

Com fer un ús responsable de la intel·ligència artificial en educació?

Aspectes tècnics i ètics de la utilització de la intel·ligència artificial a l'aula

Jordi Vitrià Marca



Títol: Com fer un ús responsable de la intel·ligència artificial en educació? Aspectes tècnics i ètics de la utilització de la intel·ligència artificial a l'aula

CONSELL DE REDACCIÓ

Directora: Núria Serrano Plana (cap de la Secció d'Universitat, IDP-UB. Facultat de Química)

Coordinador: David Bueno Torrens (Facultat de Biologia, UB)

Consell de Redacció: Direcció de l'IDP; Pilar Aparicio Chueca, Facultat d'Economia i Empresa (UB); Silvia Argudo Plans, Facultat de Biblioteconomia i Documentació (UB); Anna Forés Miravalles, Facultat d'Educació (UB); Natalia Judith Laso Martín, Facultat de Filologia i Comunicació (UB); Francesc Martínez Olmo, Facultat d'Educació (Universitat Illes Balears); Roser Masip Boladeras, Facultat de Belles Arts (UB); Antonio R. Moreno Poyato, Facultat d'Infermeria (UB); José Navarro Cid, Facultat de Psicologia (UB); Elena Iborra Ortega, Facultat de Medicina i Ciències de la Salut (UB); Max Turull Rubinat, Facultat de Dret (UB), Eva González Fernández, IDP (secretària tècnica) i l'equip de Redacció de l'Editorial OCTAEDRO.

Corrector: Xavier Torras

Primera edició: juliol de 2025

Recepció de l'original: 4/10/2024

Aceptació: 10/1/2025

© Jordi Vitrià Marca

© IDP/UB i Ediciones OCTAEDRO, S.L.

Ediciones OCTAEDRO

Bailèn, 5, pral. - 08010 Barcelona

Tel.: 93 246 40 02 - Fax: 93 231 18 68

www.octaedro.com - octaedro@octaedro.com

Universitat de Barcelona

Institut de Desenvolupament Professional

Passeig de la Vall d'Hebron, 171, 08035 Barcelona

Tel.: 934035175

Aquesta obra està sota la llicència Reconeixement-NoComercial-SenseObraDerivada de Creative Commons i la Universitat de Barcelona. Podeu reproduir, distribuir o comunicar públicament l'obra només sota els termes d'aquesta llicència. En cada còpia que reproduïu, distribuïu o comuniqueu públicament, hi heu de fer constar l'autor i la institució (IDP/UB). No podeu fer-ne un ús comercial ni tampoc obres derivades. El text complet de la llicència el podeu trobar a: <http://www.publicacions.ub.es/doi/licencia/resum-noderiv.htm>.

ISBN: 978-84-1079-135-0

Disseny i producció: Serveis Gràfics Octaedro

SOBRE L'AUTOR

Jordi Vitrià és catedràtic de la Universitat de Barcelona, on imparteix un curs d'iniciació als Usos dels Algoritmes basats en Dades i cursos avançats d'IA i Ètica en la Ciència de Dades. Actualment és membre del Departament de Matemàtiques i Informàtica de la UB. També és director del Màster de Principis Fonamentals de la Ciència de Dades, codirector del Postgrau de Ciència de Dades i Intel·ligència Artificial de la UB i director de la Càtedra d'IA i Mitjans.

ÍNDEX

Introducció	5
1. Intel·ligència artificial i educació	11
Què és la intel·ligència artificial?	13
Com es construeix un sistema d'intel·ligència artificial?	19
Usos de la intel·ligència artificial en educació	25
Riscos	26
2. Com funciona la intel·ligència artificial predictiva?	30
De les dades als models	30
El nostre primer conjunt de dades	30
El nostre primer model	31
Com millorar el model?	33
Del model a l'ús	36
3. Com funciona la intel·ligència artificial generativa?	38
Respostes plausibles i respostes correctes	40
Sobre els usos de la intel·ligència artificial generativa	43
Aspectes sociotècnics de la intel·ligència artificial generativa	45
4. Com avaluem l'ús d'un sistema d'intel·ligència artificial?	48
Pas 1: Definir el problema i el valor que aporta la intel·ligència artificial	50
Pas 2: Comprendre'n el marc legislatiu i social	51
Pas 3: Identificar-ne els riscos	55
Pas 4: Avaluar-ne la viabilitat	58
El model de Canvas EduIA	59
La decisió sobre la viabilitat	61
Conclusions	64
Quines eines i usos de la intel·ligència artificial són importants en educació?	64
Què vol dir fer un ús de la intel·ligència artificial responsable en educació?	65
Com assegurem l'ús responsable de la intel·ligència artificial en educació?	66
Bibliografia	67

INTRODUCCIÓ

L'agost del 2020, centenars d'estudiants es van congregar davant el Ministeri d'Educació britànic a Londres, corejant: «Fuck the algorithm!» ('Que el bombin, l'algoritme!') (Hao, 2020). Al cap de pocs dies, les seves protestes van portar les autoritats a fer marxa enrere i descartar les «notes» que un algoritme havia generat per als exàmens que els estudiants no van poder fer a causa de la pandèmia de la COVID-19.

Els exàmens en qüestió eren els *A-level*, un conjunt de proves que els alumnes fan al voltant dels 18 anys. Aquests exàmens són els darrers abans d'ingressar a la universitat i tenen un pes molt alt en el procés d'admissió en l'educació superior. Les universitats fan ofertes basades en aquestes qualificacions i, generalment, un estudiant ha d'assolir determinats resultats per assegurar-s'hi una plaça.

El Govern britànic va decidir que, atesa la situació creada per la pandèmia de la COVID-19, no era aconsellable fer aquests exàmens de manera presencial. En lloc d'això, va proposar que les notes es determinessin mitjançant un algoritme¹ a partir de les dades següents:

1. La *distribució de qualificacions dels centres* a què pertanyien els alumnes durant els tres anys anteriors.
2. Un *rànquing de cada alumne* dins el seu centre i per a cada assignatura concreta, basat en l'opinió del professor de l'assignatura.
3. Les *qualificacions prèvies de cada alumne* en cada assignatura del seu centre.

El resultat d'aquest algoritme va ser que gairebé el 40 % dels estudiants van rebre notes més baixes del que era habitual, mentre que només un 2 % van obtenir notes més altes, cosa que va provocar protestes públiques i accions legals. Davant aquesta situació, el Govern britànic es va retractar i va decidir que els alumnes rebrien qualificacions basades en l'estimació dels seus professors sobre quina hauria estat la seva quali-

1. En el nostre cas, l'algoritme era un algoritme que estava dissenyat amb tècniques d'intel·ligència artificial.

ficació si els exàmens s'haguessin desenvolupat segons el que estava previst.

L'algoritme en qüestió havia analitzat la distribució històrica de les notes d'un centre escolar i assignava la nota d'un alumne en funció del *rànking*. Per exemple, si un alumne estava al mig de la llista del seu centre, la seva nota era, aproximadament, igual que l'obtinguda en anys anteriors per una altra persona en la mateixa posició del *rànking*. L'objectiu dels desenvolupadors de l'algoritme en incorporar la distribució de les qualificacions va ser corregir una desviació històrica observada en alguns centres educatius, que tendien a reportar qualificacions significativament més altes en comparació amb els resultats obtinguts en els *A-level*.

Què és un algoritme?

Un algoritme és un conjunt clar i detallat de passos que se segueixen per resoldre un problema o per respondre una pregunta. Funciona com una recepta que indica què fer en cada moment, sense necessitat de creativitat o habilitats especials per seguir-lo.

Un exemple senzill d'un algoritme és el que fem servir per sumar dos nombres alts a mà:

1. Escriu els dos nombres l'un sota l'altre, alineant els dígit per columnes (unitats, desenes, centenes, etc.).
2. Comença sumant els dígit de la columna més a la dreta (unitats).
3. Si la suma és superior a 9, escriu el dígit de les unitats del resultat i porta l'1 (el ròssec) a la columna següent de l'esquerra.
4. Repeteix el procés amb cada columna cap a l'esquerra, sumant els dígit i qualsevol ròssec.
5. Un cop que hagi sumat totes les columnes, el nombre final és el resultat de la suma.

Aquest algoritme garanteix que, si segueixes els passos correctament, sempre obtindràs el resultat correcte, sense necessitat de fer cap raonament addicional.

Els algoritmes són la nostra forma principal de comunicar-nos amb els ordinadors, perquè aquests només poden processar instruccions precises, lògiques i estructurades que es puguin traduir en operacions concretes sobre la seva electrònica.

La intel·ligència artificial (IA) amplia el concepte tradicional dels *algoritmes* en introduir la capacitat de generar-los automàticament a partir de dades. Això vol dir que, en comptes que un ésser humà dissenyi explícitament cada pas d'un algoritme per resoldre un problema, la IA pot analitzar grans volums de dades, identificar-hi patrons i aprendre les regles necessàries per assolir un objectiu específic.

L'anàlisi posterior d'un grup d'experts va detectar dos problemes en les notes generades per l'algoritme: l'*escassa precisió* de l'algoritme i la *presència de biaixos injustos*.

El primer problema era un problema de «qualitat» de l'algoritme. Quan es va testejar amb alumnes d'anys anteriors, l'algoritme va predir en molts casos notes força diferents de les que aquests estudiants van acabar obtenint. Davant aquest resultat, semblava molt arriscat utilitzar els resultats de l'algoritme de manera determinant sense considerar-ne l'alt grau d'error.

El segon problema consistia en un error sistemàtic en les prediccions de l'algoritme que beneficiava els alumnes que assistien a centres amb menys de 15 alumnes per aula. El resultat de l'anàlisi va mostrar que la proporció de notes altes i molt altes que van obtenir els alumnes de nivell socioeconòmic elevat gràcies a l'algoritme es van incrementar respecte de l'any anterior més del doble que en el cas dels alumnes que assistien a escoles d'entorns de nivell socioeconòmic baix. Dit d'una altra manera, l'algoritme no només reproduïa els biaixos estructurals de la societat britànica, sinó que els amplificava.

La lliçó d'aquesta història presenta diversos punts d'interès. En primer lloc, l'ús d'algoritmes en àmbits importants de la vida de les persones, com són l'educació, el mercat laboral o la justícia, ha de complir amb un *nivell d'exigència molt elevat* en aspectes com ara la precisió i l'absència de biaixos no desitjats. Per exemple, en el camp dels recursos humans, algunes empreses han implementat sistemes d'intel·ligència artificial (IA) per filtrar currículums o avaluar candidats durant els processos de selecció. En casos notables, com el d'un sistema utilitzat per Amazon, es va detectar que els algoritmes reforçaven estereotips de gènere, desfavorint les dones que es presentaven per ocupar llocs tecnològics (Dastin, 2018). De manera similar, en l'àmbit de la seguretat pública, eines com el *software* de predicció del crim han estat objecte de crítiques perquè amplifiquen desigualtats socials i racials. A ciutats dels EUA com ara Chicago o Los Angeles, algoritmes dissenyats per identificar zones d'alt risc han donat com a resultat un control policial més gran en comunitats racialitzades, perpetuant, així, desigualtats històriques (Richardson, 2019).

En segon lloc, la detecció de problemes com els esmentats només és possible si s'exigeix un cert nivell de *transparència* que permeti a experts i ciutadans analitzar i qüestionar els resultats generats per aquests sistemes.

Finalment, la solució de problemes relacionats amb algorismes no depèn únicament d'experts, sinó també d'una ciutadania *crítica i informada* que exigeixi responsabilitat en el seu disseny i l'ús. Tal com va passar amb els estudiants al Regne Unit, altres mobilitzacions, com ara les protestes contra l'ús de reconeixement facial per part de la policia a San Francisco el 2019, han demostrat que la pressió pública pot forçar les institucions a replantejar-se l'ús d'aquestes tecnologies.

Experiències com les descrites mostren que, encara que els algorismes tenen un gran potencial, el fet d'implementar-les sense una supervisió rigorosa pot accentuar les desigualtats i perpetuar problemes estructurals. Per consegüent, adoptar-les ha d'anar acompanyat d'un marc ètic, d'una supervisió transparent i d'una participació activa de la societat a fi de garantir que té un impacte positiu.

Davant la introducció de la IA en molts aspectes fonamentals de les nostres vides, i per evitar problemes com el que s'ha exposat, el Parlament Europeu va aprovar recentment la Llei de la Intel·ligència Artificial (EUR Lex, 2024), que pretén regular l'ús d'aquesta tecnologia en diferents sectors, inclòs l'educatiu. En aquest sector, la Llei busca generar més confiança i seguretat en l'adopció de tecnologies d'IA, ja que les eines emprades s'han d'ajustar a requisits de transparència, protecció de dades i drets fonamentals estrictes que, en principi, mitigaran els seus efectes adversos, entre els quals hi ha els que hem vist en l'exemple anterior.

Aquest llibre no té per objectiu explicar l'impacte de la nova llei en el sector, ni tan sols donar un seguit de recomanacions genèriques entorn dels usos més adients de la IA en educació. La IA pot ser una eina poderosa per gestionar, donar suport i millorar l'aprenentatge de l'alumnat, les tasques dels docents i la gestió del sistema educatiu, però, com hem vist, pot ocasionar danys o perpetuar i agreujar els problemes i les desigualtats existents en l'educació si no s'apliquen adequadament. L'objectiu d'aquest llibre és metodològic: establir unes bases simples i

comprensibles que permetin a tots els interessats en temes educatius avaluar de manera crítica possibles situacions concretes amb què es trobaran en la pràctica docent. Per assentar aquestes bases, cal conèixer alguns conceptes bàsics d'aquestes tecnologies, però també la naturalesa de determinats desafiaments ètics i les conseqüències socials que s'han d'abordar i que van molt més enllà de la privadesa i la seguretat de les dades.

El capítol 1 defineix la IA i els seus usos potencials en l'educació, centrant-se en la seva capacitat per millorar la presa de decisions basada en dades, predir i modelar el rendiment dels estudiants, a més d'assistir en la creació de materials d'aprenentatge personalitzats. També examina els tipus d'IA, incloent-hi la IA simbòlica, la IA predictiva, l'aprenentatge profund i la IA generativa, tot destacant les seves característiques i aplicacions. Així mateix, aborda els reptes i els problemes que poden sorgir arran de la implementació de la IA en l'educació, com ara les desigualtats en l'accés, la discriminació i les fissures quant a la privacitat.

El capítol 2 explica el cicle de vida d'un sistema d'IA predictiva, des de la recollida de dades fins al desenvolupament de models i la seva implementació en aplicacions concretes. Il·lustra com la IA predictiva es pot fer servir per predir l'èxit dels estudiants, amb un exemple hipotètic d'una universitat que utilitza dades històriques per preveure el rendiment futur dels alumnes. A més, mostra com es creen i s'entrenen els models d'IA predictiva emprant algorismes com ara la regressió, els arbres de decisió i les xarxes neuronals.

El capítol 3 s'endinsa en el funcionament de la IA generativa, pel que fa a la manera com s'entrena amb models de llenguatge i com genera continguts nous a partir de dades existents. S'hi discuteix la paradoxa de com els models generatius poden ser competents en tasques no previstes utilitzant l'analogia de la biblioteca de Babel imaginada per J. L. Borges per il·lustrar la capacitat de produir «text plausible», però no necessàriament «text correcte». Així mateix, explora les aplicacions de la IA generativa en l'educació, entre les quals la creació de continguts, el desenvolupament d'entorns d'aprenentatge personalitzats, l'assistència en l'escriptura i la creació de continguts per part dels estudiants, l'avaluació i la retroalimentació. També toca els aspectes so-

ciotècnics de la IA generativa, com ara la concentració de poder i coneixement en mans de grans empreses tecnològiques, juntament amb les implicacions ètiques i socials d'utilitzar-la.

El capítol 4 proposa una metodologia per avaluar l'ús responsable de la IA en l'educació, composta per quatre passos: definir el problema educatiu i el valor que aporta la IA, comprendre el marc legislatiu i social, identificar els riscos associats i avaluar la viabilitat de la proposta. Descriu el model Canvas EduIA com a una eina per avaluar la viabilitat de les aplicacions de la IA en l'educació, tot destacant la importància de considerar factors com ara l'equitat, l'autonomia d'alumnes i professors, la privadesa i la transparència. També ressalta la importància de la presa de decisions grupal en l'avaluació de la IA, per mitigar els biaixos individuals i promoure una perspectiva més completa i equilibrada.

Finalment, les conclusions resumeixen les eines i els usos més destacats de la IA en l'educació, com ara la tutoria intel·ligent, l'anàlisi d'aprenentatge, l'avaluació automatitzada, l'aprenentatge adaptatiu i els assistents virtuals. Defineixen el que significa fer un ús responsable de la IA en l'educació, posant èmfasi en les consideracions ètiques, la confiança, la reputació i el compliment legal i regulatori. També s'hi proporcionen directrius per assegurar l'ús responsable de la IA en l'educació, subratllant la necessitat d'un procés deliberatiu, la participació dels actors clau i l'adaptació a les particularitats de cada context educatiu.

