Universitat Autònoma de Barcelona

2019/2020

Philosophy of Artificial Intelligence

Code: 100315
ECTS Credits: 6

Degree	Type	Year	Semester
2500246 Philosophy	OT	3	0
2500246 Philosophy	OT	4	0

Contact

Name: María Pilar Dellunde Clavé
Email: Pilar.Dellunde@uab.cat

Use of Languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

Teachers

María Pilar Dellunde Clavé

Prerequisites

None.

Objectives and Contextualisation

Can a machine be creative? Can a machine think, have emotions? Artificial intelligence is present in our daily life, in science, in art, and unfortunately also in war. What are the main issues to be addressed from a philosophical point of view regarding artificial intelligence? Do we need alternative logics for the representation of these new knowledge?

In this course we will discuss the ethical limits of the application of artificial intelligences, and the possibilities of designing a general artificial intelligence. A new society where people and agents of artificial intelligence coexist creating communities with completely different norms and potentials than the ones we have lived up to now. All the intelligence of the humanities is needed to face these new challenges. Would you like to join us in the creation of these alphabets for the future?

Competences

Philosophy

- Developing critical thinking and reasoning and communicating them effectively both in your own and other languages.
- Respecting the diversity and plurality of ideas, people and situations.
- Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
- Thinking in a critical and independent manner on the basis of the specific topics, debates and problems of philosophy, both historically and conceptually.

- Using the symbology and procedures of the formal sciences in the analysis and building of arguments.

Learning Outcomes

1. Applying philosophical rigour in a written text following the international quality standards.
2. Demonstrating a personal stance over a problem or controversy of philosophical nature, or a work of philosophical research.
3. Engaging in debates about philosophical issues respecting the other participants' opinions.
4. Establishing relationships between science, philosophy, art, religion, politics, etc.
5. Explaining the specific notions of the History of Philosophy.
6. Formulating arguments for and against an issue, using proper vocabulary, conceptual precision and argumentative coherence.
7. Identifying the main and secondary ideas and expressing them with linguistic correctness.
8. Mastering the relevant languages to the necessary degree in the professional practice.
9. Recognising and using the several forms of reasoning in the history of philosophy.
10. Regularising arguments of any source and calculating its logical correctness.
11. Rigorously building philosophical arguments.
12. Using specialized knowledge acquired in an interdisciplinary context when debating.

Content

1. What is Artificial Intelligence (AI)? Is A General AI Possible? The AI in science fiction literature and cinema.

2. Deep Learning and algorithmic biases.

3. Machines, language, creativity and emotions.

4. Representation of knowledge and reasoning. Logics for AI (modal, temporal, epistemic, deontic, and fuzzy).

5. Ethics and social robotics.

Methodology

- Combination of theoretical and practical classes.
- Joint resolution of exercises in class.
- Simbolization of arguments in natural language.
- Use of science fiction to work on philosophical creativity.
- Optional self-learning activities.
- Introduction of different levels of difficulty in the practices.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Open seminar on social robotics and ethical limits in the application of artificial intelligence	50	2	11, 10, 9

Type: Supervised			
Exercises	25	1	12, 2, 8, 4, 5, 7, 3
Reading recommended texts	25	1	12, 2, 4, 5, 7
Type: Autonomous			
Literature reading group on science fiction and artificial intelligence	25	1	11, 10, 6
Representation of knowledge in artificial intelligence	25	1	8, 4, 7, 9

Assessment

The evaluation is continuous, and with three evaluation activities: two synthesis tests, and a series of exercises within a workshop of philosophical creativity using science fiction, which will take place throughout the course. The first test will evaluate the contents of Turing machines, Deep Learning and Algorithmic Bias (points 1-3 of the program) and will be worth 40% of the note. The second test will evaluate the content of Logics for Artificial Intelligence and Social Robotics (points 4-5 of the program) and will be worth 40% of the note. The exercises made in the workshop of philosophical creativity will be worth 20% of the note.

The two tests will be carried out, one the first fortnight of November, and the other one the second fortnight of December. The third activity will be done throughout the course, it will be necessary to do at least 10 practices in class to be able to be evaluated in the workshop of creativity. At the time of carrying out each evaluation activity, the teacher will inform the students (via Moodle) of the procedure and date of review of the qualifications.

In order to participate in the recovery exam students must first be evaluated in the two synthesis tests, and have obtained a minimum of 3 points (adding the results of the two tests). "Non-evaluable" will be considered only those who have not performed any of the three activities and have not been submitted for the recovery exam.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Definitions of the main AI concepts. Non-classical logics exercises.	80%	0	0	11, 10, 9
Philosophical creativity workshop	20%	0	0	1, 11, 12, 2, 8, 4, 5, 6, 7, 3, 9

Bibliography

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.