DEF
DIGITAL EMPOWERMENT *foundation*

Advancing Data Justice
RESEARCH & PRACTICE

India Report by

Ananthu R A | Jenny Sulfath | Dr. Syed S. Kazi | Osama Manzar

Disclaimer

This booklet on the Data Justice India Report is based on research conducted by the Digital Empowerment Foundation as Policy Pilot Partners of the Alan Turing Instituite and the Global Partnership on AI, between September 2021 and March 2022.

> *I am not your data, nor am I your vote bank,*
>
> *I am not your project, or any exotic museum object,*
>
> *I am not the soul waiting to be harvested,*
>
> *Nor am I the lab where your theories are tested,*
>
> *I am not your cannon fodder, or the invisible worker,*
>
> *or your entertainment at India Habitat Centre,*
>
> *I am not your field, your crowd, your history,*
>
> *your help, your guilt, medallions of your victory,*
>
> *I refuse, reject, resist your labels,*
>
> *your judgments, documents, definitions,*
>
> *your models, leaders and patrons,*
>
> *because they deny me my existence, my vision, my space,*
>
> *your words, maps, figures, indicators,*
>
> *they all create illusions and put you on pedestal,*
>
> *from where you look down upon me,*
>
> *So I draw my own picture, and invent my own grammar,*
>
> *I make my own tools to fight my own battle,*
>
> *For me, my people, my world, and my Adivasi self!*

~Abhay Xaxa
(Adivasi rights activist and sociologist)

Preface

The Alan Turing Institute and the Global Partnership on AI was working on a project Advancing Data Justice Research and Practice, that aimed to fill a gap in data justice research and practice that provides a frame to help practitioners and users to move beyond understanding data governance narrowly as a compliance matter of individualised privacy or ethical design, while also seeking to include considerations of equity and justice in terms of access to and visibility and representation in data used in the development of Artificial Intelligence/Machine Learning (AI/ML) systems.

GPAI works on a comprehensive, objective, open, and transparent assessment of the scientific, technical, and socio-economic information relevant to understanding AI impacts, encouraging its responsible development and options for adaptation and mitigation of potential challenges.

The Alan Turing institute is UK's national institute for data sciences and Artificial Intelligence,consisting of the universities of Cambridge, Edinburgh, Oxford, UCL, Warwick, Leeds, Manchester, Newcastle, Queen Mary University of London, Birmingham, Exeter, Bristol, and Southampton, and the UK Engineering and Physical Sciences Research Council.

Along with eleven other research partners from the global south, Digital Empowerment Foundation worked with GPAI and ATI on the ADRP project, with us specifically looking into issues in India.

Digital Empowerment Foundation (DEF)'s work, since its conception, has been to digitally connect several unconnected populations. In that sense, the organisation's work has been trying to bring representation and access to the marginalised section of society. The concept of Data Justice in its work might not be related to Artificial Intelligence (AI) and Machine Learning (ML) systems per se, but the issue of access in a country where half the population is unconnected is an important part. DEF is aware of the importance of data for every human being on earth and operating in India, it understands what this means for the most unconnected, underprivileged and underserved in India- the minorities, the Dalits, the Adivasis, women who are considered to be illiterate, uneducated or without digital access. It realises and works on the importance of the internet and digitisation as a medium in making the information accessible to the communities and the critical role digitisation plays in democratising the political systems and making them more accessible. However, the agenda of the organisation and the challenges have evolved over the past twenty years. India is also going through a crisis of misinformation and disinformation, where newly digitised communities fall victims to it the most. Fighting the misinformation ecosystem has emerged as a focus of the organisation in the past few years.

Simultaneously, DEF has also been keen on providing the right information about data to the communities. It has developed a curriculum on data literacy, and it is being used as training material in its community digital centres located in rural and remote parts of India. The current study has encouraged the organisation and the team to also work on developing a 'data policy for civil society organisations.' As civil society organisations also collect data, it is important to have a set of policies on how to keep the data safe and private. Further, DEF has also realised that it is important to review various policies related to data in each state in India, to engage with the discourse on data more meaningfully in the coming days. This is a task that it sees as a preliminary exercise on data justice down the road.

More importantly, DEF also envisions a future where the marginalised communities not only have the right over their data but over how the data is collected, how the definitions of analysis are made, how the research questions are formed and how the technology is designed too.

The ADJRP project concluded in March 2022, and we have turned our findings into a report which GPAI and ATI have incorporated into their global release of findings from the project. At DEF, we have committed to ourselves to take forward the research, and further work on engaging with communities and stakeholders on the issue of data injustices. This report is a part of this commitment, trying to reach more people on what data justice is, what the six pillars of Data Justice are, and how communities can identify and tackle injustices that arise from AI/ML technologies, and try to build a discourse around it.

The insights provided in interviews and panel discussions were a learning experience for us.

We sincerely acknowledge the valuable contributions of all the participants, policy-makers, policy analysts and developers who collaborated with us on this project.

Contents

Introduction

Almost into the second quarter of the twenty first century, the widespread level of datafication in our society is very visible. Datafication is how aspects of our life are turned into data, one which has a new value in the world and economy. Massive amounts of data are being collected and processed systemically, and this has been normalised. Echoing the recently coined adage by British mathematician Clive Humby, and its reprints in The Economist[1], Indian billionaire Mukesh Ambani[2] repeated the same- data is the new oil– or rather, the world's most valuable resource is no longer oil, but data.

The primary goal of the Advancing Data Justice Research and Practice Project, as envisioned by the Alan Turing Institute and the Global Partnership on AI is to provide a broader frame, and attempt to fill a gap in existing research on the issue beyond the present narrower ones of compliance matter of individualised privacy or ethical design. [3]

From preliminary research, experts at GPAI and ATI has identified six pillars of data justice (which can be read in full in this link), which can be briefly explained like this:

Power

The concept of power helps to understand the levels at which power operates in data innovation ecosystems; to understand how power manifests and materialises in the collection and use of data in the world; and, to use this understanding to question power at its sources and to raise critical awareness of its presence and influence. While the questioning and critiquing of power are essential dimensions of data justice, its purpose of achieving a more just society demands that unequal power dynamics that harm or marginalise impacted individuals and communities must be challenged and transformed.

In short, the application of power as a pillar of data justice can be summarised into two points.

---

[1]"The World's Most Valuable Resource Is No Longer Oil, but Data," The Economist (The Economist Newspaper, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

[2]Mukesh Adhikary, "'Data Is the New Oil': Mukesh Ambani Says Global Firms Should Not Control India's Data," Business Today, January 18, 2019, https://www.businesstoday.in/latest/economy-politics/story/data-new-oil-mukesh-ambani-says-global-firms-should-not-control-india-data-159818-2019-01-18.

[3]"Advancing Research and Practice on Data Justice," GPAI, 2022, https://gpai.ai/projects/data-governance/data-justice/.

- Interrogate and critique power:

- Challenge Power: Empower People

Equity

Data equity demands the transformation of historically rooted patterns of domination and entrenched power differentials. Data innovation practices like data security, data protection, algorithmic bias, and privacy are an important subset of data equity considerations, but the transformative potential of data equity to advance social justice comes in a step earlier and digs a layer deeper: it starts with questions of how longer-term patterns of inequality, coloniality, and discrimination seep into and penetrate data innovation practices and their governance.

- Use-equity or the choice to engage

- Focus on the transformative potential of data equity

- Deploy measurement justice and statistical equity to combat discriminatory and racialised politics of data collection

Access

Access is about supporting the equitable advancement of access to research and innovation capacity. Just as equalising access to resources is important, the pillar also stresses on advancing "access to representation," access to infrastructures, equalising know-how and capability. Data justice thinking should also focus on equitably opening access to data through responsible data sharing, equitably advancing access to research and innovation capacity, equitably advancing access to the benefits of data work and equitably advancing access to capabilities to flourish.

- Prioritise the material preconditions of data justice and challenge formalist and ideal approaches

- Start from questions of access and capabilities

- Promote the airing and sharing of injustices across communities through the transformative force of data witnessing

- Focus on harms of allocation, distributive justice, and equality of opportunity as part of a wider understanding of the preconditions of equitable access

Identity

The construction and categorisation of data is shaped by the socio-cultural conditions and historical contexts from which it is derived. The social character of data coupled with the sorting and clustering that proceeds from its cleaning and pre-processing can lead to categorisations that are racialised, misgendered, or otherwise discriminatory. This can involve the employment of binary categorisations and constructions—for example, gender binaries (male/female) or racial binaries (white/non-white)—that are oriented to dominant groups and that ought to be critically scrutinised and questioned. Data justice calls for examining, exposing, and critiquing histories of racialisation and discriminatory systems of categorisation reflected in the way data is classified and the social contexts underlying the production of these classifications.

