



CONVERSATIONS ON DATA JUSTICE IN INDIA

Combatting Algorithmic Exclusions and Building Fairer Frameworks

OSAMA MANZAR, JENNY SULFATH, ANANTHU RA



All hyperboles considered, the second decade of the twenty-first century certainly sees an undeniable pervasion of algorithms into many facets of lives. Where this previously had to do with the more savvy sections of society, now it has come to impact everyone in different ways.

We are witnessing a ‘datafication’ of society, where massive amounts of data being collected and processed systemically have become normalised. This comes with its own set of new ethical issues and power dynamics that need to be analysed. Data-driven discrimination is advancing at a similar pace to data processing technologies. However, awareness and discourse around this discrimination are often limited in scope to how individual privacy is trampled on, or at max, to how data and the aforementioned lack of privacy and laws are leading to mass surveillance. While all of these are relevant issues, there is newer research that seeks to look beyond them. Data justice is defined as “fairness in the way people are made visible, represented and treated as a result of their production of digital data.” Data Justice is a broad research area that brings together social justice, technology, and the impacts of the data paradigm on various communities. This book is the output of a series of conversations held with policymakers, practitioners, impact communities and policy experts on data justice in India.

Our Editors

The Digital Empowerment Foundation helps connect and digitally enable communities that are underserved and at the margins across the country, and works in research, advocacy and policy building in digital empowerment, access, rights and justice.

Along with Osama Manzar the founder-director of DEF, Jenny Sulfath and Ananthu R A who work with the communications and research team at DEF have together conceptualised the discussions, coordinated with the participants, and conducted the pilot research that has culminated in this book.

Illustrations and Cover Design by

Sharada Kerkar

Our Authors

The authors of this book are (in alphabetic order):

Abdul Kalam Azad,
Abhishek Singh,
Bittu K,
Deepak P,
G S Naveen Kumar,
Jatin Sharma,
Jayesh Ranjan,
Nikhil Dey,
Rahul Panicker,
Rakshita Swamy,
Shaik Salauddin,
Shankar Singh,
Srinivas Kodali,
Urvashi Aneja



Conversations on Data Justice in India: Combatting Algorithmic Exclusions and Building Fairer Frameworks

Year of publication: **2023**

This work is under a creative common attribution 4.0 international licence



Edited by:

Osama Manzar, Jenny sulfath and Ananthu RA

Design and Layout by: **Satish Kumar**

Illustrations and Cover Design by: **Sharada Kerkar**

Published and Distributed by:

Centre for Development Policy and Practices

ISBN: **978-81-959145-7-9**



Scan QR Code to Read this Book

Supported by:

Digital Empowerment Foundation

Email: info@defindia.org

Website: www.defindia.org

Keywords

Algorithmic Injustice, AI for Social Good, Neutrality of Technology, Social Justice, Data Transparency, Data Access



Conversations on Data Justice in India

Combatting Algorithmic Exclusions and Building Fairer Frameworks

Editors


**Osama Manzar, Jenny Sulfath,
Ananthu RA**

Preamble:

The findings/views/opinions expressed in this book are developed from interviews and discussions with the contributors as part of a research project Digital Empowerment Foundation undertook with the support of the Alan Turing Institute and GPAI, and do not necessarily reflect the views of the publisher and the organisation(s).



Content



<u>Preface</u>	8
<u>Abbreviations</u>	10
<u>Introduction</u>	12
<i>Osama Manzar, Jenny Sulfath, Ananthu RA</i>	
Part 1: Data Injustice: Uncovering Exclusion in the Digital Age	21
1.1 Digitizing Citizenship and Measuring Genealogies: The NRC and Algorithms in Assam	22
<i>Abdul Kalam Azad</i>	
1.2 Identity, Homelessness and Machine Systems: Working with a Homeless Shelter in Delhi	36
<i>Jatin Sharma</i>	
1.3 The Moving City and Static Inequalities: Platform Economy and Transport Data in India	48
<i>Shaik Salauddin</i>	
1.4 The Decisive Machine And Algorithm Exclusions: An Overview	60
<i>Bittu K</i>	
1.5 Who Governs “Data Governance”	70
<i>Nikhil Dey</i>	


