

Titulació	Tipus	Curs
Filosofia	OT	3
Filosofia	OT	4

Professor/a de contacte

Nom: Jordi Vallverdú Segura

Correu electrònic: jordi.vallverdu@uab.cat

Idiomes dels grups

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

Prerequisits

L'assignatura no té cap prerequisit.

Objectius

"Filosofia de la Intel·ligència Artificial" és una assignatura pensada per a estudiants curiosos que volen anar més enllà dels tòpics sobre la IA. Al llarg del curs, analitzarem com van néixer les idees de computació, els algorismes i les màquines intel·ligents; com la lògica i els llenguatges formals han contribuït a donar forma al món digital; i què vol dir realment que un sistema entengui el llenguatge o simplement el simuli. També explorarem per què els grans models de llenguatge, com ChatGPT, plantegen nous desafiaments filosòfics, i com tot això afecta la societat, el coneixement i les decisions ètiques del present i del futur. No cal saber programar -només calen ganes de pensar fort i de posar la filosofia a treballar sobre la tecnologia real.

Competències

Filosofia

- 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.
- Utilitzar la simbologia i els procediments de les ciències formals en l'anàlisi i la construcció d'arguments.

Resultats d'aprenentatge

1. Construir arguments filosòfics amb rigor.
2. Establir relacions entre ciència, filosofia, art, religió, política, etc.
3. Exposar els conceptes propis de la història de la filosofia
4. Formalitzar arguments de qualsevol procedència i calcular-ne la correcció lògica.

5. Formular objeccions i contraobjeccions amb correcció lèxica, precisió conceptual i coherència argumentativa.
6. Reconèixer i utilitzar les diverses formes de raonament presents en la història de la filosofia.

Continguts

1. Mòdul I. Fonaments històrics i conceptuals

1. Càlcul, cens i control: de Leibniz a Babbage
Matemàtica com a llenguatge universal. Orígens administratius de la computació. Ada Lovelace i les primeres aportacions femenines.
2. Lògica, màquines i formalització
Frege, Boole, Peano, Hilbert. Crisi dels fonaments.
3. Turing, Church i la idea de computabilitat
Màquines de Turing, test de Turing, computació i ment.
4. Arquitectura de Von Neumann i computació clàssica
Model Von Neumann, instruccions, memòria, CPU. Comparació amb altres arquitectures.
5. Computació quàntica i implicacions filosòfiques
Qubits, superposició, porta NOT quàntica. Debats sobre significat i simulació.
6. Estadística i predicció en sistemes intel·ligents
Probabilitat, inferència bayesiana, regressió. Paper fonamental en IA moderna.

Mòdul II. Filosofia de la computació

1. Què és un algorisme?
Algorismes vs. heurístiques. Filosofia del programari.
2. Llenguatges de programació i lògica
Lisp, Prolog, llenguatge formal i semàntica.
3. Models computacionals de la ment
Funcionalisme, cognitivisme, crítiques de Dreyfus i Searle.

Mòdul III. Intel·ligència Artificial i filosofia

1. Història de la IA: de McCarthy a AlphaGo
IA simbòlica vs. subsimbòlica, IA forta i feble.
2. Ètica de la IA
Decisions automatitzades, opacitat, responsabilitat. Impactes de gènere en els sistemes d'IA.
3. Consciència i intencionalitat
Xatbots, zombis filosòfics i simulació de la ment.

Mòdul IV. Filosofia dels models de llenguatge (LLM)

1. Què és un LLM?
Arquitectura Transformer, embeddings, llenguatge com a estadística.
2. Crítiques filosòfiques als LLM
Bender, l'olor estocàstic, semàntica vs. sintaxi. Feminisme tecnocrític i representació de gènere en IA.

