Ètica i transparència en l'ús de la IA a les universitats

Autor: **Albert Sabater Coll**
Director de la Càtedra-Observatori d'Ètica en Intel·ligència Artificial de Catalunya , Departament d'Empresa de la Facultat de Ciències Econòmiques i Empresarials de la Universitat de Girona
ORCID:0000-0003-3532-6572

Revista Alma Mater
Segon Fòrum Internacional de Recerca
ISSN: XXXXXXXX Vol: 1 (2026)

Resum
Aquesta conferència aborda l'impacte de la intel·ligència artificial (IA) en l'àmbit universitari des d'una perspectiva ètica i de responsabilitat. Es defensa que la IA no és un fenomen nou, però la seva popularització recent —especialment a través dels grans models de llenguatge— exigeix una resposta formativa i institucional clara. Es plantegen els riscos de l'antropomorfització de les eines, la sobredependència tecnològica, la pèrdua de pensament crític i la insostenibilitat ambiental i social. Al mateix temps, es reconeixen beneficis com l'exploració creativa, la multiplicació d'exemples didàctics i la detecció d'errors. La conferència proposa un model d'IA responsable articulat al voltant de set principis —transparència, justícia, responsabilitat, seguretat, privacitat, autonomia i sostenibilitat—, i presenta una eina d'avaluació desenvolupada per la Càtedra-Observatori d'Ètica en Intel·ligència Artificial de Catalunya per ajudar institucions a avaluar i millorar l'ús ètic de la IA.

Paraules clau
intel·ligència artificial, ètica, transparència, universitat, responsabilitat, sostenibilitat, pensament crític, grans models de llenguatge, capacitació digital, IA responsable

Abstract
This lecture examines the impact of artificial intelligence (AI) in higher education from an ethical and responsible use perspective. It argues that AI is not a new phenomenon, but its recent popularisation—particularly through large language models—demands a clear educational and institutional response. The risks of anthropomorphising AI tools, technological over-reliance, loss of critical thinking, and

environmental and social unsustainability are discussed. At the same time, benefits such as creative exploration, multiplication of pedagogical examples, and error detection are acknowledged. The lecture proposes a responsible AI model built around seven principles—transparency, fairness, accountability, security, privacy, autonomy, and sustainability—and presents an evaluation tool developed by the Ethics in Artificial Intelligence Observatory of Catalonia to help institutions assess and improve their ethical use of AI.

Keywords

artificial intelligence, ethics, transparency, higher education, responsibility, sustainability, critical thinking, large language models, digital literacy, responsible AI

En primer lloc, gràcies per la invitació. És un plaer poder ser aquí a Andorra en un dia plenament hivernal, i en una universitat que coneixia a distància però no presencialment. Per tant, doble motiu d'agraïment.

En el meu cas, com a director de l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya, ens agrada remarcar —no només ara, sinó des de fa anys, des de la instauració de l'Observatori com a pilar de l'estratègia d'intel·ligència artificial de Catalunya— que les tecnologies d'intel·ligència artificial, no només les més recents sinó també les més tradicionals (automatització, sistemes de predicció, percepció, etc.), fa temps que són entre nosaltres. No són una novetat absoluta. Ara bé, sí que és cert que se n'ha produït una popularització, especialment a través de les darreres eines i, si voleu, dels **grans models de llenguatge**, com **ChatGPT**, etcètera.

Tot això ens porta a una reflexió central: la IA és aquí per quedar-se. Té aplicacions en molts àmbits i hem de ser capaços de treure'n el màxim profit; però no des d'una mirada ingènua, sinó amb la responsabilitat exigible davant de qualsevol tecnologia. L'objectiu és que les aplicacions d'intel·ligència artificial esdevinguin una tecnologia "normal", tan normal com una rentadora, un forn o un cotxe. Ara bé, per arribar-hi encara cal fer moltes passes. Una d'aquestes passes és desenvolupar la **capacitat sociotècnica**: com a societat, hem d'esdevenir una societat empoderada tecnològicament.