1.6 Disconnected by Design: Unraveling the Impact of Digitized Attendance on NREGA Workers <i>Shankar Singh</i>	86
Part 2: Unpacking AI Ethics	91
2.1 Humanisation of Data and the Datafication of Humans: Government, Private Entity, and the Struggle for Data Justice <i>Srinivas Kodali</i>	92
2.2 E-Governance and Digital Governance Imagination in India <i>Abhishek Singh</i>	102
2.3 Non-Neutral Technologies and AI Policies: Evaluating Biases and Algorithms <i>Urvashi Aneja</i>	112
2.4 Corporate Imaginations of Artificial Intelligence and the Need for Structural Reform <i>Deepak P</i>	118
2.5 Why Exploring the Legal Structure of Data Usage, Engagement with IT Systems and Digital Architecture is Necessary for an Ethical AI <i>Rakshita Swamy</i>	128
Part 3: Perspectives on AI Innovations: Exploring Visions and Realities	133
3.1 AI in the Global South- Gaps in Data and Possible Frameworks <i>Rahul Panicker</i>	134
3.2 Between Emerging Technologies and the Common Good: Exploring a Delicate Balance <i>Jayesh Ranjan</i>	146
3.3 Imagining Digital Governance and Health in Andhra Pradesh <i>G S Naveen Kumar</i>	154



Preface

The proliferation of digital data has benefitted various stakeholders including the state, the private sector, and even an ordinary citizen who now enjoy unprecedented ease in accessing information and services. From customised suggestions in our social media to the best song recommendations, from accurate weather forecasts prompting us to carry umbrellas to editing, typing and creating content quicker than ever, data driven technologies have revolutionised our everyday life. While data-powered technologies rule the world, there are simultaneous processes to establish the rule of law for its use. However, for the vast majority of the citizens, it is still an uncharted territory. We do not know what happens to the vast amount of data that is being collected from us, how our identities are used to profile us in specific ways, how is it restricting our social media usage, and what happens to it if it reaches the people in power who might disagree with our digital data, including our opinions. Data processing technologies are by no means unbiased or neutral; in fact, people who are already historically marginalised often end up at the receiving end of unequal or biased datafication. In light of these concerns, this edited volume seeks to assemble a multitude of perspectives and insights from diverse stakeholders on data driven technologies and systems.

It is important to note that the chapters in this volume have been drawn from personal interviews and panel discussions, which were later converted into articles. The responses and opinions of individuals working at the intersections of




technology, welfare delivery, and bureaucracy, along with the communities and their representatives impacted by the data-powered technologies, have formed the foundations of this work.

We sincerely acknowledge the partnership of the Alan Turing Institute and the Global Partnership on Artificial Intelligence in our initial research on data justice. The guidelines and documents shared by their team have helped us deepening out understanding of data justice. Apart from this, during the research and writing process, we were fortunate to receive assistance from a dedicated team. We want to express our heartfelt appreciation to Syed Kazi, whose extensive research contributions significantly enriched the content of this volume. We would also like to thank Amna Majeed and Iqbal Ahmed for their support in writing the articles.

Additionally, we extend our heartfelt gratitude to our contributors for their patience and insights in finalising the draft. We hope the opinions and insights in this volume will contribute to more research and discussions on data justice in the Indian context and will lead to a more democratic, participatory and equitable design of Artificial Intelligence and Machine Learning powered systems in the future.

ABDM - Ayushman Bharat Digital Mission
ACF - Active Case Finding
A.P - Andhra Pradesh
AI - Artificial Intelligence
AIUDF - All India United Democratic Front
ANM - Auxiliary Nurse Midwife
API - Application Program Interface
ATT - Anti-tuberculosis treatment
BDR - Bangladesh Rifles
BJP - Bharatiya Janata Party
CAA - Citizenship Amendment Act
CCC - Covid Care Centre
CES - Centre for Equity Studies
CIBIL - Credit Information Bureau (India) Limited
COWIN - Co-WIN (COVID Vaccine Intelligence Network)
CSO - Civil Society Organisation
CSOs - Civil Society Organizations
DAVP - Directorate of Advertising and Visual Publicity
DNA - Deoxyribonucleic
DTBA - Delhi TB Association
EC - Election Commission
EHCP - Education, Health and Care Plan
EHR - Electronic Health Record
EU - European Union
GDPR - General Data Protection Regulation
GOI - Government Of India
HAUSLA - National Resource Team for the Homeless
ICBG - International Cooperative Biodiversity Groups
ICP - Integrated Check Post



