Ethics

Code: 106559
ECTS Credits: 6

2024/2025

Degree	Type	Year
2504392 Artificial Intelligence	FB	2

Contact

Name: Vicente Costa Bueno

Email: vicente.costa@uab.cat

Teachers

Nuria Valles Peris

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-Political Aspects of Artificial Intelligence

1. Introduction: Why are the political and social aspects of artificial intelligence relevant?

1.1. Theory of technological mediation

1.2. Narrative around artificial intelligence and technological determinism

1.3. Responsible innovation and research (RRI)

2. Ethics and robotics

2.1. Robots and society

2.2. Ethical challenges in robotics

2.3. Applied examples of robotics in everyday life

Part II: Ethical Aspects of Artificial Intelligence

3. Introduction: Why should AI professionals study ethics?

3.1. ACM Code of Ethics and Professional Conduct

3.2. Ethical frameworks (consequentialism, theory of justice, virtue ethics...)

3.3. Ethical principles (fairness, responsibility, justice, privacy...)

4. Data collection and privacy

4.1. The importance of privacy

4.2. Main techniques for data privacy (anonymity, encryption, differential privacy...)

4.3. Privacy beyond data (in context, by design...)

5. Algorithms, decision-making, and biases

5.1. Technical definitions of bias in algorithmic outcomes

5.2. Direct and indirect algorithmic discrimination

5.3. Definition of fairness and fairness metrics

5.4. Representation of normative and ethical knowledge in AI

5.5. Ethical guidelines for reliable AI: AI-Fairness Toolkits

6. Explainability

6.1. The impact on responsibility and accountability in autonomous systems, focusing onthe case of autonomous vehicles

6.2. The importance of good explanations in AI systems

6.3. Tools for evaluating explainability

Activities and Methodology

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

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	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16
Evaluative task 2	33%	7	0.28	1, 3, 4, 5, 7, 8, 9, 10, 13, 14, 16
Evaluative task 3	33%	6	0.24	3, 5, 11, 12, 13, 15, 16

The assessment can be carried out in the two ways described below.

Continuous Assessment.

Students must take three evaluative tasks: one corresponding to Part I and two related to Part II. The task corresponding to Part I will be an exam to do in class (presumably in October). The second task will be an evaluative test to do in class, which will include short questions about Part II and an analysis of a real-world-based case of the application of an AI system. The third task will consist of the use of AI toolkits to evaluate and discuss different ethical-related metrics of an AI system.

To be eligible for continuous assessment, the student must have completed the three evaluative tasks. Furthermore, to pass the course, it is needed to obtain at least a grade of 5 in the three evaluative tasks. Otherwise, the student will need to do the recovery (see Recovery section).

The final grade for the course in this modality will be determined as the mean of the three evaluative tasks.

Single Assessment.

The student will take a final exam in January. The exam will contain three parts, and to pass the course, the student will need to obtain at least a grade of 5 (out of 10) in each part. The final grade for the course in this modality will be the mean of the grades obtained in each part of the exam.

Recovery.

To be eligible for the recovery, students must have completed the three evaluative tasks (continuous assessment) or have done the exam in January (single assessment).

Only an individual final recovery exam will be done. To pass the course in this modality, the grade of the final recovery exam must be equal to or greater than 5. The final grade will be the grade of the final recovery exam.

If the student enrolled in the continuous assessment, they only need to recover the parts in the exam corresponding to their failed evaluative tasks.

On carrying out each evaluation activity, lecturers will inform students (on Moodle) of the procedures to be followed for reviewing all grades awarded, and the date on which such a review will take place.

In the event of a student committing any irregularity that may lead to a significant variation in the grade awarded to an assessment activity, the student will be given a zero for this activity, regardless of any disciplinary process that may take place. In the event of several irregularities in assessment activities of the same subject, the student will be given a zero as the final grade for this subject.

In the event that tests or exams cannot be taken onsite, they will be adapted to an online format made available through the UAB's virtual tools (original weighting will be maintained). Homework, activities, and class participation will be carried out through forums, wikis, and/or discussions on Teams, etc. Lecturers will ensure that students are able to access these virtual tools, or will offer them feasible alternatives.

Bibliography

Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI." Berkman Klein Center for Internet & Society, 2020.

Mehrabi N., Morstatter F., Saxena N-, Lerman K., Galstyan A. *A Survey on Bias and Fairness in Machine Learning.* Association for Computing Machinery Surveys, (2021), 54(6)

Vallès-Peris N and 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.

Vallès-Peris, N., Domènech, M. (2021) Caring in the in-between: a proposal to introduce responsible AI and robotics to healthcare. AI & Society.

van de Poel, I. (2020) 'Embedding Values in Artificial Intelligence (AI) Systems', Minds and Machines, 30(3), pp. 385-409.

van Wynsberghe, A. (2013) 'Designing Robots for Care: Care Centered Value-Sensitive Design', Science and Engineering Ethics, 19(2), pp. 407-433.

Verbeek, P.-P. (2006) 'Materializing Morality: Design Ethics and Technological Mediation', Science, Technology & Human Values, 31(3), pp. 361-380.

Software

To be determined (Part II).

Language list

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