç de la ciència rau en la curiositat humana. La base del mètode científic de la filosofia i el coneixement humà és la curiositat. Per afirmar que hem arribat a l'AGI, hauríem de veure una IA curiosa.

Hi podria haver altres capacitats que s'haurien de tenir en compte, segons diferents punts de vista (purpleSlate, 2024). Per exemple, alguns autors parlen de capacitats com ara:

·La versatilitat ja és necessària en certa mesura en la IA àmplia i no garanteix una intel·ligència similar a la humana.
·L'aprenentatge i l'adaptació també són inherents a les IA àmplies fins a cert punt.
·El ronament i la resolució de problemes d'una manera avançada i no lineal, que anomenem Abstracció i Pensament Lateral.
·L'autonomia, encara que aquesta sembla més un atribut de la Super IA que de la IA general. No és, per a nosaltres, un atribut essencial per a la IA general.

Independentment d'altres punts de vista, sense almenys les nostres sis capacitats proposades, la promesa de la IA general serà només una declaració de màrqueting per atreure inversors, usuaris de pagament i altres filantrops o mecantrops. El nostre enfocament proposat confereix un enfocament sòlid i realista a la IA general sense acabar de resoldre el problema de definir la intel·ligència humana.

«El nostre enfocament proposat confereix un enfocament sòlid i realista a la IA general sense acabar de resoldre el problema de definir la intel·ligència humana.»

nodes

Tendències i futur

El paradigma de la IA aplicada

Abans d'endinsar-nos en el problema de la IA *complexa*, cal tenir una discussió prèvia sobre les IA menys complexes. El 2018, ens vam adonar que la classificació en només tres categories o nivells de capacitat (ANI / AGI / ASI) era insuficient per una raó senzilla. Solucions com IBM Watson Assistant (o Google Dialog Flow) no coincidien amb la categoria ANI o AGI. Eren massa àmplies per ser classificades com a estretes. I, sens dubte, estan lluny de ser AGI. Aquest paradigma és encara més evident ara quan mirem les IA generatives, com ChatGPT, Copilot, Gemini, etc.

Què caracteritza aquests tipus d'IA?

·Ja no són estretes, ja que no són simples algoritmes que treballen sols en entorns restringits i limitats. Els entorns sens dubte tenen més dimensions o nivells de llibertat.
·En combinar diferents tipus d'IA, són molt més flexibles i realitzen una gamma de funcions més àmplia. Veiem com ChatGPT pot (millor o pitjor) entendre el llenguatge natural i crear codi informàtic, imatges o contingut literari.
·Són els sistemes d'IA més 'aplicats': han estat transformant negocis d'una manera altament innovadora des del 2016, aproximadament.

«La GenAI és un tipus d'IA que de vegades és estreta (crea només imatges o només text) i ocasionalment àmplia o conjunta. »

Després, el 2018, va semblar completament lògic crear una nova categoria per incloure tots aquests sistemes i solucions que eren massa amplis per ser estrets. Per tant, vam començar a utilitzar una nova categoria d'IA. Inicialment vam utilitzar 'Ampli' com a terme per referir-nos-hi.

Proposem utilitzar IA àmplia, IA cooperativa o IA conjunta per a aquesta nova categoria per significar que ha augmentat les seves capacitats (sense arribar a l'AGI), però superant les capacitats de l'ANI. No ens atrau el terme GenAI per a aquest tipus d'intel·ligència, ja que altres no GenAI encaixen en aquest grup. La GenAI és

un tipus d'IA que de vegades és estreta (crea només imatges o només text) i ocasionalment àmplia o conjunta (interfícies de xat multimodals de sortida múltiple).

El nou paradigma de l'adaptació al futur

Si mirem enrere a l'actual divisió en dues *escoles de pensament* a tot el món, l'enfocament científic de la facció que dubta de la imminència de l'arribada de la IA general (AGI) es troba amb la promesa no fonamentada de la seva proximitat per part de la facció àvida de negocis.

Els autors són transparentment parcials a l'enfocament escèptic. No obstant això, hem conclòs que el nostre enfocament proposat oferirà un enfocament diplomàtic i donarà estabilitat i punts en comú a ambdues parts: les comunitats de recerca i acadèmiques.

Havent justificat primer la necessitat de crear un nou nivell d'IA per poder classificar certs tipus d'IA que eren massa amplis per ser classificats com a estrets, però lluny de ser generals, ara veiem la necessitat d'oferir una altra categoria per a sistemes d'IA que no són IAG, però que són molt més del que anomenàvem IA àmplia. Anomenem aquesta nova categoria Intel·ligència Artificial Multicapa, Abstracta o Avançada. Tenint-ho en compte, la classificació tindria cinc classes que ajudarien els experts i els profans a comprendre millor les diferents capacitats de la IA pel que fa als reptes futurs immediats. Aquesta classificació final seria:

-ANI o Intel·ligència Artificial Estreta o Feble
-ABI o Intel·ligència Artificial Àmplia, de Conjunt o Cooperativa
-AMI o Intel·ligència Artificial Multicapa, Abstracta o Avançada
-AGI o Intel·ligència Artificial General o Forta
-ASI o Superintel·ligència Artificial

«La Intel·ligència Artificial Multicapa (AMI) considera qualsevol Intel·ligència Artificial Àmplia que mostri evidència d'haver assolit almenys una de les sis capes en les capacitats d'intel·ligència humana proposades.»

nodes

Tendències i futur

Intel·ligència Artificial Abstracta o Intel·ligència Artificial Multicapa

Com ja hem assenyalat, la Intel·ligència Artificial Abstracta, Avançada o Multicapa és una categoria molt necessària, ja que cobreix la bretxa entre el que hem definit com a Intel·ligència Artificial Àmplia (ABI) i el que s'entén com a AGI. La IA Àmplia és versàtil i més complexa en arquitectura i resultats que l'ANI, però encara li falten elements essencials per 'competir' amb la intel·ligència humana. Hem proposat una llista de sis d'aquests elements.

La Intel·ligència Artificial Multicapa (AMI) considera qualsevol Intel·ligència Artificial Àmplia que mostri evidència d'haver assolit

Tipus d'IA segons la seva capacitat

[La classificació preparada per al futur]

ASI Super AI	AGI IA general	AMI IA multicapa	ABI IA àmplia	ANI IA estreta
• Escenari **Hipotètic** • Sistemes artificials amb una intel·ligència superior a la humana • Autoconsciència: són conscients de la seva pròpia existència • Són capaços de sentir, evolucionar i desenvolupar les seves pròpies emocions i sentiments	• Revolucionari, i **ficció** fins ara • Capaç d'autoaprenentatge i d'aplicar la intel·ligència abstracta a la resolució de problemes • El sistema pot pensar, entendre i actuar d'una manera que no es distingeix d'un ésser humà en qualsevol situació • També s'anomena forta o profunda.	• **Aspiracional** • Versàtil, adaptatiu, capaç d'aprendre patrons. • El sistema mostra entre 2 i 6 de les capacitats humanes proposades • **Noms proposats: IA multicapa, avançada o abstracta**	• **Real** • Capaç de fer més d'una cosa bé • Alguns tenen capacitats d'explicabilitat. • Consisteixen en diferents IA estretes que col·laboren com un conjunt. • Són menys limitats. • **Noms proposats: IA àmplia, conjunta o cooperativa**	• **Bàsica** • Fortament especialitzat: pot fer una sola cosa molt bé • Enfocament de la caixa negra: poca o poca comprensió del seu raonament intern • Funciona sota un conjunt de restriccions i limitacions • També anomenada feble

nodes

almenys una de les sis capes en les capacitats d'intel·ligència humana proposades.

