UNIVERSITY OF GOTHENBURG

DEPARTMENT OF APPLIED INFORMATION TECHNOLOGY

TIA415 Fairness, Accountability and Transparency in Artificial Intelligence, 15 credits

Rättvisa, ansvarighet och transparens i artificiell intelligens, 15 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Applied Information Technology on 2023-06-01 to be valid from 2023-06-01, spring semester of 2023.

Field of education: Technology 100%
Department: Department of Applied Information Technology

Position in the educational system

This is a compulsory course within the programme specified below.

The course can be part of the following programme: 1) Human-centered Artificial Intelligence Master's Programme (T2HAI)

Main field of studies	*Specialization*
Human-centered Artificial Intelligence	A1N, Second cycle, has only first-cycle course/s as entry requirements

Entry requirements

To be eligible for the course the student must have a Bachelor's degree (180 credits).

Applicants must prove their knowledge of English: English 6/English B from Swedish Upper Secondary School or the equivalent level of an internationally recognized test, for example, TOEFL, IELTS, or alternatively a bachelor's degree from education held in English.

Learning outcomes

On successful completion of the course the student will be able to:

Knowledge and understanding

- understand how fairness, accountability, and transparencies are theorised in different disciplines, from social science to computer science, as well as in many practical applications of AI;
- understand how values and human biases are embedded in machine learning algorithms, and critically reflect on power relationships;

Competence and skills

- analyze the design of AI systems from a fairness, accountability, transparency and ethics perspective;
- analyze the consequences and implications of AI systems for equity and power relations;
- apply case studies to analyse a design and research problem relating to AI systems and their effects;
- communicate the results of analyses in a compelling fashion;

Judgement and approach

- critique the strengths and weaknesses of models and inherent values used in machine learning research and technologies;
- identify and evaluate inequalities of resources, opportunities, and power as a result of AI, using specific domains as examples of applications;
- discuss the ethical challenges and opportunities posed in AI using arguments critically and constructively.

The course is sustainability-related, which means that at least one of the learning outcomes clearly shows that the course content meets at least one of the University of Gothenburg's confirmed sustainability criteria.

Course content

Artificial Intelligence (AI) holds the potential to transform our life positively, but can also threaten fundamental human rights and discriminate against minorities and women. This course aims to provide students with the technical and conceptual tools needed to engage deeply and critically with AI from a sociotechnical perspective. The dynamic nature of AI technologies requires a new set of skills informed by ethics, law, and policy to be applied throughout the design, development, and application of such technologies. It also involves collaboration among data scientists, computer scientists,

and professionals with knowledge of the social, ethical, and legal implications of the use of AI technologies. Tackling these challenges this course focuses on ethics, biases, regulatory issues, and responsibility attribution raised by the use of AI within three modules.

Sub-courses

1. **Ethics, Social Implications and AI** *(Etik, sociala konsekvenser och AI),* 6 credits
   Grading scale: Pass (G) and Fail (U)
   In this module, the course presents how AI systems have the potential to create new ethical challenges and reproduce and reinforce existing biases, and thus exacerbate already existing forms of discrimination and prejudice in several societal realms, such as employment and labour, social interaction, health care, education, digital divide, personal data, environment, and democracy.

   The module also explores aspects of building more ethical models, from human bias to dataset awareness. Drawing on interdisciplinary literature, the role of data privacy in AI systems, the challenges of designing ethical AI, and possible approaches to identifying sources of bias and mitigating fairness issues will be addressed.

2. **Regulation and Principles for AI** *(Reglering och principer för AI),* 3 credits
   Grading scale: Pass (G) and Fail (U)
   In this module, the course explores the Fairness, Accountability, Transparency, and Ethics (FATE) principles as concepts tied to specific socio-technical environments, not as abstract technical standards. It grounds and contextualises these principles in the experiences of individuals and communities impacted most by automated decision making, data mining, and digital technologies.

   This module also addresses the current development of regulatory frameworks to govern the development, design and application of AI. It will examine how regulation can shape AI technologies and the interplay between national and international policies and the work of other organizations, such as industry and non-governmental organisation generated guidelines.

3. **Methods for Policy Analysis and Requirement Formulation for AI** *(Metoder för policyanalys och kravformulering för AI),* 6 credits
   Grading scale: Pass (G) and Fail (U)
   In this module, the course examines methods for the analysis of policy such as content and discourse analysis and applies them to understanding the regulation of AI and its consequences for the development, procurement and implementation of AI systems. This will include approaches for the critical examination and development of policy, guidelines and recommendations, along with methods for policy informed requirements development.

Form of teaching

The course includes lectures, seminars and group work.

Language of instruction: English

Assessment

Ethics, Social Implications and AI

This module is assessed through a written exam based on the course literature.

Regulation and Principles for AI

This module is assessed through group work where students organize and lead a seminar on the regulation of a particular sector, followed by a summative presentation of their work.

Methods for Policy Analysis and Requirement Formulation for AI

This module is assessed through an analysis of policy relevant to a specific problem space and the production of requirements for a development, procurement or implementation process. This work is reported through a pitch presentation that also forms a contribution to the student's portfolio.

If a student who has twice received a failing grade for the same examination component wishes to change examiner ahead of the next examination session, such a request should be made to the department in writing and should be approved by the department unless there are special reasons to the contrary (Chapter 6 Section 22 of the Higher Education Ordinance).

If a student has received a recommendation from the University of Gothenburg for study support for students with disabilities, the examiner may, where it is compatible with the learning outcomes of the course and provided that no unreasonable resources are required, decide to allow the student to sit an adjusted exam or alternative form of assessment.

In the event that a course has ceased or undergone major changes, students are to be guaranteed at least three examination sessions (including the ordinary examination session) over a period of at least one year, but no more than two years after the course has ceased/been changed. The same applies to internships and professional placements (VFU), although this is restricted to just one additional examination session.

Grades

The grading scale comprises: Pass (G) and Fail (U).
To get a Pass grade for the whole course, a student must get a Pass grade for all parts of the examination. Detailed guidelines and assessment criteria will be made available for

students in a course guide.

Course evaluation

After completion, the course will be evaluated by the students. The results of the evaluation are reported to the director of studies and discussed with the students. A summary of the evaluation results together with the suggestions for improvement is to be made available to students and teachers.

Additional information

The course meets the following sustainability criterion confirmed by the University of Gothenburg:

5. Human rights and social equity. Distribution, discrimination, health and poverty issues and the mutual interaction between social inequality, poor health, the natural environment and people's opportunities for good living conditions.