

Degree programme	Type	Course
Philosophy	OP	3

Contact lecturer

Name : David Jorge Casacuberta Sevilla

Email : david.casacuberta@uab.cat

Teaching staff

Oscar Castro Garcia

Group languages

You can consult this information at the [end](#) of the document.

Prerequisites

None.

Objectives

Philosophy of Artificial Intelligence aims primarily to introduce students to a rigorous philosophical analysis of the foundations, implications and challenges posed by artificial intelligence as a technological, epistemic and social phenomenon. Through an itinerary combining philosophy of mind, ethics, epistemology and philosophy of technology, the course seeks to enable students to understand the central questions articulated by AI: What is intelligence? Can a machine think or understand? What are the conditions for a fair and responsible AI? What impact does it have on our conception of the subject, knowledge and society? The emphasis is not merely descriptive or technical, but genuinely philosophical: the aim is to examine arguments, concepts, and theoretical frameworks with the analytical rigor proper to the discipline.

The second part of the course will include an analysis of various texts every two sessions, as well as seminars on topics introduced by groups of students, with interaction from expert professors in the subject from the Department. The aim is to generate a circle of diverse studies-cross-cutting, interdisciplinary, and transdisciplinary-that will be part of the exam material.

Learning outcomes

- CM31 (Apply critical and reflective thinking to address contemporary problems linked to the mind, artificial intelligence, and technoscience.) Apply critical and reflective thinking to address contemporary

- problems linked to the mind, artificial intelligence, and technoscience.
- CM32 (Adopt an ethical and responsible attitude in the analysis and discussion of cognitive and technological transformations in today's society in response to situations of discrimination based on gender, race or class, such as the digital divide.) Adopt an ethical and responsible attitude in the analysis and discussion of cognitive and technological transformations in today's society in response to situations of discrimination based on gender, race or class, such as the digital divide.
 - CM33 (Integrate interdisciplinary approaches for understanding and dialogue in the face of philosophical, social and ethical implications of scientific and technological advances.) Integrate interdisciplinary approaches for understanding and dialogue in the face of philosophical, social and ethical implications of scientific and technological advances.
 - CM39 (Assess the ethical challenges posed by artificial intelligence, including issues such as privacy, discrimination, and moral responsibility.) Assess the ethical challenges posed by artificial intelligence, including issues such as privacy, discrimination, and moral responsibility.
 - KM35 (Identify the diversity of interdisciplinary approaches that integrate philosophy, neuroscience, technology, and social studies in the analysis of cognitive and technological processes.) Identify the diversity of interdisciplinary approaches that integrate philosophy, neuroscience, technology, and social studies in the analysis of cognitive and technological processes.
 - KM36 (Recognise the epistemological, ontological and ethical foundations that underlie the development of cognitive sciences, artificial intelligence and technoscience.) Recognise the epistemological, ontological and ethical foundations that underlie the development of cognitive sciences, artificial intelligence and technoscience.
 - KM37 (Explain the main contemporary philosophical theories and debates on the mind, cognition, consciousness, artificial intelligence and the relationships between science, technology and society.) Explain the main contemporary philosophical theories and debates on the mind, cognition, consciousness, artificial intelligence and the relationships between science, technology and society.
 - KM48 (Identify the different political, health, environmental, identity, and global problems of the present based on readings of reference authors and case studies.) Identify the different political, health, environmental, identity, and global problems of the present based on readings of reference authors and case studies.
 - SM36 (Critically analyse philosophical and scientific theories and models about mind, consciousness, artificial intelligence, and science and technology.) Critically analyse philosophical and scientific theories and models about mind, consciousness, artificial intelligence, and science and technology.
 - SM37 (Develop rigorous philosophical arguments that integrate ethical, social, and epistemological perspectives around technological and scientific advances.) Develop rigorous philosophical arguments that integrate ethical, social, and epistemological perspectives around technological and scientific advances.
 - SM38 (Communicate ideas related to the intersection of philosophy, science, and technology in academic and social contexts.) Communicate ideas related to the intersection of philosophy, science, and technology in academic and social contexts.