Podem distingir entre el Nivell 1 de l'AMI (AMI-1) si només s'ha assolit una capa, l'AMI-2 si se n'han evidenciat dues de manera transparent i així fins l'AMI-6 si se n'han aconseguit les sis. Aleshores, la pregunta filosòfica serà: l'AMI-6 és el mateix que l'AGI? O quins elements addicionals necessitem per cobrir aquesta bretxa?

Conclusions

Aquesta proposta ofereix un pas realista per avançar des de l'estat actual de la tècnica de la IA. Redueix l'ambigüitat i la confusió de l'ús de la IA general pel que fa a les seves expectatives, alhora que aclareix què hem d'avaluar en una IA per jutjar les seves capacitats.

També ofereix una classificació robusta que permetrà a tot el món acadèmic (investigadors, professionals del màrqueting, els Quatre Grans i altres) comunicar-se millor i gestionar les expectatives sobre la IA.

Aquesta classificació refinada ajudarà a gestionar l'expectació i les expectatives elevades d'una part de la humanitat sobre què esperar de la IA en els propers anys. Al mateix temps, pot contribuir a dissipar o moderar les pors sobre el desenvolupament real de la IA que alguns grups han mostrat als mitjans de comunicació.

Amb aquesta finalitat, al llarg d'aquest article:
-Hem explicat per què la classificació clàssica de la IA per capacitats sembla insuficient. Calen més classes per adaptar-se a alguns dels nous sistemes d'IA.

Tendències i futur

-Hem proposat refinar el sistema de classificació amb dues classes addicionals per garantir que tots els sistemes d'IA es puguin assignar adequadament a una classe de manera clara i sense ambigüitat.

-Hem definit una de les noves classes com a Intel·ligència Artificial Àmplia, de Conjunt o Cooperativa, que inclou sistemes que no són realment estrets i que normalment combinen diferents IA en un sol sistema.
-Hem definit l'altra nova classe com a Intel·ligència Artificial Multicapa, Abstracta o Avançada, que inclou sistemes que presenten una o més de les sis capacitats principals de la intel·ligència humana.

-Hem identificat les capacitats humanes que s'han de sol·licitar abans d'una IA general com:
a.Abstracció i Pensament Lateral
b.Comprensió
c.Ètica (i moral)
d.Empatia Genuïna (i intel·ligència emocional)
e.Simbolisme
f.Curiositat

-Hem justificat per què no incloem els 'sentiments' com una de les capacitats.

-Hem explicat per què aquesta nova classificació és necessària i millorarà considerablement totes les discussions presents i futures sobre la IA.

Els humans poden ser capaços de crear entitats tan intel·ligents com els humans, però citant Isaac Asimov, Ph.D., primer caldrà la capacitat de desenvolupar sistemes tan complexos com els cervells humans (Joshi, 2019).

nodes

Referències

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies. Oxford University Press.*

Banafa, A. (2024, 4 7). *From Linkedin: https://www.linkedin.com/pulse/bridging-gap-between-symbolic-subsymbolic-ai-prof-ahmed-banafa-ujkjc/*
Colomer, R. (2023, 4 1). *From Medium: https://medium.com/@rubencolomer/the-three-types-of-artificial-intelligence-ani-agi-and-asi-discovering-the-world-of-ai-36fc5fe60511*

Genesistrainingsolutions. (2023, 11 10). *From Medium: https://medium.com/@genesistrainingsolutions/lateral-thinking-unlocking-creativity-and-innovation-af8b1e6738d9*

Joshi, N. (2019, 6 19). F*rom Cognitive World: https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/*

Karthik, K. (2024, 4 25). *From Transmiter IEEE: https://transmitter.ieee.org/artificial-general-intelligence-what-is-it/*

Kharbari, V. (2019, 1 19). *From Medium: https://medium.com/acing-ai/what-is-the-chinese-room-argument-in-artificial-intelligence-d914abd02601*

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology . New York: Penguin Group.*

Niño, L. (2024, 11 6). *From Revista PYM: https://www.revistapym.com.co/articulos/digital/80535/la-inteligencia-artificial-no-es-realmente-inteligente-explica-carlos-escapa*

PurpleSlate. (2024, 6 4). *From Medium: https://medium.com/@social_65128/understanding-artificial-general-intelligence-agi-the-future-of-ai-technology-356390900e52*

Searle, J. (1980). *Minds, Brains and Programs. Behavioral Brain Sciences Vol 3 N 3, 417-457.*

Varanasi, L. (2025, 5 26). *From Busines Insider: https://www.businessinsider.com/meta-yann-lecun-ai-models-lack-4-key-human-traits-2025-5*

Sobre els autors

Joan Genís Valverde ha treballat durant 16 anys a IBM en transformació digital empresarial, en dades des del 2009 i en IA des del 2016. Ha participat en la transformació d'empreses líders a Espanya. Col·labora regularment en el món acadèmic (UNIR, EU Gimbernat, EAE Business School) creant i impartint formació en IA per a empreses. Ofereix el punt de vista aplicat d'aquest article.

Llorenç Valverde, doctor en Informàtica des del 1982 i professor emèrit de Ciències de la Computació i IA a la Universitat de les Illes Balears (UIB). Primer director de l'Institut d'Intel·ligència Artificial de les Illes Balears (IAIB) (2023-2026). Les seves activitats de recerca s'han centrat en el raonament aproximat, analògic, per defecte i les classificacions difuses. Ofereix la visió acadèmica i de recerca d'aquest article.

La intel·ligència artificial generativa

En defensa de la IA Generativa

Camilo Chacón Sartori

Podríem suposar que aquells textos que comencen amb "En defensa de X" ens conviden a sospitar que X està sota escrutini —sigui per bones o males raons—, si no, no requeriria cap defensa. Curiós seria el cas d'algú que es proposés defensar, per exemple, Internet. Encara que no seria del tot estrany. En els seus primers anys, Joseph Weizenbaum —creador d'ELIZA, el primer *bot de conversa*— va afirmar: "Internet és com un d'aquells abocadors d'escombraries als afores de Bombai. Hi ha persones, lamentablement, arrossegant-se per tot arreu, i potser troben una mica d'alumini, o potser alguna cosa que puguin vendre. Però majoritàriament, és brossa".[1]

Més tard matisà: "Això ho vaig dir, però després hi vaig afegir: 'Hi ha mines d'or i perles allà dins que una persona entrenada per formular bones preguntes pot trobar".[2] Així podem inferir que Weizenbaum no va escriure un "En defensa d'Internet". I, potser, en un dia optimista, li hauria arribat per a un "En defensa —d'alguns casos— d'Internet".