Fem servir aquestes tecnologies en l'entreteniment, la salut, els serveis socials, els serveis financers… en molts àmbits. Però això no vol dir que sempre les fem servir bé. I aquí és on cal incidir, especialment en la capacitació del nostre alumnat, que ja és professional del present i ho serà també del futur. Cal capacitar-lo en qüestions relacionades amb la predicció; en com llegir escenaris de futur amb aquestes eines; en com classificar millor la informació existent —en un context de sobresaturació informativa—; en com compartimentar-la millor i transferir-la entre nosaltres; i, com és natural, en l'ús d'eines capaces de "llegir" i interpretar el que ens envolta.

Això és especialment rellevant en el transport, i ho veiem també en l'àmbit de la salut, per exemple, amb sistemes de detecció de malalties. És aquí on la IA ha fet un salt important i on hem de maximitzar-ne el potencial.

Dit això, no hem de pensar que només el component tecnològic ens donarà totes les respostes. Hem d'estar atents a com dissenyem sistemes més eficients amb **menys dades**. Això no vol dir no utilitzar dades: se n'hauran d'utilitzar, i no només les que existeixen, sinó també les que haurem de generar. Però cal posar límits a la dependència de grans volums de dades. Avui encara fem servir, sobretot, grans models de llenguatge i altres models d'aprenentatge automàtic que necessiten masses dades per donar respostes que, a vegades, no són les millors possibles. Per tant, cal millorar la tecnologia i conèixer-ne les limitacions.

Una limitació clau és evitar **antropomorfitzar** aquestes tecnologies. Són màquines, eines i aplicacions; no són persones. I, com que no són persones, no els podem atribuir agència moral ni responsabilitat. La decisió l'ha de prendre una persona, amb l'ajuda d'una màquina, però sense confondre els rols. Això és problemàtic també jurídicament: dins del reglament d'intel·ligència artificial, per exemple, es preveu que qualsevol presa de decisions automatitzada ha de tenir una persona al darrere, perquè és la persona qui assumeix la responsabilitat legal .

També hem de pensar en **sostenibilitat**. No podem fer servir la IA per a tot: hem de fer servir el cervell. Si no, entrarem en una sobredependència tecnològica que, sobretot en etapes d'aprenentatge, pot ser perjudicial. I això no afecta només l'alumnat: també el professorat i el **PTG** , i el personal administratiu. Com qualsevol professional, si no exercitem i entrenem les nostres capacitats, s'atrofien. Cal, doncs, un equilibri i una disciplina: utilitzar aquestes eines quan realment cal, però no convertir-les en norma. En salut, per exemple, els facultatius ho expressen clarament: si la màquina ho fa tot, el professional perd capacitat. I això no és menor.

La sostenibilitat té, com a mínim, dos angles: una sostenibilitat social i professional, i una sostenibilitat ambiental. Si fem servir aquestes màquines constantment, el consum esdevé desorbitat. Ho podem comparar amb l'ús del cotxe: la gent no el fa servir per a tot; també camina. Cal ser curosos —la virtut és al mig (*"in medio virtus"*)— i, alhora, treure'n profit.

Ara bé: l'ús d'aquestes eines **no pressuposa** automàticament un coneixement ètic previ. En alguns casos, cal revisitar consideracions clàssiques i actua-

litzar-les; en altres, cal aprendre què vol dir responsabilitat quan fem servir aplicacions d'IA; què implica en termes de seguretat, privacitat o autonomia. És important no evadir responsabilitats: cal "agafar el toro per les banyes" i preguntar-nos què hem de fer. I això requereix capacitació, també en l'àmbit docent, i traslladar-la a l'alumnat.