- Interrogate, understand, and critique modes of othering

- Challenge reification and erasure

- Focus on how struggles for recognition can combat harms of representation

Participation

Prioritise meaningful and representative stakeholder participation, engagement, and involvement from the earliest stages of the data innovation lifecycle to ensure social licence, public consent, and justified public trust

- Democratise data and data work

- Understand data and data subjects relationally

- Challenge existing, domination-preserving modes of participation

- Ensure transformational inclusiveness rather than power-preserving inclusion

Knowledge

Diverse forms of knowledge, and ways of knowing and understanding, can add valuable insights to the aspirations, purposes, and justifications of data use.

- Embrace the pluralism of knowledges (semantic, epistemic, and ontological)

- Interrogate, understand, and critique the ways in which certain forms of knowledge are prioritised within decision-making relating to data.

- Challenge the presumptive authority of technical, professional or "expert" knowledge across scientific and political structures.

- Acknowledge multiple forms of knowledge

- Prioritise interdisciplinarity

- Pursue "strong objectivity"

- Cultivate intercultural sharing, learning, and wisdom

Examples and cases of data (in)justices from India

Being tasked to assess the state of data justice in India, the discourse here was also initially locked on a lot of individual privacy issues. To illustrate and better communicate the need for a discourse of data justice and the six pillars proposed by the Alan Turing Institute, several examples have been highlighted. One classic example used to show out the possible discrimination in AI is a predictive policing system that had been deployed in the US.[4] Studies show how the AI tool had used existing datasets that had been biased, influenced by several decades of systemic racism that had criminalised people from communities of Black, Indigenous, and People of Colour. Using this data, the existing biases were further reinforced.

However, in a more recent report[5], data scientists, developers and policy experts are trying to reverse this bias- precisely because such a discourse was taken forward. The new tool, as the report suggests, turns the idea around and audits police performance for biases. Of course, this is only a very early, and yet not fully analysed example, but it is a nudge in a direction that has been brought about by conversations on data unfairness.

Below are some examples we have collected from India, where hastily planned use of AI and Data driven systems have led to social injustices. These cases were taken as points to further reach out and talk to communities, developers and

[4]Will Douglas Heaven, "Predictive Policing Algorithms Are Racist. They Need to Be Dismantled.," MIT Technology Review (MIT Technology Review, December 10, 2020), https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/.

[5]Libor Jany, "Researchers Use AI to Predict Crime, Biased Policing in Major U.S. Cities like L.A.," Los Angeles Times (Los Angeles Times, July 4, 2022), https://www.latimes.com/california/story/2022-07-04/researchers-use-ai-to-predict-crime-biased-policing.

policy people who were involved in these data injustices.

Bhoomi, Karnataka

Bhoomi is a digital registry of 'rights, tenancy, and crops' produced by the state government of Karnataka, India. The registry is part of an 'open data' effort to increase both the uniformity and availability of official records including land-ownership records. While initially developed by the department of revenue for taxation purposes, the information can be viewed publicly online and at internet kiosks. It is reportedly used extensively by real estate developers. Critics have argued that the Bhoomi registry has disenfranchised members of the Dalit caste whose claims are often not documented in official records but are well supported by other means. Dalits nevertheless have long-standing land claims. However, the informal and historical knowledge that supports these claims cannot be easily accommodated in the flattened landscape of a relational database, such as the Bhoomi registry, and so are more easily dismissed or overruled. Furthermore, Bhoomi may be an example of 'open data under conditions of unequal capabilities.' Like many digital resources, the Bhoomi registry is more likely to be accessible to people with computational and interpretive skills, who are also more likely to hold greater social and political power in society. Ownership claims based on local knowledge were diminished by the Bhoomi registry.

360-degree citizen profiles and surveillance, Telangana

The Telangana state government is actively using Samagra Vedika - an integrated platform comprising a 360-degree profile of every citizen in the state. The issue that the administration came across was that, people apply for different welfare schemes or people from the same family apply for the same schemes many times. For example, people from the same family apply for housing schemes and end up owning multiple subsidised houses. Samagra Vedika gathers data from different databases regarding electricity and water connections, house and land ownership, old-age pension and vehicle ownership, among others, and assesses it to know if an applicant is truly eligible for the scheme. However, news reports show that using the Samagra Vedika system, the government initiated mass cancellations of ration cards. Several people from poor families could not access basic rations from the PDS during the pandemic as a result. In response to a petition filed by a complainant who was deemed ineligible, the civil supplies department said that: "she has not drawn ration for the previous six months, it is clearly showing that she has possessed the ration card for other purposes, and not for getting ration." However, the petitioners could not draw ration because their fingerprints were not getting detected by the machine. If the fingerprints are not working, an iris scan should authenticate the card, however, the ration shops

they went to did not have an iris scan machine. Legal redressal is inaccessible for most of the people who were denied welfare, and this exposes a lack of such options.

Development data and trans-exclusion, India

The 2011 census data categorised anyone beyond the gender category of male and female as "other". Until 2017, the application to apply for PAN did not mention a third gender category. However, transgender persons with Aadhar cards were expected to link it with PAN cards, resulting in a gender category mismatch. The Automated Decision-Making System (ADMS) is deciding based on the skewed data sets for welfare policies, which would lead to the exclusion of many transgender people from welfare policies.

OTP based inclusion in welfare schemes and the exclusion of homeless migrant workers

A lot of welfare schemes in India are OTP based. The homeless people in India mostly do not own a phone and are often under the threat of theft as they live in the street. The structure of homeless shelters is such that there is no shelf to keep their value, so a lot of migrant workers also decide to travel without mobile phones. This leads to the exclusion of this category of people from a lot of welfareand social security schemes.

Human efficiency trackers in smart cities, Maharashtra, Chandigarh, Madhya Pradesh.

In the 'smart cities' of Chandigarh, Nagpur and Indore, workers who maintain urban infrastructure are fitted with 'Human Efficiency Trackers' which automatically

deduct pay if they depart from the work schedules or routes determined by algorithmic systems, not only normalising intrusive surveillance of individuals, but also undermining worker agency and channels of negotiation and grievance redress.

Aadhaar, India

Aadhaar, the Unique IDentification project in India, planned to combat 'leakages' in welfare programs by using biometric authentication. The problem is when data

does not take into account the materialities of poverty. A common example of this is the failure of biometric authentication. Depending on age and nature of work (labour in limestone mines, for example), fingerprints and eye scans do not always work. "The system forces registrants to conform to a 'standard of normalcy' by having legible fingerprints and irises, by possessing mobile phones, by having a stable family life where the same registrant can collect rations from week to week, among other standards." When this biometric authentication is a requirement for welfare as basic as monthly rations through the Public Distribution System or subsidies on cooking fuel, it leads to exclusion. The project further plans to use the same biometric data collected to "clean out" electoral rolls. In 2014, a pilot project was done in some districts of Andhra Pradesh and Telangana, using AI software resulting in the deletion of around 4 million names from the electoral rolls. Critics predict more disenfranchisement and exclusion in the future.

AI-based discrimination against platform-workers, India

Algorithms of platform-economy apps like cab-hailing, food delivery, and home services have optimisation mechanisms that maximise profit for the platform itself while being discriminative towards the workers with demanding and exerting targets and punitive pay cuts when they fail to meet them.

Exclusion from NRC, Assam

NRC in Assam left out several people because of a mismatch in data which a software decided. "There have been hundreds of cases in which a small difference in the English spelling of a Bengali name, or a small variation in age, has been enough for the NRC authorities and the FTs to sound the death knell of dreaded 'foreignness'." The injustice here is the exclusion of already vulnerable people

Methodology and methodological challenges

The assessment of Data Justice in India was done in the following way.

It adopted a secondary approach to review and map such stakeholders in the AI space in India that would be critical to have their individual inputs as interviewees as well as to be part of the workshop in a cross-exchange and learning purpose on data justice-related aspects. This involved studying and reading emerging AI works in India at national and state levels, the initiatives involving citizens, groups and communities and looking at covert and overt ways of data injustices and the possibility and reality around it. This helped to identify, sorting respondents and participants at policy making, developers level and at affected community levels and approach them.

The work involved wholly a primary method of engaging, involving interviewees and workshop participants to bring out key aspects of data justice-related issues and possibilities in line with the prelim guide questions including the six pillars of data justice. The in-person engagements helped to receive direct inputs along with interesting cross-cutting aspects in data justice-related themes.

The approach of engaging the stakeholders was need and context-based. Every respondent contacted for the pilot research project was sent a brief on the six pillars of data justice, and two versions of the questions. A larger version, from the handbook to give a deeper understanding was sent. However, the basic ideas of the six pillars and the questions that come under each of them were shortened for the ease of the interviewees. The larger questions were overwhelming to most of the respondents, so a shorter set was contextualised according to the stakeholder and respondent.