IA Generativa

La IA generativa (des d'ara, GenAI) és un tipus d'IA —popularitzada per ChatGPT— que permet crear diferents tipus de contingut: text, codi, àudio, imatges i, fins i tot, vídeos. El seu èxit ha estat tal que, per primera vegada a la història, una tecnologia basada totalment en IA està en camí de tornar-se ubiqua. Això es pot comprovar amb un simple experiment: feu una volta per qualsevol biblioteca o cafeteria, i observeu les pantalles dels portàtils. Per a sorpresa d'uns i desinterès d'altres, molts tindran una finestra oberta de ChatGPT (o un model similar). I ves a saber què hi estaran escrivint!

1 https://archive.nytimes.com/www.nytimes.com/library/tech/99/04/circuits/articles/01skep.html
2 https://www.nytimes.com/1999/04/08/technology/l-everyone-s-a-skeptic-125784.html

La intel·ligència artificial generativa

«Feu una volta per qualsevol biblioteca o cafeteria, i observeu les pantalles dels portàtils. Per a sorpresa d'uns i desinterès d'altres, molts tindran una finestra oberta de ChatGPT (o un model similar). »

Ara bé, seria un error traçar un paral·lel directe entre Internet i la GenAI. Tot i que ambdues tecnologies són populars i fàcils d'utilitzar, la seva arquitectura —forma d'operar— és completament diferent. Internet és com un gran aparador: obert a qualsevol persona que vulgui compartir informació, sempre que compleixi certes condicions. Hi ha un acord tàcit entre els diferents participants (xarxes socials, motors de cerca, diaris, blocs i qualsevol altre mitjà) que sostenen i negocien certes regles de visibilitat i permanència. Si es crea una pàgina web en HTML, aquesta mostrarà la mateixa informació mentre no es modifiqui explícitament. Internet és, en certa manera, un reflex més o menys fidel de les intencions humanes.

La GenAI no funciona així. Resulta que, si interactues amb un Large Language Model (LLM), pots generar contingut incorrecte fins i tot sense proposar-t'ho. Això es deu al fet que la seva arquitectura és estocàstica, i el seu comportament depèn tant del conjunt d'entrenament (no revelat pels seus desenvolupadors) com de la manera com formulem les instruccions (*prompting*).

Aquesta mancança propicia un terreny fèrtil per a les crítiques. Els Weizenbaum contemporanis exhibeixen diverses objeccions contra la GenAI. Potser la principal no és tècnica, sinó social. Perquè, en el terreny tècnic, que una tecnologia produeixi errors (o "al·lucinacions", com els agrada dir als nous experts antropomorfitzants d'IA) no és res nou en informàtica. Tot programari pot fallar. Fins i tot els qui segueixen Dijkstra o Hoare —i somien amb la verificació formal— han d'acceptar que, per més que el programari cobreixi totes les excepcions i estigui lliure d'errors, mai no estarà a resguard dels errors del maquinari.

Computació evolutiva

No cal anar gaire lluny per trobar un paral·lelisme entre els errors de la GenAI i un altre subcamp de la IA. Els qui hem treballat en computació evolutiva sabem bé que l'aproximació, i no l'exactitud,

és la norma. Convivim amb l'error; errors que sorgeixen de generar solucions des d'operadors pseudoaleatoris. Però l'error, aquí, no és un defecte: és una manera d'explorar noves solucions. No és un problema, sinó un refinament.

La computació evolutiva no ha estat exempta de crítiques. Aquestes se centren en l'absència d'una teoria robusta, una manca que es compensa amb l'experimentació. Els experiments donen suport a la utilitat. I, encara que en la GenAI també hi ha exemples que sostenen la seva utilitat, les crítiques sorgeixen en l'entorn social en què es mou. Cosa que la computació evolutiva no pateix, perquè no és popular.[3] El que realment preocupa, quan tractem amb la GenAI, no són les seves equivocacions, sinó el lloc que ocupa —i ocuparà— a la societat. Els seus modes d'ús, sí, però també els seus modes d'engany.

«Però l'error, aquí, no és un defecte: és una manera d'explorar noves solucions. No és un problema, sinó un refinament.»

El problema prové de la seva facilitat d'ús; perquè qualsevol pot interactuar amb un LLM. Algú es llegeix una guia d'estratègies de prompts i ja se sent qualificat per parlar d'IA. La GenAI ha aconseguit crear tants experts d'IA en tan poc temps que resulta massa bo per ser veritat. Ara els programadors no volen programar. Els dissenyadors no volen dissenyar. I els investigadors no volen investigar. A aquest ritme totes les facultats cognitives seran delegades a un sistema de GenAI. Així, les objeccions socials contra la GenAI se centren en la mandra que suposa i el possible declivi cognitiu que pugui provocar.

«El que realment preocupa, quan tractem amb la GenAI, no són les seves equivocacions, sinó el lloc que ocupa —i ocuparà— a la societat. Els seus modes d'ús, sí, però també els seus modes d'engany. »

Sí; tota tecnologia ens facilita la vida i ens torna, en certa manera, mandrosos. Això és una cosa que sovint es diu per compensar el debat. No obstant això, la GenAI porta les coses a un altre nivell. Un usuari mandrós podria no tornar mai més a pensar en una idea original, ni tan sols a escriure un correu. Malgrat aquest panorama descoratjador, com podem anar en contra d'alguna cosa que cada vegada més persones utilitzen? La meva defensa és pragmàtica.

[3] En l'actualitat estem veient una fecunda col·laboració, en l'àmbit de la recerca, entre sistemes de GenAI amb computació evolutiva. Un exemple és AlphaEvolve de Google DeepMind, que dissenya nous algorismes seguint principis dels algorismes evolutius. Com poden dos sistemes estocàstics, no deterministes, integrar-se tan bé?

Una defensa pragmàtica

Nicholas Rescher, filòsof nord-americà, pel que fa al pragmatisme que ell defensava —i que jo subscric—, el definí així: «*Així com interessa a les persones utilitaristes, a qui adopta una postura pragmàtica li importen els resultats. Però mentre que els qui segueixen l'utilitarisme se centren a promoure la felicitat, els qui es guien pel pragmatisme tenen una mirada més àmplia i consideren l'eficàcia funcional en general».*

«Així, les objeccions socials contra la GenAI se centren en la mandra que suposa i el possible declivi cognitiu que pugui provocar.»