3. Epistemologia de la IA generativa
Coneixement, veritat i intenció en sistemes generatius.

Mòdul V. Perspectives contemporànies

1. Debats filosòfics sobre AGI (Artificial General Intelligence)
Possibilitat d'una intel·ligència general artificial. Arguments a favor i en contra. Posicions de Bostrom, Kurzweil, Russell, i crítica filosòfica (Floridi, Vallverdú, etc.).
2. Inspiració biològica i neurodivergència en IA
Neurodiverse AI, models híbrids d'inspiració biològica, plasticitat computacional. Intel·ligència no estàndard i inclusiva.
3. Debats sobre sistemes híbrids (neuro-symbolic AI)
Intel·ligència híbrida: cap a sistemes que combinen les fortaleces de la IA simbòlica (representació i raonament basats en regles) amb les capacitats d'aprenentatge de les xarxes neuronals i models d'aprenentatge automàtic. Debats filosòfics i tècnics sobre integració, explicabilitat, comprensió i límits epistemològics. i aprenentatge. Debats filosòfics i tècnics sobre integració i explicabilitat.
4. Debats filosòfics sobre AGI (Artificial General Intelligence)
Possibilitat d'una intel·ligència general artificial. Arguments a favor i en contra. Posicions de Bostrom, Kurzweil, Russell, i crítica filosòfica (Floridi, Vallverdú, etc.).
5. Cibercultura i tecnologies emergents
Realitat expandida, simulació, human enhancement.
6. Conclusions: cap a una filosofia crítica de la computació
Filosofia com a metatecnologia, recerques futures.

Activitats formatives i Metodologia

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Taller de creativitat filosòfica amb IA (experimentació amb prompting, rol de la IA com a autor)	40	1,6	1, 3, 5, 6
Tipus: Supervisades			
Estudi dels fonaments tècnics i filosòfics de la IA i la representació del coneixement (conceptes clau, supervisió docent)	44	1,76	1, 3, 4
Tipus: Autònomes			
Lectura de textos de Filosofia de la IA	34	1,36	1, 2, 6

La metodologia d'aquesta assignatura combina classes magistrals, seminaris participatius i activitats pràctiques guiades, amb un fort èmfasi en l'anàlisi filosòfica activa i l'ús declarat d'intel·ligència artificial com a eina epistemològica.

- Les classes magistrals introduiran els conceptes fonamentals i el context històric-filosòfic.
- A través de lectures comentades setmanals, els estudiants exercitaran la lectura crítica, l'anàlisi conceptual i el debat col·lectiu.

- El treball escrit, elaborat amb ChatGPT, fomentarà l'experimentació filosòfica amb eines d'IA, tot exigint transparència en el prompting i reflexió sobre la seva aplicació.
- Els exàmens avaluaran la comprensió profunda del contingut i la capacitat d'aplicar-lo de forma argumentada.

L'assignatura promou una participació activa, una actitud crítica i una mirada transversal sobre el paper de la computació i la intel·ligència artificial en la societat contemporània.

En aquesta assignatura es permet l'ús de tecnologies d'Intel·ligència Artificial (IA) com a part integrant del desenvolupament del treball, sempre que el resultat final reflecteixi una contribució significativa de l'estudiant en l'anàlisi i la reflexió personal. L'estudiant haurà de: (i) identificar quines parts han estat generades amb IA; (ii) especificar les eines utilitzades; i (iii) incloure una reflexió crítica sobre com aquestes han influït el procés i el resultat final de l'activitat.

La no transparència de l'ús de la IA en aquesta activitat avaluable es considerarà falta d'honestedat acadèmica i comporta que l'activitat s'avaluï amb un 0 i no es pugui recuperar, o sancions majors en casos de gravetat.

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
Examen final: síntesi crítica	25	8	0,32	1, 2, 3, 5, 6
Examen parcial: anàlisi i aplicació filosòfica	25	8	0,32	1, 2, 5
Lectures comentades i debat filosòfic	25	8	0,32	1, 2, 4, 6
Treball escrit amb IA (ChatGPT)	25	8	0,32	1, 3, 4, 5

1. Lectures comentades setmanals (25%)

- Els estudiants llegiran textos filosòfics clau (Turing, Searle, Bender, etc.) i hauran de participar activament en discussions guiades a classe.
- Es valorarà:
 - Comprensió del text
 - Capacitat d'anàlisi i argumentació
 - Participació en el debat col·lectiu

2. Treball escrit com ChatGPT (25%)

- Treball individual (2.500-4.000 paraules) fet com ChatGPT.
- Caldrà incloure el thread complet com a annex.