A total of 12 people were interviewed for the study. In this five people identified themselves as the member of the public; one identified as the member of the public as well as policymaker given the fact that she has been part of one government committee; two people identified themselves as policymakers; three identified as developers and one identified as developer and policymaker. Two of the respondents were not familiar at all with the data- and algorithm-related technologies and had no training or education on these. The familiarity varied from extremely familiar to moderately familiar for the rest of them. Except for two respondents, everybody lived in India. Everybody we interviewed had a bachelor's degree or above. Except for two of the respondents who had moderate access to the internet and equipment, the majority of them had unlimited access to the internet.

The ADJRP assessment involved interviews as a key format to engage

stakeholders. A total of 11 interviews were conducted for the ADJRP assessment in India. This included – 3 policymakers/enablers at national and state levels; 1 policy analyst; 4 from the developer's community; and 5 representing affected communities at the community level and research levels.

As mentioned earlier, the interactions with policymakers were the hardest to have. For most parts, they were either unwilling to answer in detail about the issues that impacted communities might have been facing as a result of inequitable access. Most simply refused to have conversations and the ones who did mostly agreed in words to understand the power relations and possible injustices, even as they were pointed out.

Most developers we talked to were aware of the potential that uncritical development of data collection and processing had in the past resulted in several injustices. Many were careful about what an unquestioned focus on neutrality and objectivity entailed. We had spoken to a developer who worked with speech recognition software to aid farmer interactions, and they were aware of the limitations of its inability to work with dialects and accents.

In the initial process of trying to look at case studies specific to our region and social contexts, we spoke to two contacts who were acquaintances, to get a broad understanding beyond what we already had. One of them was working with MeitY on a project on Data and AI. The other was a software developer who had worked on several FOSS projects in the past, and also writes on data policies.

For the ADJRP assessment, two workshop-cum-panel discussions were organised on digital mode (18th February and 25th February). The first discussion had one policymaker and three civil society representatives working with the impacted communities. It was attended by representatives from Community Information Resource Centres (CIRCs) from the ground, who run networks and centres that provide internet and information access to remote and otherwise unconnected areas.

From the interactions, we primarily intended to understand the following:

Implementing AI to address issues of communities (Equity, Participation, Knowledge Pillars)

- How has AI been deployed previously to address developmental issues? Can there be a use of AI that is just, equitable and participatory? (The case from Andhra Pradesh where an AI tool that detects pests, and communicates with the farmers has helped increase yields.)

- What are the practices in software/ developer communities and companies regarding inclusion? Are there mechanisms in place to have sufficient representation or inputs from women, queer, non-binary or trans folks, or other relevant marginalized communities when systems are built? Where do you think it should be used?

- Looking at issues where innovative AI powered strategies were used to intervene in community issues, were there instances where projects had to be dropped because during implementation, they realised it is ineffective as there are other social relations and institutions contributing to these problems.

- How can there be a process of communication built between the communities affected by such issues, the innovators and developers who can work on a solution, and the policy-level people who decide to prioritise or focus on a particular issue?

- Can we think about a bottom-up approach to building AI systems, where there is a way of raising issues in a participatory manner and collective solutions thought of locally?

Narratives of data-driven exclusions and invisibilisation.

(Power, Identity and Access Pillars)

- Listening to several narratives of exclusions and injustices encountered in India.

- Is it necessary that all exclusions or injustices resulting from AI be limited to training errors or faulty data? How are AI-powered systems intentionally used to exclude people of certain communities or groups?

- Who are the people who had to bear the brunt of AI-powered programs or automated programs? How do notions of objectivity and neutrality shape the narratives of these decisions?

Towards Data Justice: situating and reimagining the existing frameworks.

- What are the existing debates around data in India?

- What are the regulatory policies and frameworks in place regarding AI and data beyond privacy, security and protection?

- Looking at the injustices that have been mentioned, and the exclusions that have occurred, are the existing frameworks sufficient to redress the issues? How can they be modified, or even reimagined if necessary?

- How can systems be imagined and training given to people working on policy and software levels such that they are aware of these possibilities of injustices and could think around them when encountered?

In the next section, we detail the narratives of the people we have talked to.

NARRATIVES

Who is counted,

Who is excluded

Who does the counting?

Narratives from the impacted communities

The injustices that algorithms of platform and gig-economy apps cause has been documented previously[6]. In India, the workers in the gig-economy are counted as "clients," depriving them of many protections labour laws provide. In such an unorganised sector, **Shaik Salauddin** of the Indian Federation Of App Based Transport Workers (IFAT) is one of the leaders organising and unionising people working in ride-hailing and delivery apps. We speak to him in detail about the algorithms that cause injustices.

In December 2019, the Indian Parliament passed the controversial Citizenship Amendment Bill, along with the government's commitment to enforce a National Register of Citizenship. As Booker Prize winning author and activist Arundahti Roy put it, "Coupled with the Citizenship Amendment Bill, the National Register of Citizenship is India's version of Germany's 1935 Nuremberg Laws, by which German citizenship was restricted to only those who had been granted citizenship papers—legacy papers—by the government of the Third Reich. The amendment against Muslims is the first such amendment." Noting the use of an automated tool to decide the lineage of people in Assam, we spoke to **Abdul Kalam Azad**, a researcher from Assam, now at Vrije Universiteit Amsterdam, who had looked into detail the issues and exclusions created by the NRC in Assam. Learning of exclusions of Trans People from the same list, (already facing an undemocratic law like the Trans Act),[7] we spoke to two activists from the Trans Community, **Sai Bourothu**, who had worked with the Queer Incarcertaion Project and the Automated Decision Research team of The Campagin to Stop Killer Robots, and **Karthik Bittu**, a professor of Neuroscience at Ashoka University, Delhi and an activist who had

---

[6]Kashyap Raibagi, "The Plight of Gig Workers in an Algorithm-Driven World," Analytics India Magazine, March 12, 2021, https://analyticsindiamag.com/the-plight-of-gig-workers-in-an-algorithm-driven-world/.

[7]Sushmita Pathak, "India Just Passed a Trans Rights Bill. Why Are Trans Activists Protesting It?," NPR (NPR, December 4, 2019), https://www.npr.org/sections/goatsandsoda/2019/12/04/784398783/india-just-passed-a-trans-rights-bill-why-are-trans-activists-protesting-it.

worked with the Telangana Hijra, Intersex and Transgender Samiti.

Another exclusion we noted in our primary research was the homeless community, who are not counted in any of the data enumerations. We spoke to **Jatin Sharma** and **Gufran**, who is part of the Homeless Shelter in Yamuna Ghat on these exclusions and how it leads to the homeless people being denied basic healthcare and life-saving TB treatment.

Four researchers, activists and civil society leaders who had done considerable work on data related exclusions, surveillance, and identification software such as the Aadhar offered their perspectives on the debates, conversations and potential reimaginings of data injustices. **Srinivas Kodali**, independent activist and researcher; **Nikhil Dey**, of the Mazdoor Kisan Shakti Sangathan; **Apar Gupta**, lawyer and director of the Internet Freedom Foundation, and **Rakshita Swamy**, an NLU professor who also heads the Social Accountability Forum for Action and Research were the people who provided their insights.

The data injustices underneath India's Gig-Economy

In the Indian context, who does what work is closely connected to the caste and community history of each group. For example, a paper written by White and Prakash (2010) has pointed out how the SCs, STs and OBCs in India are disproportionately represented in the lower level jobs of the formal sector and predominantly informal sector (Harriss-White, B., & Prakash, A. (2010). Social discrimination in India: A case for economic citizenship). This sets the context of labour relations in India, which have historically remained unregulated or unaccounted. For example, the Gig workers have been considered as 'contractors' and not workers as per the Indian labour laws. It is only after a Public Interest litigation filed by the Indian Federation of the App based Transport Workers (IFAT) that the relationship between the aggregator and the driver was acknowledged as a wage worker relationship. It took several advocacy and legal efforts for the app based workers to be included in the Code on Social Security, 2020.[8] The interview with the president of IFAT also revealed how collectivisation and participatory subversion of the anti-worker practices supplemented by Algorithmic systems in the gig economy is challenging given the class background of the workers. "I am not a white collar leader" said Salauddin, indirectly referring to the structure of established trade unions and federation of trade unions in India, which predominantly unionise the formal workers. He also expressed how the union activities are limited by the lack of financial resources and time, since the union leaders themselves are drivers working full time. Consequently, the AI

---

[8]*"Petition in SC Seeking Social Security Benefits for Uber, Ola, Swiggy, Zomato Employees – the Leaflet," The Leaflet , September 21, 2021, https://theleaflet.in/petition-in-sc-seeking-social-security-benefits-for-uber-ola-swiggy-zomato-employees/.*

powered systems built around private transportation services are embedded in this historical unequal power relations. The pillars of data justice; power and equity are thus historically embedded in the negotiations the app based transport workers are engaged with both the state and the capital in India.

