



# Beyond traditional assumptions in fair machine learning

Niki Kilbertus

(HelmholtzAI, Munich)

09.12.2020, 16.00 (c.t.)

Online via Zoom

(Meeting-ID: 913-2473-4411; Password: StatsCol20)

In this talk, we will challenge three common assumptions in the development of fair machine learning algorithms and discuss possible attempts at overcoming the corresponding limitations. After a brief introduction to fairness in machine learning, we will first investigate to what extent causality may help overcome fundamental blind spots of statistical group fairness notions. Second, we show how cryptographic techniques, in particular secure multi-party computation, can partially resolve the tension between actively correcting for biased data (need to access sensitive attributes) and privacy (not wanting to record sensitive attributes). Finally, in our third example we argue that i.i.d. data are commonly not available due to “selective labels” where outcomes only exist when a certain decision is taken. We demonstrate how deterministic decision rules may fail in such settings and how we can design fair exploring policies.

## **Biography:**

Niki Kilbertus is leading a group on reliable machine learning at HelmholtzAI in Munich. He focuses on causality and socially beneficial machine learning systems. Prior, he obtained a PhD from the University of Cambridge as a fellow of the Cambridge-Tübingen program and an ELLIS student.