

# **Re-imagining Algorithmic Fairness in India and Beyond**

**Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi,  
Vinodkumar Prabhakaran**

**Reviewed by: Rishabh Devgon**

## **Critical Review:**

The paper [1] aims to decentralise the conventionally west centric concept of algorithmic fairness. Nithya Sambasivan et al. highlight the dangers of naive generalisation of AI fairness in the global context. The title of the paper seems appropriate and echos the research. The authors highlight three factors that require focus: Data and model distortions, Double standards and distance by ML makers and Unquestioning AI aspiration. The researchers provide background about the Western orientation in fair ML where they discuss philosophical routes which are of particular interest to me. The authors highlight the various aspects that could lead to Machine Learning unfairness, including Caste, Gender, Religion, Ability, Class, Gender Identity and Sexual Orientation.

A peculiar point raised in the paper which I resonate with is “merit is a function of caste privilege”. The paper touches on authority and how AI-based technology like Face Recognition systems are being used for purposes that they were not designed for. But, due to a lack of technical literacy, lack of critical approach towards emerging technology and widespread ignorance, these decisions go relatively unchallenged. The research is novel as it builds on the conventional epistemology of algorithmic fairness and explores new avenues for non-western spheres. It takes the case of India as an exploration so that the perspective of the global south can be highlighted and can be extended to other countries.

The research method used by the authors is a qualitative analysis through semi-structured interviews with people stemming from diverse fields who work with marginalised Indian communities. The sampling method used purposive sampling, which makes sense given the rigorous selection criteria with the requirement of vast grassroots experience. The authors also analysed policies and algorithmic deployments along with Indian news publications and community media.

A gap that I found was that since the paper situates itself in acknowledging the caste differences, they did not mention the caste divide of their informants or the authors. Hence, it is difficult to verify if the selected sample was representative or was it another case of cultural appropriation. So, a point could be made about how this research itself holds double standards, and I find that quite ironical. The authors have employed a feminist, decolonial and anti-caste lens while carrying out the analysis through open coding. This approach is justified, given the socio-cultural context of India and the country’s diversity, history and power dynamics. However, I did not observe the explicit incorporation of the values of these frameworks within this literature. The secondary research is not peer reviewed and hence may not be the best source to infer from and could add bias to the study.

Although the authors do touch upon caste differences, India also suffers from other cultural differences which I felt have not been addressed in the paper. These include regional (South, North, North East etc.) and linguistic differences along with certain problematic traditions and stereotypes. Another aspect that is unique to non-western society is Eurocentrism and the obsession with Western culture in general. An element that is particularly situated in India is the community organisation structure, i.e. the Panchayati system or localised governments which play a role in propagating specific values and add another layer of power dynamics. A deeper exploration into what role this plays could also benefit in gauging the Indian context. Within the axes of machine learning unfairness, I did not quite wrap my head around why was only a binary discussion employed when addressing gender.

A further exploration based on machine learning models could be to assign an index for the reservation that takes into consideration various socio-economic factors; this could help exploitation of reservation within the communities from the “creamy layer”. With respect to the discussion about informed consent, there is a gap in digital literacy among users so to enhance community agency in datasets perhaps better ways of self-disclosure could be explored as well. Although the paper does list down the approaches to take while reevaluating algorithmic fairness, it does not really offer institutional level changes which could massively benefit this area. Overall, I really liked this paper because it shed light on the Indian perspective, challenged the status quo with several brave comments and helped to develop a more global conception of algorithmic fairness.

## **References:**

1. Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining Algorithmic Fairness in India and Beyond. *arXiv preprint arXiv:2101.09995*.