Graduate School Event

Thesis Defense: Monitoring Algorithmic Fairness in Sequential Decision Making

Konstantin Kueffner (Thomas Henzinger Group)

Thomas Henzinger Group

Host: Krzysztof Pietrzak

As automated decision-makers have become ubiquitous in many domains of life, their decisions have become increasingly consequential. Recent years have shown that such systems can exhibit discriminatory behaviour against individuals and social groups alike, thereby amplifying existing biases and entrenching socio-economic disparities over time. Algorithmic fairness addresses this problem by developing methods to quantify and mitigate unfair behaviour. However, much of the existing literature studies fairness in a static pre-deployment setting and, therefore, neglects that automated decision-makers are often deployed in dynamic environments, where their behaviour and the populations they affect may change over time.This thesis addresses this gap through the lens of runtime verification. Instead of treating fairness as a property of a classifier together with a fixed input distribution, it reframes fairness as a property of the interaction trace between the decision-maker and its deployment environment. To evaluate such sequential fairness properties, the thesis develops runtime monitors that observe the evolving interaction between the system and the environment and issue verdicts after each new observation. Because, these monitors are designed to detect unfair behaviour during deployment, they complement fair training, auditing, verification, and enforcement by providing an additional layer of mathematically rigorous fairness assurance.In summary, the thesis develops quantitative, trace-based analogues of classical group and individual fairness measures and constructs monitors for them. This includes monitors for long-run group fairness over Markovian traces, for the time-varying welfare of a changing population in a dynamical system, and for the individual fairness of an arbitrary system generating a trace of inputs and outputs. To achieve this, the monitors combine ideas from runtime verification, sequential statistics, and nearest-neighbour search. In the group-fairness settings, monitoring is primarily a sequential statistical estimation problem: the monitor must construct statistically sound interval estimates of fairness values from dependent and partially observed interactions. In the individual-fairness setting, the main challenge is computational efficiency: the monitor must detect individual fairness violations by efficiently comparing the current decision with all previously observed decisions.

Tuesday, June 16, 2026 01:00pm - 02:00pm

Central Bldg / O1 / Mondi 3 (I01.O1.010) and Zoom

This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station. Please find a schedule of the ISTA Shuttle on our webpage: https://ista.ac.at/en/campus/how-to-get-here/ The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.