Erased Identities: Trans Community answers to Data Justice

Something that was pointed out to us was how institutional governance never actually recognized 'transgender' as an entity or as a biometric marker for any individual to have until 2014. As Sai told us, the queer and trans community has "been forcefully disappeared for 70 years of independence." As data is the primary marker for public policies and public welfare, the Trans community has been largely invisibilized, gentrified or ghettoised in the past 70-80 years of India's nation-building process. She explained how national statistical systems such as the National Sample Survey Organisation (NSSO) contributed to this. One key example, quoting Sai: "in the 2011 national census that captured approximately 4,11,000 trans persons exist in India. But these numbers were grossly underrepresented because even in small areas, there exist community groups. When this data was shared for cross verification, and they were divided district wise, it came out that there is very clear evidence from community estimation that there is average of 200 people and yet the census data pointed out that there is average of 4 transgender persons in one district. So there has been a clear disparity in how the community is represented. What is even more difficult is that this is going to inform policy. If a welfare scheme even tomorrow were to come up which determines some kind of aid for trans persons, it's going to grant that aid with the assumption that [there are only] 4 lakh in the entire country. [It] does not take into cognizance the fact that there might be so many more who have not yet been recognized or who have not been able to go through the governmental red-tapism yet to identify as such in some places."

Surveillance systems, like the model one being set up in Hyderabad[9] creates another issue for the already invisibilized and ghettoised trans-community. Being historically criminalised, and therefore having criminal records for most acts of survival like begging or sex work, any large scale predictive policing is going to be balanced against the community. Whatever existing redressal mechanisms isolate the issue into making it the responsibility of the individual "to have the knowledge of the harm [they're] going through, and then necessitated to also be able to understand the process of seeking redressal." Some of these are institutional, but some need not be intentional. If the codes were publicly available, and more

[9]"Inside the NYPD's Surveillance Machine," Ban the Scan, 2022, https://banthescan.amnesty.org/.

people from the communities disadvantaged know and can check this code, "inclusive coding to make sure that discrimination and bias are kept in check to a certain extent through their own efforts of reviewing these codes." This effectively highlights the power and participation pillars of data justice.

One of our interactions was with a science professor at a leading university in Delhi, a trans-man who had worked actively with the community on several issues. He spoke on the NRC exclusion of trans-people, and also of ML in science to broadly give an understanding of human biases. Trans-people are excluded from the NRC list. "Trans people had a combination of either missing documents because they fled abusive homes when they were young, or documents that were inconsistent". Around 2,000 trans people were excluded as a result of this, and a legal battle is ongoing[10]. Other algorithmic exclusions that happened in the country were instances of applications to institutions, where trans-people's names were misidentified as referring to two separate people with two separate names- and then summarily rejected. The same respondent also explains how ML tools work in some of the other projects he is working and collaborating in. As he explained, the 'science' of personality research has a long classist and racist history- a pseudoscience where workers are analysed and decided which role to be given based on personalities. When these ML tools were fed with datasets from classical psychology, their research has shown how the program does not provide a justifiable cut-off for saying one of these categories of personalities are more valuable than another. This helps debunk the previously held theory on the psychology of personalities.

Another aspect in relation to algorithmic injustice is how human understanding is also based on certain algorithms, and how these algorithms are also fundamentally flawed and riddled with various confirmation biases. "Human algorithms work like what we call a Bayesian learning algorithm. We see priors in how the world works, and we continue to think the world continues to work that way." AI tools can be used to show that when one feeds in datasets that don't have a bias, it shows that several things or patterns (that human beings with their cognitive biases assumed existed) do not actually exist. Race, similarly, is shown as "an arbitrary category consisting of looking at specific combinations of superficial" factors like skin or hair; when all genes are considered together, there is no consistent difference between racial categories. In this way, ML tools can challenge existing notions of power structures. Taking an example of cancer biopsies done by ML tools, the more data fed into the system can make diagnosis faster and more efficient. Of course, this has to be seen together with what the AI developers feel/need to be conscious of about working across the stack and considering other social factors as stated in examples of baby-weighing and TB samples, but unbiased,

[10]Saurav Das, "The NRC Poses a Two-Fold Predicament for Assam's Transgender Community," The Wire, 2019, https://thewire.in/rights/nrc-exclusions-assam-transgender.

centralised, anonymised records of all patients can be one such workaround in the design.

Citizenship and other Registers of Data Injustices: A Case from Assam

Interviews on the National Registry of Citizenship brought out the nuances of building an AI-powered system to determine the citizenship status of a population with a muddled history of colonialism and anti-immigrant sentiment. Wipro[11] deployed a Document Segregation and Meta Data Entry (DOCSMEN) software to digitise legacy data development of 39 million applicants in 2014. 1.9 million were excluded from the final list. The interview also pointed out that the 4 million people who did not have an Aadhar card[12] India's UID, were promised an Aadhar card after the NRC process, but continue to be excluded from all the entitlements and schemes linked to Aadhar. The government has already collected the biometric data, yet none of them knows what it is used for, nor can they reapply for a different Aadhar card as their application is "under process" for years. The interview also highlighted how the software-generated "family tree" system that verifies one's legacy data violated the basic human rights of hundreds of thousands of people who were involved in this process, either excluded from or included in the list.

An example was pointed out by a respondent who belongs to the Bengali Muslim community of Assam, seen largely scrutinised and victimised in the NRC Project. The Muslim immigrant community of Assam was brought into the State by the colonial administration as labourers to increase the revenue in 1826. They were brought from East Bengal – which later became East Pakistan and then Bangladesh[13]. The inclusion in the NRC list was based on something called the 'legacy document'. One needs to mention an ancestor who was included in an NRC done in 1951 or in the voter's list of 1966 to be in the NRC list. The legacy document should have the name and address of the ancestor, the precise address they were residing in and the precise details of everyone who is part of that family from that particular ancestors' generation. Our respondent explained the enormity of the data one had to present and how the 'family tree' algorithm excluded several in this process. One family tree will have hundreds of people

---

[11]"Wipro Digital Governance Solutions Empowered Assam Government to UPDATE NRC," Wipro, accessed November 25, 2022, https://www.wipro.com/public-sector/digital-governance--achieving-citizen-enrolment-in-record-time0/#:~:text=Wipro%20partnered%20with%20the%20Government.

[12]Aadhaar, India's UID project, assigns a twelve digit identification number to citizens based on their biometric and demographic data. Since inception and implementation, it has come under criticism for its issues of surveillance, privacy, data security, and exclusion from welfare

[13]Kaustubh Deka, "Bengali Muslims Who Migrated to Assam in 1871 Are Not 'Illegal Bangladeshis'," Scroll.in (Scroll.in, June 4, 2014), https://scroll.in/article/664077/bengali-muslims-who-migrated-to-assam-in-1871-are-not-illegal-bangladeshisAzad.

if they are basing it on their grandfather, including cousins and nephews. And each of these hundred people had to keep the matching spellings, including the spelling of the address, otherwise, the algorithm would exclude them from the list. Mild variations lead to exclusion and the grievance redressal process was reportedly even more vicious. Hundreds of these extended family members had to appear together before the tribunal to prove that they all belong to the same family. Our respondent pointed out that their plight is further complicated by the fact that the literacy rate of these regions- mostly floating islands, is as low as single digits.

Another important aspect pointed out in this is how the legacy codes given by the NRC Seva Kendra (service centres) led to the exclusion of several families. The applicants who were unsure about the address and other details of the "legacy source person" could go to the Seva Kendra to get a legacy code by providing their names and their legacy person's name. The code contains all the data about that particular person. However, if two families have the matching names of two of their ancestors, both the families would end up using the same codes for the legacy document. In the case of Assam NRC, many families had to fight each other in the tribunal to prove that the disputed ancestor was theirs. Our respondent recollected how, often, one of the families ended up losing the dispute and was excluded from the list.

According to the same respondent, the entire process of NRC citizenship contestation in the Assam State of India is built on a set of biased data: the D-voter list (the doubtful voter's list), the Assam NRC of 1951 and the 'reference cases' registered by the border police. Firstly, the 1951 Assam NRC was partial and several people were excluded from the list. The river islands of Assam that disappeared during the floods were only partially covered in the first NRC. These islands are largely populated by Bengali Muslim immigrants. Secondly, there were multiple people with the same names and ancestral names. If one of them happened to be in the reference case list or the D-voters list, all of them ended up getting excluded. The border police, deployed widely in Muslim dominant districts, has the right to search and collect the fingerprints of any 'doubtful' person.