El que no podem fer és dir als estudiants: "Això és aquí; feu-ho servir com vulgueu", sense guies, sense educació i sense delimitar què està bé i què està malament. Una de les funcions fonamentals de la universitat és transmetre una cultura de l'esforç: quan surtin al mercat laboral, no els regalaran res. Aquesta cultura de l'esforç, de l'aprenentatge i del sentit crític forma part del nou ecosistema. No volem aules on es produeixi la paradoxa que els alumnes amb pitjor base, només amb tecnologia, acabin essent els millors. El més important és que acabin essent bons amb o sense tecnologia. En una entrevista de feina, han de demostrar que saben fer servir la tecnologia, però també que poden afrontar preguntes i problemes sense ella. Si no, no accediran a les millors oportunitats, i això pot ser un problema per generacions que comencen a mostrar sobredependència tecnològica.

Això és comparable a no saber sumar, restar o multiplicar: sense una autonomia mínima, no pots operar en el món. I, en aquest marc, podem recordar —seguint una idea atribuïda a Marshall McLuhan — que res d'això és inevitable: hem d'incidir en com entomem el repte tecnològic. Hi ha actors amb interessos legítims; cal aprofitar el desplegament, però també redireccionar-lo cap on el volem i, sobretot, com el transmetem a les generacions més joves, i també a nosaltres mateixos. Això demana una reflexió profunda i una nova capacitació en aquestes tecnologies.

En el context universitari, hi ha diverses implicacions; en faré algunes pinzellades, en una sessió panoràmica i introductòria. A nivell epistemològic, hem de reflexionar sobre quins coneixements impartim i què promovem. Hi ha qüestions bàsiques que potser es mantindran, però d'altres canviaran: per accessibilitat, per formes d'accés al coneixement i per com el traslladem a l'alumnat; també per la manera com nosaltres mateixos ens recapacitem.

A nivell pedagògic, també: com utilitzem la IA? No és el mateix per a professorat i alumnat. L'alumnat pot tenir accés a determinades eines, però el seu

ús no pot ser idèntic al del professorat, que dissenya continguts pedagògics, mentre que l'alumnat és objecte d'avaluació. Per tant, la integració ha de ser diferent; no és una "barra lliure" per a tothom.

En termes de recerca, és fonamental: no només com implementem la IA en la recerca universitària, sinó també com l'estudiem; cal analitzar-ne conseqüències i efectes i fer recerca sobre el propi aprenentatge i sobre el disseny de continguts a l'aula.

Hi ha igualment una implicació personal: la relació entre alumnat i professorat, tant presencial com virtual. Cal veure com s'alteren aquestes relacions i assegurar que, encara que canviïn, siguin bones per ambdues parts. Necessitem professorat motivat i alumnat motivat; si només tenim alumnat motivat, tindrem un problema.

Això ens porta a tres reflexions mínimes. La primera: els impactes a llarg termini. És relativament senzill mirar què passa ara, observant entorns propers i internacionals; però cal pensar les derivades a llarg termini: l'impacte en la inserció laboral, en les posicions que ocuparan els alumnes i en les institucions mateixes. Això inclou també pensar conseqüències de l'absència de tecnologia en certes àrees i la necessitat de delimitar usos: com hem fet, per exemple, amb dispositius mòbils. I també cal pensar en conseqüències inesperades: imaginar contrafactuals. Què passaria si tothom fes servir grans models de llenguatge (ChatGPT o Claude) per lliurar treballs? Si això passa, no podem suspendre tothom; haurem de pensar metodologies noves. No cal canviar de cop les institucions, però sí revisar algunes regles del joc.

I, en paral·lel, cal discernir què és realment nou: una tecnologia pot ser disruptiva, però la cultura de la responsabilitat, l'esforç i la integritat acadèmica no ha canviat. Venim d'una tradició humanista i científica amb normes que ens permeten avançar lentament però de manera sostinguda. No tot el que és ràpid és necessàriament bo.