Contents

FIRST PART

1.Introduction: what do we mean by AI

1.1 Key concepts

1.2 Types of AI

1.3 Economic, social and political foundations behind AI

2.Epistemology of AI

2.1 AI as a design science

2.2 Artificial agency

2.3 Will AI replace classical science?

3.Ethics of AI

3.1 99 ethical problems associated with AI

3.2 Value alignment

3.3 The singularity

3.4 Can AI generate ethical reasoning?

4.Philosophy of Mind and AI

4.1 The misleading term "artificial intelligence"

4.2 The Turing test and its limits

4.3 Can AI be conscious?

SECOND PART

5.Philosophy of Information

5.1 Information and Intelligence

5.2 Meta Learning and Meta-Intelligence

Cybernetics and AI

6.Cybernetics and AI

6.1 From Cybersecurity to Biomechanics

6.2 The Symbiotic Human and the Cyborg Manifesto

7.Counterculture: Cyberpunk

7.1 From Virtual Reality to The Matrix: Philosophy of Science Fiction (Philosophy-Fiction)

7.2 The Cybernetic Culture Research Unit (CCRU)

8.Transhumanism and the Future of Humanity

8.1 From the Transhuman to Meta-Intelligence: Human Self-Awareness versus Artificial Metacognition

8.2 Blueprint: A Deontological Manifesto for AI in Academic Studies

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
Debates on the key issues of AI	20	0.8	CM31, CM32, CM33, CM39, KM35, KM36, KM37, KM48, SM36, SM37, SM38
Reading texts of Philosophy of the AI	34	1.36	CM31, CM32, KM35, KM36, SM37, SM38
Study of the technical and philosophical foundations of AI and knowledge representation (key concepts, faculty supervision)	64	2.56	CM33, CM39, KM37, KM48, SM36

The methodology of this course combines lectures and participatory seminars. Department professors will also take part, willingly (and where possible), in events organized by and for the course (round tables, workshops, etc.). The course promotes active participation, a critical attitude, and a cross-cutting perspective on the role of computing and artificial intelligence in contemporary society.

The use of Artificial Intelligence (AI) technologies is permitted in this course as an integral part of the development of coursework, provided that the final result reflects a significant contribution from the student in terms of analysis and personal reflection. Students must: (i) identify which parts have been generated with AI; (ii) specify the tools used; and (iii) include a critical reflection on how these tools have influenced the process and the final outcome of the activity.

Failure to disclose the use of AI in this assessable activity will be considered a breach of academic integrity and will result in the activity being graded with a 0, with no opportunity for resubmission, or more severe sanctions in serious cases.

Students are requested to use generative AI consciously in developing their activities, and only when explicitly designated by the professor.

Annotation: within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continuous assessment activities

Title	Weight	Hours	ECTS	Learning outcomes
Drafting of a collective manifesto	25	8	0.32	CM31, CM32, CM33, CM39, KM35, KM36, KM37, KM48, SM36, SM37, SM38
Final exam: critical synthesis	25	8	0.32	CM32, CM33, KM36, KM48
Midterm exam: philosophical analysis and application	25	8	0.32	CM31, CM39, KM36, KM48, SM36, SM37
Commented readings and philosophical debate	25	8	0.32	CM32, CM33, KM35, KM37

1. Weekly Guided Readings (25%)

Students will read key philosophical texts (Turing, Searle, Bender, etc.) and are expected to actively participate

in guided class discussions.

Assessment criteria:

- Text comprehension
- Analytical and argumentative skills
- Participation in collective debate

2. Written Assignment with ChatGPT (25%)

Individual paper (2,500-4,000 words) developed using ChatGPT.

The full conversation thread must be included as an appendix.

Assessment criteria:

- Critical ability to analyze, complement, or correct AI responses
- Prompting level, density, and sophistication
- Philosophical clarity and original reflection
- Transparency and justification of the interaction process with the AI

3. Midterm Exam (25%)

In-person written test held mid-course.

Open-ended questions, definitions, comparisons, and analysis of philosophical cases.

4. Final Exam (25%)

Final in-person exam with open-ended questions or critical commentary on texts.

It is possible that the Department of Philosophy will establish (as will be done during the first semester) a period dedicated to evaluative tests. The teaching staff will indicate whether such a period exists or what the test dates are at the beginning of each course.