ID - Identification
IoT - Internet of Things
IT - Information Technology
MHA - Ministry of Home Affairs
MGNREGA - Mahatma Gandhi National Rural Employment
Guarantee Act
NDA - National Democratic Alliance
NGO - Non-Governmental Organization
NPR - National Population Register
NSA - National Security Advisor
NRC - National Register of Citizens
OBC - Other Backward Class
OTP - One Time Password
PAN - Permanent Account Number
PDPA - Personal Data Protection Bill
PHC - Primary Healthcare Centre
PIP - Prevention of Infiltrators from Pakistan
RBD - Registrar of Births and Deaths
RBI - Reserve Bank of India
RTI - Right to Information Act
SC - Supreme Court
SSC - Supreme Security Council
STEM - Science, Technology, Engineering, and Medical
ST - Scheduled Tribe
STPF - Special Tiger Protection Force
TB - Tuberculosis
UID - Unique Identification
UIDAI - Unique Identification Authority of India
UN - United Nations
UNHCR - United Nations High Commissioner for Refugees
UPSC - Union Public Service Commission
US - United States
UK - United Kingdom

DATA JUSTICE IN INDIA AMIDST ALGORITHMIC EXCLUSIONS AND CHALLENGES OF FAIRER FRAMEWORK

By Osama Manzar, Jenny Sulfath and Ananthu RA

Keywords: Algorithmic Injustice, AI for Social Good, Neutrality of Technology, Social Justice, Data Transparency, Data Access

All hyperboles considered, the second decade of the twenty-first century certainly sees an undeniable pervasion of algorithms into many facets of lives. Where this previously had to do with the more savvy sections of society, now it has come to impact everyone, in different ways.

We are witnessing a ‘datafication’ of society, where massive amounts of data being collected and processed systemically have become normalised. This comes with its own set of new ethical issues and power dynamics that need to be analysed. Data-driven discrimination is advancing at a similar pace to data processing technologies.

But awareness and discourse around this discrimination is often limited in scope to how individual privacy is trampled on, or at max, to how data and the aforementioned lack of privacy and laws are leading to mass surveillance. Scholars have observed that there has been a growing disjuncture between tech justice and social justice activists where the former frames the solution of mass surveillance through a techno-legal framework, which puts the onus of protection on the ‘users’ with tools provided “expert” developers. This framework is criticised for ‘training’ the individual activists to use more secure communication channels than looking at the enabling social structures and treat it like a collective



project. Another limitation pointed out by the scholars is that this perspective may lead to characterising industry surveillance as politically benevolent as happened in the case of the confrontation between FBI and apple in early 2016 where Apple postured themselves as the protector of user interests over state surveillance. ¹While mass surveillance is pertinent issue, there are newer research that seeks to look beyond. Data justice is defined as “fairness in the way people are made visible, represented and treated as a result of their production of digital data.” Data Justice as a broad research area that brings together social justice, technology, and the impacts of the data paradigm on various communities. Data justice aims to go beyond the disconnect between tech-justice and social justice where tech-justice issues is mostly framed around the issues of mass surveillance and is limited to an expert community.

An important aspect of data justice is the adoption of decolonial theories in its analysis. Decolonialisation refers to the restoration of land and life following the end of colonialism through intellectual, political, economical and societal work. Mohammed et al (2020) uses the Algorithmic coloniality to refer to the interactions of algorithms across societies which recreate the systems of exploitations similar to colonialism; institutionalised algorithmic oppression, algorithmic exploitation and algorithmic dispossession. This framework particularly helps us to trace the historicity and geographies of digitalisation and to understand who is disposed, exploited and excluded from these systems. This framework also gives the scope looking at digital structures and algorithmic systems from the lens of prevailing social issues in India.

This edited volume seeks to delve into the multifaceted dimensions of data justice specifically in the context of

“Data-driven discrimination is advancing at a similar pace to data processing technologies.”