Pel que fa als beneficis immediats, n'hi ha diversos. El primer és la **curiositat**: una nova eina motiva conductes exploratòries, i això és fonamental a la universitat. Aquesta curiositat facilita l'adquisició de nous coneixements, també

perquè algunes d'aquestes tecnologies tenen capacitat general de transmetre coneixement. Ara bé, a mesura que les eines es normalitzen, la curiositat pot bascular entre **comoditat** (només cal un prompt) i **incertesa** (això és realment el que s'hauria de fer?). Aquesta incertesa és positiva: indica que no estem del tot segurs i que sospitem que hi ha alguna cosa més a fer.

Un altre benefici és la multiplicació d'exemples: amb un gran model de llenguatge pots generar molts exemples d'un problema o d'un concepte complex. Això és especialment útil per a l'alumnat. Ara bé, cal conduir-ho: no es tracta de tenir més exemples dels necessaris, sinó d'ajudar-los a seleccionar els millors. Les primeres cerques no les hem de fer necessàriament nosaltres; les han de fer ells, amb orientació.

Des del punt de vista d'una tecnologia experimental, també veiem que els usuaris esdevenen "tècnics de detecció d'errors". I això, en aprenentatge, és bo: podem donar la volta a la idea de les "al·lucinacions" (errors) i convertir-ho en una pràctica: "trobeu els errors". Requereix temps, però promou supervisió activa i avaluació crítica constant dels resultats. Això no és només per a sectors tecnificats: és per a tothom; i, a la universitat, és especialment valuós.

Ara bé, si mirem la piràmide de Miller i, en particular, l'ús de chatbots, això només és beneficiós si coneixem prou bé el tema. Per arribar-hi, l'alumnat necessita una base prèvia: no pot interactuar amb aquestes tecnologies sense coneixement previ. El rol acadèmic és fonamental: no podem limitar-nos a donar un chatbot i després corregir. Hem de garantir que les persones que arriben a la universitat saben prou d'un tema i, a partir d'aquí, fan servir la tecnologia per millorar l'adquisició de coneixement de manera responsable. Si no saps del tema, tens un problema. Posaré un exemple: preguntar a un chatbot si s'ha de prendre una medicació és arriscat; cal consultar un professional, i només de manera complementària mirar informació en un sistema d'aquest tipus.

A més, si fem servir aquestes eines només per curiositat, hi ha el perill de la procrastinació: perdre més temps que no pas guanyar productivitat. També hem de ser conscients del "bombardeig" comercial: hi ha companyies amb objectius comercials i avisos constants ("et puc ajudar", "sóc aquí per resoldre problemes"). Cal traslladar una cultura de "primer fes-ho tu, i després, si vols,

ho contrastes". És el mateix patró que els avisos de principi dels 2000 amb el clip de Microsoft: ara reapareix en altres formes.

La recerca i les mateixes companyies adverteixen que una sobredependència pot augmentar una certa "descàrrega cognitiva" i fins i tot fomentar ganduleria. Això no ens ho podem permetre: hem de fomentar la cultura de l'esforç. També cal considerar la difusió de desinformació, la pèrdua de creativitat i el debilitament del pensament crític independent. I, en definitiva, no volem acabar parlant com ChatGPT: ChatGPT és una màquina; nosaltres no ho som. Volem seguir sent persones ben articulades, atentes al nostre entorn i singulars. Això és rellevant no només per trobar una bona feina, sinó també per construir una bona vida.

He incorporat un gràfic d'una investigació recent de Harvard basada en l'Enquesta Mundial de Valors i indicadors multidimensionals, on situen ChatGPT "com si fos un país". L'interès és mostrar que aquestes eines tenen limitacions: s'entrenen amb dades determinades i presenten una certa homogeneïtat. En un món globalitzat, si només "parlem com ChatGPT", no serem capaços d'interactuar amb moltes realitats; i això afecta també el món dels negocis i de les relacions en general.

