

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/180036>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

## **A Global South perspective for ethical algorithms and the State**

Authors: Nandana Sengupta, Vidya Subramanian, Anwesh Mukhopadhyay, Arul George Scaria

***Standfirst:** We explore the intersection between algorithms and the State from the lenses of legislative action, public perception and the use of AI in public administration. Taking India as a case study, we contextualize the potential fallout of the absence of rigorous scholarship on such questions for countries in the Global South.*

### **Introduction**

Given the ubiquity of Artificial Intelligence (AI) algorithms in the lives of their citizens, many countries are trying to both harness the potential of such algorithms and also attempt to regulate their unchecked proliferation - consider for instance Japan's [AI Governance Guidelines](#), and the Beijing [AI Principles](#). One of the key concerns of AI proliferation that such regulations aim to address is that of algorithmic bias i.e. algorithmic output which is systematically unfair to certain groups of people. This has led to both legal and computer science scholars examining the societal ramifications of AI. Much of this conversation around “ethical algorithms” however has been restricted to economically and technologically advanced countries in the Global North, largely escaping attention in the Global South. This is despite the fact that there are potentially many different and unique sources of algorithmic bias specific to countries in the Global South with typically lower per-capita incomes and weaker institutions.

In this piece, we focus specifically on the intersection between AI and the State, for which the existing academic analysis from the Global South is grossly inadequate. We consider the question of legislative response (or lack thereof) to tackle issues relating to algorithmic bias, discuss the importance of public perception studies and highlight concerns emerging from the State utilizing AI for purported public welfare measures. We take India as a case study in this regard prompted by the following factors. First, the number of internet users in India (and their share in the

global internet user population) is rapidly [increasing](#).<sup>1</sup> Second, [reports](#)<sup>2</sup> indicate that funding for AI startups in India is on the rise. Third, the Government of India has been bullish about leveraging AI technology with significant outlays to missions like “[Digital India](#)” with potential applications ranging from policing to judicial decision making. Barring a few notable [exceptions](#),<sup>3</sup> academic scholarship on the repercussions of AI in the social context of India has been severely limited. Our piece aims to address this lacunae.

### **Policy Context and Legislative Vacuum**

Within the global context, a number of popular and readily available cross-country [assessments](#) of AI readiness have emerged and certain countries from the Global South, including India, score high in many of these assessments. However, the majority of indices reward higher proliferation of AI with little attention paid to its potential fallout. In the Indian policy context, the government’s think-tank NITI Aayog published a discussion paper titled “[National Strategy for Artificial Intelligence](#)” that offered a celebratory picture of the government’s AI aspirations but failed to address transparency, fairness and privacy concerns. NITI Aayog has since made efforts to engage with questions of responsible AI via a separate set of [discussion papers](#). There are concerns however that such ethical considerations are more of an afterthought rather than a set of founding principles, especially given that these papers came over 3 years after the national strategy document. The discussions in the papers revolve around case studies based on the Global North and not much effort has been expended to analyze challenges specific to the Indian context. The papers also allot the function of development and use of AI products almost exclusively to the private sector, and recommend the self-regulation of market entities in cases where the anticipated risk of harm from an AI system is perceived to be low.

Whereas policy documents provide an understanding of the general direction that governments plan to proceed in, laws provide the standards within which citizens and entities must operate. In India, one reason behind the lack of a legal framework to address algorithmic bias may be the legislative vacuum in data protection. A draft personal data protection bill which was under consideration since 2019 was withdrawn and a new bill titled the [Digital Personal Data Protection Bill, 2022](#) has been introduced by the government. The bill imposes some obligations, such as

notice and consent, on those processing data along with enlisting certain rights of the individual, such as the right to information about processing of personal data and right to correction and erasure. The Bill also proposes provisions for transfer of personal data outside India. [Certain provisions](#) of the bill arguably run contrary to the fundamental rights guaranteed under the Constitution. The State has been exempted from liability under broad and arguably vague headings such as public order and national security, giving rise to the possibility of unwarranted data collection and retention. Further, on provisions related to consent and storage, the bill appears to treat private and government entities engaged in similar commercial activities differently, possibly infringing on different provisions enshrined in the Constitution.