- Es valorarà:
 - Capacitat crítica per analitzar, complementar o corregir respostes de la IA
 - Nivell, densitat i refinament del prompting
 - Claredat filosòfica i reflexió pròpia
 - Transparència i justificació del procés d'interacció amb la IA

3. Examen parcial (25%)

- Prova escrita presencial a mitjan curs.
- Preguntes obertes, definicions, comparacions i anàlisi de casos filosòfics.

4. Examen final (25%)

- Prova final presencial amb preguntes obertes o comentaris crítics de textos

L'estudiant rebrà la qualificació de "No avaluable" sempre que no hagi participat en una de les 4 de les activitats d'avaluació.

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, i aquest acte d'avaluació no serà recuperable. En cas que es produeixin diverses irregularitats en els actes d'avaluació d'una mateixa assignatura, la qualificació final d'aquesta assignatura serà 0.

Recuperació

- Cal haver presentat almenys dues de les activitats per poder fer la recuperació.
- Consistirà en un examen global.

És possible que el Departament de Filosofia estableixi (tal com s'ha fet durant el primer semestre) un període de concentració de proves avaluatives. El professorat indicarà si hi ha aquest període o quines són les dates de les proves a l'inici de cada assignatura.

Avaluació única

Els estudiants que optin per l'avaluació única hauran de realitzar un examen exhaustiu al final del semestre. Aquesta prova avaluativa tindrà per objectiu mesurar la comprensió global i crítica dels continguts de l'assignatura.

L'examen inclourà:

- Diverses preguntes (tant de resposta llarga com d'anàlisi argumentativa) sobre tots i cadascun dels temes del temari.
- Es farà especial èmfasi en:
 - Els elements filosòfics (ontològics, epistemològics i metafísics vinculats a la IA).
 - Les qüestions ètiques i socials derivades de l'aplicació de sistemes intel·ligents.
 - Els aspectes lògics i computacionals rellevants, incloent-hi conceptes com algorismes, arquitectures computacionals, xarxes neuronals i lògica formal aplicats a la IA.

L'examen serà presencial, tindrà una durada de dues hores i mitja, i es durà a terme dins del període oficial d'avaluacions finals.

Per garantir l'equitat i el rigor, aquesta prova única cobrirà la totalitat de les competències i resultats d'aprenentatge establerts a la guia docent, en condicions equivalents a l'avaluació continuada.

Bibliografia

Filosofia de la computació i IA

- Boden, M. A. (2016). *AI: Its Nature and Future*. Oxford University Press.
- Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Wiley-Blackwell.
- Dennett, D. (1991). *Consciousness Explained*. Little, Brown and Company.
- Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.
- Searle, J. (1980). *Minds, Brains and Programs*. *Mind*, 89(4), 417-424.
- Smith, B. C. (1996, 2019). *On the Origin of Objects / The Promise of Artificial Intelligence*. MIT Press.
- Vallverdú, J. (2024). *Causality for Artificial Intelligence: From a Philosophical Perspective*. Springer.

Ètica i societat

- Bender, E. et al. (2021). *On the Dangers of Stochastic Parrots*. FAccT.
- Coeckelbergh, M. (2018). *Ética de la inteligencia artificial*. Cátedra.
- Coeckelbergh, M. (2022). *Robo Ethics*. MIT Press.
- Crawford, K. (2021). *Atlas of AI*. Yale University Press.

Lògica, epistemologia i llenguatge

- Smith, N. J. J. (2013). *Vagueness and Degrees of Truth*. Oxford University Press.
- Van Benthem, J. (2000). *Modal Logic for Open Minds*. CSLI.
- Van Benthem, J., Van Ditmarsch, H., Van Eijck, J., Jaspars, J. (2016). *Logic in Action*. CSLI.

Programari

Cap programari.

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	1	Català	primer quadrimestre	matí-mixt
(TE) Teoria	1	Català	primer quadrimestre	matí-mixt