

**Filosofía de la inteligencia artificial**

Código: 100315  
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2500246 Filosofía	OT	3	0
2500246 Filosofía	OT	4	0

**Contacto**

Nombre: María Pilar Dellunde Clavé  
Correo electrónico: Pilar.Dellunde@uab.cat

**Uso de idiomas**

Lengua vehicular mayoritaria: catalán (cat)  
Algún grupo íntegramente en inglés: No  
Algún grupo íntegramente en catalán: Sí  
Algún grupo íntegramente en español: No

**Equipo docente**

María Pilar Dellunde Clavé

**Prerequisitos**

Ninguno.

**Objetivos y contextualización**

¿Puede crear una máquina? ¿Puede pensar, tener emociones? La inteligencia artificial está presente en nuestra cotidianidad, en la ciencia, en el arte, y desgraciadamente también en la guerra. ¿Qué preguntas y reflexiones nos hacemos desde la filosofía? ¿Necesitamos lógicas alternativas para la representación de estos nuevos conocimientos?

En este curso reflexionaremos filosóficamente sobre la inteligencia artificial, sobre los límites éticos de su aplicación, las posibilidades de diseñar una inteligencia artificial general, y sobre sus arquitecturas cognitivas. Una nueva sociedad donde personas y agentes de inteligencia artificial convivimos creando comunidades con normas y potencialidades completamente diferentes de las que hemos vivido hasta ahora, necesita toda la inteligencia de las humanidades para hacer frente a los nuevos retos ¿Te apuntas a crear estos alfabetos de futuro?

**Competencias**

- Filosofía
- Desarrollar un pensamiento y un razonamiento crítico y saber comunicarlos de manera efectiva, tanto en las lenguas propias como en una tercera lengua.
  - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
  - Respetar la diversidad y pluralidad de ideas, personas y situaciones.

- Utilizar la simbología y procedimientos de las ciencias formales en el análisis y construcción de argumentos.
- Utilizar un pensamiento crítico e independiente a partir de los temas, los debates y los problemas que plantea la filosofía tanto histórico como conceptualmente.

## Resultados de aprendizaje

1. Aplicar el rigor filosófico en un texto escrito siguiendo los estándares internacionales de calidad.
2. Construir argumentos filosóficos con rigor.
3. Debatir a partir del conocimiento especializado adquirido en un contexto interdisciplinar
4. Demostrar una posición propia ante un problema o controversia de relevancia filosófica, o en un trabajo de investigación filosófica.
5. Dominar en el grado necesario los idiomas relevantes en la práctica profesional.
6. Establecer relaciones entre ciencia, filosofía, arte, religión, política, etc.
7. Exponer los conceptos propios de la historia de la filosofía
8. Formalizar argumentos de cualquier procedencia y calcular su corrección lógica.
9. Formular objeciones y contraobjeciones con corrección léxica, precisión conceptual y coherencia argumentativa.
10. Identificar las ideas principales y las secundarias y expresarlas con corrección lingüística.
11. Participar en debates sobre cuestiones filosóficas respetando las opiniones de otros participantes
12. Reconocer y utilizar las diversas formas de razonamiento presentes en la historia de la filosofía.

## Contenido

1. ¿Qué es la Inteligencia Artificial (IA)? Es posible una IA General? La IA en la literatura y el cine de ciencia ficción.
2. Deep Learning y los sesgos algorítmicos.
3. Máquinas, lenguaje, creatividad y emociones.
4. Representación del conocimiento y razonamiento. Lógicas para la IA (modal, temporal, epistémica, deóntica y difusa).
5. Ética y robótica social.

## Metodología

- Combinación de clases teóricas y prácticas.
- Resolución conjunta de ejercicios en clase.
- Simbolización de argumentos en lenguaje natural.
- Uso de la ciencia ficción para trabajar la creatividad filosófica.
- Actividades de autoaprendizaje opcionales.
- Introducción de diferentes niveles de dificultad en las prácticas.

## Actividades

Título	Horas	ECTS	Resultados de aprendizaje
--------	-------	------	---------------------------

Tipo: Dirigidas

Seminario abierto sobre robótica social y límites éticos en la aplicación de la inteligencia artificial	50	2	2, 8, 12
Tipo: Supervisadas			
Ejercicios	25	1	3, 4, 5, 6, 7, 10, 11
Lectura de textos recomendados	25	1	3, 4, 6, 7, 10
Tipo: Autónomas			
Grupo de lectura sobre ciencia ficción e inteligencia artificial	25	1	2, 8, 9
Representación del conocimiento en IA	25	1	5, 6, 10, 12

## Evaluación

La evaluación se considera de forma continua y con tres actividades de evaluación: dos pruebas de síntesis y una serie de ejercicios dentro de un taller de creatividad filosófica que utilizará la ciencia ficción, teniendo lugar a lo largo de todo el curso. La primera prueba evaluará el contenido de máquinas de Turing, Deep Learning y el sesgo algorítmico (puntos 1-3 del programa) y tendrá un valor del 40% de la nota; La segunda prueba evaluará el contenido de lógicas para la inteligencia artificial y robótica social (puntos 4-5 del programa) y valorará el 40% de la nota. Los ejercicios realizados en el taller de creatividad filosófica valdrán el 20% de la nota.

Ambas pruebas se llevarán a cabo, una en la primera quincena de noviembre, y la otra en la segunda quincena de diciembre. La tercera actividad se llevará a cabo a lo largo de todo el curso, siendo necesario realizar al menos 10 prácticas en clase para poder ser evaluados en el taller de creatividad. En el momento de realizar cada actividad de evaluación, el profesor informará a los estudiantes (a través de Moodle) del procedimiento y la fecha de revisión de las calificaciones.

Para poder participar en la recuperación, se deben presentar a evaluación en las dos pruebas de síntesis y obtener un mínimo de 3 puntos (añadiendo los resultados de las dos pruebas). "No evaluable" se considerarán sólo aquellas personas que no hayan realizado ninguna de las tres actividades y no se han presentado a la recuperación.

## Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Definiciones de los conceptos fundamentales en IA. Ejercicios de lógicas no clásicas.	80%	0	0	2, 8, 12
Taller de creatividad filosófica	20%	0	0	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12

## Bibliografía

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.
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.
8. Carme Torras, *The Vestigial Heart*, MIT Press, 2018.
9. Alan Turing, *Computing Machinery and Intelligence*, *Mind*, Issue 236, pp. 433-460, 1950.
10. Toby Walsh, *Machines That Think: The Future of Artificial Intelligence*. Prometheus Books, 2018.