Modeling Epistemic Fairness in Networked Opinion Dynamics: A Socio-Technical Approach

Camilla Quaresmini

Algorithmic fairness is a broad area, which is receiving increasing attention in recent years. It examines forms of discrimination that can arise in algorithmic systems making or aiding human decisions across multiple facets of life and application domains. In technical practice, algorithmic fairness is often reduced to a distributive problem expressed through optimization or parity metrics. This reduction obscures harms that arise when credibility, knowledge, and understanding are unevenly distributed - who is believed, whose perspectives shape understanding, and how adoption spreads.

This dissertation argues that fairness analysis must be anchored in a broader theory of justice in order to make such harms intelligible, and develops a formal extension of algorithmic fairness in that direction. Building on Miranda Fricker's account of epistemic (testimonial and hermeneutical) injustice, we articulate an epistemic conception of fairness suited for networked settings, where influence, uptake, and adoption are mediated by social power asymmetries. We formalize this within opinion dynamics models by extending the Linear Threshold Model to encode two epistemic dimensions: reliability (the influence an agent ought to have) and credibility (the influence actually afforded, potentially distorted by identity-based prejudice). Credibility is modeled both as an individual trait and as a relational property of speaker-hearer pairs, revealing conditions under which credibility deficits or excesses reconfigure cascades.

From this we derive Epistemic Fairness (EF) as a design principle that treats credibility distortions as structural features to be corrected so that afforded influence tracks warranted reliability, thereby targeting the testimonial pattern of epistemic injustice. Embedding EF in closed-loop control (via a Linear Quadratic Regulator formulation) shows that accounting for an epistemic dimension not only exposes structural harms overlooked by standard metrics but also improves policy efficiency by leveraging social contagion, often achieving targets with fewer external incentives. Numerical experiments and a data-driven study of sustainable-mobility adoption (constructed from EU survey data) demonstrate that correcting epistemic distortions can reduce the external incentives required to achieve policy goals by exploiting social contagion more efficiently.

Finally, the framework is extended to hermeneutical injustice in advertising environments, where targeting can impose epistemic exclusion (systematic under-exposure) or epistemic bombing (systematic over-exposure). We propose a Hermeneutical Fairness (HF) objective that penalizes such distortions in conceptual access and implement it in an operational algorithm for exposure adjustment.

Overall, the contribution is dual. Philosophically, the work reframes algorithmic fairness as an epistemic problem rooted in the distribution of credibility and interpretive resources. Technically, it translates this reframing into formal models and control-based interventions that can be deployed as decision-support tools, showing that incorporating an epistemic dimension changes not only how fairness is defined but also how policy efficiency is assessed.