¹ Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*, 3(2), 205395171667967. doi:10.1177/2053951716679678



emerging Artificial Intelligence and Machine Learning (AI/ML) systems and their governance. By examining the experiences and perspectives of scholars, practitioners, activists, technologists and bureaucrats, we aim to shed light on the existing realities of data injustice faced by vulnerable populations in India.

Artificial Intelligence and Machine Learning (AI/ML) systems have already been used for governance and policies in the country. In fact, the government policy think tank (and the successor to the Planning Commission), NITI Aayog, has several strategy documents² on using AI in a beneficiary way. Several local governments have partnered with AI software developers to enforce welfare schemes³ and improve crop yields⁴.

And in most of these cases, present inequalities that exist have been further reinforced and exacerbated. For example, while the availability of data on previously invisibilised communities are taken advantage of by policy makers, there are several communities who are still excluded from the systems of measurement. For example, the circular migrants, a category of migrant who move from one place to another and back to the origin for shorter period of time is not accounted as migrants by two important population surveys in India, the national census and the National Sample Survey (NSSO) because the data collection methods are designed to capture population mobility and not labour mobility⁵. They are subsequently excluded from

2 NITI Aayog. (2021, February). Responsible AI. Retrieved from <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>

NITI Aayog. (2022, November). Ai for All 2022. Retrieved from https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf

3 Caravan. (2020, August 21). Cancelled ration cards deprived Telangana's poor of food rations amid lockdown. Caravan Magazine. Retrieved from <https://caravanmagazine.in/government/cancelled-ration-cards-deprived-telanganas-poor-of-food-rations-amid-lockdown>

4 Wadhvani AI. (2021, October). Pest Management in Cotton Farms. Retrieved from https://www.wadhwaniai.org/wp-content/uploads/2021/10/13_Pest-Management-in-Cotton-Farms-KDD.pdf

5 Srivastava, R. (2011). Internal migration in India. Human Development in India.



the automated systems designed for entitlement delivery because their access barriers are unaccounted for, or the lack of evidence to prove their vulnerability. Homeless population is another such category that is excluded from the data enumeration systems and falls through the gaps of policy decisions. Jatin Sharma's chapter titled "Identity, Homelessness and Machine Systems: Working with a Homeless Shelter in Delhi" sheds light on the intersection of identity, homelessness, and machine systems and the subsequent exclusions. Sharma's chapter presents a compelling exploration of his experiences working with a homeless shelter in Delhi, delving into the challenges individuals without stable housing face in accessing the basic entitlements required to survive in a city.

The chapter titled "The Decisive Machine And Algorithm Exclusions: An Overview" by Bittu K provides another compelling exploration of how algorithms can perpetuate exclusionary practices. It provides an example of the transgender persons excluded from the National Register of Citizenship (NRC) Assam due to the absence of specific criteria for transgender individuals. The NRC demanded many documented proofs of one's legacy data which was difficult for the transpersons who fled their homes at a young age to gather. Bittu K's analysis demonstrates how data sets are framed with cultural biases and a subjective, often heteronormative worldview. Abdul Kalam Azad, in his chapter titled "Digitizing Citizenship and Measuring Genealogies: The NRC and Algorithms in Assam," offers an in-depth understanding of Assam NRC itself and the far-reaching consequences of algorithmic decision-making in a community that is unlettered, geopolitically vulnerable and often bear the burden of proving that they are not foreigners in a location that has a complicated colonial past. Azad's analysis unveils the intricacies of the NRC exercise, particularly emphasising the significance of the family tree algorithm and its impact on the community in question. Notably, the chapter underscores the aspect of power within

“Data justice aims to go beyond the disconnect between tech-justice and social justice where tech-justice issues is mostly framed around the issues of mass surveillance and is limited to an expert community.”




