Towards Visualising Procedural Fairness in Automated Decision Making

Fairness in Automated Decision Making (ADM) is a pivotal component of trustworthy artificial intelligence. Most of the existing work on algorithmic fairness has focused on distributive fairness, i.e., to guarantee similar outcomes for similar individuals from a decision making mechanism. However, this understanding of fairness can be reductionist, as it does not consider how the decision was reached. A more holistic approach would consider both the outcome and the process that led to it; procedural fairness focuses on this second aspect, as well as in the stakeholders' perception of fairness. Most of the fair AI research has focused on distributive concerns, with little attention paid to its procedural counterpart.

In 1980, Leventhal characterised fair procedural decision making through six *justice rules*: consistency, accuracy, ethicality, representativeness, bias suppression, and correctability. In this talk, we will define these rules and propose an updated version of them in order to make these extended ruleset more relevant to the ADM context. We will also categorise ADM into five broad phases (data generation, data preparation, modelling, model deployment, and decision challenging), and explore factors for each of these phases that are likely to impact the aforementioned justice rules.

We aim at changing the reductionism (or the perception of it) of ADM through visual explanations and/or metrics that certifies whether a process is procedurally fair or not. Specifically, we will propose a visualisation tool similar to Stoyanovich and Howe's 2019 *Nutritional Labels*, with sections for each ADM phase. These sections would address the (extended) justice rules in an understandable way, i.e., as visualisations, metrics, summaries, or simplified if-then clauses. We also need to design and perform a fairness-perception study to validate our solution, as well as to identify key areas of concern. This project is still in an early stage, and feedback from the Systems community would be enriching and welcome.