Una primera interpretació errònia seria assumir que el resultat justifica qualsevol cost, fins i tot derrocar tota ètica. Però el pragmatisme no és utilitarisme. Més aviat l'eficàcia funcional comporta un respecte a certes normes. I s'adopta perquè, al final, els beneficis són més grans que els perjudicis. Una visió pragmàtica incorpora la gradualitat de les coses, afegint-hi matisos que s'avaluen en els resultats, evitant un joc de suma zero.

Vist així, la GenAI no és només una tecnologia respectable sinó un esglaó inevitable cap al següent: una nova arquitectura que supleixi (o redueixi) els errors amb altres estratègies, més segures i menys perjudicials per als humans. Però cal no descartar-la sense abans estudiar-la, analitzar-la, criticar-la i investigar-la. Perquè, com va dir Weizenbaum: *"Hi ha mines d'or i perles allà dins que una persona entrenada per formular bones preguntes pot trobar"*.

Qui s'encarregarà d'ensenyar a formular aquestes bones preguntes? Se suposa que els qui estem involucrats en la investigació d'IA. Però alguns estan enfocats a presentar els fets negatius, passant per alt els positius.

Una objecció a aquesta visió pragmàtica seria afirmar "Que una cosa sigui àmpliament utilitzada, no necessàriament vol dir que li hàgim de donar suport". Això, és clar, ens sembla lògic. Però hi ha una confusió: quan tenim la mala sort d'escoltar la música de

moda, o de llegir algun llibre que és a la prestatgeria dels *best-sellers*, tendim a aplicar aquesta objecció; sobretot si abans hem conegut altres plaers. Assumim de seguida que una cosa popular no és sinònim de bona. De fet, ens inclinem a pensar el contrari. La confusió és en fer servir els adjectius 'bo' o 'divertit' amb una tecnologia (en aquest cas, la GenAI), però no té sentit. Això és utilitarisme. Una tecnologia no és bona ni dolenta, ni divertida ni avorrida. Una tecnologia podria ser eficaç a la pràctica o no; aquest és l'enfocament pragmàtic. A continuació mostraré dos exemples dels usos de la GenAI que, actualment, no tenen una alternativa superior.

GenAI i l'escriptura

Qualsevol que escrigui sap què significa tenir un bon editor. Algú que pugui veure les teves mancances des de la distància, i que sobretot t'aprecïi i vulgui que l'obra sigui millor. Proposo que l'escriptor actual col·labori amb aquest tipus d'eines per construir un nou assistent. No per substituir un editor, sinó per construir un altre tipus de col·laborador, que l'ajudi a millorar la seva escriptura, sempre que no oblidi que el més important per a un escriptor és escriure, no deixar d'escriure per guanyar velocitat.[4]

I el rebuig que ara genera la GenAI no és aliè a les tecnologies que han transformat l'escriptura. Quan van aparèixer els primers llibres impresos, hi va haver qui els va menysprear perquè tornarien més mandroses les persones, ja que anirien en detriment de l'oralitat i ningú recordaria res. I, encara que és cert que es va perdre, parcialment, la capacitat de memoritzar —que es reforça amb l'oralitat—, el llibre va afegir noves possibilitats que ho van compensar.

Alguna cosa similar passa amb la traducció. Abans teníem Google Translate, i després DeepL, que va significar un avenç. No obstant això, actualment els LLM són capaços de traduir amb més precisió, mantenint no només la semàntica, sinó el canvi d'estil que implica passar del castellà a l'anglès, per exemple. I no només la traduc-

[4] He escrit un llibre sobre aquest tema: Camilo Chacón Sartori. *Palabras y Algoritmos. Cómo la IA transformará la escritura*. 2025. Editorial Marcombo.

ció escrita, ara, gràcies a la multimodalitat, podrem comunicar-nos amb persones d'altres cultures, mantenint, fins i tot, el nostre to de veu original, traduït a un altre idioma.

GenAI i la programació

És recurrent somiar amb automatitzar els programadors. Fins i tot jo, com a programador, ho he pensat. No ho puc negar; potser perquè programar, encara que plaent, es torna dur quan s'ha d'escriure codi trivial o el problema creix en complexitat. La GenAI ha vingut a recordar-nos una cosa important relacionada amb la programació: no tot el codi és necessàriament igual. Hi ha codi rutinari: invocar una API, fer una consulta a una BD, crear una simple pàgina web. Avui, tot això es pot fer amb qualsevol LLM i sense errors. I està millorant per a tasques complexes: com generar algoritmes sofisticats d'optimització o un primer projecte de programari que inclogui *frontend*, *backend* i desplegament al núvol.

«L'activitat de programar es dirigeix cap a una on convisquin tant codis creats per humans com per màquines.»

L'activitat de programar es dirigeix cap a una on convisquin tant codis creats per humans com per màquines. I encara que els errors de codi d'ambdós puguin semblar similars, el seu origen és diferent. Això és clau per construir una integració correcta.[5]

És clar, un crític de la GenAI podria al·legar que aquests casos d'ús no són un benefici, sinó perjudicials per a la societat. D'una banda, tothom tendirà a escriure igual perquè els LLM tendeixen a 'estandarditzar' l'escriptura i, de l'altra, perquè el codi generat amb LLM produirà més bugs que solucions. Però no vull pensar així, ja que seria una postura que provoca inacció. I, si la GenAI fos perjudicial per a les nostres vides, a més de criticar-la, hi ha alguna cosa que podríem fer per millorar-la o superar-la?

Comparteixo el que va dir el filòsof espanyol, José Antonio Marina, referint-se a aquesta postura: «El pessimisme té un prestigi intel·lectual que no es mereix».[6]

[5] Vegeu la meva recerca relacionada: Camilo Chacón Sartori 2025. *Architectures of Error: A Philosophical Inquiry into AI-Generated and Human-Generated Code. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5265751*

[6] *La vacuna contra la insensatez,* pág. 21. 2025. Editorial Ariel.

Prendre acció. Proposar solucions. Programar. Segueixen la petjada del pragmatisme: no seure només a criticar mentre el caos succeeix sinó intentar sortir al món a combatre, no amb les armes del pessimisme, sinó amb aquelles que rescaten allò positiu i busquen una alternativa, encara que sigui equivocada. Perquè tots cometem errors; siguin agents biològics o artificials.

Espero que davant la GenAI siguem escèptics sense caure en el pessimisme, realistes sense caure en la resignació i curiosos sense caure en la ingenuïtat.

La intel·ligència artificial generativa

Què en pensen?