“This edited volume seeks to delve into the multifaceted dimensions of data justice specifically in the context of emerging Artificial Intelligence and Machine Learning (AI/ML) systems and their governance

the discourse of data justice, continuing the question of how communities fall through the gap of ‘measuring’. It highlights the historical powerlessness experienced by a marginalised community and the mounting pressure placed upon them to meticulously preserve and present all relevant documents that align with the requirements of the new Machine Learning system employed by the NRC. This scrutiny reveals how the state can effectively exploit the veneer of machine objectivity to impose decisions that are steeped in deeply ingrained biases. Nikhil Dey in his chapter titled “Who governs data governance” delves into the complex relationship between the state and data governance, drawing from his experience with the right to information movement in India. He raises concerns about the current digital data protection framework in the country, arguing that it may do more harm than good due to a limited understanding of individual privacy. Dey warns of a troubling scenario where the state can label important public information as “private,” withholding it despite the arduous efforts of people to make it accessible.

Similarly, Shankar Singh, representing the trade union of MNREGA (Mahatma Gandhi National Rural Employment Guarantee Act) workers, contributes a chapter that highlights the gaps between the digital vision and ground reality. Specifically, he focuses on the challenges faced by workers participating in a poverty eradication employment guarantee scheme. These workers often lose wages when their attendance is digitized in areas with unreliable internet connectivity.

The initial chapters of this edited volume aim to move beyond sensationalized and speculative science fiction scenarios about the impact of Machine Learning on humanity. Instead, they shed light on the existing realities of data injustice faced by the most vulnerable populations in a post-colonial country.




While this chapter serves as a critical reminder of the need to examine algorithmic systems within the broader social, political, and historical contexts in which they operate, the next session in the book discusses the broader political discourses on data, data ownership and data governance framework reflecting on what is an ethical AI. Srinivas Kodali's chapter in this section builds upon the discussions of exclusion and power dynamics in the context of data justice. Focusing on the humanisation of data and the datafication of humans, Kodali examines the control and access to datasets by the government and private entities. Highlighting the unequal treatment of data and its implications for transparency and privacy, the chapter emphasizes the need to extend individual rights to privacy towards community-led collection, access, and control of data. By shedding light on examples such as the Aadhaar project and the Fintech industry.

Abhishek Singh's chapter presents a distinct perspective on the matter of data and data governance, drawing from his association with the MyGov division of the Government of India. He distinguishes between personal data and open data, providing readers with a framework that explores the existing policy landscape concerning AI ethics.

In Urvashi Aneja's chapter, she emphasizes that although there are robust frameworks for the responsible use of AI, the discussion largely remains at the level of principles. She highlights the inadequacy of state mechanisms to keep pace with the evolving technologies and their associated challenges. Urvashi echoes Srinivas' call for community participation in governance and encourages the exploration of data trusts, community-based models, and cooperatives.


Deepak P sheds light on the adverse implications of AI development on the working class and the unfortunate reality of machine learning-driven decision-making in welfare programs. He highlights the eventual deskilling of



interactive workforces, such as call centers. Deepak also underscores that these systems are often created by an elite workforce that is detached from social realities, be it caste in India or race in the Western context.

Rakshita Swamy's chapter raises thought-provoking questions about the applicability of a general framework to govern diverse contexts of digital technologies and AI-powered systems. She challenges the fundamental assumption that the state should be the sole custodian of data, advocating for people's right to access information about developmental data. She proposes the concept of a "data statement," similar to a bank account statement, which would track how individuals' data is utilized by the government.

The third part of the book contains articles that emerged out of the conversations with people working at the intersections of technology, welfare service delivery and and bureaucracy in India who have used AI technology to address the developmental issues. Rahul Panicker's chapter sheds light on a critical aspect often overlooked in technological interventions: the significant role of social issues in hindering their progress. While advancements in technology offer solutions, Panicker highlights the complex social factors that can impede the effectiveness and adoption of these interventions. One example he shares is a project aimed at efficiently testing TB. In the TB testing project, the author worked on utilizing computer vision to detect TB-positive samples. However, they discovered that the main challenges did not stem from the computer vision aspect but rather from issues related to sample collection and slide preparation. Merely fixing AI components did not lead to significant improvements in the overall scheme of things. Panicker also underscores a significant difference in the way policies are framed in the global north and south. He argues that while in the global north, the regulations are in place to control the "bad uses of AI" in countries like India, the



focus has to be on the “good uses of AI” where the resource scarcity is an issue in addressing health and welfare issues. Jayesh Ranjan, while acknowledging the potential for biases and lack of accuracy in data however emphasises the need to view data and emerging technologies in a positive light and highlight their potential for social good as Panicker. Naveen Kumar’s chapter focussing on the digital health initiatives in Andhra Pradesh, also recognises the importance of data security and privacy in the digital health ecosystem. He emphasises the need for a Chief Security Officer of an independent authority to protect data and information.