1. INTEL·LIGÈNCIA ARTIFICIAL I EDUCACIÓ

La digitalització progressiva del nostre entorn personal, familiar, d'estudi i de treball ha facilitat el desenvolupament d'un conjunt de tecnologies, basades en la ciència de dades i la intel·ligència artificial (IA), que permeten automatitzar algunes tasques que fins ara estaven relacionades exclusivament amb la intel·ligència humana i que obren la porta a un nou model per a la presa de decisions i la resolució de problemes. El món de l'educació no n'ha estat una excepció i, durant aquests darrers anys, hem vist múltiples propostes per transformar la manera com s'ensenya i s'aprèn, que van des de la personalització de l'aprenentatge i el desenvolupament de tutors virtuals o assistents d'ensenyament fins a la gestió i l'administració educativa.

L'ús de les tecnologies d'IA actualment es divideix en dues àrees principals (Kennedy *et al.*, 2023). La primera és la generació de coneixement a partir de dades, que permet extreure informació valuosa i identificar patrons significatius en grans volums de dades, cosa que facilita la presa de decisions fonamentades. Això té aplicacions en anàlisis predictives, identificació de tendències, classificació i segmentació de dades, així com en l'automatització de descobriments científics. Aquest enfocament és especialment rellevant en sectors com ara la medicina, amb diagnòstics assistits per IA; les finances, amb la detecció de frau; i la logística, amb l'optimització de rutes, entre d'altres. La segona àrea és la facilitació de la comunicació amb els ordinadors mitjançant interfícies naturals, que possibilita que les persones interactuïn de forma més intuïtiva i fluida amb la tecnologia. Això s'aconsegueix mitjançant el reconeixement de veu i llenguatge natural, recurrent a assistents virtuals del tipus Alexa, Siri o Google Assistant; el reconeixement d'imatges i gestos, utilitzat en sistemes de seguretat, videojocs i interfícies accessibles; el processament i generació de text, com en *chatbots*, traductors automàtics i eines de redacció; i la interacció multimodal, que combina diverses formes de comunicació, com ara veu i gestos, per controlar dispositius. Aquestes dues àrees es complementen, ja que milloren la nostra capacitat d'anàlisi i maneig de dades alhora que eliminen barreres entre humans i màquines, fet que redunda en una tecnologia més inclusiva i eficient.

El valor fonamental de la IA en l'àmbit educatiu rau en la diversitat de les seves aplicacions, que es poden classificar en tres sectors principals. En primer lloc, la presa de decisions basada en dades permet analitzar grans volums d'informació per identificar patrons i tendències útils que milloren la presa de decisions en àrees com les polítiques educatives, l'administració escolar, la selecció de materials i mètodes d'ensenyament, l'avaluació de l'aprenentatge i les adaptacions curriculars per satisfer diferents necessitats educatives. Tot i que sempre hi hauria d'haver un ésser humà darrere de cada decisió, l'ús adequat de les dades pot ajudar a minimitzar els errors que solen tenir lloc quan les decisions es prenen confiant massa en la intuïció o basant-se únicament en pràctiques prèvies que han donat bons resultats.

En segon lloc, la predicció i el modelatge, a través de models estadístics i algorismes d'aprenentatge automàtic, faciliten l'anàlisi de l'historial acadèmic dels estudiants, la participació en l'aula i altres factors per preveure'n el rendiment futur. Això permet als educadors intervenir de manera precoç amb els estudiants que presenten risc, proporcionant-los el suport addicional necessari. A més, poden ajustar el contingut, les avaluacions i els desafiaments d'acord amb les necessitats d'aprenentatge individuals de cada alumne, optimitzant, així, la seva trajectòria educativa.

Finalment, la IA assisteix en l'educació per mitjà de la generació de materials d'aprenentatge personalitzats que s'adapten a les necessitats, els nivells d'habilitat i els interessos de cada persona. Per exemple, la IA pot crear textos, problemes i exemples en temps real per a estudiants que avancen a ritmes diferents. També permet als educadors dissenyar preguntes d'examen i exercicis de pràctica amb una àmplia varietat i nivells de dificultat adaptats a les habilitats dels estudiants, millorant la seva preparació per als exàmens. Així mateix, la IA pot ajudar a superar les barreres de comunicació i llenguatge en entorns educatius multilingües gràcies a la generació de traduccions en temps real i materials d'aprenentatge en diversos idiomes.

La implementació d'aquestes tecnologies en l'educació, malgrat que té un potencial immens, també pot comportar desafiaments i problemes: desigualtats en l'accés, riscos de discriminació, problemes de privadesa, etc. Abordar aquests problemes exigeix una planificació acurada,

inversió en infraestructura i formació, una comprensió clara de les limitacions de la tecnologia i polítiques sòlides per minimitzar els perills associats a aquesta tecnologia. La col·laboració entre desenvolupadors de tecnologia, educadors, psicòlegs, filòsofs, alumnat, famílies i legisladors és crucial si es volen maximitzar els beneficis de la IA en l'educació, alhora que se'n minimitzen els riscos i se superen els desafiaments.

Què és la intel·ligència artificial?

No hi ha una definició acceptada universalment de la paraula *intel·ligència*, però podem elaborar-ne una que ens serveixi de referència:

La intel·ligència és una facultat natural que permet a certs organismes vius assolir objectius complicats en situacions complexes de manera racional, especialment quan s'enfronten a circumstàncies noves i s'hi han d'adaptar.

La *intel·ligència* no és una característica exclusiva dels humans, fins al punt que podríem dir que qualsevol organisme viu mínimament complex posseeix un cert nivell d'intel·ligència. Tenint en compte això, podem afirmar que la facultat de la intel·ligència es manifesta de moltes maneres possibles, i la intel·ligència humana és una de les manifestacions més complexes i àmplies d'aquesta facultat.

Com hem dit, atesa la seva complexitat i varietat, el concepte de *intel·ligència* no té una definició precisa i els seus límits no estan clarament establerts. No obstant això, hi ha un cert consens al voltant de la idea que, en el cas dels éssers humans, es pot descompondre en una sèrie de competències.

Una *competència* és una habilitat específica, o un conjunt d'habilitats, que permet a un individu o sistema resoldre determinats tipus de problemes o adaptar-se a situacions noves mitjançant el processament d'informació. En el context de la intel·ligència humana, les competències poden ser múltiples i variades, i abastar àrees cognitives, emocionals, socials i pràctiques.

Les *competències cognitives* són fonamentals per al funcionament humà, ja que permeten processar informació, comprendre-la, emmagatzemar-la i utilitzar-la.

mar-la en la memòria i aplicar-la en la resolució de problemes i la presa de decisions. Entre les més importants hi ha la percepció, la capacitat d'aprenentatge, el raonament, la planificació i la presa de decisions, cadascuna amb aplicacions pràctiques en la vida diària.

La *percepció* és l'habilitat per captar i interpretar informació a través dels sentits i fa possible una interacció efectiva amb l'entorn. Per exemple, quan conduïm un cotxe, la percepció ens ajuda a identificar senyals de trànsit, detectar altres vehicles i avaluar la velocitat i la distància per evitar accidents.

La *capacitat d'aprenentatge* fa referència a la facultat d'adquirir i retenir coneixements o habilitats mitjançant l'experiència o l'ensenyament. Un exemple quotidià seria aprendre un nou idioma: les persones processen informació nova, com ara vocabulari i gramàtica, i l'apliquen en situacions comunicatives.

El *raonament* implica analitzar informació i arribar a conclusions lògiques. Això es pot observar quan un individu resol un problema matemàtic o decideix com s'ha de reparar un electrodomèstic avariament, analitzant-ne les possibles causes i escollint la millor solució.

La *planificació* permet organitzar activitats o recursos per assolir un objectiu específic. Per exemple, planificar unes vacances implica establir un pressupost, triar destinacions, reservar allotjament i organitzar itineraris per optimitzar l'aprofitament del temps disponible.

Per acabar, la *presa de decisions* es refereix a escollir la millor opció entre diverses alternatives, considerant la informació disponible i la incertesa associada a l'efecte de la decisió. Un exemple seria decidir quina carrera estudiar, avaluant factors com ara els interessos personals, les perspectives laborals i els costos educatius, a fi d'adoptar una decisió informada.

Aquestes competències treballen de manera interconnectada, facilitant l'adaptació i l'èxit en un entorn complex i en canvi constant.

El terme *artificial intelligence* es va fer servir per primera vegada en una escola d'estiu, la *Dartmouth Summer Research Project on Artificial Intelligence*, realitzada el 1956 al Dartmouth College, una universitat d'elit dels EUA. La conferència va aplegar vint de les ments més brillants en

informàtica i ciències cognitives de l'època per debatre sobre la conjectura següent (McCarthy *et al.*, 1955):

[...] que cada aspecte de l'aprenentatge o qualsevol altra característica de la intel·ligència, en principi, es pot descriure amb tanta precisió que es pot fer una màquina per simular-lo. S'intentarà trobar com fer que les màquines facin servir el llenguatge, formin abstraccions i conceptes, resolguin diversos tipus de problemes que ara estan reservats per als humans i es millorin a si mateixes. Creiem que es pot aconseguir un avenç significatiu en una o més d'aquestes qüestions si un grup de científics seleccionat amb cura hi treballa durant un estiu.

L'objectiu plantejat per a un estiu encara està pendent de solució i la seva consecució és l'objectiu d'una nombrosa comunitat científica que actualment inclou milers de científics d'arreu del món.

Durant les dècades següents, la IA es va desenvolupar a batzegades, amb períodes de ràpid progrés intercalats amb els anomenats «hiverns» de la IA, temporades durant les quals el finançament de la investigació en IA va ser escassa (Crevier, 1993). Avui dia, i molt especialment d'ençà del 2023, ens trobem en un «estiu» de la IA, que alguns creuen ja definitiu per al pas de la IA del món de la investigació al món de les aplicacions. Això s'ha produït com a resultat de la confluència de tres elements crítics: el creixement exponencial de les dades existents gràcies a internet i la Web, la generació de nous algorismes d'aprenentatge automàtic i el desenvolupament de processadors amb una potència de càlcul elevada en paral·lel.

Des dels inicis, definir el concepte d'IA ha resultat un tema ardu, com és natural quan parlem d'un fenomen que no acabem d'entendre. En aquest llibre assumim una posició pragmàtica i la definirem tal com ho fa la Llei europea d'intel·ligència artificial (EUR Lex, 2024):

[...] un sistema basat en màquines que està dissenyat per funcionar amb diversos nivells d'autonomia i que pot mostrar capacitat d'adaptació després que s'hagi desplegat, i que, per a objectius explícits o implícits, infereix, a partir de l'entrada que rep, com generar sortides com ara prediccions, continguts, recomanacions o decisions que poden influir en entorns físics o virtuals.

La definició que ens dona la Llei europea d'intel·ligència artificial ressalta alguns punts clau que podem desglossar en un llenguatge més accessible.

Primer, la IA *pot crear sistemes que operen de manera (parcialment) autònoma*, la qual cosa significa que poden operar per si mateixos sense necessitat que algú li indiqui constantment què fer. Per exemple, un assistent virtual com ara Alexa o Siri pot respondre les teves preguntes i prendre decisions sense que li hakis d'explicar com fer-ho cada vegada.

A més, té *capacitat d'adaptació*, és a dir, pot aprendre de la seva experiència o de la informació que rep durant el funcionament. Això li permet millorar el seu exercici amb el temps. Un exemple quotidià seria un sistema de recomanació, com el de Netflix, que aprèn quines sèries o pel·lícules t'agraden segons allò que has vist i et suggereix noves opcions.

La IA també *pot prendre decisions o fer prediccions* basant-se en la informació que se li dona. Per exemple, un sistema de navegació GPS pot utilitzar dades en temps real sobre el trànsit i les preferències de l'usuari per suggerir-li la millor ruta.

Finalment, cal considerar que aquestes decisions o recomanacions poden tenir un impacte tant en el món físic com en el virtual, afectant altres persones o la societat en conjunt.

Actualment, hi ha un consens entre la majoria dels investigadors de l'àrea sobre el fet que encara estem lluny no només de construir artefactes amb intel·ligència general, sinó, fins i tot, de comprendre tots els aspectes de la intel·ligència. Però això no vol dir que no hi hagi determinades àrees de recerca en la IA que hagin assolit resultats parcials força interessants. Les àrees més importants són:

- *L'aprenentatge automàtic* a partir de dades, conegut en anglès com a *machine learning*, és una branca de la IA que se centra en la construcció d'algoritmes que permeten a les màquines aprendre i fer prediccions o prendre decisions basades en dades. Els sistemes d'aprenentatge automàtic construeixen models matemàtics que s'ajusten o s'entrenen a partir de les dades d'entrada perquè puguin, sense estar explícitament programades per a cada cas, prendre decisions raonables quan s'exposen a dades noves.

Exemple: Imagina que tens una màquina que intenta endevinar si una fruita és una poma o una taronja basant-se en algunes dades. Per fer-ho, utilitzes l'aprenentatge automàtic (o *machine learning*). Primer de tot, proporciones a la màquina dades d'exemple sobre diferents fruites, com ara:

- *Color*: És vermella o taronja?
- *Grandària*: És petita com una poma o més gran com una taronja?
- *Textura*: És llisa o rugosa?

A partir d'aquestes dades, la màquina aprèn a identificar els patrons. Per exemple:

- Les pomes solen ser vermelles, petites i llises.
- Les taronges solen ser taronges, grans i rugoses.

Ara, quan li mostres una fruita que no ha vist abans, la màquina n'analitza les característiques (color, grandària i textura) i fa servir el que va aprendre dels exemples previs per dir-te si es tracta d'una poma o d'una taronja.

Aquest és un exemple senzill de com funciona l'aprenentatge automàtic: la màquina aprèn a prendre decisions basant-se en les dades que li proporcionen, sense que tu li diguis explícitament totes les regles.

- La *percepció visual de l'entorn* és una font d'informació que ens permet explorar el nostre entorn i fer-nos una idea de l'estat del món. Aquest camp combina mètodes de l'aprenentatge automàtic, el processament d'imatges i la visió artificial per permetre a les màquines veure i processar imatges i vídeos de manera similar a com ho fan els humans, fins i tot a escales i velocitats que, de vegades, superen les capacitats humanes.

Exemple: Un exemple addicional de percepció visual de l'entorn és utilitzar aplicacions per al diagnòstic mèdic basat en imatges. Aquestes aplicacions fan servir algorismes de visió artificial i aprenentatge automàtic per analitzar imatges mèdiques com ara radiografies, tomografies o ressonàncies magnètiques, així com per detectar possibles problemes de salut. Per exemple, un sistema d'IA pot examinar imatges de la retina, el sistema pot identificar condicions com ara retinopatia diabètica o degeneració macular. Aquests sistemes no només ajuden els especialistes a prendre decisions més ràpides i precises, sinó que també poden processar grans volums de dades en poc temps, facilitant el cribratge de malalties en poblacions àmplies i millorant l'accessibilitat al diagnòstic en àrees amb pocs recursos mèdics.

- El *reconeixement i producció de la parla* a partir dels sons és un element clau com a forma de comunicació amb els humans i obre la

porta a un estil de col·laboració força més natural entre humans i ordinadors.

Exemple Un exemple simple és l'ús d'assistents de veu com ara Alexa, Siri o Google Assistant. Aquests sistemes poden reconèixer ordres de veu de l'estil: «Quin temps farà demà?» o: «Posa una alarma a les 7 del matí», processar el llenguatge parlat i respondre amb informació o accions específiques, com ara proporcionar el pronòstic del temps o configurar l'alarma. Això permet una interacció més natural, ja que pots parlar amb la tecnologia en lloc d'escriure o prémer botons.

- L'àrea de *la percepció, l'anàlisi, la generació i la traducció del llenguatge natural* s'ocupa de la interacció entre els ordinadors i els humans mitjançant el llenguatge humà. L'objectiu és fer possible que els ordinadors compreguin, interpretin i produeixin llenguatge humà de manera útil.

Exemple: Un exemple simple és l'ús de *chatbots* en serveis d'atenció al client. Així, quan escrius: «Quin és l'estat de la meva comanda?» en el xat d'una botiga en línia, el *chatbot* entén la teva pregunta en llenguatge natural, cerca la informació en la base de dades i contesta amb alguna cosa de l'estil: «La teva comanda arribarà demà a la tarda». Això facilita una interacció ràpida i eficient, sense necessitat d'intervenció humana directa.

- El *raonament i la planificació* són àrees de recerca cabdals en el camp de la IA, enfocades a dotar les màquines de la capacitat per prendre decisions lògiques i elaborar plans d'acció per assolir objectius específics. Aquestes àrees aborden alguns dels desafiaments més complexos i fascinants en la creació de sistemes intel·ligents que poden actuar de manera autònoma.