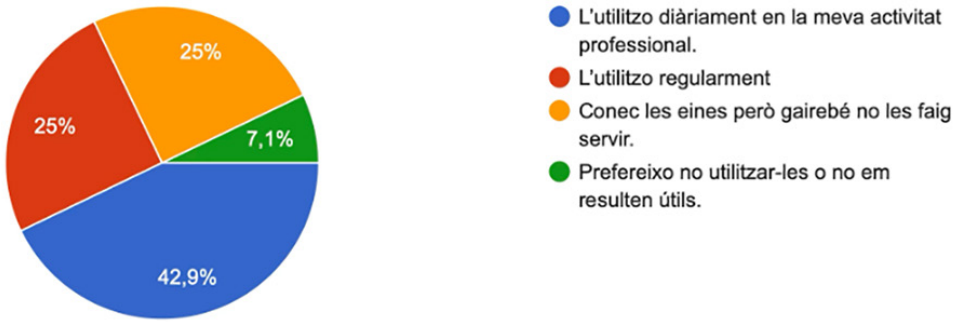
Lucia Roda, Vicent Costa

A l'Associació Catalana d'Intel·ligència Artificial (ACIA) vam dissenyar una petita enquesta informal amb l'objectiu de conèixer l'estat d'opinió de la nostra comunitat respecte al tema monogràfic d'aquest número 61 de *NODES*: la intel·ligència artificial generativa (IA generativa). Aquesta tecnologia, que ha irromput amb força en els àmbits professional, creatiu i educatiu, planteja reptes i oportunitats que mereixen una reflexió col·lectiva.

L'enquesta s'iniciava amb dues preguntes destinades a conèixer la freqüència i els usos que les persones enquestades fan de la IA, amb la possibilitat d'afegir comentaris per matisar o ampliar la informació. A continuació, se'ls demanava que valoressin l'impacte de la IA generativa en el seu àmbit professional mitjançant una escala de l'1 al 5. Tot seguit, es presentaven quatre preguntes amb diverses afirmacions sobre l'impacte professional i les preocupacions ètiques que es podrien derivar de l'ús d'eines d'IA generativa, en les quals calia indicar el grau d'acord (des de 'molt d'acord' fins a 'gens d'acord'). Finalment, una darrera pregunta oferia l'oportunitat d'incloure observacions generals sobre qualsevol de les qüestions tractades en el qüestionari.

Vam rebre 28 respostes, i aproximadament el 39 % de les persones participants vau afegir algun comentari a la resposta estàndard. Tal com el nostre benvolgut Josep Puyol ja advertia en les anàlisis d'enquestes que va escriure, l'anàlisi que presentem tampoc no té cap pretensió especial: ens limitarem a comentar els resultats i a destacar algunes de les vostres opinions que considerem rellevants, sense arribar a conclusions.

1. Quina d'aquestes afirmacions descriu millor la teva relació amb la IA generativa?

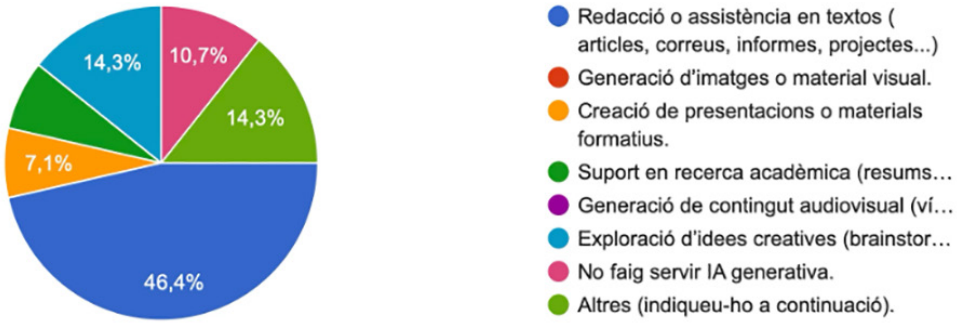


Aquesta pregunta tracta sobre la relació amb les eines d'IA generativa en l'activitat professional. A primera vista, és prou evident que l'ús d'aquestes eines és elevat, ja siga un ús regular (el 25 % de les persones enquestades) o diari (al voltant del 43 %). Algú remarca que els models generatius formen

part de la seua recerca, mentre que una altra persona fa una distinció molt interessant: si bé fa servir aquestes eines en la seua activitat professional, sempre ho fa per a tasques que, en el context de la seua disciplina, poden ser considerades auxiliars: "Per exemple, [per a] millorar algunes expressions d'un text ja redactat (normalment en anglès) o ajudar-me a trobar l'error en algun codi si a mi m'està costant. Mai l'utilitzo per a generar el text o el codi directament."

Una de cada quatre persones enquestades gairebé no fa servir aquestes eines, mentre que només un 7 % declara una preferència per no fer-les servir. Trobem també un comentari molt interessant pel que fa al paper de la IA generativa en la disciplina científica de la IA: "Crec que des del punt de vista científic la IA generativa té poc interès. Ja no és ciència, és marketing i lluita pel poder."

2. Per a quines activitats acostumes a utilitzar IA generativa?



nodes

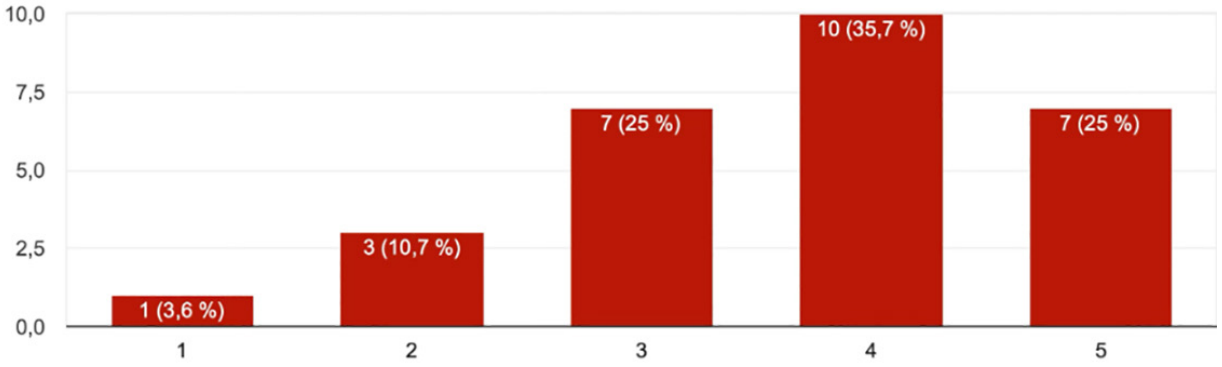
Tal com es recull en tres comentaris afegits a la qüestió i en algun correu que ens heu fet arribar, aquesta pregunta presentava el problema de només deixar marcar una opció, quan potser hi havia més d'una possible resposta que s'adaptés a l'activitat de les persones enquestades. Demanem disculpes per això. Així doncs, en aquest sentit, els percentatges no són massa acurats, tot i que hem entès que l'opció marcada a les respostes correspon a l'activitat més comuna de la persona que ha respost.