Dit això, la IA pot ajudar en àmbits com la identificació de patrons (una gran aportació des de l'aprenentatge automàtic) i, combinant aprenentatge automàtic, llenguatge natural i generativa, pot facilitar l'observació de tendències a partir de milions de dades. També permet generar continguts, informes personalitzats, etc. Ara bé, cal control: verificar resultats, considerar drets d'autor i limitar la generació indiscriminada. Una aportació molt nova és la possibilitat de fer pluja d'idees i abstracció sense tenir moltes persones reunides: podem explorar escenaris hipotètics i descontextualitzats ("fes-me un poema", "fes-me un codi", "què pensaria un filòsof al Japó?"), i això pot obrir camins creatius que, a priori, no exploraríem.

També hem de saber quan evitar fer-la servir. En relació amb els grans models de llenguatge, encara estem lluny d'una precisió realment alta: parlem d'errors que, en alguns casos, poden arribar al 30–40%. Pot ser d'ajuda, però exigeix verificació constant. A classe es veu fàcilment: mateixa pregunta en

dispositius diferents i respostes no idèntiques. Això passa perquè no són sistemes deterministes com una calculadora: són models probabilístics. L'alumnat no està preparat per entomar-ho si no el formem perquè pensi també en termes probabilístics.

Un altre criteri bàsic: no compartir dades sensibles. Molts models són en línia i la privacitat no està garantida; hi pot haver actors maliciosos. Per això, institucions financeres i públiques prohibeixen compartir informació personal. Hem d'educar l'alumnat perquè no hi aboqui dades personals.

I un principi central: la decisió no la pot prendre un gran model de llenguatge; l'hem de prendre nosaltres. Pot ajudar amb escenaris i informació, però la decisió crítica és humana i no es pot delegar. Legalment és així, però també ho hem de recordar des d'un punt de vista ètic: decideixen les persones, no les màquines.

En aquest estadi, alguns consells bàsics són: no utilitzar la IA per resoldre problemes senzills (per evitar dependència i per sostenibilitat); no caure en el "el xat m'ha dit això" com a argument; no donar confiança excessiva als resultats; no antropomorfitzar les eines; no tenir expectatives irreals. Són tecnologies de propòsit general, però orientades a tasques específiques: no serveixen per a tot. Encara som en un model d'"IA estreta", que resol tasques concretes. I no hem de sobrevalorar les dades d'entrenament: no ho són tot; hi ha molt que encara no està incorporat, etiquetat o utilitzat.

Tot això requereix **transparència**. La volem per garantir que els sistemes siguin comprensibles per usuaris i parts interessades, i per obtenir explicacions clares. Hem de poder entendre què demanem i què obtenim: la cadena de raonament, com s'han fet servir les dades, i distingir si un resultat és acceptable o no, encara que internament el model sigui complex. També hem d'utilitzar els millors models disponibles segons la tasca, no una sola eina per a tot; documentar processos; comunicar limitacions i incerteses. Això forma part de la metodologia universitària. Si abans referenciàvem bibliografia, ara també cal documentar com s'ha fet servir una eina: quines instruccions s'han donat, quan ha fallat, quan ha funcionat, i què se n'ha après. En generativa, transparència també vol dir comprendre com es genera el contingut, quines fonts de

dades hi ha i quines limitacions té el sistema. I això afecta universitat, ciutadania, empreses i administració: hem de pensar en termes d'**IA responsable**.

Què vol dir "IA responsable"? Hi ha tres ingredients fonamentals. Primer, no som només usuaris o clients: hem de conèixer aspectes ètics i legals, com a mínim els bàsics; no són un complement, són part del funcionament responsable, com mirar instruccions d'una rentadora o saber canviar una roda. Segon, cal un seguiment constant: potser no el faran els alumnes, però sí les institucions, internament i externament, respectant acords entre institucions i empreses. Tercer, cal aquesta visió crítica socràtica: "per a què serveix?", "què ens aporta?", i documentar eleccions i opcions, en estudiants, professorat i administració.