Exemple: Un exemple simple és el funcionament d'un robot aspirador. El robot analitza el disseny d'una habitació, identifica obstacles com ara mobles i decideix el millor camí per netejar tot l'espai de manera eficient. A més, si la bateria està baixa, planifica tornar a la base de càrrega abans de continuar la tasca. Aquest procés combina raonament i planificació per assolir el seu objectiu: netejar la casa de manera autònoma.

Aquests resultats s'han anat produint al llarg dels darrers setanta anys de forma progressiva, amb molts alts i baixos que han fet que la IA es vagi incorporant de manera gradual a les nostres vides.

Com es construeix un sistema d'intel·ligència artificial?

Les primeres aplicacions d'IA, desenvolupades als anys setanta, es van basar en l'enfocament *simbòlic*, també anomenat «basat en regles» (Boden, 2014). Aquest enfocament consistia a definir un conjunt de regles clares i precises que indiquen a la màquina, de manera explícita, com actuar en diferents situacions. Per fer servir aquestes regles, s'utilitzava un motor d'inferència, que és com el «cervell» del sistema: un programa informàtic que analitzava les dades d'entrada, buscava les regles rellevants i les aplicava per produir un resultat o prendre una decisió. És a dir, el motor d'inferència connectava les regles amb dades per resoldre problemes o respondre preguntes.

Les regles s'implementen a través de la programació i sorgeixen de l'experiència humana en la matèria. Un cop definides les regles, la IA s'encarrega d'aplicar-les a les dades d'entrada. Per exemple, un sistema basat en regles pot ser utilitzat per classificar correus electrònics com a correu brossa o *spam* basant-se en la presència de determinades paraules clau o patrons en el correu.

Com funciona un sistema basat en regles per detectar *spam*?

Un sistema basat en regles per detectar *spam* funciona aplicant un conjunt predefinit de regles o criteris per identificar correus electrònics no desitjats. El procés podria ser el següent:

1. Els experts en seguretat i correus electrònics defineixen un seguit de regles basades en característiques comunes de l'*spam*. Aquestes regles poden ser simples o complexes, depenent del sistema. Per exemple:
 - a) Correus amb certes paraules clau com ara «gratis», «guanyar» o «oferta limitada».
 - b) Presència d'enllaços a dominis sospitosos.
 - c) Ús excessiu de majúscules i signes d'exclamació.
2. Cada correu electrònic entrant s'analitza per veure si compleix alguna de les regles definides. Això inclou revisar el cos del missatge, les capçaleres, els enllaços i els fitxers adjunts.
3. A cada regla se li assigna un pes o una puntuació. Si un correu compleix una regla, se li sumen els punts corresponents a la puntuació total.
4. S'estableix un llindar de puntuació. Si la puntuació total del correu supera aquest llindar, el correu es marca com a *spam*.

Durant dècades es van desenvolupar aquesta mena de «sistemes experts» d'IA, que podien estar basats en centenars de regles, per a una vasta gamma d'aplicacions, com ara diagnòstics mèdics, aplicacions fi-

nanceres i processos de fabricació (Jackson, 1990). Independentment del nombre de regles, per a un humà és fàcil entendre la lògica de les decisions d'aquests sistemes.

Els sistemes basats en regles encara són útils i utilitzats en certs àmbits, però, a mesura que la quantitat de regles creix i que les seves interaccions es multipliquen, el desenvolupament i el manteniment es compliquen extraordinàriament. Una altra de les seves limitacions és que els experts han d'expressar de manera explícita les regles, un aspecte que no és factible en molts àmbits de la intel·ligència com ara el diagnòstic a partir d'imatges mèdiques.

Molts dels avenços actuals de la IA se sustenten en un model alternatiu a les regles: les tècniques d'aprenentatge automàtic o *IA predictiva* (Alpaydin, 2021). Aquestes tècniques es basen en l'anàlisi de dades per extreure'n automàticament les regles de decisió. En aquest cas, el paper de l'expert no és enunciar les regles, sinó que es redueix a etiquetar les dades de manera que el sistema d'IA sigui capaç de generar les regles a partir d'aquesta informació implícita.

Com funciona un sistema basat en IA predictiva per detectar *spam* ?

Un sistema basat en IA predictiva per detectar *spam* utilitza tècniques d'anàlisi de dades i models predictius per identificar correus electrònics no desitjats. El procés podria ser aquest:

- Es recopila un gran nombre de correus electrònics etiquetats com a *spam* i no *spam*. Aquestes dades serviran per entrenar el model.
- S'extreuen característiques rellevants dels correus electrònics com ara paraules clau, freqüència de paraules, longitud del correu, presència d'enllaços, adreces de remitent, etc.
- Se seleccionen un o més algorismes d'aprenentatge automàtic. Alguns algorismes comuns per a la detecció de *spam* inclouen *Naive Bayes*, regressió logística, arbres de decisió, màquines de suport vectorial (SVM) o xarxes neuronals.
- El conjunt de dades es divideix en un conjunt d'entrenament i un conjunt de proves.
- El model s'«entrena» utilitzant el conjunt d'entrenament, ajustant els paràmetres de l'algorisme seleccionat per minimitzar l'error en la classificació de correus electrònics. El resultat és un model matemàtic que permet analitzar un correu i determinar-ne la classe a partir de les característiques rellevants.
- S'avalua el model entrenat utilitzant el conjunt de proves per mesurar-ne la precisió.

En el cas d'un sistema d'aprenentatge automàtic aplicat a l'educació, les dades podrien estar formades per la informació sobre el rendiment acadèmic dels estudiants, com ara qualificacions en exàmens, treballs que s'han entregat, assistència a classes i participació en activitats extracurriculars.

La tasca del sistema d'IA seria, llavors, processar aquestes dades per inferir automàticament la relació entre les dades d'entrada (rendiment i activitats dels estudiants) i l'etiqueta de sortida (resultats acadèmics).

Cal assenyalar que, en aquest cas, la relació entre les dades d'entrada i les dades de sortida no estarà especificada en forma de regles de lectura fàcil per a un ésser humà. En lloc d'això, s'expressa mitjançant complexes relacions matemàtiques implementades en algorismes, com ara les xarxes neuronals. Això fa que sigui impossible, per a un ésser humà, entendre la lògica de les decisions del sistema simplement inspeccionant l'algorisme resultant.

Els models basats en la IA predictiva constitueixen la base de la segona onada d'aplicacions d'IA que va arrencar als anys noranta. Aquests models s'han fet servir en àmbits tan dispars com les finances, la banca, la salut, la medicina, el comerç electrònic o el transport i la logística.

El 2012 va aparèixer un nou tipus de models, els models d'*aprenentatge profund*, que van reforçar aquesta tendència i van eixamplar la gamma d'aplicacions (Zhang *et al.*, 2023). L'aprenentatge profund és un subcamp de la IA predictiva que usa xarxes neuronals per analitzar altres classes de dades (imatges, vídeos, textos) amb una elevada precisió.

Els avantatges de l'aprenentatge profund inclouen la capacitat per processar i aprendre a partir de grans quantitats de dades no estructurades i la seva habilitat per dur a terme tasques de classificació i predicció amb molta precisió. Tanmateix, també s'associa a desafiaments, com ara la necessitat de grans conjunts de dades per entrenar els models de manera efectiva, el seu alt cost computacional i la dificultat per interpretar els models, en el que es coneix com el problema de la «caixa negra» (Castelvecchi, 2016).

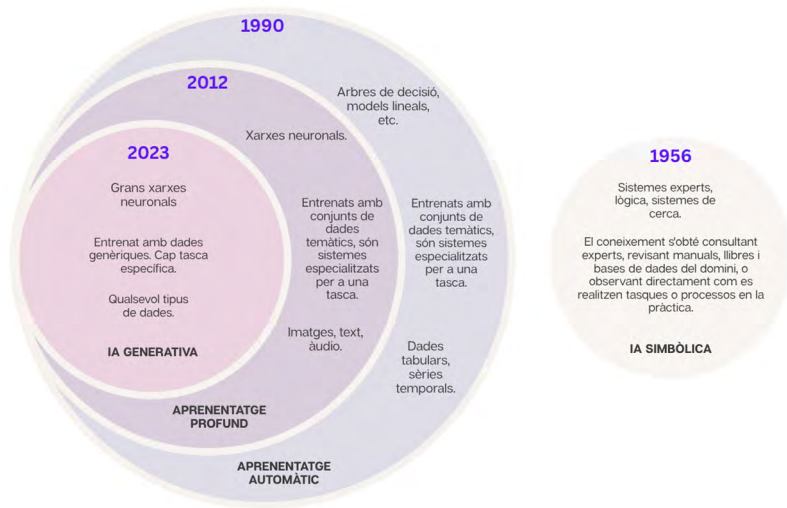


Figura 1. Les eines d'IA que podem utilitzar actualment s'han desenvolupat majoritàriament dins de quatre marcs conceptuals: el simbòlic, el predictiu, l'aprenentatge profund i el generatiu. Aquesta figura mostra els quatre marcs, les eines principals, com creen els seus models i, finalment, en quina mena de dades es poden aplicar.

Les xarxes neuronals profundes van millorar significativament el processament del llenguatge natural respecte a les tecnologies basades en els models predictius o les basades en el model simbòlic, fet que va permetre la representació del significat del text de manera més precisa i útil per a moltes aplicacions. Això va fer possible l'existència dels assistents virtuals nascuts la primera dècada del segle XXI, capaços d'interpretar alguns aspectes del llenguatge humà com ara identificar una necessitat i fer una acció per satisfer-la, o respondre amb un guió predefinit a una determinada pregunta.

Però el 2017 es va produir un avenç fonamental en el camp de l'aprenentatge profund que va transformar radicalment el processament del llenguatge natural: la introducció d'una nova classe de xarxa neuronal profunda basada en mecanismes d'atenció (Vaswani *et al.*, 2017). Aquest disseny va fer possible la construcció dels grans models de llenguatge –*large language models* (LLM)– (Mann *et al.*, 2020), assentant, així, els fonaments del que avui coneixem com a *intel·ligència artificial generativa*.

Com aprèn un sistema d'IA generativa?

Un sistema d'IA generativa aprèn a partir de volums enormes de dades utilitzant tècniques avançades d'aprenentatge automàtic, particularment, xarxes neuronals profundes. Aquí es mostra com funciona el procés d'aprenentatge amb un exemple específic: un model de generació de text com GPT (*generative pre-trained transformer*).

1. Es recopilen ingents quantitats de dades de text de diverses fonts, com ara llibres, articles, pàgines web, etc. Aquestes dades han de ser variades i representatives de la llengua (i dels registres d'aquesta llengua) que es vol modelar.
2. El text es converteix en seqüències d'elements numèrics (que corresponen a paraules o subparaules) que la xarxa neuronal pot processar.
3. S'utilitza una arquitectura de xarxa neuronal profunda, com ara un *transformer*, per entrenar el model. Un *transformer* té múltiples capes d'atenció que li permeten entendre les relacions contextuais entre les paraules en una seqüència.
4. El model s'entrena en dues fases: preentrenament i ajustament fi (*fine-tuning*).
5. El model s'entrena en una tasca de llenguatge general, com és predir la paraula següent en una oració. Per exemple, donat el text: «El gat està sobre la», el model aprèn a predir «teulada» a partir del context. Això es fa emprant grans volums de text sense etiquetes, on el model aprèn les estructures del llenguatge i les relacions entre paraules.
6. Després del preentrenament, el model s'ajusta utilitzant un conjunt de dades específic per a la tasca desitjada, com ara generació de text coherent o resposta a preguntes.
7. Un cop entrenat, el model pot generar text nou. Se li proporciona una entrada inicial (*prompt*) i el model genera text basat en allò que ha après:
 - **Entrada** (*prompt*): «Una vegada hi havia, en un petit poble,».
 - **Sortida** (resposta del model): «Una vegada hi havia, en un petit poble, envoltat de turons verds i boscos densos. Els habitants vivien en harmonia, i cada primavera celebraven una gran festa en honor de la natura. Un dia, un estrany visitant arribà al poble, portant històries de terres llunyanes i aventures extraordinàries...».

El llançament del LLM més famós, ChatGPT (OpenAI, 2022), va obrir la porta d'aquesta tecnologia a qualsevol persona amb accés a internet, que pot interactuar amb una interfície web o una aplicació al mòbil de forma senzilla. Això va posar en relleu els impressionants avenços dels LLM, ja que anteriorment els LLM més potents només estaven disponibles per a investigadors amb molts recursos i per als que tenien uns coneixements tècnics molt profunds.

Posteriorment, aquests models s'han generalitzat i són capaços d'acceptar entrades multimodals (text, imatges, vídeo), i també de generar sortides multimodals, constituint l'element principal de l'anomenada IA generativa.

El paper dels experts en el desenvolupament de la IA generativa encara es redueix més que en el cas de la IA predictiva, atès que són capaços d'aprendre patrons complexos partint de les dades sense necessitat d'intervenció o etiquetatge manual per part d'experts. Els models s'entrenen per predir o reconstruir parts de les dades d'entrada a partir d'altres parts de les mateixes dades. Aquest enfocament aprofita l'estructura inherent de les dades per aprendre representacions útils i generals. Per exemple, un model podria aprendre a predir la pròxima paraula en una oració o completar una imatge parcialment oculta utilitzant només el context proporcionat per les dades mateixes. Aquest mètode redueix la necessitat d'elevades quantitats de dades etiquetades, que poden ser costoses i laborioses de produir.

La IA generativa permet als ordinadors crear contingut per si mateixes. Imaginem que, en lloc de simplement respondre a preguntes concretes, per a la qual cosa ha estat «entrenat» amb una base de dades específica, un ordinador pot interpretar imatges, resumir textos, compondre música o generar idees sobre qualsevol temàtica. I no sols això, sinó, inclús, arribar a comunicar-se amb els usuaris en llenguatge natural.

Taula 1. Característiques principals dels diferents tipus d'IA

Tipus d'IA	Necessitat de dades	Tipus de dades	Rol d'un expert
IA simbòlica	Molt pocs. Només serveixen per validar el sistema, ja que la informació que volem capturar és al cap de l'expert.	Dades tabulars.	Definir de manera explícita les regles que es poden aplicar a un concepte.
IA predictiva	Centenars o milers de dades, per descobrir-hi patrons que representin els conceptes que volem aprendre.	Dades tabulars. Sèries temporals. Imatges o àudios molt simples.	Etiquetar cadascuna de les dades amb el concepte que li correspon.
Aprenentatge profund	De desenes de milers a milions de dades, per descobrir-hi patrons que representin els conceptes que volem aprendre.	Imatges, àudio, vídeos, textos.	Etiquetar cadascuna de les dades amb el concepte que li correspon.
IA generativa	Molts milions de dades.	Imatges, àudio, vídeos, textos.	Cap.

Avui dia hi ha aplicacions pràctiques de les quatre classes d'IA que hem vist (taula 1), però la majoria corresponen a la IA predictiva o a l'aprenentatge profund, i la IA generativa, que ha generat potencialment més expectatives, és encara minoritària.

Usos de la intel·ligència artificial en educació

De manera general, els usos de la IA en l'entorn educatiu es poden caracteritzar en funció del receptor dels beneficis potencials de l'ús d'aquesta tecnologia.



Figura 2. Usos de la IA en educació

La *IA centrada en l'estudiant* té com a objectiu millorar l'experiència individual i personalitzar l'educació segons les necessitats específiques de cada alumne. Això inclou l'ús de tutors intel·ligents que proporcionen tutories personalitzades i adaptatives, ajudant els estudiants a aprendre al seu propi ritme. També es fan servir plataformes d'aprenentatge adaptatiu que ajusten el contingut educatiu en funció del progrés i el rendiment de l'alumne, presentant nous reptes només quan està preparat per avançar. Així mateix, els assistents virtuals basats en IA poden respondre preguntes, oferir explicacions addicionals i guiar l'alumnat a través de materials d'estudi complexos, facilitant-ne la comprensió i l'aprenentatge.

Quant a la *IA centrada en el professor*, l'objectiu és donar suport als educadors millorant l'eficiència i l'efectivitat dels mètodes d'ensenyament.

Això s'aconsegueix per mitjà d'eines que analitzen el rendiment dels estudiants durant les activitats formatives, identificant àrees en què necessiten més suport, cosa que permet als professors ajustar les seves estratègies pedagògiques. D'altra banda, aplicacions d'automatització de tasques administratives alliberen els docents d'activitats repetitives com ara qualificar exercicis o gestionar l'assistència, i gràcies a això es poden centrar en l'ensenyament. També s'utilitzen plataformes de desenvolupament professional que ofereixen recursos i formació personalitzada adaptats a les necessitats i els interessos de cada educador.

Finalment, la *IA centrada en la institució* pot optimitzar la gestió i les operacions de les institucions educatives mitjançant diverses aplicacions. Els *sistemes de gestió de l'aprenentatge* (LMS) integren IA per millorar l'administració de cursos, la comunicació entre alumnat i docents, i el seguiment del rendiment estudiantil. A més, eines basades en IA analitzen grans volums de dades per predir tendències futures en matriculació, identificar necessitats de recursos o proposar millores en el desenvolupament curricular. També s'empren solucions d'IA per incrementar la seguretat en el campus per mitjà de sistemes avançats de vigilància i per facilitar l'accés inclusiu a recursos educatius destinats a estudiants amb discapacitats.

