

Titulació	Tipus	Curs
Intel·ligència Artificial / Artificial Intelligence	FB	2

Professor/a de contacte

Nom: Maria Pilar Dellunde Clave

Correu electrònic: pilar.dellunde@uab.cat

Equip docent

Nuria Valles Peris

(Extern) Roger Deulofeu Batllori

Idiomes dels grups

Podeu consultar aquesta informació al [final](#) del document.

Prerequisits

Aquesta assignatura no té prerequisits

Objectius

A partir casos d'estudi reals, aquest curs està dissenyat per introduir als estudiants les implicacions ètiques i socials de la intel·ligència artificial (IA). L'assignatura promourà la incorporació d'estratègies i l'ús d'eines per minimitzar els riscos ètics alhora desenvolupar sistemes d'IA en el marc d'una IA responsable.

Competències

- Actuar amb responsabilitat ètica i amb respecte pels drets i deures fonamentals, la diversitat i els valors democràtics.
- Actuar en l'àmbit de coneixement propi avaluant les desigualtats per raó de sexe o gènere.
- Comunicar-se de manera efectiva, tant oralment com per escrit, utilitzant adequadament els recursos comunicatius necessaris i adaptant-se a les característiques de la situació i de l'audiència.
- Concebre, dissenyar, analitzar i implementar agents i sistemes ciberfísics autònoms capaços d'interactuar amb altres agents o persones en entorns oberts, tenint en compte les demandes i necessitats col·lectives.

- Desenvolupar pensament crític per analitzar de manera fonamentada i argumentada alternatives i propostes tant pròpies com alienes.
- Identificar, analitzar i avaluar l'impacte ètic i social, el context humà i cultural i les implicacions legals del desenvolupament d'aplicacions d'intel·ligència artificial i de manipulació de dades en diferents àmbits.
- Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes destacats d'índole social, científica o ètica.
- Treballar de manera autònoma, amb responsabilitat i iniciativa, planificant i gestionant el temps i els recursos disponibles i adaptant-se a les situacions imprevistes.

Resultats d'aprenentatge

1. Analitzar críticament els principis, valors i procediments que regeixen l'exercici de la professió.
2. Analitzar les desigualtats per raó de sexe o gènere i els biaixos de gènere en l'àmbit de coneixement propi.
3. Comprendre les implicacions socials, ètiques i legals de la pràctica professional en IA.
4. Comunicar-se de manera efectiva, tant oralment com per escrit, utilitzant adequadament els recursos comunicatius necessaris i adaptant-se a les característiques de la situació i de l'audiència.
5. Desenvolupar pensament crític per analitzar de manera fonamentada i argumentada alternatives i propostes tant pròpies com alienes.
6. Explicar el codi deontològic, explícit o implícit, de l'àmbit de coneixement propi.
7. Identificar els biaixos socials, culturals i econòmics dels algoritmes.
8. Identificar les principals desigualtats i discriminacions per raó de sexe o gènere presents en la societat.
9. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes destacats d'índole social, científica o ètica.
10. Saber analitzar casos d'aplicació de la IA des d'un punt de vista ètic, legal i social.
11. Saber treballar en equip en el disseny de projectes interdisciplinaris. Ser capaç de col·laborar amb no professionals i professionals d'altres sectors.
12. Ser capaç d'incorporar els principis de la recerca i innovació responsable en els desenvolupaments basats en la IA.
13. Ser capaç d'incorporar valors adequats a les necessitats de les persones en el disseny de dispositius dotats d'IA.
14. Treballar de manera autònoma, amb responsabilitat i iniciativa, planificant i gestionant el temps i els recursos disponibles i adaptant-se a les situacions imprevistes.
15. Valorar com els estereotips i els rols de gènere incideixen en l'exercici professional.
16. Valorar les dificultats, els prejudicis i les discriminacions que poden incloure les accions o projectes, a curt o llarg termini, en relació amb determinades persones o col·lectius.