At the Margins of Urban and Data: Homeless Population

From the interaction on the homeless shelters in New Delhi, the depth of a digitalised system of governance and health was revealed. Almost every health service, from following up on Tuberculosis (TB) Treatment (India is the highest in TB incidence statistics, with over 2.64 million cases), accessing vaccination, or even simply getting admitted to the hospital requires one to have identification

like the Aadhaar, and at times even a mobile phone where verification OTPs are sent. A health scheme named Nikshay[14] was designed by the government to cater to the nutritional requirements of recovering TB patients. As per the scheme a sum of Rs. 500 (~$6.5) is transferred to the bank account of TB patients under treatment. Despite the high occurrence of TB in the homeless population, many of them can't avail of this scheme due to the lack of a Nikshay ID and bank account. The homeless community does not have addresses, and therefore no IDs. The majority of them do not have a mobile phone as keeping them safe is difficult. In this case, they are dependent on the shelter staff for all OTP-based ID authentication systems. This is further complicated by their status as migrant workers who travel from one place to another and can't come back to a single shelter to avail of any entitlement service.

"There's almost 50-60 percent difference between the enumeration done by the civil society groups and the state census. Because the police are the ones who are doing the data collection." "The issue here is that with technological interventions, there is this assumption going around that you'll get a holier truth now, you'll get an unquestionable truth of some kind through the technology without the realisation that the theory laden ness of the data that technology will gather, that is coming from the human values. And what kind of human values are encoded in that algorithm will determine what kind of data you're collecting, and so exclusion is happening at that level. Now technology or no technology, you'll have outliers and exclusions. But once this is institutionally acknowledged, and recognized that the problem lies here, there'll at least be scope for challenging, negotiating and so on."

Another researcher on data policy, Srinivas Kodali, who had worked on Aadhaar and disenfranchisement mentions the inevitability of it. Data is being collected, and it has as good as become an inevitability. But when the state says they are collecting health data of citizens, they mean it only from an economic point of view that benefits a very few such that you can be sold products. Will the health data collected find out that platform workers have declining health, or are susceptible to accidents because of the incentive goals and therefore needed to be covered under benefits? This is another analysis that can come off from the data, depending on how it is looked at. "I don't think [the workers] are saying that we don't collect our data. Do actually collect it, but you're not doing that justice part," says Srinivas. "I'm just hoping that the AI revolution that the Indian government speaks about actually does take note of the ground workers on the street," he says, pointing towards an inclusive use of data. In the smart city mission the government had launched, a lot of the money that was given out was for data collection- building dashboards of data of cities. What gets counted in these is another question of exclusion. Slums are part of the data so much that

---

[14]NI-KSHAY-(Ni=End, Kshay=TB) is the web enabled patient management system for TB control under the National Tuberculosis Elimination Programme (NTEP)

funding drives the number of slums counted. Homeless or slum populations are counted as nuisances to traffic or the gentrified city. A comparison is when AI systems detected people without masks and fined them, instead of spreading or supplying masks. We "essentially have to give control to the people, you have to let them access these datasets, use it for their own good but what actually is happening is that the government has taken control of people's data and if we've seen all the national programmes or data collection programmes of the government, they are essentially claiming our bodies as government property." But it matters how the government is identifying or classifying these sections of people: "[we] don't know how the government is going to classify you. When we talk about the identity petitioners who were in front of the supreme court, saying we don't want to be identified as these kinds of people forever. We don't want to be identified as slum dwellers, we don't want to be identified as some kind of different people, we don't want that stamp to exist." This, he says, is the kind of profiling that Aadhaar was going to create - and this was known. There was no participation in the project, no means to ensure equity or question existing power structures. India's UID does not ensure there is no harm. The amount of data collected without transparency is disempowering the people, perhaps restricting them from accessing finance because of the categorisation that happens with data. Srinivas agrees how the government claims it has opened up data with new policies like Open Data[15]. But will that qualify as data justice? The issue, he says, "is that without the knowledge, without further internal data sets that the government possesses, you can't use these data sets. You don't have the computational infrastructure, you don't have the skill set of people to do anything with it. So data justice, without actually giving out the funds, giving out the infrastructure, giving out the manpower, is not going to be enough." When it is just people with access that gains, those gains are asymmetry.

Data Protection, Surveillance and Privacy: Legal Perspectives and Community Resistances

In a larger group conversation we had, we brought in experts from civil society, law and movements, who flagged several issues regarding the state of data in India. Apar, lawyer and director of the Internet Freedom Foundation mentioned the lack of cohesiveness in the policy or strategy documents that have been developed by various government departments and states. There are missing social or independent audits that demonstrate the effectiveness of outcomes in AI-based applications and deployments, as in the example of Aarogya Setu App (India's only App to deal with Covid-19 at pan India level). Most of these technologies become vaporware, basically something announced promisingly, but never really delivered. Utility audits are therefore necessary for AI-based systems, as with any

---

[15]See Annexure on policies and laws

technology that requires a lot of public expenditure, and collects a lot of data on the public. There is also the larger issue of a lack of a strong data protection law[16], leading to the possible deployment of AI-based systems utilising personal data for targeting. Without legal limits defined, there is no way to enforce data-related concerns, and these need to be done via an independent body. As Rakshita Pointed out, there should also be a legal audit that goes beyond a broad ethical checklist. As Nikhil of MKSS had explained, when large data systems are being built with public money, what will ensure they are used to proactively provide entitlements to people instead of dealtying and leading to exclusion? But another question that Rakshita raises is regarding making a conversation around data justice, and all the issues regarding its impacts on developments and democracy, go beyond domain experts, policy makers and lawyers, and get grassroots civil society groups, grassroots movements, trade unions and people themselves to be part of this conversation. The imagination required there will be different, but crucial. There must be a fine balance between where data can be provided to supplement decision making and where data actually makes the decision itself. There has to be clear safeguards, clear processes in place marking where data only facilitates and where it actually takes all the role for itself and makes a decision. It is very hard to separate data from 'the right to information'[17]; it's a subset of the other that is getting increasingly larger. There is a need to look at how the decision making vis-a-vis data takes place, in the gathering stage, the aggregation and amalgamation stage and also in the stage of use.

The primary points that these law and policy experts bring in are regarding the pillar of participation and power. Without a conversation where the people at the grassroots are involved, there is a violation of participation, where data work needs to be democratised to ensure inclusiveness of all voices, and an understanding of the pillar of power is a deeper critique of embedded power structures of the state.

---

[16]See appendix on laws

[17]Right to Information Act, see Annexure.

Conversing with the policymakers and developers

While people from the community who are impacted by AI and Data driven systems are the least heard narratives who need to have a more direct say in the implementation of these systems, it is also important to talk to the policymakers who make the policies and the developers who code life into the systems to see if they are aware of the potential inequalities that may be propagating. An aware developer can implement measures to counterbalance at least some of the negative effects of the AI-system, and a policymaker who understands the nuances of power or equity that make the policies they frame would be impacting, they can work on building software and policies that are more equal.

It is not the case, however, that all stakeholders understood the broad concerns that were raised in light of data justice. For India's developers, although part of formal employment, are often overworked and lacking strong unionisation. Attempts at unionising are relatively recent.[18] Policymakers were asked of their awareness of data injustices when deciding on policy that mandates AI and data systems.

Here, we spoke to **Jayesh Ranjan**, IAS and special IT Secretary of Telangana State, one of the few states that have their dedicated AI missions, and is itself an IT and tech powerhouse. **Naveen Kumar**, another IAS officer who is the special Health Secretary of Andhra Pradesh had worked on integrating data systems with health, and was therefore a potential stakeholder we could converse with. **Abhishek Singh**, IAS is the CEO of MyGov (a citizen engagement platform by the government) and the National E-Governance Division (NeGD). **Parminder Jeet Singh,** head of the non-profit IT for Change, was also part of the expert committee on the Non-Personal Data Governance Framework. **Rahul Panicker**, a developer, now principal technologist to a robotics firm in San Francisco, was previously the head of Wadhwani AI in Mumbai who worked on several AI for Good initiatives. Rahul, as both a developer and a contributor to the national policy document on AI was another insightful interaction we had in the course of our research. **Deepak Padmanabhan**, a developer now teaching Computer Science at Queens University, Belfast, and an activist himself had different perspectives to share which were resourced with his experiences as a developer and also his activism. Two other **developers** who have requested anonymity and worked on AI tools that improve customer support and language processing have also shared their narratives from the side of developers.

---

18 "Unions in It Sector?," The Hindu BusinessLine, April 12, 2019, https://www.thehindubusinessline.com/opinion/letters/letters-to-the-editor/article26822393.ece.

We chose to speak to these policymakers because of the key role they played in the country's or their state's AI and Data policies, and coders who worked on projects that they believed could have been implemented better.

As one developer we talked to mentioned, and a science professor from a prestigious university also added, there is an evident lack of social science training for India's STEM graduates. One developer who had been working on a language processing AI tool confirmed the same from his experiences in his workplaces. While he was open to conversation on the issue, we also had another interaction that shut down such a possibility. He alleged that understanding how AI tools work is something not to be left to the people to understand. It is a matter of technical, subject expertise, and those without it wouldn't really care, just like how they 'don't understand why fuel prices fluctuate, or how a medicine works.' In this sense, we discovered that the developers were mostly marked by the power, knowledge and participation pillars, but more so a lack of awareness and action on the line of these three pillars. Without the necessary training in ethics, social sciences or understanding of how the technologies can have impacts on communities, how they are excluding people, or how everyone is not represented in the processes of assessing these impacts, we see a need to reorient these along the lines of specifically the pillars of power, knowledge and participation.