Cadascun dels models d'IA que hem vist pot tenir un ús en aquestes àrees, depenent del cas concret.

Riscos

L'anàlisi de riscos és un pas essencial per garantir un ús responsable i ètic de la IA en l'educació. Aquest procés implica identificar, avaluar i mitigar els possibles impactes negatius que poden sorgir a l'hora d'implementar sistemes d'IA en contextos educatius. En aquest àmbit, en què la interacció amb estudiants, docents i comunitats és altament sensible, és fonamental basar-se en principis clars que guiïn el desenvolupament, la implementació i l'avaluació d'aquestes tecnologies.

La Unesco, en la seva *Recomanació sobre l'Ètica de la Intel·ligència Artificial* (Unesco, 2022), proposa un conjunt de principis que ofereixen un marc ètic per abordar els riscos associats a l'ús de la IA, també

dins l'àmbit educatiu. Aquests principis són clau per analitzar i mitigar els riscos vinculats amb l'equitat, la privadesa, la transparència i la sostenibilitat, entre d'altres. Aquests principis es resumeixen en els punts següents:

- *Proporcionalitat i innocuïtat*: els sistemes d'IA han de ser proporcionals als objectius educatius i no causar cap dany als estudiants. Cal avaluar els riscos potencials de la IA en l'educació i prendre mesures per mitigar-los.
- *Equitat i no discriminació*: els sistemes d'IA han de promoure la justícia social i la igualtat d'oportunitats per a tot l'alumnat, sense discriminació per raça, gènere, origen o capacitat. Els estats membres han de fomentar l'accés inclusiu a la IA, adaptant els continguts i els serveis al context local i respectant el multilingüisme i la diversitat cultural.
- *Sostenibilitat*: la IA ha de contribuir a una educació sostenible que tingui en compte les necessitats de les generacions presents i futures.
- *Dret a la privacitat i protecció de dades*: la privacitat dels estudiants ha de ser respectada, protegida i promoguda. És essencial protegir les dades personals dels alumnes i utilitzar-les de manera responsable i ètica.
- *Supervisió i decisió humana*: els docents han de mantenir el control sobre el procés educatiu. La IA ha de ser una eina de suport, no pas un substitut dels docents.
- *Transparència i explicabilitat*: els sistemes d'IA han de ser transparents i explicables, de forma que estudiants i docents comprenguin per què han pres una decisió determinada. Això inclou informar els usuaris quan un producte o servei ha estat proporcionat mitjançant IA.
- *Responsabilitat i rendició de comptes*: s'han d'establir mecanismes clars de responsabilitat i rendició de comptes per a l'ús de la IA en l'educació. Això implica investigar i reparar els danys ocasionats per sistemes d'IA, a més d'establir mecanismes d'aplicació i mesures correctives per garantir els drets humans i l'estat de dret.

Més enllà d'aquests principis, la Unesco destaca la importància de la sensibilització i l'educació sobre la IA. Això inclou promoure programes de sensibilització sobre els avenços de la IA i les seves repercussions en

els drets humans. També és cabdal fomentar la investigació sobre la IA, incloent-hi la investigació sobre l'ètica de la IA.

Passar dels principis ètics a l'ús pràctic constitueix una tasca complexa que comporta fer front a diversos desafiaments tant tècnics com socials i educatius. Per més que els principis com l'equitat, la transparència i la sostenibilitat ofereixin una guia de pes, la seva implementació en contextos reals presenta nombroses dificultats.

En primer lloc, els principis ètics acostumen a ser amplis i abstractes, cosa que deixa marge per a interpretacions diferents depenent del context. Per exemple, garantir l'equitat en l'ús de la IA pot implicar accions molt diferents en una escola rural sense accés a internet i en un entorn urbà amb recursos tecnològics avançats. Aquesta manca de claredat sobre com interpretar i aplicar aquests principis pot dificultar la forma com es tradueix en normes i pràctiques concretes.

Un altre repte que cal ressaltar és l'absència de recursos i capacitats en molts sistemes educatius. La implementació de la IA demana una infraestructura tecnològica adequada, com ara dispositius i connectivitat, que no sempre està disponible en totes les regions. Encara més, els docents i administradors educatius sovint no compten amb la formació necessària per utilitzar eines d'IA de manera efectiva. Fins i tot quan hi ha recursos tècnics, el cost elevat del desenvolupament i manteniment d'aquestes tecnologies pot representar una barrera, especialment en sistemes educatius amb pressupostos limitats.

La complexitat tècnica dels sistemes d'IA també implica un obstacle. Molts models d'IA, i més encara els més avançats, com són els basats en aprenentatge profund, són difícils d'entendre fins i tot per part dels experts. Això complica la transparència i l'explicabilitat, dos principis essencials per generar confiança en aquestes tecnologies. Així mateix, els sistemes d'IA depenen de les dades amb què s'entrenen. Si aquestes dades contenen biaixos històrics o desigualtats, la IA pot perpetuar o fins i tot amplificar aquests problemes, contradient principis com ara l'equitat.

D'altra banda, l'absència d'una regulació adequada agreuja la situació. En molts països, les normatives sobre l'ús de la IA en l'educació són inexistents o estan poc desenvolupades, la qual cosa provoca incertesa

sobre com protegir la privacitat dels estudiants o com garantir la rendició de comptes. Així mateix, les diferències entre marcs reguladors de diferents països fan que costi més implementar principis ètics de manera uniforme a escala global.

A això s'hi sumen els conflictes d'interessos. Un gran nombre d'eines d'IA són desenvolupades per grans empreses tecnològiques l'objectiu principal de les quals és el benefici econòmic. Això pot entrar en conflicte amb principis com ara la sostenibilitat o l'equitat, sobretot si les tecnologies es dissenyen pensant més en mercats lucratius que no pas en contextos educatius diversos. A més, els responsables dels sistemes educatius poden prioritzar resultats ràpids, com seria l'augment de les puntuacions en els exàmens, sobre una implementació reflexiva i ètica d'aquestes eines.

Finalment, la diversitat dels contextos educatius afegeix un altre nivell de complexitat. Les escoles i els estudiants tenen necessitats i condicions molt diferents, cosa que significa que no hi ha una solució única que funcioni per a tothom. Adaptar els sistemes d'IA a aquests contextos diversos requereix temps, esforç i recursos addicionals, una circumstància que en pot endarrerir o dificultar la seva implementació.

2. COM FUNCIONA LA INTEL·LIGÈNCIA ARTIFICIAL PREDICTIVA?

La comprensió del cicle de vida d'un sistema d'intel·ligència artificial predictiva és un dels elements bàsics per tenir una visió realista i crítica de l'aplicació de la IA en l'educació.

El cicle consta de quatre passos: la recollida d'un conjunt de dades rellevant per resoldre un problema, la creació d'un model que codifica el coneixement implícit en les dades, la creació d'una aplicació informàtica que implementa el model i, finalment, l'ús d'aquesta aplicació en un entorn concret.

De les dades als models

Per entendre aquests passos, pot ser útil pensar en un exemple hipotètic.² Posem per cas que una universitat vol utilitzar dades històriques d'estudiants per construir models per predir *l'èxit futur dels estudiants que comencen un grau*. Aquests models poden ser útils per donar suport a canvis en polítiques i pràctiques de la institució. Per exemple, aquest model permetria identificar l'alumnat que pot estar en risc de rendiment acadèmic baix o d'abandonament, a fi d'intervenir-hi amb recursos de suport com ara tutories o assessorament acadèmic des del principi del recorregut acadèmic.

El nostre primer conjunt de dades

Les *dades* de cada alumne que inicialment es poden veure com a necessàries per a l'objectiu definit poden incloure elements molt diversos: variables sociodemogràfiques (gènere, raça, país d'origen, nombre de germans, barri i ciutat de residència, etc.), notes de batxillerat, notes de selectivitat, llistat d'activitats extraescolars que ha fet, esports que practica, etc. En el cas d'antics alumnes, també disposem de la nota mitjana de totes les convocatòries d'examen a què s'han presentat en

2. Aquest exemple està inspirat en el cas proposat per Google com a exemple de biaix algorítmic a: <https://pair.withgoogle.com/explorables/hidden-bias>

el transcurs del seu grau, que pot representar l'èxit de l'estudiant a la universitat. En termes d'un sistema d'IA, el primer conjunt de dades constitueix l'*entrada* del sistema (el conjunt de dades que volem interrogar, que normalment es denoten amb la lletra X) i el segon, la nota mitjana de totes les convocatòries d'examen a les quals s'han presentat en al llarg del grau, són les dades de *sortida* o etiquetes (les dades que constitueixen la resposta a l'entrada, que s'acostumen a denotar amb la lletra Y).

El nostre primer model

El pas següent és la selecció del tipus de *model* matemàtic que farem servir per representar el coneixement implícit que creiem que contenen les dades. Els models creats mitjançant tècniques d'IA solen expressar-se de forma matemàtica.

En el nostre exemple, podríem començar definint manualment, sense recórrer a cap mena d'anàlisi de les dades, un model d'èxit d'un estudiant, que anomenarem Y :

$$Y = \text{Nota mitjana del batxillerat}$$

La interpretació de l'expressió matemàtica és ben simple: l'èxit d'un estudiant que comença un grau, representat pel valor numèric Y , és igual a la nota mitjana de les assignatures que va cursar durant el batxillerat, representat pel valor numèric X .

A causa de la limitació de la dada d'entrada X que s'ha usat, la nota mitjana que va obtenir en el batxillerat, no podem esperar que aquest model sigui gaire precís. Per avaluar aquesta precisió, podem fer servir una base de dades d'antics alumnes amb les característiques abans esmentades. Per a això, n'hi ha prou amb comparar el valor que el model prediu per a cada alumne de la base de dades i la nota mitjana real del seu grau.

En la figura 3 podem veure el resultat per a un conjunt de 100 estudiants. Els punts blaus corresponen als valors (x,y) dels alumnes disponibles en el conjunt de dades i la línia vermella correspon a la predicció que hauria fet el model Y per a aquestes dades.

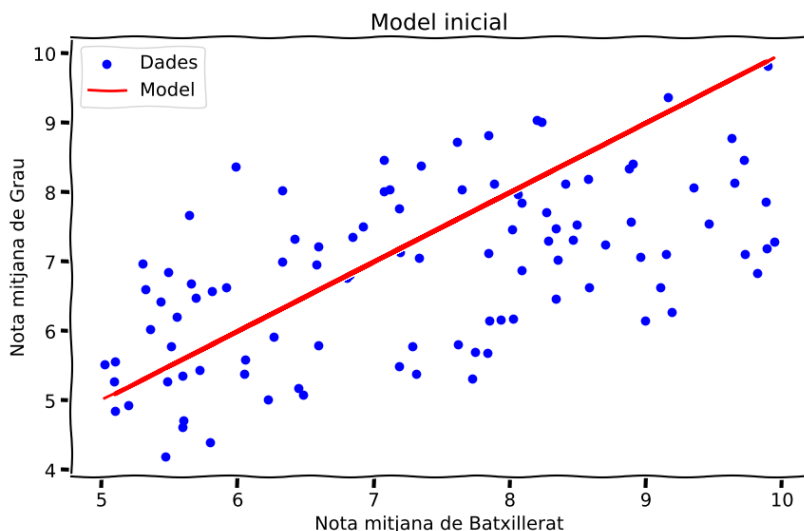


Figura 3. Exemple del resultat d'aplicar un model simple (sense cap mena de tècnica d'IA)

Suposem el cas d'un estudiant que va obtenir una mitjana de 7,7 al batxillerat i una nota final de 5,2 al grau universitari. Si el nostre model hagués estat utilitzat per predir la nota del grau, hauria estimat un 7,7. L'error de predicció en aquest cas hauria estat de +2,5 punts, el que representa una clara sobreestimació del rendiment real de l'estudiant.

En general, podem observar que el model sobreestima de manera notable la nota de la majoria dels estudiants (la qual cosa és normal, perquè els resultants mitjans d'un estudiant en la universitat són, en general, més baix que en el batxillerat) i la subestima en alguns casos. Encara més, només és capaç de predir la nota més o menys bé en molt pocs casos (els punts blaus que estan a prop de la línia vermella).

Si volem construir un model més precís, amb menys error de predicció, podem utilitzar eines d'IA, que ens permeten descobrir de forma automàtica com podem dissenyar el model Y a partir de la combinació de les característiques disponibles de cada estudiant.

Com millorar el model?

Encara dins el nostre exemple, el pas següent envers un model predictiu més precís podria ser estimar un model una mica més complex,³ en termes matemàtics, a partir de les dades.

És a dir, si denotem la nota mitjana del batxillerat com a x_1 , passar del model que teníem,

$$Y = x_1$$

a un model d'aquest tipus:

$$Y = ax_1 + b,$$

en què a i b representen dos paràmetres numèrics desconeguts. En aquest cas, la tasca de l'aprenentatge predictiu és determinar els valors a i b tals que minimitzen l'error de predicció del model.

Si apliquem un d'aquests algoritmes d'IA disponibles per a aquesta tasca a un conjunt de dades històriques (que anomenarem *conjunt d'aprenentatge*), obtenim el model de la figura 4. El procés que ens permet ajustar els valors a i b s'anomena *procés d'entrenament*.

Com funciona un procés d'entrenament?

El procés d'entrenament s'inicia dividint les dades d'aprenentatge en dos conjunts: conjunt d'entrenament i conjunt de prova. El conjunt d'entrenament s'utilitzarà per ajustar el model, mentre que el conjunt de prova s'emprarà per avaluar el seu rendiment.

El segon pas és triar un algoritme d'aprenentatge adequat per al problema. Això pot incloure models de regressió, arbres de decisió, xarxes neuronals, entre d'altres (Alpaydin, 2021).

Tot seguit, el model s'entrena fent servir el conjunt d'entrenament. L'algoritme ajusta els paràmetres interns per minimitzar l'error de predicció mitjançant tècniques com ara el descens de gradient.

Per acabar, el model final s'avalua amb dades que no ha vist abans, per assegurar que no ha sobreajustat i que generalitza correctament dades noves.

3. A fi de no complicar la comprensió del rol dels models en IA, en l'explicació farem servir models matemàtics força simples, tot i que, en la realitat, aquests models serien molt més complexos.

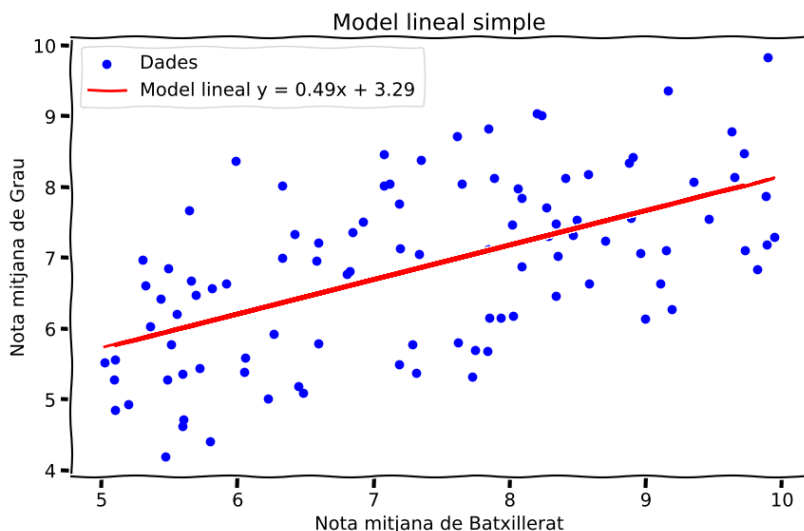


Figura 4. Model lineal simple obtingut amb tècniques d'IA predictiva

El mètode d'aprenentatge predictiu ha determinat, a partir de l'anàlisi de les dades històriques disponibles, que si $a = 0,49$ i $b = 3,29$, obtenim un model una mica millor, representat per la línia vermella. Aquest model és millor que l'anterior, ja que no sobreestima ni subestima de forma sistemàtica, però l'error global que comet és, de totes maneres, molt gran (la majoria dels punts blaus estan lluny de la línia vermella).

Per disminuir encara més l'error, podem fer servir la resta de les característiques disponibles⁴ (com ara el nombre d'activitats extraescolars) i no limitar-nos a la nota de batxillerat. Si tenim n característiques diferents, podem representar cada estudiant i per aquest conjunt de característiques: $(x_1^i, x_2^i, \dots, x_n^i)$.

En el nostre exemple, si fem servir totes les característiques disponibles de cada estudiant per construir un model més complex,

$$Y = a_1 x_1 + \dots + a_n x_n + b$$

podríem obtenir un model amb uns errors com els que es mostren a la figura 5.