It is important to acknowledge that this collection of essays are developed from interviews and discussions and has its inherent limitations in terms of representativeness and comprehensiveness. The view points shared in this volume is also emerging from a diverse perspective of impacted communities, civil society representatives, policy makers and developers often with varying degrees of conflicting opinions. One of the objectives of this volume is to lay open these diverse opinions and perspectives and highlight the complex landscape of data in the time of rapid technological advancement. In a particular point of history, where crucial legislations are being developed on data, it is important to have a public discussion around various aspects of it. By presenting these diverse perspective, we aim to foster a deeper understanding of the complexities and trade offs involved in shaping the data justice framework from a post colonial country. The insights shared in this volume discusses the power dynamics inherent in the deployment of AI/ML systems, the adverse inclusions, absolute exclusions, the lack of representativeness in the development of these systems, absence of grievance redressal mechanisms and the potential of these technologies in achieving a more equitable access to resources.

While the chapters in this volume do not provide definitive

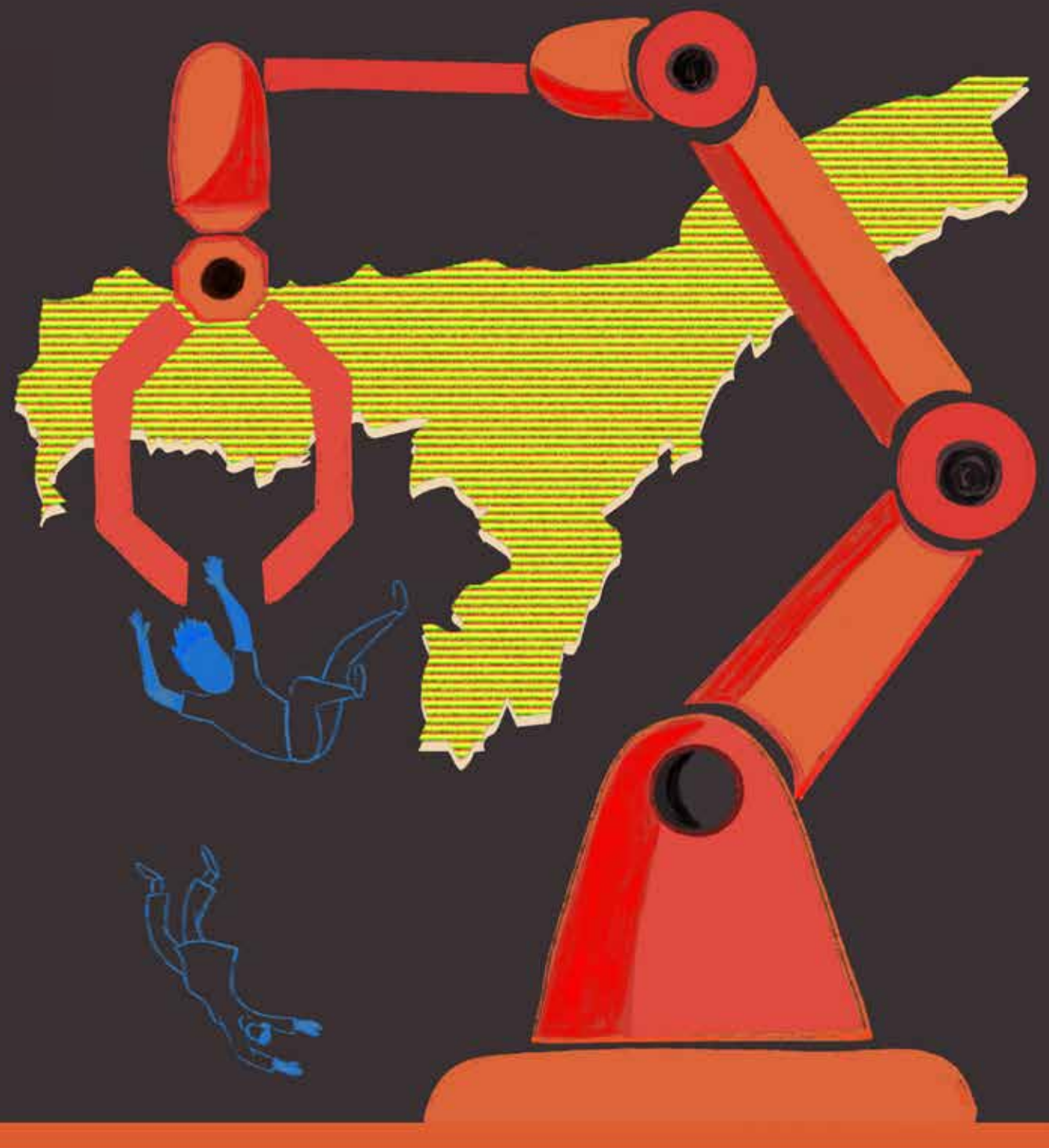


solutions or unified perspective, it underscores the relevance of constantly scrutinising the impact of advanced technology in different social, political and historical contexts. It challenges the assumption that, despite the claims of efficiency and objectivity, different power relations influence its implications from the moment a decision is made to use it. It also reminds us that while building a system of grievance redressal is important, it is also crucial to foresee the impact and decide when an advanced technology should be used. Moving forward, it is imperative to look at data governance from a critical lens and ensure that the advancement of data powered systems are in line with the larger agenda of social justice. We hope this volume will foster a meaningful dialogue in this direction.




Data Injustice

**Uncovering Oppression
and Exclusion in the
Digital Age**



**DIGITIZING CITIZENSHIP AND MEASURING
GENEALOGIES: THE NRC AND ALGORITHMS IN
ASSAM**

ABDUL KALAM AZAD



In this article, the author brings together his experiences and observations of the ground reality in Assam following the attempted implementation of the National Register of Citizens and our conversations with him regarding the use of large datasets and automation in this process. He examines the role of Aadhaar and explains how the algorithms that seem to be tracing lineages have been causing exclusions of people and effectively causing them to lose citizenship.

Keywords: Citizenship, NRC, Statelessness, Refugees, Aadhaar, Lineage, Family Documents

Abdul Kalam Azad