However, the interactions with other developers generally did receive positive responses.

On Existing Frameworks for Data

The President of the National e-governance department was another of our respondents to talk about data at the level of policies. His opinion was that the way to go forward in cases of possible exclusions due to the implementation of AI-powered systems in governance is not to discard the programmes but to build support systems around them to avoid exclusions. For example, responding to the question of the data injustices that take place in India due to AI-powered systems in governance such as Samagra Vedika,[19] He said: "I can think of an analogy. Very often we make highways, for the convenience of citizens, people move and that leads to economic growth and all. [But] accidents also happen. People do lose their lives. But what do you do? You try to ensure that road safety measures are taken up, people are made more aware, you ensure that if certain bottlenecks result in more accidents, you try to make them safer, you bring in

---

[19]Samagra Vedika is an integrated platform comprising a 360-degree profile of every citizen in the state of Telangana using Big data, ML and Graph database. Tushar Dhara, "Cancelled Ration Cards Deprived Telangana's Poor of Food Rations amid Lockdown," The Caravan, August 21, 2020, https://caravanmagazine.in/government/cancelled-ration-cards-deprived-telanganas-poor-of-food-rations-amid-lockdown.

more safety measures, cars are made safer, they are, people are taught about using seat belts, so these are the measures that you do to ensure that when people are using a highway, they remain safe on the highway, speed limits are prescribed. So similarly when we are building a data-based IT system for the larger public good, there might be some elements who'd try to gain the system, who'd try to subvert the system, there will be risks."

Parminder Jeet Singh, of IT for Change was one of the discussants in our workshop on Data Justice. Along with his work with his organisation, he was also part of the expert committee on the Non-Personal Data Governance[20] Framework (NPD), and he was of the opinion that the problems of homeless people, people from the transgender community and the gig workers have a political context and these problems are magnified with the use of technology. While these injustices need to be resisted, the concern about the fact that these arguments often extend towards a standpoint that is essentially anti-technology. In the midst of some dilemmas on the collection of health data in the background of the health data retention policy, though poor people and civil society are concerned about their health data being collected, ultimately multinational hospitals would build technologically advanced health systems using the data of the rich for the rich and the poor and the under-privileged will be left out of it due to the lack of representational data.

"Now the same people who do not want so-called poor people's data to be collected, go to an AI conference and they're constantly complaining about data bias in AI. Of course, there will be data bias if you don't allow data collection of certain kinds of people. I know the problem with blacks in the US, with Dalits in India, but this is the reality of the two sides of the problem. On one side the need for inclusion of the data, because everything in the world, education, health, agriculture, everything is going to become data-based and if data is not there, one will be finding deficient services." He added to substantiate. He also spoke about how the NPD framework goes beyond the two usual policy frameworks which either advocates market innovation within a regulatory framework or a model of benevolence which advocates philanthropy. In the right-based approach of the NPDG framework, the data subjects, individual and collective have a right over their data and the value of all the derivative data would be with the collective.

'AI for Social Good': Experiences from Rural India.

There are several AI-powered projects that are in development and under execution right now, that are run in collaboration with various government departments. One such AI-powered tech was an app designed for ASHA workers that helped them provide accurate, timely, geo-tagged and tamper-proof weight estimation of new-born under a month of age.

The World Health Organization classifies newborn infants with weight less than 2500g as Low Birth Weight (LBW) infants.[21] It is also estimated that a quarter of Indian newborns are LBW[22], and this is directly linked to the baby's survival chances. The WHO's goals are also to cut the incidence of LBW infants. Therefore, accurately identifying low birth weight babies is an important first step to providing them with further healthcare, and decreasing child mortality rates. This proved to be a challenge in practice, because when records were examined, almost all the babies noted were recorded being *exactly* 2.5 kgs, the minimum healthy weight to not fall under the LBW criteria. It was clear that there was some fudging of numbers going on at some levels of the data collection and recording process.

One planned solution consisted of software that converted a video taken with the smartphone the ASHA[23] workers are provided with into a 3D mesh of the baby, which the software can then use to accurately estimate the weight of the baby[24]. Here, AI solution is one part of the technology stack that has to fit into the workflows of everyone involved.

Designing solutions requires a different approach to avoid exclusions. 'Product innovation is about working with the users and identifying the market gaps,' the developer mentioned. Designing an AI solution demands one to look into other parts of the societal chain- the example of the anthropometry solution ran into very different sets of errors that did not have to do with data gaps or biases. For one, even as the health system tried out the solution, it could not take into account the lighting conditions of rural Indian homes- which are not ideal for mobile cameras to measure such particular detail. Data collected under ideal conditions to build the software, say, from hospitals, would have much better lighting conditions. However, because of a very stark existence of caste- certain parts of a village are caste ghettos- if the health worker doesn't visit the place, none of these data factors would apply in the otherwise ideal AI-based datasets.

This was one of the issues with a software developed to improve TB detection. While the solution effectively detected TB samples with a high rate of precision, the problem with India's TB infrastructure does not actually lie in the detection part. Connecting this to the narratives from the representatives of the Hausla homeless shelter on India's TB crisis, one could understand how India's problem with TB is more social- it is a lack of policies and welfare benefits that directly help the

[21]World Health Organization, "Low birth weight Country, regional and global estimates: Department of Reproductive Health and Research (RHR), 2005." https://www.who.int/publications/i/item/9280638327

[22]UNICEF. State of the World's Children: Celebrating 20 Years of the Convention on the Rights of the Child. Unicef, 2009

[23]ASHA (Accredited Social Health Activist) workers are India's frontline health workers instituted by the Ministry of Health and Family Welfare as a part of India's National Rural Health Mission.

[24]"Newborn Anthropometry." Wadhwani AI, November 3, 2021. https://www.wadhwaniai.org/programs/newborn-anthropometry/.

patients continue their course of treatment and provide them with nutritious food that needs to be looked at, much more than the stage of testing and detecting the disease.

Taking these examples on AI systems, human workflows have to be appropriately modified so all of this forms part of the larger solution, and this requires a multidisciplinary approach. The institute Rahul worked for had to work with "agricultural experts, with people who have social sector background in deploying programmes, doctors, product designers, engineers" to work between these workflows. AI can reflect the creators' intent as "technology is fundamentally an amplifier of human intent'. This points to the problem of weak institutions. There cannot just be 'unintentional bad,' but also 'intentional bad'. The developers have to be conscious of this. "There are significant power disparities and these power disparities also apply across communities, across religion, across castes, across social-economic strata, gender, age, education levels," and the solutions cannot be for just the literate or digitally literate people. While designing any AI-based solution, the developers' community is and should be engaged in this understanding. There cannot be Business to Consumer approaches without human intermediaries which in these cases, are the agricultural extension workers or the ASHA workers.

For developers, there is a need to be sensitive 'about preventing unintentional bad,' and need to involve the community in the AI based data processes. India faces an acute shortage of doctors and agricultural scientists in relation to the population. In addition to talking about minimising damage from AI based data distortions and 'blind automation', there is a real need to understand how it can be used for good. By trying to combat the statistical injustices and challenge presumed structures of authority of knowledge, focusing on the transformative potentials, and trying to look at the subjects of the AI tool in the relational sense, the pillars of equity, participation and knowledge can be best related with the conversation we had.

Labourers' Dignity and Linguistic Diversity: Narratives on inclusive Coding

India is one of the most linguistically diverse nations in the world. According to the 2001 census, there are 30 languages in India, it is estimated that there are 1599 dialects within these main languages.[25] India also has a complicated history of linguistic politics, where the official languages are the tongues spoken by the dominant communities and several regional dialects are considered inferior. One developer we interacted with was working on an AI tool that helped farmers in

[25]"Census Data 2001: Government of India." Census Data | Government of India. Accessed November 25, 2022. https://censusindia.gov.in/census.website/data/

markets by converting speech to text, for better communication between different language-speakers. Such a tool would require vast amounts of data to train the algorithm properly to convert speech accurately. With a lack of commercial interest, the project would fall short of including several of these languages, like those spoken in North-East India, a region and communities that have been facing systemic neglect, racism and exclusion. It is also moreover very rare that farmers, or any community for that matter, speak a formal and standardised dialect. Following a top-down approach in developing, and having little consultations with the community, the software is intended to leave issues of representation in the data. Existing power structures are not properly engaged with, and the questions of access and participation are not addressed.

