KURSPLAN
Etik i Artificiell Intelligens, 7,5 högskolepoäng

Ethics of Artificial Intelligence, 7.5 credits

Kurskod:	TAIR22	**Utbildningsnivå:**	Avancerad nivå
Fastställd av:	VD 2021-03-01	**Utbildningsområde:**	Tekniska området
Reviderad av:	Utbildningschef 2023-10-25	**Ämnesgrupp:**	DT1
Gäller fr.o.m.:	2025-01-01	**Fördjupning:**	A1N
Version:	4	**Huvudområde:**	Datavetenskap

Lärandemål

After a successful course, the student shall

Kunskap och förståelse
- display knowledge of relevant concepts, theories and problems within the ethics of AI
- demonstrate comprehension of relevant ethical concerns in practical application areas that use AI-based technology

Färdighet och förmåga
- demonstrate the ability to explain the ways some of the ethical concerns of AI are being addressed, both practically and theoretically
- demonstrate the ability to reflect on and explain the practical implications of ethical aspects of AI for the development of AI-based technology

Värderingsförmåga och förhållningssätt
- demonstrate the ability to develop a theoretically informed argumentation related to relevant ethical challenges of AI and its use
- demonstrate the ability to express a critical, sound and open position and stance to ethical concerns of AI

Innehåll

Artificial Intelligence is increasingly playing an integral role in our daily lives. AI-based technology is used more and more in areas such as criminal justice, healthcare, transportation, education, etc. and such technologies will probably have a significant impact on the development of humanity in the near future. As designers and developers of such technologies, it is mandatory that we consider and reflect on the legal and moral repercussions of AI, and even on the fundamental principles of ethical life.

Thus, this course gives an introduction to the ethics of AI, discussing ethical concerns that arise from the use and development of AI. We organize the course content by themes, following Müller (2020)[1] division in objects (AI-systems as tools made and used by humans) and subjects (AI systems as subjects).

The theme AI-systems as objects includes learning and discussing issues such as:
- Privacy and manipulation
- Opacity (black-box machine learning) and bias
- Human-robot interaction
- Employment
- The effects of autonomy

AI systems as subjects include topics such as:
- Ethics for the AI systems themselves in machine ethics
- Artificial moral agency

We explore tools, methods and policies to address some of these aspects, e.g., addressing opacity through interpretable machine learning and explainable AI solutions.

Additionally, this course introduces and reviews Fairness, Accountability, and Transparency (FAT) aspects in Machine Learning (ML), in particular:
- Potentially discriminatory effects of using AI/ML
- The dangers of inadvertently encoding bias into automated decisions
- Solutions that account for fairness in algorithm development

Finally, we will discuss the problem of a possible future AI superintelligence leading to a singularity.

Undervisningsformer

Lectures, seminars, group discussions and assignments.

Undervisningen bedrivs på engelska.

Förkunskapskrav

The applicant must hold the minimum of a bachelor's degree (i.e the equivalent of 180 ECTS credits at an accredited university) with at least 90 credits in Computer Engineering, Computer Science or Electrical Engineering (with relevant courses in computer engineering), or equivalent, or passed courses at least 150 credits from the programme Computer Science and Engineering. The bachelor's degree should comprise a minimum of 15 credits in mathematics. Proof of English proficiency is required.

Examination och betyg

Kursen bedöms med betygen 5, 4, 3 eller Underkänd.

The examination consists of active participation in seminars and group discussions, and three assignments. The first assignment corresponds to investigating a particular ethical challenge related to fairness in a practical case and the second and third focuses on debating (for or against) a relevant ethical topic, where one is a group presentation, and one is a written assignment.

The final grade for the course is based upon a balanced set of assessments. The final grade will only be issued after satisfactory completion of all assessments.

Poängregistrering av examinationen för kursen sker enligt följande system:

Examinationsmoment	Omfattning	Betyg
Seminarier och gruppdiskussioner	1,5 hp	U/G
Inlämningsuppgift 1	3 hp	U/G
Inlämningsuppgift 2	1,5 hp	5/4/3/U
Inlämningsuppgift 3	1,5 hp	5/4/3/U

Kurslitteratur

The literature list for the course will be provided 8 weeks before the course starts.

The course literature comprises relevant papers from the area of ethics of AI.

Bibliography (examples of papers that provide an overview of the field):

Bostrom, N., and Yudkowsky, E. (2014). The ethics of artificial intelligence. The Cambridge handbook of artificial intelligence, 1, 316-334.

EU Commission. High-Level Expert Group on AI. (2019). Ethics guidelines for trustworthy AI. URL: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Müller, Vincent C. (2020). Ethics of Artificial Intelligence and Robotics. The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), Edward N. Zalta (ed.)

Russell, S., Hauert, S., Altman, R., and Veloso, M. (2015). Ethics of artificial intelligence. Nature, 521(7553), 415-416