Continguts

Part I: Aspectes ètics de la intel·ligència artificial

1. Introducció: Per què els professionals de la IA hauríem d'estudiar ètica?
 - 1.1. Codi Ètic i de Conducta Professional de l'ACM.
 - 1.2. Marcs ètics (conseqüencialisme, teoria de la justícia, ètica de la virtut...).
 - 1.3. Principis ètics (equitat, responsabilitat, justícia, privacitat...).
2. Recollida de dades i privadesa
 - 2.1. La importància de la privadesa
 - 2.2. Principals tècniques per a la privadesa de dades (anonimat, xifrat, privadesa diferencial...).

2.3. Privadesa més enllà de les dades (en el context, per disseny...).

3. Algorismes, presa de decisions i biaixos

3.1. Definicions tècniques de biaix en resultats algorísmics.

3.2. Discriminació algorísmica directa i indirecta.

3.3. Definició d'equitat i mètriques d'equitat.

3.4. Representació del coneixement normatiu i ètic en IA.

3.5. Directrius ètiques per a una IA fiable: AI-Fairness Toolkits.

4. Explicabilitat

4.1. L'impacte sobre la responsabilitat i la rendició de comptes en els sistemes autònoms, centrant-nos en el cas dels vehicles autònoms.

4.2. La importància de les bones explicacions en els sistemes d'IA.

4.3. Eines per a avaluar l'explicabilitat.

Part II: Aspectes ètico-polítics de la intel·ligència artificial

5. Introducció: per què són rellevants els aspectes polítics i socials de la IA.

5.1. Teoria de la mediació tecnològica.

5.2. Narrativa al voltant de la IA i determinisme tecnològic

5.3. Innovació i recerca responsable (RRI)

6. Ètica i robòtica

6.1. Robots i societat.

6.2. Reptes ètics en robòtica.

6.3. Exemples de robòtica en l'àmbit quotidià.

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Assistència i participació activa a classe	30	1,2	3, 4, 5, 9, 10, 12, 13, 14
Casos d'estudis	50	2	1, 3, 5, 7, 8, 12, 15
Pràctiques i exercicis	50	2	2, 5, 6, 8, 9, 10, 13, 14, 16

L'orientació del curs és predominantment pràctica. Cada classe començarà generalment amb la presentació d'un cas d'estudi real, que donarà lloc a una discussió grupal. A continuació, s'introduiran i explicaran els conceptes, els mètodes o els sistemes d'IA relacionats amb els reptes ètics plantejats pel cas d'estudi.

Finalment, l'alumnat farà pràctiques individuals o grupals per reforçar el seu aprenentatge del contingut de la classe. En algunes sessions es reserva temps per repassar i corregir aquestes pràctiques. Algunes classes consistiran en visites a centres de recerca d'IA.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, perquè els alumnes completin les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura.

Avaluació

Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Pràctica avaluativa 1	34%	7	0,28	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16
Pràctica avaluativa 2	33%	7	0,28	1, 3, 4, 5, 6, 7, 8, 9, 10, 14, 16
Pràctica avaluativa 3	33%	6	0,24	4, 5, 9, 11, 12, 13, 14

L'avaluació de l'assignatura es pot dur a terme mitjançant dues modalitats: avaluació continuada o avaluació única.

Avaluació continuada

L'estudiant haurà de completar individualment tres activitats d'avaluació: dues corresponents a la Part I i una a la Part II.

- Activitat 1: prova escrita a classe amb preguntes curtes sobre la Part I i anàlisi d'un cas real d'aplicació d'un sistema d'IA.
- Activitat 2: ús de kits d'eines d'IA per avaluar i discutir mètriques ètiques d'un sistema d'IA.
- Activitat 3: prova escrita a classe sobre la Part II.

