Universitat Autònoma de Barcelona

2022/2023

Ethics

Code: 106559
ECTS Credits: 6

Degree	Type	Year	Semester
2504392 Artificial Intelligence	FB	2	1

Contact

Name: David Jorge Casacuberta Sevilla
Email: david.casacuberta@uab.cat

Use of Languages

Principal working language: english (eng)
Some groups entirely in English: Yes
Some groups entirely in Catalan: No
Some groups entirely in Spanish: No

Teachers

Miquel Domenech Argemi

Prerequisites

No prerequisites

Objectives and Contextualisation

The aim of this course is to familiarize undergraduate students with the ethical, social, and political issues that may arise when using artificial intelligence algorithms to make decisions that affect people and how they can be detected in time.

The subject thus avoids problems that are still far away at the technological level such as the idea of singularity or superintelligence and focuses more on real ethical issues that affect us as people here and now.

We will explore issues such as biases in algorithms, the erosion of privacy caused by the search for custom profiles, and how they can be used to manipulate our decisions, but we will also analyze how these problems can be detected and resolved by presenting frameworks such as ethical audits or exploring metrics in algorithms that include concepts of equity and justice.

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 independently, with responsibility and initiative, planning and managing time and available resources, and adapting to unforeseen situations.

Content

1. Introduction to ethics and its theoretical frameworks

1.1 Utilitarianism

1.2 Deontological ethics

1.3 Theory of justice.

1.4 Virtue ethics

2. Privacy and personalized profiles

2.1 Principles for data protection

2.2 Anonymization and re-identification

2.3 Differential privacy

2.4 profiles and manipulation

2.5 the filter bubble

3. Algorithmic biases

3.1 Causes of biases in a database

3.2 How we can detect bias

3.3 Accuracy versus Justice

4. Ethical audits

4.1 Basic principles for an ethical audit

4.2 The ethical matrix

4.3 The pre-mortem

4.4 Equity metrics for algorithms

5. Science, Technology and Ethics

5.1. Social impact of innovations

5.2. Technological mediation

5.3. Materialized morality

5.4. Ethics and design

6. RRI and AI

6.1. What is RRI?

6.2. RRI applied to AI

7. Ethics and Robotics

7.1. Robots and society

7.2. Ethical concerns in robotics

7.3. Care robots/killer robots

Methodology

The course will combine lectures with debates and discussion exercises in the classroom. We will work on a specific project of application of AI to the human sphere and will study progressively how its deployment can generate different ethical and social problems, looking for both the causes of these problems and possible solutions.

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.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Lesson attendance and active participation	22	0.88	14, 3, 5, 13, 1, 11, 12, 15
Master classes	64	2.56	2, 5, 8, 9, 13, 1, 12, 15, 7
Seminars	42	1.68	4, 14, 5, 10, 9, 11, 6

Assessment

The evaluation is organized around three tests

A case study where the ethical implications of a specific artificial intelligence project will be examined as well as the possible solutions to the problems that its application may entail (60%).
A group work. Students will have to write a text that shows that they have acquired the concepts worked on in the second part of the subject. (30%)
Class participation in group discussion and debate activities (10%)

Students who either failed or didn't present items 1 or 2 (or both) are eligible for reassesment. A test is approved with a minimum grade of 5.

Students who ultimately didn't present items 1 nor 2 will be counted as non-assessable.

Note: 15 minutes of a class will be reserved, within the calendar established by the center / degree, for the completion by the students of the surveys of evaluation of the performance of the teaching staff and of evaluation of the subject. / module.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Case studies	50%	10	0.4	14, 8, 10, 13, 1, 11, 12
Lesson attendance and active participation	20%	2	0.08	3, 5, 1, 12, 15, 6, 7
Written assignment in gropus	30%	10	0.4	4, 2, 3, 5, 9, 15, 6, 7

Bibliography

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

Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI*. Oxford Handbooks.

Latour, B. (1999) La esperanza de Pandora. Ensayos sobre la realidad de los estudios de la ciencia. Barcelona: Gedisa, 2022.

O'neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway books.

Mepham, B. (2013). Ethical principles and the ethical matrix. Practical Ethics for Food Professionals: Ethics in Research, Education and the Workplace, 52.

Pariser, E. (2011). The filter bubble: How the new personalized web is changing what we read and how we think. Penguin.

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of internal medicine, 169(12), 866-872.

Sparrow, R. (2007) 'Killer robots', Journal of Applied Philosophy, 24(1), pp. 62-77.

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

Python programming language