

Filosofia de la intel·ligència artificial

Codi: 100315
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2500246 Filosofia	OT	3	0
2500246 Filosofia	OT	4	0

La metodologia docent i l'avaluació proposades a la guia poden experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries.

Professor/a de contacte

Nom: Maria Pilar Dellunde Clavé
Correu electrònic: Pilar.Dellunde@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: català (cat)
Grup íntegre en anglès: No
Grup íntegre en català: Sí
Grup íntegre en espanyol: No

Equip docent

Maria Pilar Dellunde Clavé

Prerequisits

L'assignatura no té cap prerequisit.

Objectius

Pot crear una màquina? Pot pensar, tenir emocions? La intel·ligència artificial és present en la nostra quotidianitat, en la ciència, en l'art, i malauradament també en la guerra. Quines preguntes i reflexions ens fem des de la filosofia? Necessitem lògiques alternatives per a la representació d'aquests nous coneixements? En aquest curs reflexionarem filosòficament sobre la intel·ligència artificial, sobre els límits ètics de la seva aplicació, les possibilitats de dissenyar una intel·ligència artificial general, i sobre les seves arquitectures cognitives. Una nova societat on persones i agents d'intel·ligència artificial convivim creant comunitats amb normes i potencialitats completament diferents de les que hem viscut fins ara, necessita tota la intel·ligència de les humanitats per fer front als nous reptes. T'apuntes a crear aquests alfabets de futur?

Competències

- Filosofia
- Desenvolupar un pensament i un raonament crítics i saber comunicar-los de manera efectiva, tant en les llengües pròpies com en una tercera llengua.
- Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
- Respectar la diversitat i la pluralitat d'idees, persones i situacions.

- Utilitzar la simbologia i els procediments de les ciències formals en l'anàlisi i la construcció d'arguments.
- Utilitzar un pensament crític i independent a partir dels temes, els debats i els problemes que planteja la filosofia tant històricament com conceptualment.

Resultats d'aprenentatge

1. Aplicar el rigor filosòfic en un text escrit seguint els estàndards internacionals de qualitat.
2. Construir arguments filosòfics amb rigor.
3. Debatre a partir del coneixement especialitzat adquirit en un context interdisciplinari
4. Demostrar una posició pròpia davant d'un problema o una controvèrsia de rellevància filosòfica, o en un treball de recerca filosòfica.
5. Dominar en el grau necessari els idiomes rellevants per a la pràctica professional.
6. Establir relacions entre ciència, filosofia, art, religió, política, etc.
7. Exposar els conceptes propis de la història de la filosofia
8. Formalitzar arguments de qualsevol procedència i calcular-ne la correcció lògica.
9. Formular objeccions i contraobjeccions amb correcció lèxica, precisió conceptual i coherència argumentativa.
10. Identificar les idees principals i les secundàries i expressar-les amb correcció lingüística.
11. Participar en debats sobre qüestions filosòfiques respectant les opinions dels altres participants
12. Reconèixer i utilitzar les diverses formes de raonament presents en la història de la filosofia.

Continguts

1. Què és la Intel·ligència Artificial (IA)? És possible una IA General? La IA en la literatura i el cinema de ciència ficció.
2. Deep Learning i els biaixos algorítmics.
3. Màquines, llenguatge, creativitat i emocions.
4. Representació del coneixement i raonament. Lògiques per a la IA (modal, temporal, epistèmica, deontica i difusa).
5. Ètica i robòtica social.

Metodologia

- Combinació de classes teòriques i pràctiques.
- Resolució conjunta d'exercicis.
- Simbolització d'arguments en llenguatge natural.
- Ús de la ciència-ficció per treballar la creativitat filosòfica.
- Activitats d'autoaprenentatge.
- Introducció de diferents nivells de dificultat en les pràctiques.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Representació de coneixement en intel·ligència artificial	50	2	2, 8, 12
Tipus: Supervisades			
Grup de lectura de literatura de ciència ficció i intel·ligència artificial	25	1	3, 4, 5, 6, 7, 10, 11
Seminari obert sobre robòtica social i límits ètics en l'aplicació de la intel·ligència artificial	25	1	3, 4, 6, 7, 10