El més destacable és que quasi la meitat de l'ús sembla estar relacionat amb la redacció de textos. L'exploració d'idees i un ús més heterogeni de les eines d'IA generativa representen cadascuna poc menys del 15 % de les respostes. Així, algú destaca l'ús «per a buscar paraules clau donades descripció o camps de recerca similars, així com ajuda per a la cerca de referències en el camp», mentre que un altre comentari destaca les «consultes sobre llenguatges de programació». Cap resposta ha destacat la generació de contingut audiovisual, i això potser és un biaix, tenint en compte que la comunitat de l'ACIA se centra, evidentment, en la recerca en IA. Finalment, quasi un 11 % de les persones enquestades no fan servir la IA generativa.

«Crec que des del punt de vista científic la IA generativa té poc interès. Ja no és ciència, és marketing i lluita pel poder.»

La intel·ligència artificial generativa

3. Com valores l'impacte de la IA generativa en el teu àmbit professional? (1 = gens rellevant, 5 = profundament transformador.)

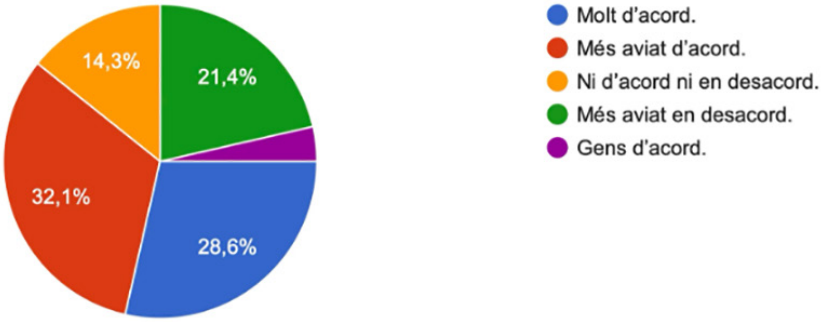


Basant-nos en les dades d'aquesta pregunta (on 10 de les 28 persones enquestades consideren la IA generativa molt rellevant en el seu àmbit professional, amb un valor de 4), l'opinió mitjana se situa en 3,68, fet que confirma que les persones enquestades perceben la IA generativa amb un impacte significatiu, clarament superior a la valoració mitjana, que seria el valor 3. Ara bé, tot i que l'opinió majoritària apunta a una gran rellevància, la dispersió d'1,07 (desviació típica) indica que no hi ha un consens total, sinó que existeix un rang divers de percepcions que s'estenen des del «gens rellevant» (1 resposta) fins al «profundament transformador» (7 respostes), reflectint la varietat d'experiències i criteris dins de l'associació.

A continuació, es van presentar quatre afirmacions i es va sol·licitar a les persones enquestades que indiquessin el seu grau d'acord o desacord, mitjançant una escala qualitativa que oscil·lava entre 'molt d'acord' i 'gens d'acord'.

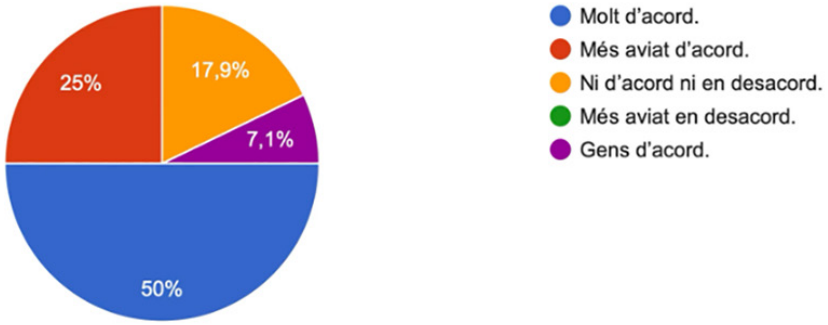
nodes

4. El valor del treball intel·lectual i creatiu es diluirà progressivament si no es redefineixen els criteris d'autenticitat, autoria i aportació humana.



A la primera pregunta, referida al valor del treball creatiu i intel·lectual humà davant el progrés de la IA generativa, més del 60 % de les persones enquestades consideren que aquest tipus de treball es veurà devaluat si no es redefineixen els criteris d'autenticitat, autoria i aportació humana. Concretament, més del 28 % de les respostes manifesten estar molt d'acord amb aquesta idea, mentre que poc més del 32 % mostren estar més aviat d'acord. Un 14,3 % de les respostes són neutrals, i aproximadament una quarta part expressa desacord.

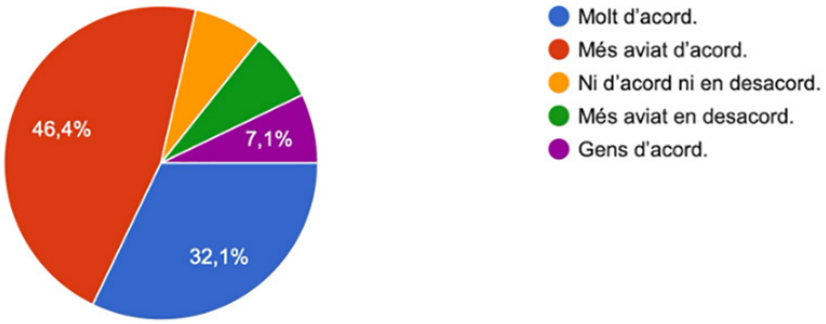
5. La IA generativa pot reforçar els imaginaris dominants —masculins, eurocentrats, capacitistes— si no es qüestionen els seus models i les dades amb què han estat entrenats.



La majoria de les respostes (75 %) coincideixen en què, si no es qüestionen els models i les dades amb què han estat entrenats, podria produir-se un reforçament d'imaginaris esbiaixats amb l'ús de la IA generativa. Al voltant d'un 18 % de les persones enquestades es mostren neutrals, mentre que poc més d'un 7 % no hi està gens d'acord.

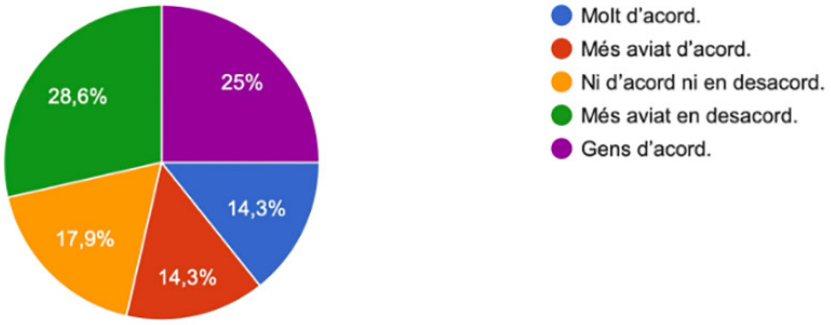
nodes

6. L'ús de la IA generativa hauria de ser incorporat en els processos educatius, artístics i professionals, perquè representa un progrés inevitable.