Even assuming consensus on data protection may take longer to achieve and legislation may be years out, it is unclear why more specific measures cannot be considered for AI governance, independent of data protection. For instance, the US has taken active steps to bring forth legislation for general governance of algorithms through the [Algorithmic Accountability Act of 2022](#). The bill lays out a framework directing companies to conduct an impact assessment for the use of automated decision-making systems to enhance transparency and directs the Federal Trade Commission (FTC) to draft the relevant guidelines. The bill also provides consumers with a mechanism to obtain information regarding the specifics of automation of critical decisions by companies. The White House has drafted a blueprint for an AI [Bill of Rights](#) for safeguarding society from threats arising as a consequence of automated systems encroaching on public rights. The European Commission (EC) has also come up with a [proposal](#) for stitching a legal framework for the regulation of AI which envisions a risk-based approach classifying AI risks into three categories - unacceptable risk, high-risk uses and minimal or low risk. High-risk uses are specifically those which may potentially infringe on the fundamental rights of citizens.

## **Public Perceptions**

A possible reason behind the varied response of lawmakers to AI governance across countries could be differing public perceptions. There is some indication that potentially adverse outcomes of AI might be considered to be of more relevance by the public in the Global North. A

perception [study](#) in the UK revealed that almost half of the respondents were convinced that AI would have a negative impact on society and more than two-thirds opined in favor of AI regulation.<sup>4</sup> Another [study](#) in the US revealed that most respondents found it unacceptable to use algorithms for recidivism prediction, screening for jobs, and assigning credit scores.<sup>5</sup> In a cross country [public perception study](#),<sup>6</sup> respondents of countries with worse HDI had more positive expectations from AI. Close to 40% of those surveyed in Brazil and Nigeria and over 50% of respondents from India believed AI would be a net positive for society. This figure hovered around 20% for respondents from countries like Australia, Canada, US and France.

An important gap in the literature on AI and society within the Global South is the lack of rigorous evaluation of public awareness and attitudes carried out by independent researchers. A preliminary search reveals that of the 14 survey studies matching the keywords “public perceptions” and “Artificial Intelligence”, 12 referred to perceptions of countries from the Global North and advanced economies like Japan. Given that there is evidence to suggest that awareness of potential consequences of AI adoption (and consequently policy response) is unequally spread across nations, there is an urgent need for carrying out public awareness and attitude surveys focussed on the Global South to push the legislative dialogue forward in these countries.

### **Algorithms in Public Administration**

Algorithmic bias is concerning at all levels, but becomes critically important when associated with actions of the State. In India, two major areas where algorithmic decision-making is being adopted by the State are predictive policing and the use of facial recognition technology (FRT).

Algorithms have been employed by the police to proactively counter criminal activities as well as purportedly establish a link between perpetrators and crimes based on past criminal history. Techniques such as hotspot policing, regression analysis and data mining are on the rise, with Delhi police’s Crime Mapping, Analysis and Mapping System (CMAPS) being a prime example. There is a possibility that data fed into

such algorithms carry the prejudice of police officers against marginalized sections of society such as Dalits and other [minorities](#).<sup>7</sup> Conscious or subconscious bias may also enter during the coding process. There is by now [well-documented](#) precedents of such discrimination within the United States where the use of algorithms in policing has been shown to lead to exacerbated racial bias in policing and sentencing.<sup>8</sup>

Similar concerns exist with regard to FRT which is being increasingly utilized by law enforcement agencies across the globe, including in India. The Indian state of Telangana, for instance, with around 6 million cameras, has been [branded](#) as ‘one of the most surveilled places in the world’ by Amnesty International. This concern is further exacerbated by [invitation of bids](#) by the National Crime Record Bureau for the creation of a National Automated Facial Recognition System, a step that would facilitate the use of FRT to aid in criminal investigation by matching photographs of suspects against image databases. The Bangalore Metro Rail Corporation has recently partnered with Google to incorporate FRT in a bid to replace tokens and smart cards at metro stations for [automatic fare collection](#). Other [states](#) in the country are also developing facial recognition capabilities for policing and surveillance. This is even as the first [case](#) against the employment of FRT on the grounds of infringing one’s right to privacy is pending before the Telangana High Court against the State and City Police Commissioner.