Per poder ser avaluat per aquesta modalitat, cal haver realitzat les tres activitats. Per superar l'assignatura, cal obtenir una nota mínima de 5 sobre 10 en cadascuna d'elles. En cas contrari, l'estudiant haurà de presentar-se a la recuperació. La nota final serà la mitjana aritmètica de les tres activitats d'avaluació.

Avaluació única

L'estudiant realitzarà un examen final individual al gener, estructurat en tres parts equivalents a les activitats de l'avaluació continuada, tant pel tipus de prova com en pes. Per aprovar, cal obtenir una nota mínima de 5 sobre 10 en cadascuna de les tres parts. La nota final serà la mitjana aritmètica de les tres qualificacions.

Recuperació

Podran accedir a la recuperació:

- Estudiants que hagin completat les tres activitats d'avaluació (avaluació continuada)
- Estudiants que s'hagin presentat a l'examen de gener (avaluació única)

La recuperació consistirà en un examen final individual. Per aprovar, cal obtenir una nota mínima de 5. Aquesta serà la nota final de l'assignatura.

Revisió de qualificacions

Després de cada activitat d'avaluació, el professorat informará a través de Moodle sobre les notes obtingudes i el procediment i la data per a la revisió.

Matrícula d'honor

S'atorgaran a estudiants amb una nota final de 10. Si hi ha més estudiants amb aquesta nota que matrícules d'honor assignades a aquesta assignatura, es realitzarà una prova addicional per determinar a qui s'atorguen.

No avaluable

L'estudiant rebrà la qualificació de "No avaluable" si no es presenta a més d'una activitat d'avaluació (avaluació continuada) o si no es presenta a l'examen de gener (avaluació única).

Estudiants repetidors

No es preveu cap tractament diferenciat per a estudiants repetidors.

Ús de la Intel·ligència Artificial (IA)

En aquesta assignatura, no es permet l'ús de tecnologies d'IA en cap de les seves fases. Qualsevol treball que inclogui fragments generats amb IA serà considerat una falta d'honestedat acadèmica i pot comportar una penalització parcial o total en la nota de l'activitat, o sancions majors en casos de gravetat.

Irregularitats

Qualsevol irregularitat que pugui alterar significativament la qualificació d'una activitat comportarà una nota de zero en aquesta activitat. En cas de múltiples irregularitats, la nota final de l'assignatura serà zero, independentment de qualsevol procés disciplinari.

Adaptació a format en línia

En el cas que les proves o exàmens no es puguin fer presencialment, s'adaptaran a un format en línia posat a disposició a través de les eines virtuals de la UAB (es mantindrà la ponderació original). Els deures, activitats i la participació a classe es realitzaran mitjançant fòrums, wikis i/o discussions a Teams, etc. Els professors vetllaran perquè els estudiants puguin accedir a aquestes eines virtuals o els oferiran alternatives factibles.

Bibliografia

1. Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
2. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center for Internet & Society.
3. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35.
4. van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, 30(3), 385-409.
5. van Wynsberghe, A. (2013). Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics*, 19(2), 407-433.
6. Vallès-Peris, N., & Domènech, M. (2020). Robotists' imaginaries of robots for care: The radical imaginary as a tool for an ethical discussion. *Engineering Studies*, 12(3), 156-176.
7. Vallès-Peris, N., & Domènech, M. (2021). Caring in the in-between: A proposal to introduce responsible AI and robotics to healthcare. *AI & Society*.
8. Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. *Science, Technology & Human Values*, 31(3), 361-380.

Programari

Per determinar (Part I).

Grups i idiomes de l'assignatura

La informació proporcionada és provisional fins al 30 de novembre de 2025. A partir d'aquesta data, podreu consultar l'idioma de cada grup a través daquest [enllaç](#). Per accedir a la informació, caldrà introduir el CODI de l'assignatura

Nom	Grup	Idioma	Semestre	Torn
(PAUL) Pràctiques d'aula	711	Anglès	primer quadrimestre	tarda
(TE) Teoria	71	Anglès	primer quadrimestre	tarda