Single Assessment

Students who choose the single assessment option will be required to complete a comprehensive final exam at the end of the semester. This exam is designed to assess their overall understanding and critical engagement with the course content.

The exam will include:

- Multiple questions (both long-form and analytical) covering each and every topic from the syllabus.
- Particular emphasis will be placed on:
 - Philosophical dimensions (ontological, epistemological, and metaphysical issues related to AI),
 - Ethical and social questions arising from the deployment of intelligent systems,
 - Logical and computational aspects, including concepts such as algorithms, computational architectures, neural networks, and formal logic as they relate to AI.

The exam will be in person, will last two and a half hours, and will be held during the official final assessment period.

This single assessment will cover the full range of competencies and learning outcomes stated in the course guide, under conditions equivalent to those of continuous assessment.

Bibliography

Bostrom, N. (2016). Superinteligencia: Caminos, peligros, estrategias. Teell Editorial.

Chang, R. (Ed.). (1997). Incommensurability, Incomparability, and Practical Reason. Harvard University Press.

Coeckelbergh, M. (2021). Ética de la inteligencia artificial. Cátedra.

Copeland, J. (2007). Inteligencia artificial: Una introducción filosófica. Alianza Editorial.

Crawford, K. (2022). Atlas de la IA: Poder, política y costes planetarios de la inteligencia artificial. Agua Clara.

Crawford, K. (2023). Atlas de Inteligencia Artificial Poder, Política y Costos -- Kate Crawford; Francisco Diaz Klassen.

Davis, E. (2023). Tecgnosis. Caja Negra Editora.

Domingos, P. (2015). El algoritmo maestro: Cómo el aprendizaje automático está cambiando el mundo. Planeta.

Floridi, L. (2013). The philosophy of information. OUP Oxford.

Floridi, L. (2022). Ética de la inteligencia artificial: Principios, desafíos y oportunidades. Alianza Editorial.

Giordano, F. F. (ed.) (2023). Cultura Cibernética y Otros Escritos del CCRU. Holobionte Ediciones.

Giordano, F. F. (ed.) (2024). Más Allá de las Pantallas. Y otros ensayos de Sadie Plant 1990-2000. Holobionte Ediciones. Colección Antifuturos.

Haraway, D. (2025/1991). Manifiesto Ciborg (5.ª ed.). Kaotica Libros.

Hofstadter, D. R. (2008). Gödel, Escher, Bach: Un eterno y grácil bucle. Tusquets.

Ireland, A. (2022). Filosofía-Ficción: Inteligencia Artificial, tecnología oculta y el fin de la humanidad.

Holobionte Ediciones.

Jarrahi, M. H., et al. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. Business Horizons, 66(1), 87-99.

Kurzweil, R., & Galmarini, M. A. (2000). La era de las máquinas espirituales. Planeta.

León XIV. (2026). Magnifica Humanitas: Sobre la custodia de la persona humana en el tiempo de la inteligencia artificial. Vaticano.

Russell, S. (2022). Compatible: Inteligencia artificial y el futuro del ser humano. Debate.

Russell, S. J., & Norvig, P. (2022). Artificial Intelligence: A Modern Approach. Pearson Education Limited.

Searle, J. R. (1992). El redescubrimiento de la mente. Crítica.

Simon, H. A. (1996). Las ciencias de lo artificial (3.ª ed.). Comares.

Tegmark, M. (2018). Vida 3.0: Ser humano en la era de la inteligencia artificial. Paidós.

Vallverdú, J. (2024). Causality for Artificial Intelligence: From a Philosophical Perspective. Springer.

Wooldridge, M. (2009). An Introduction to MultiAgent Systems (2nd ed.). Wiley.

Zuboff, S. (2020). La era del capitalismo de la vigilancia. Paidós.

Software

None.

Course groups and languages

The information provided is provisional until November 30. After this date, you will be able to consult the language of each group through this [link](#). To access the information, you will need to enter the course CODE

Type of teaching	Group	Language	Semester	Shift
(TE) Theory	1	Catalan	second	morning-mixed

semester				
(PAUL) Classroom practices	1	Catalan	second semester	morning-mixed