Amb aquesta pregunta volíem identificar quantes de les persones enquestades consideren que la IA generativa ha d'incorporar-se en àmbits de clara importància social, com ara l'educació, en la mesura en què aquesta tecnologia representa un progrés inevitable. El 32,1 % hi ha estat 'molt d'acord', mentre que el 46,4 % hi està 'més aviat d'acord', sumant en total un 78,5 %, això és, més de tres quartes parts, 22 de les 28 respostes. La resta de respostes es reparteixen equitativament entre 'ni d'acord ni en desacord', 'més aviat en desacord' i 'gens d'acord', amb un 7,1 % (2 respostes) cadascun.

7. He començat a delegar a la IA generativa tasques que abans assumia com a pròpies —contestar correus o missatges, formular idees o buscar consells personals— i, tot i que m'és útil, em pregunto si estic cedint alguna cosa que abans definia la meva manera de pensar o crear.



Les respostes han estat lleugerament més heterogènies que en els altres casos, tot i que més de la meitat de les persones enquestades es mostren en desacord amb l'afirmació. Així mateix, una mica més d'una quarta part de les respostes indiquen que sí que hi ha delegació i també reflexió sobre aquesta cessió. Finalment, prop d'un 18 % adopta una posició neutra.

Per cloure l'enquesta, vam fer una pregunta oberta per tal de recollir idees sobre el tema del monogràfic i donar l'opció de matisar les respostes que creguéssiu convenient: «Tens algun comentari que vulguis afegir sobre les qüestions tractades en aquest qüestionari?» Així doncs, per exemple, un dels comentaris rebuts esclareix, molt encertadament, que l'avaluació a la qüestió 3 només tracta de l'impacte de la IA generativa en l'àmbit professional dels enquestats, sense entrar en la na-

turalesa, positiva o negativa, d'aquest impacte. Així mateix, algunes respostes suposadament neutres («ni d'acord ni en desacord»), com ara a la pregunta 6, el que poden indicar és que s'admeten casos positius («per exemple, pluja d'idees per a nous projectes contextualitzats») i casos negatius («per exemple, estudi de coneixements fonamentals»).

Altres comentaris s'han centrat en el tema dels límits en l'ús d'eines d'IA generativa. A tall d'exemple, trobem un cas clar de limitació absoluta («No li delego cap tasca important o creativa»), una proposta interessant que no aprova l'ús en tasques com ara la redacció de correus electrònics («Hauríem de triar no involucrar la IA en la nostra interacció social amb altres persones, com ara els correus electrònics» —traducció de l'anglès), o un altre suggeriment que insisteix a veure la IA generativa com una eina de feina que no pot substituir el treball: «S'ha d'utilitzar com a eina, mai com a substituta de la teva feina.»

Així mateix, hi trobem una menció als biaixos que pot contenir el qüestionari. Tot i que no se'n presenten més detalls, podem entendre que la crítica va dirigida a la manca d'una o més preguntes on es tracti en positiu l'ús d'eines d'IA generativa (hi coincidim). Finalment, es recull un recordatori: delegar no significa traslladar la responsabilitat del material generat per una eina d'IA generativa; en última instància, som nosaltres els responsables.

Moltíssimes gràcies a totes les persones que vau col·laborar amb l'enquesta.

Premis Marc Esteva Vivanco

Innovació en la detecció 3D: una tesi que impulsa les aplicacions reals de núvols de punts

Thanasis Zoumpekas

Directores: Anna Puig, Maria Salamó
Universitat de Barcelona

La innovació en la detecció 3D, com es descriu en la tesi titulada en anglès "Advancing 3D Point Cloud Understanding in Real-World Applications" (https://tinyurl.com/zoumpekasp-hdthesis), impulsa la comprensió avançada dels núvols de punts 3D en aplicacions reals, millorant la capacitat d'interpretació i ús pràctic d'aquesta tecnologia en entorns del món real. La recerca presentada en aquesta tesi aborda els reptes tecnològics associats a l'ús de dades tridimensionals, o 'núvols de punts', que representen amb precisió objectes i entorns en 3D, i en desenvolupa solucions per millorar tant l'eficiència com la robustesa en camps com la detecció de riscos naturals o la identificació d'eines en processos industrials (veure Foto 1). Les seves contribucions ofereixen un impacte directe en sectors clau, amb aplicacions pràctiques que donen resposta a necessitats urgents i faciliten una implementació més eficient de tecnologies intel·ligents. L'objectiu principal d'aquesta tesi ha estat estudiar com interpretar la geometria continguda en els núvols de punts 3D, analitzant tant conceptes teòrics com pràctics que permetin la presa de decisions en aplicacions que depenen d'una interpretació geomètrica precisa.

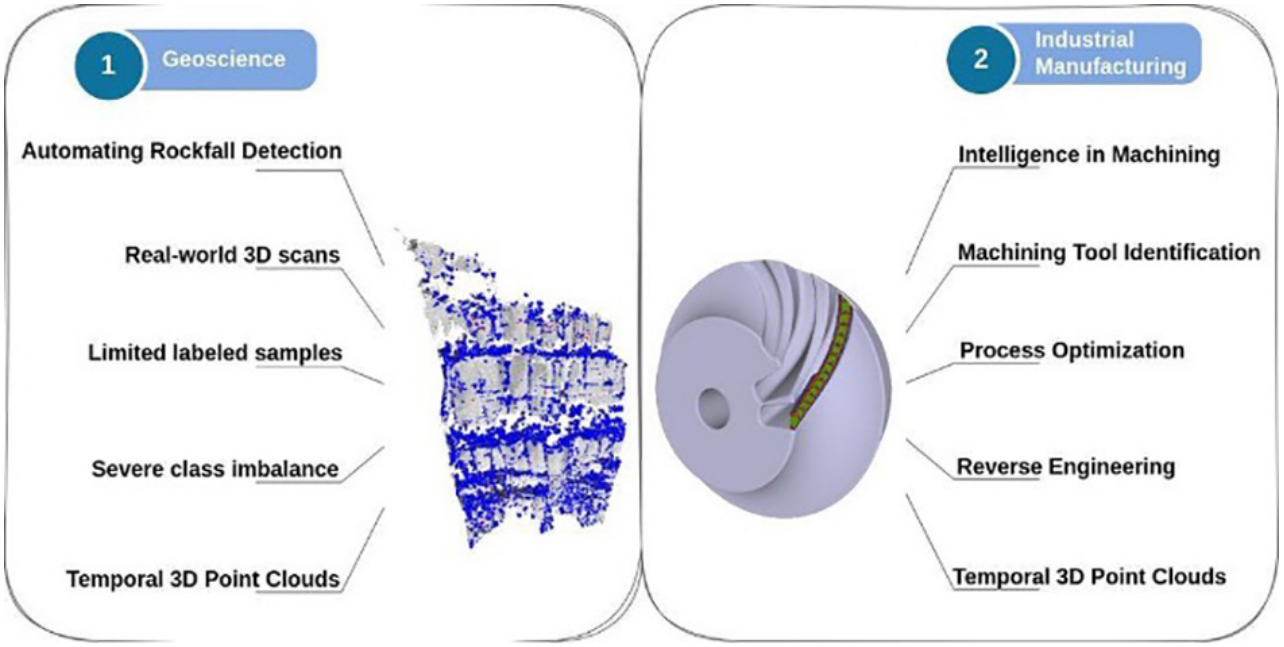


Foto 1: Aplicacions en la detecció de riscos naturals i la identificació d'eines en processos industrials