4. Això ens obliga a «complicar» el model matemàtic que utilitzem.

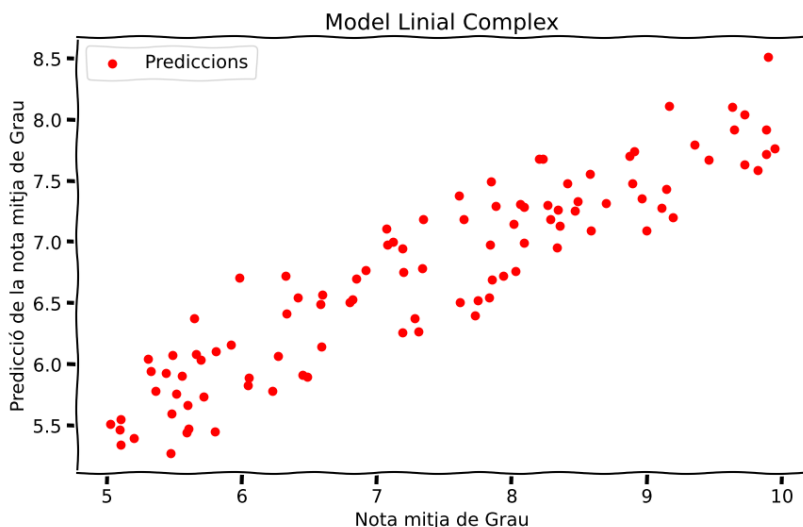


Figura 5. Error per a cadascuna de les dades d'un model més complex

En aquesta gràfica podem apreciar la relació entre la nota de grau real i la que ha predit el model complex.⁵ El nou model és molt millor que els anteriors des de tots els punts de vista: no sobreestima ni subestima, i l'error global és petit.

Finalment, si en comptes d'un model lineal fem servir l'aprenentatge automàtic per generar un model molt més flexible,⁶ amb més capacitat per modelitzar les relacions entre les característiques, podríem arribar a models amb un error de predicció força petit, com el de la figura 6.

L'obtenció de models predictius precisos que es nodreixen de grans volums de dades depèn, en gran manera, de la pertinència de les característiques seleccionades per abordar el problema específic. La qualitat de les dades té un paper essencial en l'eficàcia d'aquests models. Fins i tot els algoritmes d'aprenentatge més avançats poden produir resultats

5. En aquest gràfic, un model predictiu perfecte generaria una línia vermella diagonal.

6. No és objectiu d'aquest llibre fer un repàs dels algoritmes d'IA que es poden fer servir per a aquest fi, i que van des dels arbres de decisió fins a les xarxes neuronals, ja que els problemes que volem discutir sobre l'ús de models predictius en educació són independents de la forma concreta del model.

inexactes o esbiaixats si les dades no són d'alta qualitat o si les característiques no estan seleccionades adequadament.

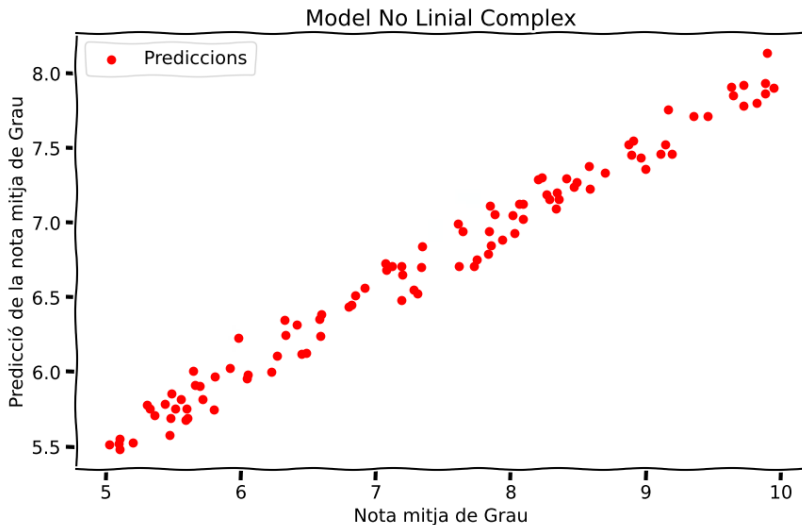


Figura 6. Error per a cadascuna de les dades d'un model no lineal complex

Del model a l'ús

Un cop disposem d'un model específicament desenvolupat per a aquesta tasca, el camí fins als possibles usos ja és directe.

A l'inici del curs següent, la universitat pot construir una aplicació informàtica tal que, ateses les dades d'entrada X dels nous alumnes, faci la predicció de Y . I pot posar aquesta aplicació a disposició dels membres de les seves facultats.

Fins a aquest punt, el model que hem construït encara no té un ús definit: serà cadascuna de les facultats la que el determinarà. Per exemple, una facultat (ús 1) pot aplicar aquest model a un estudiant que acaba d'ingressar-hi i predir la nota mitjana que traurà al cap de quatre anys per determinar la seva idoneïtat per rebre una beca de suport a l'estudi. Altres poden aplicar-lo a tota la cohort d'entrada (ús 2) i determinar quins recursos docents caldran per mantenir un nivell de coneixement determinat a la sortida dels estudis.

Com hem vist en aquest exemple, no hi ha una correspondència unívoca entre la tasca que s'ha fet servir per entrenar el model predictiu i els seus usos finals. Un mateix model pot ser utilitzat per a múltiples usos, uns de més justificats i d'altres menys. És en aquest punt on queda palesa la necessitat d'una anàlisi de l'ús dels models predictius.

3. COM FUNCIONA LA INTEL·LIGÈNCIA ARTIFICIAL GENERATIVA?

Així com en el cas de la IA predictiva és fàcil per a qualsevol persona forjar-se una idea intuïtiva de la seva funcionalitat (atesa una dada d'entrada X , predir el valor de sortida Y més probable en funció de les dades amb què ha estat entrenat), el cas de la IA generativa és un xic més complex, per l'aparent manca de correspondència entre la tasca amb què s'entrena i els usos que se li adjudiquen en el món real. De fet, el més curiós és que la IA generativa basada en text no és sinó un cas particular d'IA predictiva, per bé que els seus usos són molt diferents.

La paradoxa de la IA generativa rau en la relació entre el seu funcionament intern i els usos que se li assignen en món real. Si bé pot semblar que la IA generativa fa tasques del tot diferents de les de la IA predictiva, en essència, totes dues es basen en principis similars de predicció.

Com ja hem vist, la IA predictiva és relativament fàcil d'entendre: donat un conjunt de dades d'entrada X (per exemple, l'historial d'un pacient), el sistema prediu el valor de sortida Y (com seria la probabilitat de desenvolupar una malaltia). Aquest enfocament reflecteix una relació directa entre l'entrenament del model (aprendre patrons a les dades) i la seva aplicació (predir valors futurs).

En el cas de la IA generativa, com són els models de generació de text, les tasques que du a terme semblen menys directament connectades amb el procés d'entrenament. Per exemple, un model com GPT s'entrena per predir la paraula que segueix en una seqüència de text donada. No obstant això, en el món real, s'utilitza per generar assajos, escriure poesia, programar o respondre preguntes complexes. Això pot semblar desconnectat, perquè el model no va ser explícitament entrenat per «escriure assaigs» o «resoldre problemes de programació».

La paradoxa es resol en comprendre, com apuntàvem, que la IA generativa basada en text no és res més que un cas particular d'IA predictiva. Durant el seu entrenament, el model aprèn a predir la paraula o el fragment de text més probable en funció del context proporcionat (el text previ). Aquesta tasca predictiva bàsica, repetida a gran escala, permet

al model construir seqüències completes que semblen creades intencionalment per a tasques específiques.

Per exemple:

- En escriure poesia, el model simplement prediu paraules que siguin coherents amb l'estil líric i el ritme de les paraules prèvies.
- En generar codi, el model prediu fragments de programes vàlids i contextualment rellevants en funció del llenguatge i de les instruccions donades.

L'aparent salt entre l'entrenament (predir la paraula que segueix) i els usos reals (escriure assaigs o resoldre problemes) s'explica tenint al cap que la tasca de predir paraules encapsula una riquesa implícita de coneixement sobre el llenguatge, el context i el significat.

La relació entre les dades d'entrada i la sortida en un model d'IA generativa no s'especifica en forma de regles clares i llegibles per als éssers humans. En comptes d'això, el model aprèn aquestes relacions complexes a través de patrons en les dades durant l'entrenament, utilitzant tècniques avançades basades en xarxes neuronals profundes (*transformers*).

Això permet que la IA generativa produeixi text, imatges, música i altres classes de dades que semblen creades per persones, però amb una lògica subjacent de generació que pot costar d'interpretar.

Comprendre a fons la connexió entre la tasca simple per a la qual s'han entrenat aquests models i la seva elevada competència en altres tasques complexes no és el propòsit d'aquest llibre. Amb tot, cal desenvolupar una idea intuïtiva sobre els models generatius que ens permeti avaluar-ne la utilitat per resoldre determinades tasques i, sobretot, per preveure'n els riscos. Amb aquesta finalitat, farem servir una analogia proposada per Léon Bottou i Bernhard Schölkopf basada en un relat curt de J. L. Borges titulat *El jardí dels senders que es bifurquen* (Bottou et al., 2023).

En el relat es descriu una biblioteca que no només conté tots els textos produïts pels humans, sinó que, molt més enllà del que ja s'ha escrit en la realitat, abasta també tots els textos que un ésser humà podria llegir i comprendre, si més no, superficialment. Aquesta col·lecció de textos «plausibles» podria contenir llibres, diàlegs, articles, oracions, pàgines web, programes d'ordinador... en qualsevol idioma, en qualsevol forma.

Ara imaginem un sistema d'IA que és capaç de memoritzar tots aquests textos i de generar, a partir d'una seqüència de paraules concreta, la llista de paraules que la segueixen en algun llibre de la biblioteca, ordenades per la seva freqüència.

Si apliquem aquest procés de forma recursiva, la seqüència de paraules generada per la màquina a partir d'una seqüència d'entrada es troba en algun lloc de la nostra col·lecció infinita de tots els textos plausibles i, per consegüent, forma part d'un dels llibres de la nostra biblioteca.

Respostes plausibles i respostes correctes

Els usos i els riscos de la IA generativa es desprenen d'una simple observació basada en l'analogia entre els grans models de llenguatge i la biblioteca de Borges: els textos que un individu podria llegir i, com a mínim, comprendre de manera superficial no només estan formats per relats relacionats amb fets «reals», sinó que qualsevol idea fantasiosa que qualsevol humà pugui entendre, o simplement qualsevol idea que sigui falsa però comprensible, formarà part de la biblioteca. La descripció del sistema solar basada en la mecànica de Newton formarà part de la biblioteca, però també les teories terraplanistes. A partir d'aquí és fàcil entendre que, davant el desconeixement del que és veritat i del que no ho és, el programa d'IA que genera el text només és capaç de generar text plausible, però no text correcte. En conseqüència, la sortida d'un model d'IA generativa sempre té una certa probabilitat de ser incorrecta.

Els grans models de llenguatge han estat entrenats amb dades textuais que no s'han revisat des del punt de vista de la seva realitat o factualitat. Des d'aquesta perspectiva, podem considerar que són dades que podrien formar part de la biblioteca de Borges. Per això, les dades produïdes pels grans models de llenguatge formen part de la biblioteca de Borges i no és possible determinar a priori si seran reals o falsos. Vist així, podem dir que les dades que produeix una IA generativa quan un usuari la interroga amb una frase en llenguatge natural són:

- Útils per resoldre alguna tasca, si són correctes i tenim alguna forma de verificar-ne la correcció.
- Inútils, si són incorrectes i ho sabem.
- Perilloses, si són incorrectes i no ho sabem.

Si pensem en la IA generativa d'aquesta manera i apliquem un principi de precaució, la decisió sobre l'adequació d'aquesta tecnologia a un ús determinat, maximitzant-ne la utilitat i minimitzant-ne el risc, es torna clara i senzilla. La IA generativa és una eina magnífica quan és utilitzada per un usuari expert que la pot emprar com una ajuda eficaç en la seva feina. Els experts disposen del coneixement necessari per avaluar i contextualitzar les respostes generades, cosa que els permet identificar possibles errors o inconsistències, així com espremer el potencial de la tecnologia per millorar-ne la productivitat i la creativitat.

Però la situació canvia dràsticament quan la IA generativa la fan servir usuaris que no tenen la capacitat per destriar la veracitat d'una resposta. En aquests casos, l'eina esdevé perillosa, ja que pot generar informació errònia o enganyosa que l'usuari podria prendre com a certa si no té l'habilitat de qüestionar-la o de verificar-la com cal. Això pot portar a la presa de decisions incorrectes, a la producció i difusió d'informació falsa, i a conseqüències negatives en diversos àmbits, des de l'educació i la investigació fins als mitjans de comunicació i les xarxes socials. Per tant, és crucial implementar mesures que assegurin l'ús responsable de la IA generativa, fomentant la capacitació dels usuaris i establint mecanismes de control i verificació de la informació.

Per acabar, és interessant adonar-se que la IA generativa proposa un canvi en el rol de les dades, tal com es mostra en la figura 7. Les dades i les seves etiquetes (produïdes per éssers humans i relacionades amb una tasca concreta) han d'estar disponibles per poder entrenar un model d'IA predictiva i, a partir d'aquell moment, el model produït és útil per respondre a preguntes relacionades amb una determinada tasca, i amb cap altra.

En el cas de la IA generativa, amb la mera existència de dades genèriques n'hi ha prou per entrenar un model, i és en el moment que fem una pregunta quan especifiquem la tasca que volem resoldre i quines dades calen per respondre. En la figura 8 podem veure com ChatGPT contesta una pregunta per a la qual no ha estat explícitament entrenat si especifiquem les dades d'entrada (anomenades *prompts*) i la tasca que volem resoldre.

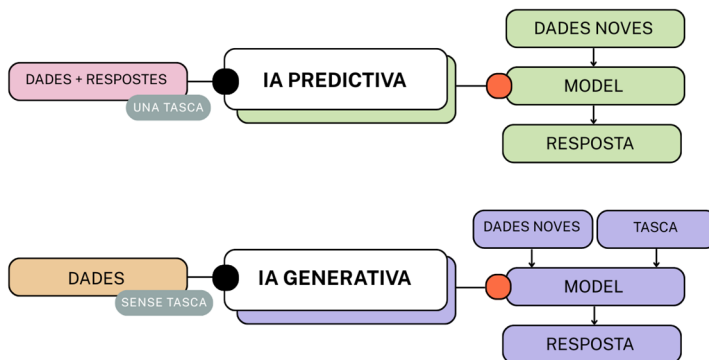


Figura 7. Relació entre dades, etiquetes i models en la IA predictiva i la IA generativa

Primera notícia d'avui: "Afectacions en la circulació de trens a la R1 a causa d'un atropellament entre Badalona i Montgat"

Segona notícia d'avui: "Un 81,51% dels estudiants preinscrits per accedir als estudis de grau en algun dels set centres públics catalans ja té plaça assignada"

Pots qualificar aquestes notícies en una escala de 0 a 10 pel que fa al nivell de preocupació que poden generar en la població?

Sí, podem establir una escala de preocupació del 0 al 10, on:

- 0 representa una notícia sense cap implicació preocupant,
- 10 representa una notícia d'altíssima gravetat o preocupació generalitzada.

1. "Afectacions en la circulació de trens a la R1 a causa d'un atropellament entre Badalona i Montgat"

Valoració de preocupació: 7/10

Raó:

- Té una afectació immediata i directa sobre un nombre significatiu de persones (usuaris del transport públic).
- L'atropellament implica un possible incident greu o fatal, que pot generar alarma o empatia.
- Afecta la mobilitat metropolitana, especialment en hora punta.
- És una preocupació **local i temporal**, però intensa.

2. "Un 81,51% dels estudiants preinscrits per accedir als estudis de grau en algun dels set centres públics catalans ja té plaça assignada"

Valoració de preocupació: 2/10

Raó:

- És una notícia **principalment informativa i positiva** per a la majoria.
- El 18,49% restant pot experimentar preocupació, però es tracta d'un procés amb més fases (assignacions posteriors, reclamacions).
- No genera una alarma generalitzada, però pot ser rellevant per a famílies i estudiants afectats.

Figura 8. ChatGPT contesta una pregunta (predir el nivell de preocupació que suscita una notícia entre els ciutadans) per a la qual no ha estat explícitament entrenat

Enginyeria de prompts

Un *prompt* és una instrucció o conjunt d'instruccions proporcionades a un sistema d'IA, com ara un model de generació de text o d'imatges, per guiar la sortida o la resposta del sistema. En el context de la IA generativa, un *prompt* pot ser una frase, una pregunta, una descripció o qualsevol mena d'entrada textual o visual que s'utilitza per sol·licitar una resposta específica del model.

