

Ètica

Codi: 106559

Crèdits: 6

Titulació	Tipus	Curs	Semestre
2504392 Intel·ligència Artificial / Artificial Intelligence	FB	2	1

Professor/a de contacte

Nom: David Jorge Casacuberta Sevilla

Correu electrònic: david.casacuberta@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: anglès (eng)

Grup íntegre en anglès: Sí

Grup íntegre en català: No

Grup íntegre en espanyol: No

Equip docent

Miquel Domenech Argemi

Prerequisits

Aquesta assignatura no té prerequisits

Objectius

Aquesta assignatura té com a objectiu familiaritzar a l'alumnat del Grau amb aquells problemes ètics, socials i polítics que poden sorgir al emprar algorismes d'intel·ligència artificial per prendre decisions que afecten a les persones.

L'assignatura defuig així de problemes que són encara ben lluny a nivell tecnològic com la idea de la singularitat o la superintel·ligència i es centra més en problemes ètics reals que ens afecten com a persones aquí i ara.

Explorarem així qüestions com els biaxios en algorismes, l'erosió a la privacitat causada per la cerca de perfils personalitzats i com poden utilitzar-se per manipular les nostres decisions, però també analitzarem com aquests problemes poden detectar-se i solucionar-se presentant així marcs de referència com les auditories ètiques o explorar mètriques en algorismes que incloguin conceptes d'equitat i justícia.

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. Ser capaç d'incorporar els principis de la recerca i innovació responsable en els desenvolupaments basats en la IA.
12. Ser capaç d'incorporar valors adequats a les necessitats de les persones en el disseny de dispositius dotats d'IA.
13. Treballar de manera autònoma, amb responsabilitat i iniciativa, planificant i gestionant el temps i els recursos disponibles i adaptant-se a les situacions imprevistes.
14. Valorar com els estereotips i els rols de gènere incideixen en l'exercici professional.
15. 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

1. Introducció a l'ètica i els seus marcs teòrics

1.1 L'utilitarisme

1.2 L'ètica deontològica

1.3 Teoria de la justícia

1.4 Ètica de les virtuts

2. Privacitat i perfils personalitzats

2.1 Principis per la protecció de dades

2.2 Anonimització i reidentificació

2.3 Privacitat diferencial

2.4 perfils i manipulació

2.5 la bombolla dels filtres

3. Biaixos algorítmics

3.1 Causes dels biaixos en una base de dades

3.2 Com podem detectar biaixos

3.3 exactitud versus justícia

4. Auditories ètiques

4.1 Principis bàsics per una auditoria ètica

- 4.2 La matriu ètica
- 4.3 El pre-mortem
- 4.4 Algorismes i mètriques equitatives
- 5. Ciència, tecnologia i ètica
 - 5.1. Impacte social de les innovacions
 - 5.2. La mediació tecnològica
 - 5.3. La moral materialitzada
 - 5.4. Ètica i disseny
- 6. RRI i IA
 - 6.1. Què és la RRI?
 - 6.2. La RRI aplicada a la IA
- 7. Ètica i robòtica
 - 7.1. Els robots i la societat
 - 7.2. Preocupacions ètiques a la robòtica
 - 7.3. Robots assistencials/robots assassins

Metodologia

L'assignatura combinarà classes magistrals amb exercicis de debat i discussió en l'aula. Es treballarà des d'un projecte concret d'aplicació de la IA a l'esfera humana i s'estudiaran progressivament de quina forma la seva aplicació pot generar diferents problemes ètics i socials, cercant tant les causes d'aquests problemes com les possibles solucions.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Assistència i participació activa a classe	22	0,88	3, 4, 5, 9, 10, 11, 12, 13
Classes Magistrals	64	2,56	2, 5, 6, 8, 9, 10, 12, 13, 15
Seminaris	42	1,68	1, 3, 5, 7, 8, 11, 14

Avaluació

L'avaluació s'organitza al voltant de tres proves

1. Un estudi de cas on s'examinaran les implicacions ètiques d'un projecte específic de intel·ligència artificial així com les possibles solucions al problema que la seva aplicació pugui comportar (60%) .
2. Un treball grupal. Els alumnes hauran d'escriure un text que evidenciï que han adquirit els conceptes treballats a la segona part de l'assignatura. (30%)
3. La participació en classe en activitats de discussió i debat en grup (10%)

Podran presentar-se a reevaluació aquells estudiants que hagin suspès o no s'hagin presentat ni a la prova 1 o la prova 2. No cal haver-se presentat a cap prova per poder optar a la reevaluació. Una prova està aprovada amb una nota mínima de 5.

Els estudiants que finalment no es presentin ni a la prova 1 ni a la prova 2 constaran com a no avaluable.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Assistència i participació activa a classe	20%	2	0,08	4, 5, 10, 12, 13, 14, 15
Estudi de casos	50%	10	0,4	3, 6, 7, 9, 10, 11, 12
Treball en grup	30 %	10	0,4	1, 2, 4, 5, 8, 13, 14, 15

Bibliografia

Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI*. Oxford Handbooks.

Latour, B. (1999) La esperanza de Pandora. Ensayos sobre la realidad de los estudios de la ciencia. Barcelona: Gedisa, 2022.

O'neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway books.

Mepham, B. (2013). Ethical principles and the ethical matrix. Practical Ethics for Food Professionals: Ethics in Research, Education and the Workplace, 52.

Pariser, E. (2011). The filter bubble: How the new personalized web is changing what we read and how we think. Penguin.

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of internal medicine*, 169(12), 866-872.

Sparrow, R. (2007) 'Killer robots', *Journal of Applied Philosophy*, 24(1), pp. 62-77.

Vallès-Peris N and Domènech M (2020) Roboticians' Imaginaries of Robots for Care: The Radical Imaginary as a Tool for an Ethical Discussion. *Engineering Studies*, 12 (3): 156-176.

Vallès-Peris, N., Domènech, M. (2021) Caring in the in-between: a proposal to introduce responsible AI and robotics to healthcare. *AI & Society*.

van de Poel, I. (2020) 'Embedding Values in Artificial Intelligence (AI) Systems', *Minds and Machines*, 30(3), pp. 385-409.

van Wynsberghe, A. (2013) 'Designing Robots for Care: Care Centered Value-Sensitive Design', *Science and Engineering Ethics*, 19(2), pp. 407-433.

Verbeek, P.-P. (2006) 'Materializing Morality: Design Ethics and Technological Mediation', *Science, Technology & Human Values*, 31(3), pp. 361-380.

Programari

Llenguatge de programació Python