Dignity of labour, as Marx argues,[26] is essential for both the individual worker as well as the idea of labour taken entirely. In our free and conscious activity, the products and the processes of labour, we express our species character. It is not something that is simply related to wages or better conditions of work- both something lacking in most of India's outsourced 'call-centre' economies. As one of our conversations pointed out, there is clear pyramidal division within the ICT sector, where the workers on the helplines and call-desks unofficially fall into a different tier. It is only recently- in 2017, to be exact- that unionisation has started to slowly pick up.[27] One of our conversations was with a developer who worked with AI systems on call centres, and now teaches and researches on AI both technically and in the social sciences. He explained his experiences, saying how "injecting A.I. into the customer care centres makes a statement that an employee's work can be automated." Ironically, the workers would themselves provide the data for the AI to learn their work, leading to displacing themselves, the human workers, from their own jobs. He maintains that we need an interdisciplinary approach that works out the sociological aspects of human-technology interactions as well, because right now, it is just the engineers who play any role. Their lack of training in these social aspects reflets in the designs. They rarely have space for contestability, for dispute or dialogue, given the real impacts the software seems to be making. During his design of software, he recalled how the team built an option where the call centre worker had an option to override an AI suggestion on his gut-feeling and provide an alternative solution. The respect and importance given to the human contribution, he says, had also "contributed to the efficiency improvement and considering the worker as a respectable person." While these best align with the pillars of knowledge and participation, one key point is the dignity of a worker, something beyond the six-fold categorisation of pillars.

---

[26]Karl, Marx, "Economic and philosophic manuscripts of 1844," In Social Theory Re-Wired, pp. 152-158, Routledge, 2016.

[27]Monalisa Das, "TN Gets India's First Trade Union for IT Sector: What's in It for Employees," Newsminute, 2017,https://www.thenewsminute.com/article/tn-gets-indias-first-trade-union-it-sector-whats-it-employees-72073.

Building State-Level Policies and Planning for Accountability

The Indian State of Telangana is probably the only state that has leaped ahead in terms of AI research[28] and deployment having an AI strategy already in place, and having recognised 2020 as the 'Year of AI,[29]' making several efforts around these. There are definitely scopes of misuse and error around AI deployment, and the state's IT Secretary Jayesh Ranjan is confident that mechanisms to fix these have been taken. The state's strategy document on AI[30] has, according to him, "very consciously put together a section which speaks about governance, ethics and privacy." He says it will be ensured that the "AI procurement guidelines are commensurate with the standards of ethics and governance," and that they are also planning to introduce AI into the curriculum. "We have also promised that we will run data ethics courses in our institutions, particularly those institutions where AI is not an independent subject." The IT Secretary acknowledges how the systems in place are not 100% accurate. He claims, however, that the percentage of accuracy has gone up from 73 to 87 per cent. On Samagra Vedika, he says that it 'is a pro-poor initiative,' whose beauty is that it can process applications "without relying on Aadhaar, phone numbers or anything else which is of an extremely personal nature". The Supreme Court, on their historic September 2018 verdict, struck down Section 57 of Aadhaar act, which had allowed using Aadhaar "for establishing the identity of an individual for any purpose.[31] " He justifies this by saying how one cannot refuse to subject oneself to police enquiry in a passport application saying it violates one's privacy. "The government has all the rights to inquire about your eligibility. Instead of doing it manually, I'm doing it partly using technology and partly manually."

Now, the extent to which redressal of grievances will be handled is uncertain, but the general promises of inclusion of ethics and injustices in the syllabi, and having a strategy document that has similar guidelines is along the pillars of participation, access and power pillars. The establishment of an ombudsman to interrogate such instances also runs along these lines of questioning power structures, and enabling access to infrastructure and ensuring public trust. And this was something that came up in our conversation with another administrator, an IAS officer from Andhra Pradesh. When the Right to Information act was passed back

---

[28]"To Make Hyderabad amongst Top 25 Global AI Innovation Hubs," Telangana AI Mission (T-AIM), accessed November 25, 2022, https://ai.telangana.gov.in/.

[29]"Telangana's Year of AI- 2020 and Beyond," Govt. of Telangana. https://invest.telangana.gov.in/wp-content/uploads/2021/01/AI_2020-and-Beyond.pdf

[30]"Telangana Artificial Intelligence Framework 2020," Govt. of Telangana, https://it.telangana.gov.in/wp-content/uploads/2020/07/Govt-of-Telangana-Artificial-Intelligence-Framework-2020.pdf

[31]"What has been changed in Aadhaar Amendment Bill," SFLC, 2019. https://sflc.in/what-has-been-changed-aadhaar-amendment-bill

in 2005, there were independent statutory bodies set up as part of it, to enforce accountability.[32] Speaking on how Andhra Pradesh plans to digitalise health services, he suggested that there should be an independent Data Ombudsman similar to how the Information Commission is statutory. The independent or judiciary based system of an ombudsman can help resolve issues or complaints of data security, breaches, and misuse that may arise.

Recommendations

From our conversations and research, and taking into account the narratives and suggestions from the people we talked to, we came up with some general recommendations regarding the landscape of data justice in India, and possibilities to combat data injustices.

- An ethics committee should be in place within IT companies to ensure inclusion, prevent biased data collection, ensure representation.
Several of the developers we talked to, and other communities who have been following them, mentioned the issue of the demographic composition of the workforce and decision-makers. Focussing on the pillars of Participation and Power, meaningful participation with representation can lead to a transformative inclusion.

- Courses on AI should have modules on data justice:
Another point raised across the workshops and interviews both were regarding the general lack of social science education and understanding of the ethical, social, political implications of the tools they are designing and the power it holds. While the several strategy and policy documents for AI mention about the social impacts, on the ground, such discourses are not part of the curriculums in technical education institutions yet. It is imperative, therefore, that developers who design AI systems should have a broad understanding of the dimensions of social justice their systems will interplay with.

- Vernacular resources should be made available for the common citizen to understand what is data and what is data justice.

- The Aadhaar, or India's national project to build a UID for all its citizens have been flagged on multiple levels in the past for the power such data concentration holds. As the Data Justice project looks beyond these questions of data protection and surveillance, our interactions unearthed possible social exclusions the identification entails. Using AI to look into

[32]"Central Information Commission," CIC. http://cic.gov.in/index.html

datasets the Aadhaar is linked to, enforcing it to avail welfare benefits in a system where access itself is an issue, has led to several exclusions. With new similar identification programs are also in the works, the entire project needs to be audited and assessed, as well as redressal mechanisms put in place that do not put the burden on the welfare beneficiaries. There should be transparency in such processes, and sufficient consultations with the impacted community before being put into place.

- Issue of access infrastructure:
  A general point from our interactions that are relevant to India, and our own organisation's work on the ground bringing internet connectivity to communities is the issue of access infrastructure. One of the fundamental questions of data justice in India actually goes back to a fundamental question of access to data, information, digital infrastructure, digital knowledge and literacy. To be counted as part of the process is an important factor. As the transgender activist we talked to mentioned,

  > "most of the arguments on sex workers rights or trans rights aren't fundamentally dependant on numbers, because anything that is stigmatised will anyway show up in smaller numbers. Our argument is not around numbers. Even if there is one person, what policy should be in place. But to the extent that policymakers want numbers, it matters"

  This is the same for most marginalised identities that aren't counted as part of the system. Also, as our interactions from the homeless shelter showed, most technologies or solutions are designed counting the privileged in mind- and this is a complicated term because, in most of the global south, access to these technologies or the know-how puts one in the relatively privileged sections of society.

- There should be a comprehensive social auditing and policy analysis of the different AI frameworks and strategy documents that have been made by several individual states (like Telangana) and other central government institutions (like NITI Aayog) in India.

  - The dominant, existing research on data justice points out to the aspects on how data is biased, technology amplifies the biased data and therefore impinges on social justice. In the context of the global south, however, there are some differences in the social and economic scenarios. With a doctor ratio of 1:1511 and a nurse ratio of 1:650 compared to the UN mandates of 1:1000 and 1:300 respectively, India faces a shortage of doctors and skilled medical experts. This is where the points raised by our interviewees are important. Unbiased medical data and automation might help in diagnosis. Just as the pest detection algorithms have worked in the scenario of shortage of experts.

Future Work

Digital Empowerment Foundation is actively working to break down the data discourses and data rights to the communities. We are now implementing a "Talking Data to The Fourth Pillar" program, focussing on hundred journalists in India and Bangladesh in partnership with the Association of Progressive Communication and VOICE Bangladesh. The objective of this program is to encourage reporting on data and privacy-related issues in the subcontinent and engage journalists in these discourses. As a continuation of the Data Justice project, we have also collected the visual narratives of gig and platform workers. These narratives are being developed as part of a documentary on data justice. We are also developing interviews and interactions we had for this project into a collection of articles in collaboration with the respondents. Further, we are actively incorporating data rights for rural communities in all our programs to educate and learn from them about the day-to-day issues with data protection and privacy.