Tot i que pugui semblar que, per aconseguir la resposta desitjada d'un sistema d'IA generativa, només calgui escriure una pregunta o alguna classe de *prompt*, en la pràctica no sempre és fàcil que l'usuari obtingui exactament el resultat que espera. En alguns casos, això pot representar la feina de dies o setmanes. Això ha ocasionat un augment en la demanda de treballs relacionats amb l'enginyeria de *prompts*, la qual es refereix als processos i tècniques per crear entrades que generin un resultat d'IA generativa que s'acosti el màxim possible a la intenció original de l'usuari.

Sobre els usos de la intel·ligència artificial generativa

La IA generativa gaudeix d'un ampli ventall d'aplicacions en l'educació, en permetre crear continguts, personalitzar experiències d'aprenentatge i donar suport tant a estudiants com a professors (Ribera *et al.*, 2024). Seguidament, es descriuen alguns dels usos més destacats de la IA generativa en l'àmbit educatiu:

Creació de continguts educatius:

- *Generació de materials d'estudi*: donat un context especificat en un *prompt*, pot ser útil per ajudar el professor en la creació de textos, imatges, resums, explicacions i materials didàctics, sempre que aquests materials siguin validats pel docent.
- *Ajuda en l'elaboració d'exercicis i problemes*: de la mateixa manera, pot ajudar en la proposició de problemes, qüestionaris i exercicis pràctics adaptats a diferents nivells de dificultat.

Desenvolupament d'entorns personalitzats d'aprenentatge:

- *Tutories virtuals*: es poden crear tutors virtuals que poden interactuar amb els estudiants, respondre preguntes i proporcionar explicacions addicionals en temps real. L'ús de la IA generativa per part de l'alumne ha d'estar molt ben estudiat i monitorat, atès que aquest no és capaç de discernir entre respostes correctes i respostes plausibles però incorrectes.

- *Simulacions i entorns d'aprenentatge immersius*: es poden generar simulacions realistes i entorns virtuals per a pràctiques en camps com ara la ciència, la medicina i l'enginyeria.

Suport en l'escriptura i creació de continguts per part dels estudiants:

- *Assistència en l'escriptura*: la IA generativa ens permet anar un pas més enllà del corrector ortogràfic per a l'assistència en l'escriptura i crear eines que ajuden l'alumnat a redactar assaigs, corregir errors gramaticals, millorar l'estil d'escriptura i proposar idees per als seus treballs. Aquest ús de la IA generativa sempre ha d'estar subjecte a una sèrie de normes relacionades amb la integritat acadèmica i el seu compliment efectiu.
- *Creació de materials per a projectes multimèdia*: és possible generar imatges, vídeos i presentacions a partir de descripcions textuais proporcionades pels estudiants.

Avaluació i *feedback*:

- *Generació de comentaris personalitzats*: permet l'anàlisi de la feina dels estudiants per proposar al professor comentaris detallats i personalitzats.
- *Creació d'exàmens i avaluacions*: ajuda a dissenyar exàmens i avaluacions que s'ajusten al nivell de coneixement de l'estudiant.

Suport a la inclusió i accessibilitat:

- *Conversió de text a veu i viceversa*: es genera contingut en diferents formats per a estudiants amb discapacitats visuals o auditives.
- *Traducció automàtica*: es tradueixen en temps real materials educatius per a estudiants que fan servir altres idiomes.

Creació de recursos per a professors:

- *Disseny de plans de classe*: es poden generar de manera automàtica plans de classe i materials de suport basats en els objectius educatius i el currículum.
- *Elaboració de recursos didàctics*: es poden crear diapositives, fulls de treball i guies d'estudi que els docents poden utilitzar a les classes.

Motivació i compromís dels estudiants:

- *Generació d'històries i narratives*: és possible la creació d'històries i continguts narratius que poden fer que l'aprenentatge sigui més atractiu i fàcil de memoritzar.
- *Jocs educatius i gamificació*: es poden desenvolupar jocs i activitats interactius que incentivin la participació activa de l'alumnat.

El cas de l'avaluació basada en IA

L'avaluació dels alumnes mitjançant IA és una de les aplicacions més prometedores en educació. Tot i això, no hi ha un consens general sobre la concepció pedagògica de l'avaluació amb IA. Aquest problema s'ha exacerbat en els darrers temps, amb l'aparició de la IA generativa, que ha resultat ser un element disruptor de moltes activitats d'avaluació utilitzades actualment.

El valor de la IA en els processos avaluatius rau en el suport als docents i als estudiants, en integrar millor l'avaluació en el procés d'aprenentatge i oferint una retroalimentació personalitzada, més justa i inclusiva. Aquest model d'avaluació, però, hauria d'estar fonamentat per un model pedagògic diferent, prioritzant la personalització per damunt dels models basats en qualificacions. Tampoc no podem menysprear la reducció de la burocràcia i del temps dedicat a la correcció per part del professorat ni el fet que ens obre la possibilitat de fer un seguiment gairebé continu dels estudiants que pot permetre al professorat conèixer els seus estils d'aprenentatge, identificar les dificultats i gestionar els suports necessaris.

Entre els riscos i les dificultats, hi ha la possibilitat que aquesta tecnologia generi manca de confiança en els resultats per part de les persones avaluades o les seves famílies, i també que es produeixi un empobriment si l'avaluació perd el factor humà que sempre té en consideració el context i la situació de l'alumne.

Aspectes sociotècnics de la intel·ligència artificial generativa

La IA generativa requereix enormes quantitats de dades i una gran potència computacional, i això sense comptar innovacions contínues en els mètodes d'entrenament. Aquestes capacitats, en la majoria dels casos, només estan disponibles per a les grans empreses tecnològiques internacionals i unes poques economies. Això implica que la capacitat de crear i controlar la IA generativa està fora de l'abast de la major part de les empreses i països, cosa que dona lloc a una sèrie de dubtes sobre certs aspectes sociopolítics relacionats amb el seu ús, que van des de la concentració de poder i coneixement en mans d'unes quantes enti-

tats fins a l'eixamplament de la bretxa digital entre estats (Narayanan, 2024).

Les empreses tecnològiques no només tenen la capacitat de desenvolupar sistemes d'IA avançada, sinó també els recursos necessaris per fer-ho a una escala que la major part de la comunitat investigadora no pot igualar. Això inclou l'accés a infraestructures computacionals potents, enormes volums de dades i talent especialitzat. Aquesta concentració de poder, però, planteja uns quants desafiaments significatius. Per una part, aquestes empreses tenen una influència directa en la direcció que segueix el desenvolupament de la IA i en les aplicacions que es prioritzen, fet que pot conduir a un monopoli de la innovació. En aquest context, les decisions sobre quines àrees s'han d'explorar i quines tecnologies s'han d'implementar tendeixen a estar condicionades per objectius comercials o interessos específics, passant per alt les necessitats de comunitats menys representades o les regions amb recursos limitats.

Per altra part, aquesta dinàmica pot perpetuar desigualtats existents, ja que els beneficis derivats de la IA generativa –com són la creació d'eines avançades, millores en la productivitat o solucions personalitzades– es dissenyen des de perspectives limitades. Això, a banda de limitar la seva utilitat a contextos diversos, amplia la bretxa entre els qui tenen accés a aquestes tecnologies i els qui no.

Un altre aspecte crític té relació amb la privadesa i l'ètica. El desenvolupament de sistemes d'IA generativa acostuma a requerir la recopilació i el processament de grans quantitats de dades, moltes de les quals són personals i sensibles. La concentració d'aquesta informació en mans d'unes poques empreses posa sobre la taula riscos evidents: des de possibles abusos de poder en forma de vigilància massiva fins a la manipulació d'informació amb finalitats polítiques o comercials.

En aquest escenari, és indispensable establir marcs reguladors que fomentin un equilibri entre innovació tecnològica i responsabilitat social, promovent l'accés equitatiu a la tecnologia i assegurant que el desenvolupament respecti els principis ètics fonamentals. En cas contrari, hi ha el risc que la IA, en lloc de democratitzar les oportunitats, esdevingui un catalitzador de noves desigualtats i tensions globals.

Aquestes reflexions també han d'acompanyar l'ús de la IA generativa en el camp de l'educació. Els docents i els estudiants han de ser capaços de tenir una visió crítica respecte als valors, les normes culturals i els costums socials que estan integrats en els models d'entrenament de la IA generativa. És crucial que es qüestionin i analitzin a fons aquests assumptes, ja que poden influir en els resultats generats per la IA i reforçar biaixos i prejudicis. Per exemple, si les dades utilitzades per entrenar aquests models provenen predominantment d'una cultura específica, és probable que els resultats facin prevaldre valors i perspectives que no són universals, cosa que podria portar a l'exclusió o a la representació incorrecta d'altres cultures i contextos.

Un altre punt que convé tenir present és que, tal com hem esmentat, els models d'IA generativa es desenvolupen a partir de quantitats ingents de dades, sovint obtingudes d'internet i generalment sense el permís dels propietaris dels continguts esmentats. Com a resultat, aquests sistemes poden haver violat drets de propietat intel·lectual, i per aquest motiu ja hi ha diverses demandes internacionals en curs relacionades amb aquest tema. Els docents i els estudiants han de ser conscients que les imatges o codis generats amb IA generativa poden infringir drets de propietat intel·lectual de tercers i que les imatges, sons o codis que creen i comparteixen a internet poden ser utilitzats per les companyies que desenvolupen aquesta tecnologia sense el seu permís.

Un exemple clar d'aquesta problemàtica és el cas d'eines com ara Stable Diffusion o DALL-E, que són models de generació d'imatges entrenats amb enormes conjunts de dades recopilades d'internet. Aquests conjunts inclouen imatges disponibles públicament, moltes de les quals estan protegides per drets d'autor, com ara obres d'artistes o fotografies comercials. Tanmateix, durant l'entrenament, aquestes imatges són utilitzades sense el consentiment explícit dels propietaris dels drets. Un altre cas concret és el de GitHub Copilot, una eina de generació de codi basada en IA entrenada amb repositoris públics de codi de plataformes com ara GitHub. Tot i que l'eina és útil per a programadors, en alguns casos pot generar fragments de codi gairebé idèntics als originals amb què va ser entrenada, una circumstància que fa inevitable preocupar-se per eventuais violacions de llicències de codi obert.

4. COM AVALUEM L'ÚS D'UN SISTEMA D'INTEL·LIGÈNCIA ARTIFICIAL?

En capítols anteriors hem vist algunes de les capacitats dels sistemes d'intel·ligència artificial (IA), però també algunes de les seves limitacions i riscos (escassa precisió, manca de veracitat en les respostes, etc.). Fer un ús responsable de la IA en educació, vetllant perquè els beneficis de la IA es maximitzin alhora que es minimitzen els riscos i els possibles danys, és essencial per assegurar que aquestes tecnologies milloren l'ensenyament i l'aprenentatge, respectant sempre els drets i el benestar de totes les persones involucrades.

L'ús responsable de la IA en l'educació no es garanteix simplement seguint un seguit de recomanacions genèriques preestablertes. És un procés que ha d'involucrar diversos actors i ser portat a terme en cada aula, escola o institució educativa quan s'hi introdueixen noves aplicacions. Dit d'altra manera, es tracta més d'un procés per trobar respostes a un conjunt de preguntes específiques que d'un mer procés consistent a completar una llista de verificació predefinida. Amb una metodologia adequada, la implementació responsable de la IA en l'educació pot millorar els resultats d'aprenentatge, enfortir el rol dels docents i promoure l'equitat educativa. Sense aquesta metodologia, els riscos poden ser significatius.

En aquest capítol es proposa una metodologia que es pot adaptar a un munt de casos. Cadascun dels passos d'aquesta metodologia vol respondre una pregunta, tal com es pot apreciar en la figura 9.

En aquest llibre es proposa un procés simple que comprèn quatre passos: la definició del problema educatiu que volem tractar, l'anàlisi del marc legislatiu i social en què es produeix aquesta aplicació, l'anàlisi dels riscos associats i, finalment, la determinació de la viabilitat de la proposta en un context donat.



Figura 9. Procés de validació per assegurar l'ús responsable de la IA en educació

Pas 1: Definir el problema i el valor que aporta la intel·ligència artificial

El primer pas del procés és definir de manera clara i transparent allò que es pretén assolir amb la IA: quins són els objectius que es volen assolir, mitjançant quin procés i quins són els resultats esperats.

Per exemple, si l'objectiu és millorar l'eficiència administrativa de les tasques del professor per mitjà d'una eina de suport al disseny d'activitats durant el *procés* de creació de materials formatius, el *resultat* esperat podria definir-se com una reducció efectiva del temps en tasques administratives, mantenint o millorant la qualitat del resultat de la tasca. Amb vista a fases posteriors d'aquest procés, és interessant expressar el resultat esperat en uns termes que es relacionin amb el valor de l'ús de la IA, sigui per la capacitat per personalitzar l'aprenentatge, per optimitzar processos, per millorar la presa de decisions o per fomentar la inclusió o promoure la innovació.

També és cabdal tenir en compte que, atès l'estat actual de la IA en l'educació i les moltes incògnites que sorgeixen de la seva aplicació, és útil conceptualitzar l'ús d'una eina d'IA en un procés educatiu com una intervenció amb un impacte esperat, que demana validació sota condicions experimentals adequades. Visualitzar l'impacte esperat i pensar detalladament com es mesurarà aquest impacte de la manera més objectiva possible és central, si volem disposar d'evidències que justifiquin l'ús futur de l'eina.

Amb tot, definint clarament l'objectiu i el resultat esperat no n'hi ha prou. En passos posteriors caldrà comprovar-ne la viabilitat en un context determinat. En conseqüència, és interessant, en aquest pas, mapejar els processos involucrats, les persones que tenen algun paper en el procés i identificar els punts en què la IA pot afegir valor.

En el nostre exemple caldria identificar les subtasques que fa el professor i identificar aquelles en què un sistema d'IA predictiva o generativa poden aportar valor: en la planificació de lliçons?, en l'avaluació?, en la comunicació amb els estudiants?

També s'han de comunicar aquests objectius i resultats a totes les parts interessades, tot assegurant que hi hagi una comprensió compartida

i un compromís amb la visió del projecte. Això ajudarà a alinear els esforços i a facilitar la col·laboració entre les diferents persones i els equips implicats.

Per acabar, cal establir un pla de revisió i ajustament continu. La tecnologia i les necessitats organitzatives canvien amb el temps, per la qual cosa és crucial revisar periòdicament l'exercici de la IA i efectuar els ajustaments necessaris per assegurar que es continuen complint els objectius establerts i que s'hi aporta valor.

Pas 2: Comprendre'n el marc legislatiu i social

El primer element que s'ha de considerar després de definir un ús de la IA en educació és analitzar el marc legislatiu i els límits que aquest ens imposa. Alguns dels usos de la IA en educació han estat definits pel Parlament Europeu com a usos d'alt risc (EUR Lex, 2024):

La implementació de sistemes d'IA en l'educació és important per ajudar a modernitzar completament els sistemes educatius, augmentar la qualitat de l'educació, tant en línia com fora de línia, i accelerar l'educació digital, posant-la, així, també a disposició d'un públic més ampli. S'han de classificar com d'alt risc els sistemes d'IA que s'utilitzen en l'educació o la formació professional, i, en especial, els que determinen l'accés o influeixen substancialment en les decisions sobre admissió o distribueixen les persones entre diferents institucions educatives i de formació professional, o els que avaluen les persones a partir de proves fetes en el marc de la seva educació o com a condició necessària per accedir-hi o per avaluar el nivell apropiat d'educació d'una persona i influir substancialment en el nivell d'educació i formació que rebran les persones o al qual podran accedir, i supervisar i detectar comportaments prohibits d'una persona i, en conseqüència, afectar la seva capacitat per assegurar la seva subsistència. Quan no es dissenyen i utilitzen correctament, aquests sistemes poden envair especialment i violar el dret a l'educació i a la formació, i el dret a no patir discriminació, a més de perpetuar patrons històrics de discriminació, per exemple, contra les dones, els grups d'una edat determinada, les persones amb discapacitat o les perso-

nes amb un determinat origen racial o ètnic o amb una determinada orientació sexual. (*AI Act, 2024*)

La implementació de sistemes d'IA en l'educació es considera capital per ajudar a modernitzar els sistemes educatius, augmentar la qualitat de l'educació i accelerar l'educació digital, posant-la, així, a disposició d'un públic més ampli. Però, quan no es dissenyen ni s'utilitzen correctament, aquests sistemes poden envair greument i violar el dret a l'educació i la formació, i el dret a no patir discriminació, a més de perpetuar patrons històrics de discriminació, per exemple, contra les dones, els grups d'una certa edat, les persones amb discapacitat o les persones d'un origen racial o ètnic determinat o amb una orientació sexual concreta.

Actualment, la Llei se centra específicament en l'ús de la IA per prendre decisions en l'educació que impacten en la participació de l'alumnat en els programes. No aborda directament l'ús de la IA en l'aprenentatge. No obstant això, aquests aspectes no tenen unes fronteres ben definides i no es pot descartar que en un futur pròxim els criteris s'ampliïn o es modifiquin.