Abdul is a scientific researcher at Amsterdam's Vrije University, where he has also completed his PhD. He has previously worked with Center for Equity Studies, New Delhi and has done extensive research and writing on the NRC in Assam and the resulting statelessness.



Introduction

The discourse on the National Register of Citizens (NRC) in Assam must be oriented toward the questions of human rights and statelessness and should be a cause of concern at an international level. The NRC is a register of “legitimate citizens” and was implemented in Assam in 1951 and 2019 (Jayal, 2019). The Indian government also plans to implement it in the whole country (Roy, 2020). The motivation behind the NRC is to identify and deport “illegal” immigrants (Sufian, 2020). On August 31 2019, the NRC was released, where out of 31 million, 19 lacks were excluded (Karmakar, 2019). The NRC authorities have asserted that they will provide opportunities for those left out to contest the exclusion and prove their citizenship (Raturi, 2020). NRC is one of the processes in Assam’s citizen regime through which citizenship is contested. Two other parallel processes include the election commission’s doubtful voter’s system and Assam police border organisations reference cases. The contested cases of citizenship are then determined by the foreigners’ tribunal, and those who are rendered stateless or declared non-citizens are provisioned to be kept in detention camps (Azad et al., 2020). Assam’s citizenship regime is inhuman and produces statelessness on an industrial scale by deploying digital infrastructures and algorithms.

NRC, Assam and Longue Duree


It is important to understand how an exercise like the NRC is *implemented* and how within its myriad and intricate demands and procedures, automated systems and algorithmic mechanisms enable exclusion and



discrimination. In order to understand this, let us first understand the historical and political context of the NRC in Assam. The history of the NRC can be traced back to the mid-19th century when the British government brought in a large number of persons to resettle in Assam in order to work and labour on the fertile land under colonial projects such as “Grow More Food” (Guha, 2013). Bengali Muslim peasants constituted a large part of this resettled population. Today these descendants of Bengali Muslim peasants are known as Miya, often weaponised against the community (Azad, 2018; Dutta et al., 2021). The resettled population adopted the local language and culture and also appealed to the colonial administration to register them as Assamese-speaking Muslims in the 1941 census. However, the local population felt anxiety regarding large-scale immigration. Thus, the colonial government introduced policies such as The Line System in the 1920s and prohibited Muslims from settling down in certain areas. A ‘Development Scheme’, introduced in the 1940s, withheld land rights from Muslims who came to Assam after January 1 1938 (Guha, 2013).

