Fairness in Algorithmic Decision Making: A Domain-Concrete Approach
21-25 March 2022, Lorentz workshop @ Snellius

Description and aims

Algorithmic decision-making powered by modern machine learning and artificial intelligence (AI) can improve our society in many ways, but it can also have discriminatory effects. For instance, the state can use AI to detect welfare fraud, and insurers can use AI to predict risks. AI decision-making presents our society with serious challenges, which can be divided into two broad categories. *First*, AI can lead to harm to, or discrimination of, people with a certain ethnicity, gender, or another characteristic protected by non-discrimination law. However, such AI-driven discrimination can remain hidden, among other reasons because many AI systems are opaque. *Second*, AI can be unfair in other ways. For example, AI-driven differentiation could reinforce socio-economic inequality, or AI could incorrectly predict that somebody will not be able to repay a mortgage loan.

While many aspects of (un)fairness cut across domains, others are specific to certain domains. Also, re-designing algorithmic systems always requires a careful analysis not only of abstract goals, but also of the specific needs of the direct and indirect stakeholders, as well as of the knowledge, methods, rules, infrastructure, etc., in the specific domain. To understand and process these domain specifics, interdisciplinary collaboration is key.

Therefore, in this workshop, unlike in much other work especially in AI Fairness, we did not focus on metrics and methods that are largely understood to be context-free. Instead, our starting points were concrete domains and case studies, a commitment to interdisciplinary collaboration from the outset, with participation of different stakeholders.

Program, preparations and how it unfolded, Organization (before the workshop)

Our team worked together in the months previous to the workshop in online meetings and via emails, primarily to figure out the best format for our goals, and to establish the participant group that would best bring the topic forward.

In putting together the participant list, we ensured balance along three dimensions: work sector, academic field, and seniority. We invited people from academia, industry and NGOs, experts from computing sciences, social sciences, and legal science in equal proportions, including established professors and early career researchers such as PhD candidates.
Our only (conscious) bias was towards scholars demonstrating their willingness to conduct interdisciplinary research in fairness and AI.

Our original plan was to have an in-person workshop. Due to the developments both of the COVID-19 pandemic itself and of the restrictions on travel and meetings induced by it, we reconsidered this multiple times, and even though we then had to postpone, we maintained our position on a primarily in-person format. In the end, a few people - including two presenters - participated online, which worked reasonably flawlessly both from the technical and the social perspective. (We express extra thanks to the Lorentz organization for their help with enabling online participation.) While the shrinking and expanding of the workshop during the COVID-induced uncertainties was a challenge, the Lorenz team was guiding us through the process very well.

Format of the workshop

The first day of the workshop centered on getting to know each other and everyone's backgrounds. We asked everyone to state, in their intro, what they considered to be the most important open questions and challenges around algorithmic fairness. In addition, we had a poster session where participants could show ongoing research.

The next three days each focused on a specific domain area important for automated decision making: recruitment and employment, fraud detection in the welfare state, and banking and insurance. On the last day, we engaged in discussions to consolidate insights, reflect on the workshop with its outcomes and limitations, and discuss future ideas and directions.

Outcomes and take-aways

In the workshop, based on sector-specific applications and use cases, we took steps in connecting technical with non-technical elements of AI and the role of data and algorithms in the creation or reinforcement of already existing biases. During the week-long workshop, we worked on technical and legal challenges, insights and methods from the social sciences and use cases from various sectors where data and AI are used but where the risk of discrimination and unequal treatment is high (banking, HR and recruitment and social services). Besides project content sessions, the workshop aimed to organise a community of researchers around this theme, and to connect different insights and parties (academic and non-academic) working on data-fication, AI and equality.

The goal of the workshop was to bring a diverse set of people into the same room and get them to share their perspectives, listen to each other, and reflect on most pressing issues around algorithmic decision-making. Accordingly, the outcomes were related to building the common language for conversation, argumentation, etc. First, the introduction day allowed us to understand each others' backgrounds and focus areas, and the three days of presentations and discussions with concrete domain focus brought up the differences in thinking about and tackling the same issues. While participants all agreed that the topic of the workshop is important, it was more difficult to agree on which areas are most pressing, which tools

we should use to move forward, and which research directions are most meaningful. Some discussions triggered strong emotional reactions, related to old discipline-based frustrations or feeling misunderstood. The organizers agree that these five days together were a good reminder of how much work is yet to be done to truly understand each other's perspectives, language, and practices in the different fields.

During the workshop, a new community was built of scientists interested in fairness in algorithmic decision making. Friendships were built between researchers from various disciplines, and from various levels of seniority. The researchers from Belgium and the Netherlands (and others who are able to join future meetings in this geographic area) will meet more often from now on.

Online, in-person, and hybrid events: some thoughts

We are grateful to the Lorentz Center for the support and valuable nudges towards thinking deeply about our concrete and detailed plans for the workshop, and for this being a continuous process in the run-up of the event. Both the Lorentz Center and we have found ourselves in a learning process that (not only the scientific) world is undergoing at the moment: the question from where people are participating in meetings. We would like to share some thoughts on how future organizers and the Lorentz Center can further improve the support processes for events.

We recommend that the Lorentz Center and future organizers work towards considering the option to create hybrid events as an 'enriching feature' rather than an 'afterthought plan B'. The first question that such an attitude would require organizers to answer is: do they believe that online participation can be enriching and an opportunity to do more than without the online participants? If their honest answer is 'no', it may still be worthwhile to keep online participation as an option (travel and meeting restrictions as well as rapid changes to them may stay with us for some time to come), but the preference for in-person attendance should be clearly communicated. If their honest answer is 'yes', they should also communicate this clearly and think of creative measures for making online participation fun and worthwhile. Social and rhetorical protocols for how to signal these two options are evolving. Interestingly, in our specific team we have had proponents of both the 'yes' and the 'no' options, and we believe that much creativity and new developments for continuous improvements in hybrid meetings can also arise out of this tension.

In both cases, the organisers should make sure they develop, before the workshop, a 'script' (even if merely a loose one) for which actions take place when and how online and in-person participants can be activated and brought together, and that they test this if logistically possible. Throughout the workshop, one person who is on-site during the workshop should be responsible for the practical measures: checking that presentations are seen and mirrored to both audiences, that cameras and microphones work well, that online participants are not 'lost', etc. This does not have to be the same person throughout, and the

load is not prohibitive, so this 'hybrid-responsible' will still be able to participate fully. However, other roles such as session chair should not be performed simultaneously.

Lastly, we thank the Lorentz center again for their help. The whole experience, before, during, and after the workshop was great - for the organizers and the participants. We also express our thanks to the anonymous reviewers of our original proposal; they gave useful hints.

Mykola Pechenizkiy, Eindhoven University of Technology
Tjerk Timan, TNO
Bettina Berendt, TU Berlin
Anikó Hannák, University of Zürich
Frederik Zuiderveen Borgesius, Radboud University