Annexure

Overview of the legal framework in India

National Data Sharing and Accessibility Policy, 2012 and 2022

The governments in state and centre, have a lot of data collected and compiled under various projects and activities under them. This data is of relevance to research and policy, and could be made public just like other information that should be public under the RTI. In 2012, the National Data Sharing and Accessibility Policy was published by the Indian Government, "to increase the accessibility and easier sharing of non-sensitive data amongst the registered users and their availability for scientific, economic and social developmental purposes."

Some of the critiques that were raised regarding this, as mentioned in one of our conversations as well, were regarding the quality of the data.

In February 2022, the Ministry of Electronics and Information Technology (MEITY) released the draft India Data Accessibility and Use Policy 2022 (or Draft Policy) for public consultation. With this, the government hopes to eliminate the issues of the old OGD policy, and open this data up for automation. However critics still point out that it proposes to price data on the basis of a 'data ownership' mode,lacks a grievance redressal mechanism and needs a robust Data Protection policy for proper functioning.[33]

The National Strategy on Artificial Intelligence, 2018

Outlining the nation's plans and ambitions for utilising AI, the premier think tank born out of the Planning Commission, the NITI Aayog, released two strategy documents on AI in 2018 and 2021, identifying five focus areas where AI could have a positive impact. The 2018 document, National Strategy for AI (AI4ALL), demonstrates how AI can be applied to India's healthcare, agriculture, education, smart cities and infrastructure, smart mobility and transportation such that it benefits the people. The 2021 document further builds on it and "aims to establish broad ethics principles for design, development, and deployment of AI in India"; the document from August 2021 gives recommendations to the private sector, research and academia in their use of AI. The paper even addresses ethical concerns regarding the use of AI, like fairness, privacy and transparency. Critics have however pointed out that there have been little to no public consultation or

[33]Anita Gurumurthy and Nandini Chami, "Forging a Social Contract for Data," IT for Change, March 1, 2022, https://itforchange.net/forging-a-social-contract-for-data.

responses sought unlike policies usually do[34]. Also, these documents, however, beyond framing broad policies for AI such that it balances opportunities for innovation and the possibilities of risks, does not delve deep into the broader concept of data justice.

The Personal Data Protection Bill, 2019

In 2019, the Indian Parliament constituted a Joint Parliamentary Committee, to look into drafting, taking public and expert opinion, and finalising a Personal Data Protection Bill. In December 2021, it was finally tabled, and renamed the Data Protection Bill. It was only in 2017 that the Puttaswamy judgement in the Indian Supreme Court defined the right to privacy as a fundamental right, and asked the government to draft a data protection law. In 2019, the Personal Data Protection Bill was introduced in the Lok Sabha. Although it did take hints from the EU's GDPR, it was severely limited in scope and handed over sweeping powers to the state. The renamed bill tabled in 2021, as critics point out, enables personal data to be surveilled and commercialised under the terminology of "data economy" [35]

[34]"Niti Aayog Discussion Paper: An aspirational step towards India's AI policy," CIS-India, https://cis-india.org/internet-governance/files/niti-aayog-discussion-paper

[35]Rohin Garg, "Key Takeaways: The JPC Report and the Data Protection Bill, 2021 #Saveourprivacy," Internet Freedom Foundation (Internet Freedom Foundation, December 21, 2021), https://internetfreedom.in/key-takeaways-the-jpc-report-and-the-data-protection-bill-2021-saveourprivacy-2/.

Bibliography

Data Justice India Report, DEF & Alan Turing Institute

"Central Information Commission." CIC. http://cic.gov.in/index.html

"Niti Aayog Discussion Paper: An aspirational step towards India's AI policy." CIS-India. https://cis-india.org/internet-governance/files/niti-aayog-discussion-paper

"Telangana Artificial Intelligence Framework 2020." Govt. of Telangana. https://it.telangana.gov.in/wp-content/uploads/2020/07/Govt-of-Telangana-Artificial-Intelligence-Framework-2020.pdf

 "What has been changed in Aadhaar Amendment Bill." SFLC, 2019. https://sflc.in/what-has-been-changed-aadhaar-amendment-bill

"Advancing Research and Practice on Data Justice." GPAI, 2022. https://gpai.ai/projects/data-governance/data-justice/.

"Census Data 2001: Government of India." Census Data | Government of India. Accessed November 25, 2022. https://censusindia.gov.in/census.website/data/.

"Inside the NYPD's Surveillance Machine." Ban the Scan, 2022. https://banthescan.amnesty.org/.

 "Newborn Anthropometry," Wadhwani AI, November 3, 2021, https://www.wadhwaniai.org/programs/newborn-anthropometry/.

"Petition in SC Seeking Social Security Benefits for Uber, Ola, Swiggy, Zomato Employees – the Leaflet." The Leaflet , September 21, 2021. https://theleaflet.in/petition-in-sc-seeking-social-security-benefits-for-uber-ola-swiggy-zomato-employees/.

"The World's Most Valuable Resource Is No Longer Oil, but Data." The Economist. The Economist Newspaper, 2017. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

"To Make Hyderabad amongst Top 25 Global AI Innovation Hubs." Telangana AI Mission (T-AIM). Accessed November 25, 2022. https://ai.telangana.gov.in/.

"Unions in It Sector?" The Hindu BusinessLine, April 12, 2019. https://www.thehindubusinessline.com/opinion/letters/letters-to-the-editor/article26822393.ece

"Wipro Digital Governance Solutions Empowered Assam Government to UPDATE NRC." Wipro. Accessed November 25, 2022. https://www.wipro.

com/public-sector/digital-governance--achieving-citizen-enrolment-in-record-time0/#:~:text=Wipro%20partnered%20with%20the%20Government

Das, Monalisa. "TN Gets India's First Trade Union for IT Sector: What's in It for Employees." Newsminute, 2017. https://www.thenewsminute.com/article/tn-gets-indias-first-trade-union-it-sector-whats-it-employees-72073.

Das, Saurav. "The NRC Poses a Two-Fold Predicament for Assam's Transgender Community." The Wire, 2019. https://theleaflet.in/petition-in-sc-seeking-social-security-benefits-for-uber-ola-swiggy-zomato-employees/.

Deka, Kaustubh. "Bengali Muslims Who Migrated to Assam in 1871 Are Not 'Illegal Bangladeshis'." Scroll.in. Scroll.in, June 4, 2014. https://scroll.in/article/664077/bengali-muslims-who-migrated-to-assam-in-1871-are-not-illegal-bangladeshisAzad.

Dhara, Tushar. "Cancelled Ration Cards Deprived Telangana's Poor of Food Rations amid Lockdown." The Caravan, August 21, 2020. https://caravanmagazine.in/government/cancelled-ration-cards-deprived-telanganas-poor-of-food-rations-amid-lockdown.

Garg, Rohin. "Key Takeaways: The JPC Report and the Data Protection Bill, 2021 #Saveourprivacy." Internet Freedom Foundation. Internet Freedom Foundation, December 21, 2021. https://internetfreedom.in/key-takeaways-the-jpc-report-and-the-data-protection-bill-2021-saveourprivacy-2/.

Gurumurthy Anita, and Nandini Chami. "Forging a Social Contract for Data." IT for Change, March 1, 2022. https://itforchange.net/forging-a-social-contract-for-data.

Heaven, Will Douglas. "Predictive Policing Algorithms Are Racist. They Need to Be Dismantled." MIT Technology Review. MIT Technology Review, December 10, 2020. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Jany, Libor. "Researchers Use AI to Predict Crime, Biased Policing in Major U.S. Cities like L.A." Los Angeles Times. Los Angeles Times, July 4, 2022. https://www.latimes.com/california/story/2022-07-04/researchers-use-ai-to-predict-crime-biased-policing.

Marx, Karl. "Economic and philosophic manuscripts of 1844." In Social Theory Re-Wired, pp. 152-158. Routledge, 2016.

Mukesh Adhikary, "'Data Is the New Oil': Mukesh Ambani Says Global Firms Should Not Control India's Data," Business Today, January 18, 2019, https://www.businesstoday.in/latest/economy-politics/story/data-new-oil-mukesh-ambani-

says-global-firms-should-not-control-india-data-159818-2019-01-18.

Pathak, Sushmita. "India Just Passed a Trans Rights Bill. Why Are Trans Activists Protesting It?" NPR. NPR, December 4, 2019. https://www.npr.org/sections/goatsandsoda/2019/12/04/784398783/india-just-passed-a-trans-rights-bill-why-are-trans-activists-protesting-it.

Raibagi, Kashyap. "The Plight of Gig Workers in an Algorithm-Driven World." Analytics India Magazine, March 12, 2021. https://analyticsindiamag.com/the-plight-of-gig-workers-in-an-algorithm-driven-world/.

World Health Organization. "Low birth weight Country, regional and global estimates: Department of Reproductive Health and Research (RHR), 2005." https://www.who.int/publications/i/item/9280638327