Tipus: Autònomes			
Lectura textos recomanats	25	1	5, 6, 10, 12
Resolució de problemes	25	1	2, 8, 9

Avaluació

L'avaluació es planteja de forma continuada, i amb tres activitats avaluatives: dues proves de síntesi, i una sèrie d'exercicis en el sí d'un taller de creativitat filosòfica utilitzant la ciència ficció, que tindrà lloc al llarg de tot el curs. El primer test avaluarà els continguts sobre màquines de Turing, Deep Learning i biaixos algorítmics (punts 1-3 del programa) i valdrà un 40% de la nota; el segon test avaluarà els continguts de lògiques per a la intel·ligència artificial i robòtica social (punts 4-5 del programa) i valdrà un 40% de la nota. El exercicis fets al taller de creativitat filosòfica valdran un 20% de la nota.

Els dos tests es realitzaran, un al novembre, i l'altre al desembre. La tercera activitat es farà al llarg de tot el curs. En el moment de realització de cada activitat avaluativa, la professora informará l'alumnat (via Moodle) del procediment i data de revisió de les qualificacions.

Per participar a la recuperació l'alumnat haurà d'estar prèviament avaluat en les dues proves de síntesi, i haver obtingut un mínim de 3 punts (sumant els resultats de les dues proves). Es consideraran com 'No avaluables' només les persones que no hagin realitzat les tres activitats i tampoc s'hagin presentat a la recuperació.

En cas que l'estudiant realitzi qualsevol irregularitat que pugui conduir a una variació significativa de la qualificació d'un acte d'avaluació, es qualificarà amb 0 aquest acte d'avaluació, amb independència del procés disciplinari que s'hi pugui instruir. En cas que es produeixin diverses irregularitats en els actes d'avaluació d'una mateixa assignatura, la qualificació final d'aquesta assignatura serà 0.

En cas que les proves no es puguin fer presencialment s'adaptarà el seu format (mantenint-ne la ponderació) a les possibilitats que ofereixen les eines virtuals de la UAB. Els deures, activitats i participació a classe es realitzaran a través de fòrums, wikis i/o discussions d'exercicis a través de Teams, etc. El professor o professora vetllarà perquè l'estudiant hi pugui accedir o li oferirà mitjans alternatius, que estiguin al seu abast.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Definicions dels conceptes fonamentals.	40%	0	0	2, 8, 12
Exercicis lògiques no clàssiques	40%	0	0	2, 8
Taller de creativitat filosòfica	20%	0	0	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12

Bibliografia

1. Johan van Benthem, *Modal Logic for Open Minds*, Center for the Study of Language and Information, 2000.
2. Johan van Benthem, Hans van Ditmarsch, Jan van Eijck, Jan Jaspars, *Logic in Action*, Center for the Study of Language and Information, 2016.
3. Margaret A. Boden, *AI: Its nature and future*, Oxford University Press, 2016
4. Jack Copeland, *Artificial Intelligence: A Philosophical Introduction*, Wiley-Blackwell, 1993. Hi ha traducció al castellà a Alianza Editorial (1996).

5. Pedro Messeguer, Ramón López de Mantaras, *Inteligencia Artificial*, Editorial CSIC, 2017.
6. Nick Smith, *Vagueness and degrees of truth*, Oxford University Press, 2013.
7. Stuart Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd edition, Prentice Hall Press (2009) Traducció al castellà: *Inteligencia Artificial: Un enfoque moderno*, 2nd edition.
8. Carme Torras, *The Vestigial Heart*, MIT Press, 2018. La versió original catalana (sense els recursos sobre ètica i robòtica social) es: *Mutació Sentimental*, Pagès Editors, 2008.
9. Alan Turing, *Computing Machinery and Intelligence*, *Mind*, Issue 236, pp. 433-460, 1950. Hi ha traducció al castellà: *¿Puede pensar una máquina?*, KRK Ediciones, 2012.
10. Toby Walsh, *Machines That Think: The Future of Artificial Intelligence*. Prometheus Books, 2018.