

Degree	Type	Year
Artificial Intelligence	FB	2

Contact

Name: Maria Pilar Dellunde Clave

Email: pilar.dellunde@uab.cat

Teachers

Nuria Valles Peris

(External) Roger Deulofeu Batllori

Teaching groups languages

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

Prerequisites

No prerequisites

Objectives and Contextualisation

Drawing from real-world case studies, this course is designed to instill in students awareness of the ethical and societal implications of artificial intelligence (AI). It provides comprehensive instruction on incorporating strategies and utilizing tools to minimize ethical risks while fostering the development of AI systems within the framework of responsible AI.

Competences

- Act with ethical responsibility and respect for fundamental rights and duties, diversity and democratic values.
- Act within the field of knowledge by evaluating sex/gender inequalities.
- Communicate effectively, both orally and in writing, adequately using the necessary communicative resources and adapting to the characteristics of the situation and the audience.
- Conceive, design, analyse and implement autonomous cyber-physical agents and systems capable of interacting with other agents and/or people in open environments, taking into account collective demands and needs.

- Develop critical thinking to analyse alternatives and proposals, both one's own and those of others, in a well-founded and argued manner.
- Identify, analyse and evaluate the ethical and social impact, the human and cultural context, and the legal implications of the development of artificial intelligence and data manipulation applications in different fields.
- Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
- Work independently, with responsibility and initiative, planning and managing time and available resources, and adapting to unforeseen situations.

Learning Outcomes

1. Analyse AI application cases from an ethical, legal and social point of view.
2. Analyse sex/gender inequalities and gender bias in the field of knowledge.
3. Communicate effectively, both orally and in writing, adequately using the necessary communicative resources and adapting to the characteristics of the situation and the audience.
4. Critically analyse the principles, values and procedures that govern the practice of the profession.
5. Develop critical thinking to analyse alternatives and proposals, both one's own and those of others, in a well-founded and argued manner.
6. Evaluate how stereotypes and gender roles affect the professional exercise.
7. Evaluate the difficulties, prejudices and discriminations that can be found in actions or projects, in a short or long term, in relation to certain people or groups.
8. Explain the code of ethics, explicit or implicit, that pertains to the field of knowledge.
9. Identify the main sex- and gender-based inequalities and discrimination present in society today.
10. Identify the social, cultural and economic biases of certain algorithms.
11. Incorporate the principles of responsible research and innovation in AI-based developments.
12. Incorporate values appropriate to people's needs when designing AI-enabled devices.
13. Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
14. Understand the social, ethical and legal implications of professional AI practice.
15. Work in teams to design interdisciplinary projects. Be able to collaborate with non-professionals and professionals from other sectors.
16. Work independently, with responsibility and initiative, planning and managing time and available resources, and adapting to unforeseen situations.

Content

Part I: Ethical Aspects of Artificial Intelligence

1. Introduction: Why should AI professionals study ethics
 - 1.1. ACM Code of Ethics and Professional Conduct
 - 1.2. Ethical frameworks (consequentialism, theory of justice, virtue ethics...)
 - 1.3. Ethical principles (fairness, responsibility, justice, privacy...)
2. Data collection and privacy
 - 2.1. The importance of privacy
 - 2.2. Main techniques for data privacy (anonymity, encryption, differential privacy...)
 - 2.3. Privacy beyond data (in context, by design...)
3. Algorithms, decision-making, and biases

- 3.1. Technical definitions of bias in algorithmic outcomes
- 3.2. Direct and indirect algorithmic discrimination
- 3.3. Definition of fairness and fairness metrics
- 3.4. Representation of normative and ethical knowledge in AI
- 3.5. Ethical guidelines for reliable AI: AI-Fairness Toolkits
- 4. Explainability
 - 4.1. The impact on responsibility and accountability in autonomous systems, focusing on the case of autonomous vehicles
 - 4.2. The importance of good explanations in AI systems
 - 4.3. Tools for evaluating explainability

Part II: Ethical-Political Aspects of Artificial Intelligence

- 5. Introduction: Why are the political and social aspects of artificial intelligence relevant?
- 5.1. Theory of technological mediation
- 5.2. Narrative around AI and technological determinism.
- 5.3. Responsible innovation and research (RRI)
- 6. Ethics and robotics
 - 6.1. Robots and society
 - 6.2. Ethical challenges in robotics
 - 6.3. Applied examples of robotics in everyday life

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Case studies	50	2	4, 14, 5, 10, 9, 11, 6
Lesson attendance and active participation	30	1.2	14, 3, 5, 13, 1, 11, 12, 16
Practices and exercise	50	2	2, 5, 8, 9, 13, 1, 12, 16, 7

The course's orientation is predominantly practical. Each class will typically commence with the presentation of a real-world case study, fostering a subsequent group discussion. Following that, concepts, methods, or AI systems related to the ethical concerns raised by the case will be introduced and explained. Finally, students will engage in individual or group practices to reinforce their learning of the lecture. In some classes, time will be kept for reviewing and correcting these practices. Few classes will consist of visits to AI research centers.