Així mateix, ens hem d'atenir a les obligacions que ens imposa el marc legislatiu. En cas que l'ús previst d'un sistema d'IA entri en la categoria d'alt risc, les organitzacions responsables hauran de disposar d'un sistema de gestió de la qualitat específic que asseguri els punts següents:

1. L'existència d'un sistema de governança de les dades.
2. Un nivell elevat de transparència del sistema d'IA.
3. Uns nivells elevats de robustesa i precisió en els models d'IA usats.
4. Uns nivells adequats de ciberseguretat que protegeixin els sistemes d'IA.
5. L'existència d'un sistema de gestió de riscos.
6. L'existència d'una documentació tècnica que reflecteixi el desenvolupament i el manteniment del sistema d'IA.
7. L'existència de manteniment de registres i traces del funcionament del sistema d'IA que en permeti l'anàlisi.
8. Un nivell adequat de supervisió humana en la presa de decisions.

La presència d'aquest sistema de gestió de la qualitat només és el primer element, ja que, abans de desplegar el sistema, l'organització res-

ponsable haurà de demostrar la conformitat de la seva aplicació davant les autoritats pertinents. Això vol dir que:

- Una organització o entitat acreditada certifiqui que el sistema d'IA compleix els principis ètics i els estàndards tècnics establerts per assegurar la confiabilitat de la IA.
- El sistema d'IA ha estat registrat en una base de dades centralitzada de la Unió Europea, cosa que facilita la supervisió i el compliment de les normatives.
- El sistema ha rebut la marca CE, que assegura que es compleixen les regulacions europees aplicables.
- S'ha fet una avaluació sobre com el sistema d'IA podria afectar els drets fonamentals de les persones, assegurant que no es violin drets com ara la privadesa, la no discriminació i altres drets humans essencials.

Cal destacar que el sistema de gestió de la qualitat no només afecta el desenvolupament del sistema, sinó que s'aplica en tot el seu cicle de vida.

En cas que l'ús previst d'un sistema d'IA no entri en la categoria d'alt risc, les obligacions són menors, però no menyspreables. L'organització responsable no haurà de demostrar la conformitat de la seva aplicació davant les autoritats pertinents abans del seu desplegament, però haurà de ser capaç de passar una auditoria en qualsevol moment del cicle de vida de l'aplicació. Aquesta auditoria se centrarà molt especialment en el vessant de la transparència.

El marc legislatiu no és l'únic punt en què ens hem de fixar quan analitzem el marc d'aplicació de la IA. Hi ha una segona qüestió, centrada en els aspectes socials i que és previ al marc legislatiu, que s'ha de respondre: és la IA el mètode més adient, des d'un punt de vista social, per assolir els objectius educatius proposats?⁷

En l'àmbit educatiu, la IA s'ha d'utilitzar per donar suport i enriquir l'experiència d'aprenentatge dels alumnes, promoure el seu benestar i el dels docents, i millorar els processos administratius necessaris per a

7. Aquesta pregunta adquireix la seva dimensió real si pensem en el problema dels exàmens *A-level* al Regne Unit, ja que va ser un problema d'aquesta classe el que va obligar a fer marxa enrere el Govern britànic.

aquesta finalitat. Però, més enllà d'aquestes consideracions genèriques, cada cas concret ha d'estar ben justificat. Reprenent el cas dels exàmens *A-level* del Regne Unit: era l'ús d'un sistema predictiu com el que s'ha descrit el mètode més adequat per assignar els estudiants a les universitats sense traïr els principis que han de regir el sistema educatiu?

Per respondre aquesta pregunta, podem fer ús d'un concepte essencial en qualsevol àmbit relacionat amb el bé comú: la *legitimitat*. La legitimitat és el criteri que garanteix que una institució, persona o sistema té dret a exercir el poder que se li ha atorgat d'acord amb les lleis vigents o els principis morals socialment acceptats. La reacció dels estudiants al Regne Unit va ser un cas clar de discussió de la legitimitat de l'ús de la IA per part del Govern per a un tipus determinat de decisió.

En general, podem considerar que l'ús de processos de presa de decisions automatitzats en un escenari específic és legítim si compleix diversos criteris en dos plans diferents (Barocas, 2023):

Els resultats de les decisions són:

- *Precisos*: el sistema ofereix resultats que són correctes en la majoria dels casos o molt propers al resultat ideal.
- *Fiabls*: el sistema ofereix resultats estables i coherents en diferents escenaris possibles.
- *Efícaços*: el sistema ofereix resultats que generen els efectes esperats en el món real.
- *Compatibles amb els valors* dels individus afectats per les decisions.

El procés de presa de decisions està:

- *Ben executat, de forma coherent i correcta*: quan la presa de decisions és arbitrària en aquest sentit del terme, els individus senten que estan subjectes a diferents esquemes de presa de decisions i reben decisions distintes simplement perquè passen pel procés de presa de decisions en moments diferents. Aquest principi es basa en la creença que les persones tenen dret a decisions similars, llevat que hi hagi raons per ser tractats de manera diferent.
- *Ben justificada*: sobre la base d'uns principis compartits i racionals. En cas contrari, direm que la presa de decisions és arbitrària. Les societats modernes no admeten la presa de decisions arbitràries sobre

temes importants per a la vida de les persones, ja que això posa de manifest una manca de respecte per les persones que les reben.⁸

Els anys anteriors a la COVID-19, la decisió sobre l'accés en l'educació superior al Regne Unit s'implementava mitjançant un procés burocràtic, que incloïa un examen i una avaluació manual, que estava legítim des del punt de vista de la ciutadania. Però l'ús de l'algoritme d'IA no va ser percebut com a legítim per la mateixa ciutadania. L'origen de l'absència de legitimitat de l'algoritme era la *manca de precisió* (un nivell d'error molt elevat en les prediccions), agreujat per un biaix negatiu respecte als estudiants de nivell socioeconòmic baix, que va ser percebut com a injustificat i, per tant, *arbitrari*.

Pas 3: Identificar-ne els riscos

La necessitat d'un pas centrat en l'anàlisi de les implicacions ètiques de la IA en l'educació i els riscos associats es fa evident si es du a terme una simple cerca bibliogràfica (Holmes, 2022). Quan s'exploren les notícies o la literatura actual, s'observa una preocupació creixent per com la IA està transformant l'àmbit educatiu. Aquestes transformacions presenten una sèrie de desafiaments ètics que no poden ser ignorats. Per exemple, si bé és cert que l'ús d'algoritmes per personalitzar l'aprenentatge pot millorar l'experiència educativa de l'alumnat, també planteja qüestions al voltant de la privadesa de les dades, el biaix algorítmic i l'equitat en l'accés a aquestes tecnologies.

Així mateix, la integració de la IA en l'educació pot alterar la relació entre docents i estudiants, així com els mètodes pedagògics tradicionals. Les decisions automatitzades que afecten l'avaluació i el progrés dels alumnes requereixen una supervisió ètica rigorosa a fi d'evitar injustícies i assegurar que els processos siguin transparents i equitatius. També és essencial considerar com la dependència de la IA podria influir en

8. Quan la presa de decisions està relacionada amb qüestions que tenen poques conseqüències en la vida de les persones, es poden acceptar processos de presa de decisions arbitràries. Per exemple, decidir mitjançant un sorteig quin serà l'ordre d'intervenció dels alumnes en un examen oral grupal és un procés de decisió arbitrari que pot ser acceptat socialment.

el desenvolupament d'habilitats crítiques i el pensament independent de l'alumnat.

Sense ànim de ser exhaustius, els riscos induïts per un determinat ús de la IA en educació estan relacionats amb les preguntes següents:

Privadesa i seguretat de les dades

- Com es protegeix la informació personal dels estudiants i se n'evita l'accés no permès?
- Es manegen de manera segura i ètica les dades personals recollides? Com s'assegura que la informació personal dels estudiants es fa servir només per als usos en què és imprescindible?

Biaix algorítmic

- Quins factors rellevants, com ara el gènere, la raça o la cultura, s'han de tenir en compte quan pensem en possibles problemes de discriminació?
- S'han avaluat els algoritmes que volem fer servir des del punt de vista de la discriminació algorítmica?
- Quines estratègies preventives i de mitigació es poden engegar?

Equitat en l'accés

- Hi ha desigualtats en l'accés a les eines que farem servir?
- Com s'aborda la bretxa digital entre diferents grups socioeconòmics?
- Com s'assegura que les persones amb discapacitats també se'n beneficiïn?
- Són les tecnologies d'IA inclusives i accessibles per a tots els estudiants?

Transparència en decisions automatitzades

- Si l'algoritme serveix com a sistema d'ajuda en la presa de decisions, es comprèn amb claredat quins criteris fa servir per justificar les decisions que proposa?
- Podem assegurar que tots els processos de decisió estan subjectes a la supervisió humana?

Impacte en la relació docent-estudiant

- Com afecten als objectius educatius els canvis introduïts per la IA en la dinàmica d'interacció i comunicació dins l'aula?
- Com es redefineix el paper del docent en un entorn en què la IA té un paper important?

Mètodes pedagògics tradicionals

- Com s'adapten els mètodes d'ensenyament existents a aquesta nova tecnologia?
- Es preserven i potencien les habilitats crítiques i el pensament independent?
- Hi ha una potencial disminució de les habilitats socials?
- Com es fomenta un aprenentatge equilibrat entre tecnologia i habilitats humanes?

Avaluació i progrés dels estudiants

- Són fiables i equitatives les avaluacions automatitzades?
- Quin és l'impacte de l'avaluació automatitzada en la motivació i el desenvolupament dels estudiants?

Responsabilitat i rendició de comptes

- Com s'educa l'alumne sobre els aspectes de plagi i de deshonestat acadèmica?
- Qui es responsabilitza de les decisions proposades per la IA a l'alumne?
- S'han establert i comunicat a tots els professors els marcs legals i ètics per a la implementació d'IA a l'organització?

Capacitació i formació d'usuaris

- És adequada la formació perquè els educadors utilitzin la IA de manera efectiva i ètica?
- S'estan desenvolupant competències digitals en estudiants i docents?

Algunes d'aquestes implicacions ètiques es poden resoldre amb solucions tecnològiques (per exemple, proporcionant un entorn segur que previngui la fugida de dades personals o els ciberatacs que manipulin els sistemes). D'altres, però, requereixen una anàlisi ètica per part del docent i l'organització.

Pas 4: Avaluar-ne la viabilitat

Un cop identificat el valor de l'ús de la IA, la seva seguretat jurídica i social, la seva alineació amb les persones involucrades i també els seus riscos potencials, cal avaluar la viabilitat de la proposta tant des d'un punt de vista pràctic com ètic, i prendre una decisió final.

La presa de decisions és el procés cognitiu que permet seleccionar un curs d'acció particular entre múltiples alternatives. Aquest procés comporta avaluar els pros i els contres de diferents opcions, considerar els possibles resultats i, tot seguit, triar una opció en particular.

Els éssers humans estem inherentment subjectes a una àmplia gamma de biaixos cognitius que poden ennuvolar el nostre judici i, per aquest motiu, s'han desenvolupat una sèrie d'eines i enfocaments estructurats que poden ajudar a generar una perspectiva més objectiva a través de la qual avaluar una decisió, minimitzant-ne els possibles biaixos. L'ús d'aquestes eines estandarditzades pot conduir a decisions més consistents, millorant la qualitat dels resultats amb el temps.

La millor eina o metodologia per a la presa de decisions és aquella amb la qual estem familiaritzats i que sabem fer servir: anàlisi DAFO, matrius de decisió, etc., però, per il·lustrar aquest pas, farem servir un model de Canvas específicament dissenyat per a aquesta tasca: l'Edu-IA Canvas. Els models de Canvas són eines visuals que es fan servir per dissenyar, descriure, analitzar i gestionar models de negoci⁹ i altres classes de decisions estratègiques. El terme Canvas (que en anglès significa 'llenç') al·ludeix a un quadre o plantilla que s'utilitza per estructurar i organitzar idees i conceptes de manera clara i concisa.

L'ús d'un model de Canvas segueix aquests passos:

1. *Selecció de la classe de Canvas*: triar el model de Canvas que més bé s'ajusti a les teves necessitats (per exemple, Business Model Canvas per a un negoci complet, Lean Canvas per a empreses emergents, etc.). En el nostre cas, definirem un nou model de Canvas adaptat a la nostra problemàtica.

9. Un dels models de Canvas més coneguts és el Business Model Canvas, creat per Alexander Osterwalder, que ajuda les empreses a delinear tots els aspectes clau del seu model de negoci.

2. *Treball en equip*: involucrar les parts interessades rellevants, com ara membres de l'equip docent, responsables de l'organització o familiars dels alumnes.
3. *Materials necessaris*: preparar un llenç gran o una pissarra, *post-it* i marcadors. També pots fer servir eines digitals com ara Miro, Lucidchart o Canva per crear Canvas virtuals.
4. *Emplenament dels blocs*: el model de Canvas es compon de diversos blocs o seccions que s'han d'emplenar amb informació clau durant un procés deliberatiu de l'equip.
5. *Revisió i validació*: has de revisar cada bloc amb el teu equip i ajustar la informació com calgui.

El model de Canvas EduIA

Generalment, un Canvas es divideix en blocs o seccions que representen diferents aspectes clau de la problemàtica que s'ha de tractar.

El model de Canvas que proposem comença amb un bloc simple que ens permet classificar el projecte docent en una sèrie de classes predefinida segons el seu objectiu genèric: millorar els resultats d'aprenentatge dels alumnes, enfortir el rol dels docents, promoure l'accés o l'equitat en l'educació, o reduir la càrrega burocràtica per concentrar-se en activitats de valor afegit.

EduIA Canvas			
Objectiu La IA ha de ser utilitzada per assolir objectius educatius ben definits, basats en una evidència social, educativa o científica vàlida que demostrï que això beneficia l'estudiant. S'ha identificat clarament l'objectiu educatiu que es pretén assolir mitjançant l'ús de la IA?		Valor S'ha identificat un recurs d'IA alineat amb els objectius educatius? Es pot explicar per què aquest recurs en concret té la capacitat d'assolir l'objectiu educatiu específic anteriorment? De quina informació es disposa per afirmar que l'ús de la IA proposada està recolzat per evidència social, educativa o científica?	
Dades Quines dades són necessàries per assolir el recurs d'IA? És viable disposar d'aquestes dades? Quina és la qualitat de les dades disponibles?	Processos S'han identificat els processos educatius en què s'utilitzarà aquest recurs d'IA? Com s'utilitzarà la IA per millorar aquests processos? S'ha dissenyat una estratègia de gestió del canvi perquè la IA s'utilitzi de manera efectiva?	Intervenció Es pot conceptualitzar l'ús del recurs amb una intervenció en els processos educatius?	Impacte Com es mesurarà, monitoritzarà i avaluarà l'impacte de la intervenció? El càlcul que els impactes previstos de l'ús de la IA no són satisfactoris, com s'identificarà si això es deu al disseny del recurs, a la seva aplicació o a una combinació de tots dos factors?
Risques Equitat. Com es garantirà que tots els estudiants puguin accedir i beneficiar-se de la IA? S'ha tingut en compte que el recurs pugui ser accessible i adequat a les necessitats dels estudiants amb necessitats cognitives o físiques? S'ha valorat el risc de discriminació directa o indirecta derivada de l'ús d'aquest recurs? Autonomia de l'alumne i del professorat. Quins resultats desfavorables podria predir un sistema d'IA? Quina acció perjudicial es podria prendre potencialment a partir d'aquesta predicció? Quines mesures positives es podrien adoptar per prevenir que es produïssi el resultat predit? Es pot afirmar que el recurs no temeraria l'addicció entre els estudiants ni els fomentaria a utilitzar-lo més enllà d'un punt que sigui beneficiós per al seu aprenentatge? En quines activitats curriculars o extracurriculars s'informa als alumnes sobre els beneficis i riscos de la IA? Com es capacitaran els educadors i/o altres professionals rellevants? Com es garantirà que l'autoritat dels educadors no es vagi atorbida com a resultat de l'ús de la IA? Privacitat. Es pot confirmar que l'ètic i el seu ús compliran amb tots els marcs legals rellevants? Quins usos de l'informació podrien considerar-se com a violació dels estudiants i dels docents, directament o indirectament? En cas d'utilitzar dades personals, com es justifica que aquestes dades són necessàries per a finalitats educatives i que el seu tractament beneficiarà els estudiants? Transparència i responsabilitat. Es pot definir una estructura de responsabilitat clara sobre l'ús de la IA a nivell dels docents i de l'organització educativa? Es pot implementar una estratègia de transparència adequada per al recurs d'IA seleccionat?			

Figura 10. El model EduIA Canvas

Nota: El canvas es pot obtenir escrivint a l'autor: jordi.vitria@ub.edu

Tot seguit, tenim el bloc format per l'objectiu del nostre projecte i el valor esperat.

