ID

VP053_INF

Visiting Professor Program
Academic Year 2024/2025

TEACHING COMMITMENT: 16 hours

COURSE TITLE

Fairness in AI

TEACHING PERIOD

2nd term

SCIENTIFIC AREA

Computer Science

LANGUAGE USED TO TEACH

English, Italian

COURSE SUMMARY

While the new wave of so-called "deep learning" systems displays impressive performance in various tasks, these models are very hard to understand in various ways. To cite one issue among others, the GPT family of models employs 175 billion learnable parameters. When something goes wrong, it is almost impossible to understand "why" a model has made a particular decision.

The technical optimism and excitement around Machine Learning has also pushed businesses to apply it in situations where it may impact people's well-being directly, such as loan applications, candidate selection for job offers and evaluating the chance of re-offending for people who commited crimes. Computer vision applications based on neural networks have even been employed to judge beauty contests.

In such contexts, opaque models are particularly problematic as there is a concrete risk for discrimination against certain groups of people. The existence of the gender pay gap, for example, shows how there are complex correlations between law-protected attributes (gender) and other, non-sensitive attributes such as a person’s yearly salary.

If models are learning from biased data, it follows that they will learn to output biased decisions; if we are unable to explain those decisions, we are left with very little human control over what ultimately is a software process. On top of being philosophically troubling and unethical, recent legislation might see these methodologies as unlawful.

LEARNING OBJECTIVES

During this class, you will learn about the current discussion in the AI and ML literature about how to control AI and ML algorithms so that they are trustworthy. We will be focusing on the idea of fairness and egalitarianism as a possible answer to the problem of biased data. The learning objectives are both theoretical and practical. From a theory perspective, students will be able to understand the philosophical and ethical principles motivating the study of fairness in AI and Computer Science in general. From a practical standpoint, students will get exposure to state-of-the-art methodologies for fairness interventions at the data and model level (pre-processing and in-processing, respectively).

VISITING PROFESSOR PROFILE

The instructor holds a distinguished position as either a Lecturer, Researcher, or Professor at a reputable university, with a record of research within the realm of Artificial Intelligence and Fairness. They are currently teaching or they have taught in the past courses focused on the critical issue of "Fairness in AI."

CONTACT REFERENT

Roberto Esposito
roberto.esposito@unito.it