La darrera part de la presentació trasllada aquesta idea d'IA responsable a través d'eines que estem desenvolupant. No és fàcil; cal metodologia per entomar reptes del desplegament recent de tecnologies de generació de continguts. Però també calen mètriques i indicadors per valorar fins a quin punt complim expectatives. Des de l'Observatori, i amb l'ajuda de l'ecosistema, hem desenvolupat una eina: un **model de principis i indicadors observables** (accessible al web de l'Observatori) que gira al voltant de set principis. L'objectiu és interrogar fins a quin punt la interacció amb una aplicació és transparent, justa, responsable i segura, i si incorpora privacitat, autonomia i sostenibilitat. Si tenim aquest coneixement, moltes coses no se'ns escaparan i la interacció es normalitzarà.

Aquest treball no es fa només des de principis ètics: també en relació amb el desplegament normatiu. Igual que amb el Reglament de Protecció de Dades, hem d'aprendre què es pot fer i què no es pot fer, i quin és el marc jurídic. Per normalitzar-ho, necessitem eines que ens permetin navegar aquests aspectes. El model inclou una preavaluació de categories de risc (des del "risc" fins a l'"alt risc"; els usos "inacceptables", que ja estan prohibits, no s'hi discuteixen). Això ajuda a conèixer millor quines eines es poden fer servir, de quina manera, segons el sector i el nivell de risc, i a acotar usos per evitar riscos innecessaris.

Es tracta d'un model seqüencial: una part inicial de preavaluació per evitar entrar en usos prohibits; després, una recollida d'informació per obtenir mè-

triques i resultats; i una síntesi final en tres resultats globals per veure fins a quin punt avancem en IA responsable en cadascun dels principis (amb una visualització tipus semàfor; com més "verd", millor). Però la idea no és només "sortir verd", sinó conèixer preguntes clau i avançar.

També és important, de cara al marc jurídic, situar la interacció en una **matriu de riscos**, tal com preveu la legislació d'IA. Això serveix per decidir quins canvis cal fer, tant en la nostra interacció com en relació amb proveïdors i intermediaris (els qui ofereixen serveis d'IA). Encara que una institució "no faci res", la normativa genera expectatives de canvi. Per tant, cal promoure IA responsable també exigint a proveïdors: "no comprarem aquest programari si no feu canvis", perquè, si no, el problema no és només del proveïdor: també és nostre com a usuaris. Utilitzar tecnologies experimentals no és un problema si milloren; si ens hi quedem instal·lats de manera confortable, llavors sí que hi ha un problema, perquè encara hi ha molt camí per recórrer.

Després de l'autoavaluació, proporcionem informació perquè es pugui utilitzar internament i externament; oferim catàlegs legals i exemples per mostrar d'on surt tot, perquè sovint, més enllà d'escenaris hipotètics, tothom vol veure casos reals. A més, des d'una perspectiva institucional, per aterrar el reglament d'IA i altres normatives, també oferim orientacions sobre la relació entre institucions i proveïdors: com hauria de ser contractualment, i com connectar punts de millora amb clàusules contractuals que després cada institució pot adaptar.

Un resultat fonamental de la IA responsable és la **confiança**. Sense confiança, tenim un problema: hem de poder confiar en aquestes tecnologies, i també confiar que en fem un bon ús.

Per acabar: aquestes tecnologies són aquí per quedar-se. En molts casos ens estalviaran feina i poden incrementar la productivitat. Ara bé, caldrà veure què fem amb el temps que guanyem: si el dediquem a revisar què hem fet malament o a altres activitats. I, en aquest marc, serà important continuar plantejant reflexions de mig i llarg termini sobre conseqüències inesperades de l'ús d'aquestes aplicacions d'intel·ligència artificial. I fins aquí la meva presentació. Moltes gràcies.