La part de l'objectiu té com a funció respondre les preguntes següents: atès que la IA ha de ser utilitzada per assolir objectius educatius ben definits, basats en una evidència social, educativa o científica sòlida del fet que això és en benefici de l'estudiant, s'ha identificat amb claredat l'objectiu educatiu que es pretén assolir mitjançant l'ús de la IA?

En la part del valor, les preguntes són: s'ha identificat un recurs d'IA que sigui útil per als objectius educatius? Es pot explicar per què aquest recurs en particular té la capacitat d'assolir l'objectiu educatiu especificat anteriorment? De quina informació es disposa per afirmar que l'ús de la IA proposada està recolzat per evidència social, educativa o científica?

El bloc següent té per objectiu diagnosticar la maduresa del nostre entorn per adoptar la solució basada en IA.

En la primera part s'han de diagnosticar les dades necessàries (si és que són necessàries): quines dades calen per aplicar el recurs d'IA? És viable tenir aquestes dades? Quina és la qualitat de les dades de què es disposa?

En la segona, cal identificar els processos educatius involucrats: s'han identificat els processos educatius en què es farà servir aquest recurs d'IA? Com s'usarà la IA per millorar aquests processos? S'ha dissenyat cap estratègia de gestió del canvi perquè la IA s'utilitzi de manera efectiva?

A continuació, es defineix l'efecte esperat de la intervenció proposada i es proposa una avaluació del seu impacte.

Pel que fa a la intervenció, hem de ser capaços de respondre aquesta pregunta: es pot conceptualitzar l'ús del recurs com una intervenció en els processos?

En el cas de l'impacte, les preguntes són: com es mesurarà, monitorarà i avaluarà l'impacte de la intervenció? Si s'aprecia que els impactes de l'ús de la IA no són satisfactoris segons allò que s'havia previst, com identificaràs si això és degut a com es va dissenyar el recurs, a com s'està aplicant o a una combinació d'ambdós factors?

Finalment, tenim l'anàlisi de riscos, que ha de cobrir tots aquests aspectes:

- *Equitat*
Com es garantirà que tots els estudiants puguin accedir a la IA i beneficiar-se'n? S'ha tingut present que el recurs pugui ser accessible i adequat a les necessitats dels estudiants amb necessitats cognitives o físiques especials? Discriminació?
- *Autonomia de l'alumne i del professor*
Quins resultats desfavorables podria predir un sistema d'IA? Quina acció perjudicial es podria prendre potencialment a partir d'aquesta predicció? Quines mesures positives es podrien adoptar per prevenir que es doni el resultat predit? Es pot afirmar que el recurs no fomenta l'addicció entre els estudiants o que no els forçarà a estendre l'ús d'un recurs més enllà d'un punt que sigui beneficiós per al seu aprenentatge? En quines activitats curriculars o extracurriculars s'informa els alumnes sobre els beneficis i els riscos de la IA? Com es capacitaran els educadors o altres professionals rellevants? Com es garantirà que l'autoritat dels educadors no es vegi soscavada com a resultat de l'ús de la IA?
- *Privadesa*
Es pot confirmar que l'eina i el seu ús compleixen tots els marcs legals rellevants? Quins usos de l'eina podrien considerar-se com a vigilància dels estudiants i com podrien beneficiar els estudiants, sigui directament o indirectament? En el cas d'usar dades personals, com es justifica que aquestes dades són necessàries per a fins educatius i que processar aquestes dades beneficiarà els estudiants?
- *Transparència i responsabilitat*
Es pot definir una estructura de responsabilitat clara sobre l'ús de la IA a nivell dels docents i l'organització? Es pot implementar una estratègia de transparència adient per al recurs d'IA seleccionat?

La decisió sobre la viabilitat

Un cop tenim un Canvas amb un contingut consensuat, podem passar a la fase final de decisió. La presa de decisions grupals ofereix diversos avantatges importants tant per a les organitzacions com per als individus involucrats. Un dels principals és la diversitat de perspectives

que aporta la participació de persones amb antecedents i experiències dispars. Això no sols genera una àmplia gamma d'idees i enfocaments, sinó que també permet una millor solució de problemes, ja que els grups poden identificar i encarar-se als desafiaments de manera més efectiva i innovadora.

Un altre avantatge significatiu és l'augment de la creativitat i la innovació. La sinergia d'idees que es produeix a través de la interacció entre els membres del grup fomenta el pensament col·laboratiu, permetent que es construeixin solucions noves a partir de les contribucions de tots. A més, la presa de decisions de grup estimula l'acceptació i el compromís, atès que els individus implicats tendeixen a sentir-se més compromesos en les decisions finals i més disposades a acceptar-les, després que tothom hagi pres part activa en el procés.

Per acabar, l'avaluació d'opcions en un entorn grupal tendeix a ser més completa i equilibrada. La diversitat d'opinions permet una anàlisi més exhaustiva de les opcions disponibles, fet que ajuda a identificar riscos potencials i desavantatges. Així mateix, la presa de decisions grupals pot mitigar els biaixos individuals, oferint una visió més objectiva i reduint la influència de prejudicis personals en el procés de decisió.

Quan els membres d'un grup tenen antecedents similars, és més probable que sobreestimin les seves habilitats i cometin errors de manera similar, cosa que pot comportar riscos injustificats. Augmentar la diversitat del grup disminueix aquests riscos, perquè els membres es basen en experiències diferents per recopilar informació i resoldre problemes (Page, 2008; Hong, 2004).

Per a una bona presa de decisions grupals, primer hem d'acordar una regla de decisió, és a dir, el mètode que es farà servir per prendre la decisió final. És clau triar l'enfocament abans de començar i assegurar-se que totes les persones involucrades saben quin és el seu rol. Les opcions més comunes són:

- *Consens*: tots els membres del grup hi estan d'acord.
- *Democràtic*: la majoria dels membres del grup hi estan d'acord.
- *Consentiment*: cap membre del grup s'hi oposa.
- *Consulta*: ens basem en opinions de diversos experts externs. Durant la consulta, l'assessorament generalment s'obté de manera indepen-

dent i després s'integra, però també pot implicar convidar experts a proporcionar assessorament consensuat en un entorn de petit grup.

- *Agregació:* s'hi afegeixen múltiples judicis (mitjançant enquestes, entrevistes, etc.) de moltes persones externes diferents.

L'elecció de la regla de decisió depèn de les prioritats del grup, així com del temps i dels recursos disponibles.

CONCLUSIONS

Quines eines i usos de la intel·ligència artificial són importants en educació?

La intel·ligència artificial (IA) està transformant l'àmbit educatiu, atès que proporciona eines i aplicacions que ofereixen noves possibilitats per millorar tant l'ensenyament com l'aprenentatge. Entre les més rellevants hi ha els *sistemes de tutoria intel·ligent*, que brinden instrucció personalitzada adaptada al ritme, l'estil d'aprenentatge i les necessitats de cada estudiant. Aquestes eines no només ajuden a identificar àrees de dificultat, sinó que també proporcionen recursos addicionals per facilitar l'aprenentatge, fent que l'experiència educativa resulti més eficient i efectiva.

Una altra aplicació clau és l'*anàlisi d'aprenentatge*, implementada, generalment, en *sistemes de gestió d'aprenentatge* (LMS). Aquestes plataformes permeten als educadors recopilar i analitzar dades sobre el rendiment dels estudiants, identificar patrons i tendències, i aplicar ajustaments basats en l'evidència per optimitzar els mètodes d'ensenyament. Aquesta anàlisi propicia un aprenentatge més informat i centrat en els resultats.

A més, l'*avaluació automatitzada* ha emergit com a una eina essencial, ja que, gràcies a aquesta, es pot avaluar el rendiment de l'alumne de manera ràpida i objectiva. Això alleuja la càrrega administrativa dels docents i assegura que l'alumnat rebi retroalimentació immediata, una cosa vital per al seu progrés acadèmic.

L'*aprenentatge adaptatiu* també té un paper crucial en ajustar automàticament el contingut i les activitats educatives en funció del rendiment i les necessitats individuals de cada estudiant. D'aquesta manera, es propicia una experiència d'aprenentatge més inclusiva i personalitzada. Així mateix, els assistents virtuals i els *chatbots* faciliten tant el suport acadèmic com l'administratiu, ajudant l'alumnat a navegar pels cursos, resoldre dubtes freqüents i accedir a recursos addicionals de forma ràpida i eficient.

Aquestes eines es basen en models d'IA predictiva o d'IA generativa, la comprensió dels quals és fonamental per avaluar-ne els beneficis i els riscos potencials. És essencial considerar com aquestes tecnologies poden influir en els processos educatius a llarg termini i garantir que s'implementin de manera eficaç i responsable.

Què vol dir fer un ús de la intel·ligència artificial responsable en educació?

Fer un ús responsable de la IA en l'àmbit educatiu demana dissenyar, construir i implementar sistemes amb un enfocament ètic, transparent i orientat a l'impacte positiu en estudiants, docents i la societat en general. Aquest enfocament no només és un imperatiu ètic, sinó que també porta beneficis tangibles, com ara accelerar la innovació, millorar la competitivitat o enfortir la confiança en les institucions educatives.

En aquest context, establir estàndards d'IA responsable en l'educació és indispensable per diverses raons. En primer lloc, les consideracions ètiques garanteixen que els sistemes d'IA es desenvolupin sense pràctiques perjudicials, com ara la discriminació, els biaixos algorítmics o la invasió de la privadesa. La implementació d'estàndards ètics minimitza el risc de conseqüències negatives i assegura que les eines d'IA es fan servir per beneficiar tots els actors involucrats.

D'altra banda, prioritzar la responsabilitat en la IA enforteix la confiança en les institucions educatives i la seva reputació. Les organitzacions que adopten pràctiques responsables generen credibilitat entre els alumnes, els docents, els pares i la societat en general. També és rellevant el compliment legal i regulador, ja que els marcs normatius estan evolucionant ràpidament per abordar les implicacions ètiques i socials de la IA. Complir aquestes normatives, més enllà de prevenir sancions legals, també posiciona les institucions com a líders en innovació responsable.

Com assegurem l'ús responsable de la intel·ligència artificial en educació?

Garantir l'ús responsable de la IA en l'àmbit educatiu requereix un enfocament integral i participatiu. Aquest procés no es pot limitar a seguir una llista estàtica de recomanacions, sinó que s'ha d'adaptar a les particularitats de cada aula, institució o context educatiu.

El punt de partida és un procés deliberatiu que aborda determinades preguntes clau: quins objectius educatius es volen assolir? Quin valor aportarà la IA a l'ensenyament i l'aprenentatge? És viable en el pla tècnic i de procediment? Quins riscos ètics o socials podrien sorgir? Aquestes qüestions ajuden a guiar la implementació de la IA i a anticipar possibles desafiaments.

L'ús del model EduIA Canvas pot ser especialment útil en aquest procés, ja que facilita la planificació estructurada i col·laborativa de projectes d'IA, en permetre a les institucions dissenyar estratègies que considerin tant els beneficis com els riscos associats.

Amb una metodologia adequada, la IA té el potencial de transformar l'educació, en millorar els resultats d'aprenentatge, donar suport als docents en la seva tasca i promoure l'equitat educativa. Per contra, l'absència d'un enfocament ben estructurat pot derivar en riscos significatius, com ara l'exclusió digital, la perpetuació de desigualtats i la dependència excessiva de tecnologies no provades.

En conseqüència, el compromís amb un ús responsable de la IA ha de ser continu, tot integrant principis ètics, avaluacions rigoroses i adaptacions constants per assegurar que aquestes tecnologies compleixin la promesa d'enriquir l'educació de manera inclusiva i sostenible.

BIBLIOGRAFIA

- Alpaydin, E. (2021). *Machine learning*. MIT Press.
- Barocas, S., Hardt, M. i Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.
- Boden, M. A. (2014). GOF AI. A: Frankish, K. i Ramsey WM (ed.). *The Cambridge Handbook of Artificial Intelligence* (p. 89-107). Cambridge University Press.
- Bottou, L. i Schölkopf, B. (2023). *Borges and AI*. arXiv preprint arXiv: 2310.01425
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, 538(7623), 20.
- Crevier, D. (1993). *AI: the tumultuous history of the search for artificial intelligence*. Basic Books.
- Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG>
- EUR Lex (2024). Regulation (EU) 2024/1689 de la European Parliament and of Council of 13 June 2024 Laying down harmonised rules on artificial intelligence and amending Regulations. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Hao, K. (2020). *The UK exam debacle reminds us that algorithms can't fix broken systems*. MIT Technology Review.
- Hastie, T., Tibshirani, R., Friedman, J. H. i Friedman, JH (2009). *The elements of statistical learning: data mining, inference, and prediction* (vol. 2, p. 1-758). Springer.
- Holmes, W., Persson, J., Chounta, I. A, Wasson, B. i Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Consell d'Europa.
- Hong, L. i Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385-16389.
- Jackson, P. (1990). *Introduction to expert systems*. Addison-Wesley Longman.
- Kennedy, B., Tyson, A. i Saks, E. (2023). *Public awareness of artificial intelligence in everyday activities*. <https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities>

- Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Amodei, D. *et al.* (2020). *Language models are few-shot learners*. arXiv preprint arXiv:2005.14165
- McCarthy, J., Minsky, M. L., Rochester, N. i Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12-12.
- Narayanan, A. i Kapoor, S. (2024). *AI snake oil: What artificial intelligence can do, what it can't, and how to tell the difference*. Princeton University Press.
- OpenAI (2022). *Introducing ChatGPT* [Large language model]. <https://openai.com/ca-ES/index/chatgpt>
- Page, S. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition*. Princeton University Press.
- Prats, M. A. i Sintes, E. (2021). *Com impulsar la transformació digital de l'escola*. <https://fundaciobofill.cat/publicacions/educacio-hibrida>
- Ribera, M. i Díaz, O. (2024). *ChatGPT y educación universitaria: posibilidades y límites de ChatGPT como herramienta docente*. Universitat de Barcelona. IDP/ICE i Octaedro.
- Richardson, R., Schultz, J. M. i Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev.*, en línia, 94, 15.
- Unesco (2021). *Recomanació sobre l'ètica de la Intel·ligència Artificial*. Unesco.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. i Polosukhin, I. (2017). Attention is all you need. Dins: I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan i R. Garnett (ed.). *Advances in Neural Information Processing Systems* (vol. 30). Curran Associates.
- Zhang, A., Lipton, Z. C., Li, M. i Smola, A. J. (2023). *Dive into deep learning*. Cambridge University Press.

NORMES PER ALS COL-LABORADORS

<https://bit.ly/3DKFESh>

EXTENSIÓ

Les propostes del Quadern no poden excedir **l'extensió de 50 pàgines (en Word)**, uns 105.000 caràcters, espais, referències, quadres, gràfiques i notes incloses.

PRESENTACIÓ D'ORIGINALS

Els textos han d'incloure, en format electrònic, un **resum** d'unes deu línies i tres paraules clau, no incloses al títol. Igualment, han de contenir el **títol**, un **abstract** i tres **keywords** en anglès.

Per a les **formes de citar i referències bibliogràfiques**, cal remetre's a les utilitzades en aquest *Quadern*.

AVALUACIÓ

L'acceptació d'originals es regeix pel **sistema d'avaluació externa per experts**.

Els originals són llegits, en primer lloc, pel **Consell de Redacció**, que valora l'adequació del text a les línies i objectius dels *Quaderns* i si compleix els requisits formals i els mínims de contingut científic exigits.

Els originals són sotmesos, en segon lloc, a **l'avaluació de dos experts**, especialistes en la temàtica de la qual tracta l'original i l'àmbit disciplinari corresponent. Els autors reben els comentaris i suggeriments dels avaluadors i la valoració final amb les esmenes i canvis que cal fer, si és el cas, abans de ser acceptat per a la seva publicació.

Si els canvis exigits són significatius o afecten bona part del text, el nou original és sotmès a l'avaluació de dos experts externs i d'un membre del Consell de Redacció. El procés es duu a terme com a «doble cec».

REVISORS

<https://bit.ly/3oF4izw>

L'Institut de Desenvolupament Professional (IDP/ICE) de la Universitat de Barcelona inicià fa uns anys la publicació dels **QUADERNS DE DOCÈNCIA UNIVERSITÀRIA** amb l'objectiu de posar a l'abast del professorat universitari documents i materials de treball referits a temes relacionats amb la docència superior que facilitessin la seva formació, l'intercanvi d'experiències i la difusió de «bones pràctiques» docents. Amb aquests *Quaderns* pretenem estar atents als temes nous i emergents en l'actual conjuntura universitària, per tal de donar a conèixer i difondre iniciatives innovadores en el camp de la docència universitària, que responguin a les línies següents:

- Propostes de marcs de referència rigorosos i generals que ajudin a clarificar conceptes clau.
- Estratègies docents i bones pràctiques de planificació, metodologia i avaluació de l'ensenyament-aprenentatge, desenvolupades en contextos acadèmics específics i diversos.
- Tècniques i tàctiques, de marcat caràcter didàctic, presentades en materials i propostes concretes de treball i reflexió sobre la pràctica d'equips docents disciplinaris o interdisciplinaris.