Un dels avenços principals de la tesi es centra en la millora dels models d'aprenentatge automàtic aplicats a la segmentació i identificació d'objectes en núvols de punts 3D. A través d'una rigorosa anàlisi de les arquitectures d'aprenentatge profund que permeten la segmentació de núvols de punts 3D i proposta de nous índexs de rendiment, aquesta recerca ha aconseguit optimitzar models de segmentació 3D per garantir-ne la precisió i la robustesa, però també la seva eficiència en termes de temps i ús de recursos. Aquesta millora és fonamental per a aplicacions que requereixen gran capacitat computacional, com ara la conducció autònoma o l'anàlisi de grans estructures arquitectòniques. A més, s'ha desenvolupat una eina interactiva de visualització per a l'anàlisi dels models de segmentació de núvols de punts 3D

«Un dels avenços principals de la tesi es centra en la millora dels models d'aprenentatge automàtic aplicats a la segmentació i identificació d'objectes en núvols de punts 3D.»

nodes

que permet als usuaris professionals observar els resultats del procés d'aprenentatge i prendre decisions més informades.[1] Amb aquestes propostes, la tesi no només contribueix a l'optimització tecnològica sinó que ofereix eines efectives per a un ús més eficient de la segmentació en núvols de punts 3D.

Una segona contribució d'aquesta investigació es dirigeix al camp de les Ciències de la Terra, concretament en la detecció de despreniments de roques, on s'han proposat vàries estratègies per a identificar i anticipar riscos geològics. En concret, s'han utilitzat conjunts de dades reals obtinguts amb làser terrestre (TLS) en diversos terrenys muntanyosos (e.g., Castellfollit i Montserrat). En un primer sistema desenvolupat amb la col·laboració del grup RISKNAT (http://www.ub.edu/risknat/), validat en el massís de Montserrat, s'utilitza un model de detecció que aplica diferents algorismes d'aprenentatge automàtic i tècniques de remostreig espacial per equilibrar el nombre limitat de mostres de casos reals de despreniments. Amb aquesta nova estratègia, de la qual ha sorgit un entorn intel·ligent que permet detectar zones candidates a produir despreniments, s'ha aconseguit millorar la precisió i la robustesa en la detecció de zones de risc, que ha contribuït a la seguretat de les infraestructures i a la protecció de les poblacions. Com a segona estratègia, en lloc de centrar-se a aplicar estratègies de remostreig de les escasses zones candidates, l'enfoc ha estat entendre els escanejos 3D reals i proposar un procés per augmentar mostres limitades de dades etiquetades que integressin veïnatges espacio-temporals de punts 3D. La nova estratègia de com estem abordant aquest problema de desequilibri de classes inherent a la detecció de despreniments, ens ha conduït a tenir resultats que augmenten la precisió en identificar àrees candidates. Val a dir que la investigació s'alinea amb els Objectius de Desenvolupament Sostenible de les Nacions Unides, ODS11 i ODS15, que promou la resiliència davant riscos naturals i la conservació del territori, alhora que facilita eines per a la gestió sostenible d'espais naturals.

«La investigació s'alinea amb els Objectius de Desenvolupament Sostenible de les Nacions Unides, ODS11 i ODS15, que promou la resiliència davant riscos naturals i la conservació del territori, alhora que facilita eines per a la gestió sostenible d'espais naturals.»

[1] *https://github.com/thzou/CLOSED_dashboard*

Finalment, la tesi aborda la innovació en el camp de la fabricació industrial, on s'ha desenvolupat nous models intel·ligents per a identificar eines de mecanitzat en el procés de manufactura d'una peça industrial. Aquesta aplicació té un impacte directe en l'optimització de la producció i l'ús eficient de recursos, estalvia material de mecanitzat i millora la precisió de l'enginyeria inversa en peces mecàniques complexes. Primer, s'ha proposat un nou procés i s'han desenvolupat processos per a la tasca d'identificació intel·ligent d'eines de mecanitzat a partir de dades de núvols de punts 3D temporals d'una peça en procés. A més, s'han creat i proporcionat dos conjunts de dades nous de núvols de punts 3D que s'han compartit amb la comunitat científica. Segon, s'ha experimentat amb la xarxa neuronal PointNet i s'han proposat tres noves variants. Els models creats han demostrat assolir una precisió del 95% en la detecció i identificació d'eines, que ha permès una optimització efectiva dels processos de fabricació i s'ha alineat amb els principis de sostenibilitat i producció responsable. Finalment, s'ha proposat un sistema intel·ligent extrem a extrem (end-to-end) per realitzar aquesta identificació. Aquesta innovació, realitzada amb col·laboració de l'empresa RISC Software GmbH (https://www.risc-software.at/en/), no només millora la productivitat i redueix costos, sinó que també minimitza els residus industrials i fomenta un ús més conscient dels recursos materials, d'acord amb els Objectius de Desenvolupament Sostenible, en concret ODS9 i ODS12.

Premis Marc Esteva Vivanco

Aquests resultats posen les bases per a noves aplicacions en gestió de riscos i optimització de processos industrials, i contribueixen a fer-los més segurs i sostenibles. Tots els avenços presentats en aquesta tesi, així com les dades de suport i els recursos, estan disponibles per a la comunitat investigadora en repositoris públics, amb la finalitat d'afavorir la recerca i el desenvolupament futurs. Finalment, destacar que aquesta tesi ha estat finançada pel programa Horizon 2020 de la Unió Europea, sota la iniciativa Marie Sklodowska-Curie, i ha rebut el Premi Marc Esteve Vivancos de l'Associació Catalana d'Intel·ligència Artificial (ACIA) l'any 2024.

Lliurament del Premi Marc Esteva Vivanco per l'Associació Catalana d'Intel·ligència Artificial.

nodes


IA
Amb el suport de:
i2cat
FUNDACIÓ
bigml
IDEAI
esade
RAMON LLULL UNIVERSITY
Institute for Data Driven Decisions (EsadeD3)
PAL
UOC
Universitat Oberta de Catalunya
CVC
Centre de Visió per Computador
UAB
Universitat Autònoma de Barcelona
laSalle
UNIVERSITAT RAMON LLULL
INSTITUT D'INVESTIGACIÓ EN INTEL·LIGÈNCIA ARTIFICIAL - IIIA
BIOINFORMATICS BARCELONA
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Politècnica Superior d'Enginyeria de Manresa
torre juana
Open Space Technology
EIAC
Observatori d'Ètica en Intel·ligència Artificial de Catalunya
Universitat de Girona