There is a large and growing global [literature](#) that evaluates algorithmic bias in FRTs, with the largest prediction errors being associated with women and minorities.<sup>9</sup> The ACM U.S Technology Policy Committee (USPTC) has [concluded](#) FRT to be unreliable and capable of negatively impacting the liberty and privacy of minority populations. It is unclear if such evidence of algorithmic bias is being considered before the widespread application of FRT across domains in India. A cautious approach by the State is imperative before extending the use of algorithms to diverse areas of decision-making, given that it carries important implications for social justice. Rigorous scientific studies set in India and other countries in the Global South that measure and contextualize potential bias in algorithmic policing and FRT are needed to provide direction to the State when designing and implementing such systems for widespread use.

## **Way forward**

AI and algorithms have come a long way and so has the conversation surrounding them. Any competent evaluation of AI systems, the policies surrounding them, governments' preparedness etc. must prominently feature studying the social consequences of their adoption. While such consequences can range from shifts in employment patterns to political polarization, the subject of this article is the interplay between algorithmic proliferation and the State.

We consider this intersection by looking through the lenses of legislative action, public attitudes and use of algorithms by the State itself, with a specific focus on the Indian experience. We find that the policy discourse is skewed towards favoring proliferation, with the discussion on the management and mitigation of potential biases coming in more as a footnote. We posit that the lack of rigorous evaluations of public attitudes towards the potential use (or misuse) of algorithms by the State may be a roadblock towards development of meaningful legislation. Finally, we are concerned about the use of algorithmic systems for public administration without sufficient public discourse around embedded biases.

We find that the academic literature focussing on the Global South experience in general and the Indian experience in particular is scant. The histories, economics, institutions and social contexts of such countries are uniquely different from those of the countries where the current academic literature on AI and society is centered. We hope this piece plays a part in increased academic engagement on these issues in the Global South.

## References

1. Internet adoption in India, IAMAI-Kantar ICUBE 2020 report.
2. Mani, Deepa., K, Monish. & Mohan, Madhumitha. The AI Index 2021 Annual Report, Indian School of Business.
3. Sambasivan, Nithya., Arnesan, Erin., Hutchinson, Ben., Doshi, Tulsee. & Prabhakaran, Vinodkumar. *ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (2021).
4. Holder, C., Khurana, V. & Watts, M. (2021).
5. Smith, A. Pew Research Center (2018).

6. Kelley, P. G. et al., *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (2021).
7. Marda, V. & Narayan, S., *FAT\* '20: Proceedings of ACM Conference on Fairness, Accountability, and Transparency* (2020).
8. Selbst, Andrew D. 52 *Georgia Law Review* 109 (2017).
9. Buolamwini, Joy, & Timnit Gebru. In *Conference on fairness, accountability and transparency*, pp. 77-91. PMLR. (2018).

## **Competing Interests Statement**

The authors declare that they have no competing financial interests or personal relationships that could influence the work reported in this article.

## **Acknowledgements**

The authors are grateful for financial support for their research from a collaborative research grant awarded jointly by the Indian Institute of Technology Delhi and National Law University Delhi.

## **Author Affiliations**

### **Nandana Sengupta (corresponding author)**

Assistant Professor, School of Public Policy, Indian Institute of Technology Delhi

Address: Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

Email: [nandana@iitd.ac.in](mailto:nandana@iitd.ac.in)

### **Vidya Subramanian**

Affiliation: Research Fellow, National Law University, Delhi

Address: National Law University, Sector-14, Dwarka, Delhi -110078

Email: [vidya.subramanian@nludelhi.ac.in](mailto:vidya.subramanian@nludelhi.ac.in)



**Anwesh Mukhopadhyay**

Affiliation: MRes student, Department of Economics, University of Warwick

Address: Department of Economics, University of Warwick, Coventry, CV4 7AL, United Kingdom

Email: [Anwesh.Mukhopadhyay@warwick.ac.uk](mailto:Anwesh.Mukhopadhyay@warwick.ac.uk)

**Arul George Scaria**

Affiliation: Associate Professor, National Law School of India University

Address: National Law School of India University, Nagarbhavi, Bengaluru, India 560072

Email: [arul.scaria@nls.ac.in](mailto:arul.scaria@nls.ac.in)