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

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Evaluative task 1	34%	7	0.28	4, 2, 14, 3, 5, 8, 10, 9, 13, 1, 11, 12, 16, 6, 7
Evaluative task 2	33%	7	0.28	4, 14, 3, 5, 8, 10, 9, 13, 1, 16, 7
Evaluative task 3	33%	6	0.24	3, 5, 13, 15, 11, 12, 16

The assessment of the course can be carried out through two modalities: continuous assessment or single assessment.

Continuous Assessment

The student must individually complete three assessment activities: two corresponding to Part I and one to Part II.

- Activity 1: written test in class with short questions on Part I and analysis of a real case involving the application of an AI system.
- Activity 2: use of AI toolkits to evaluate and discuss ethical metrics of an AI system.
- Activity 3: written test in class on Part II.

To be eligible for this modality, all three activities must be completed. To pass the course, the student must obtain a minimum grade of 5 out of 10 in each activity. Otherwise, the student will have to take the resit exam. The final grade will be the arithmetic mean of the three assessment activities.

Single Assessment

The student will take an individual final exam in January, structured in three parts equivalent to the activities of the continuous assessment, both in type and weight. To pass, the student must obtain a minimum grade of 5 out of 10 in each of the three parts. The final grade will be the arithmetic mean of the three scores.

Resit Exam

Students eligible for the resit exam are:

- Those who have completed the three assessment activities (continuous assessment)
- Those who have taken the January exam (single assessment)

The resit will consist of an individual final exam. To pass, a minimum grade of 5 is required. This will be the final grade for the course.

Review of Grades

After each assessment activity, the teaching staff will inform students via Moodle about the grades obtained and the procedure and date for the review.

Honours Distinction

Honours distinctions will be awarded to students with a final grade of 10. If there are more students with this grade than the number of honours distinctions available for the course, an additional test will be held to determine the recipients.

Not Assessable

The student will receive the qualification "Not assessable" if they do not attend more than one assessment activity (continuous assessment) or if they do not attend the January exam (single assessment).

Repeat Students

No differentiated treatment is foreseen for repeat students.

Use of Artificial Intelligence (AI)

In this course, the use of AI technologies is not permitted at any stage. Any work that includes AI-generated content will be considered a breach of academic integrity and may result in partial or total penalties on the activity grade, or more severe sanctions in serious cases.

Irregularities

Any irregularity that may significantly alter the grade of an activity will result in a zero for that activity. In the case of multiple irregularities, the final grade for the course will be zero, regardless of any disciplinary proceedings.

Adaptation to Online Format

If tests or exams cannot be held in person, they will be adapted to an online format made available through the UAB's virtual tools (the original weighting will be maintained). Homework, activities, and class participation will be carried out via forums, wikis, and/or discussions on Teams, etc. The teaching staff will ensure that students can access these virtual tools or will offer feasible alternatives.

Bibliography

1. Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
2. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center for Internet & Society.
3. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35.
4. van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, 30 (3), 385-409.
5. van Wijnsberghe, A. (2013). Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics*, 19(2), 407-433.
6. Vallès-Peris, N., & Domènech, M. (2020). Roboticists' imaginaries of robots for care: The radical imaginary as a tool for an ethical discussion. *Engineering Studies*, 12(3), 156-176.
7. Vallès-Peris, N., & Domènech, M. (2021). Caring in the in-between: A proposal to introduce responsible AI and robotics to healthcare. *AI & Society*.
8. Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. *Science, Technology & Human Values*, 31(3), 361-380.

Software

To be determined (Part I).

Groups and Languages

Please note that this information is provisional until 30 November 2025. You can check it through this [link](#). To consult the language you will need to enter the CODE of the subject.

Name	Group	Language	Semester	Turn
(PAUL) Classroom practices	711	English	first semester	afternoon
(TE) Theory	71	English	first semester	afternoon