“
Post the
Bangladesh
liberation war,
anti-Muslim
prejudices
rose as the
suspicion
of ‘illegal’
Bangladeshi
immigrants
became more
intense.

India became divided and gained independence from British colonial rule in 1947. During this period, Assam experienced various tussles for power between the Congress and the Muslim League. Anti-muslim sentiments and the fear of losing land, language and culture among the Assamese community led to several instances of state-sponsored violence and persecution of Muslims (Azad, 2018). Some examples are the ‘rioter bosor’ (in 1950, when thousands of Muslim families fled to East Pakistan from fear of persecution) and the eviction of thousands of Muslims by Gopinath Bordoloi in 1946. Between 1961-69, under a scheme called Prevention of Infiltrators from



Pakistan (PIP), more than 198000 Muslims were forcibly deported by the Assam border police to East Pakistan, and the unofficial figures are estimated to be much higher (Hussain, 1993).

Post the Bangladesh liberation war, anti-Muslim prejudices rose as the suspicion of “illegal” Bangladeshi immigrants became more intense, even though only 3% of Bangladeshi refugees settled in Assam. In 1983, allegedly, nearly 3000 Muslims were massacred on a single day, and this is infamously known as the Nellie Massacre (Gill, 2014). It is not only the Muslims of Bengal origin but also other marginalised communities, like linguistically minority Bengali Hindus, who have also been facing the wrath of the citizenship regime in Assam. The current ruling dispensation has brought in a new amendment to the Citizenship Act to protect the Hindus and made the Muslim prime target of the citizenship regime (Hashmi, 2022). The notions of a legal citizen of India and a legitimate resident of Assam are loaded with a history of colonial laws and motives, power struggles and deep anti-Muslim sentiments. These form the historical and conceptual frame for understanding the NRC exercise, its implementation and its consequences.

Proofs Legacy Documents and Family Tree

A principal feature of the NRC exercise is the digitising of existing documents. One important document in the NRC process is the legacy document. The legacy document consists of proof that a person’s ancestor was on a voter’s list from 1966 to 1971 or required proof of ancestors being enrolled in the 1951 NRC (Roy, 2016). However, producing this legacy document can be difficult for many reasons. One, either the NRC was not conducted in many places in 1951 or a person may have a receipt of participating in the 1951 NRC, but the government department did not have



the data. The government explained on several occasions that data may have been lost due to floods or destroyed. Therefore, in many cases, the data that the people possessed was not considered, and automatically, they were excluded. In such cases, the government said that if the excluded persons submitted their documents, the administration would provide legacy documents later; however, this did not happen for even a single person.

The NRC undertaking, thoroughly dependent on machines and algorithms, made peculiar demands and excluded many people for perplexing reasons. One such demand was the requirement to provide an exact and detailed account of a family tree. This produced several barriers and proved to be far from reasonable. Let us take the example of a family where all members use the same legacy document and also provide their family tree. The necessity of the legacy document from the 1951 NRC meant that multiple persons (descendants of one ancestor) would be using the same document to prove their legacy. In approximately 70 years (between the 1950s to 2010s), four generations of children, grandchildren, nieces and nephews would have been born and would be citing the same document. These elaborate requisites, which were verified by machines and software, were supplemented by the necessity that the names, spellings, ages and addresses provided by each family member should accurately match the ones provided by all other family members.

“Similarly, the NRC authority allowed anyone to file an objection against any ‘doubtful’ exclusion without providing the basis of such doubts.


This stipulation proved to be quite impractical considering that ordinarily, one may need to learn the details of a distant cousin or nephew with infallible exactness. In one case, 100 family members were excluded because of spelling errors. In other cases, families had to travel several hundred kilometres to their distant families in order to verify intricate details of identification. If the machine detected an error, then 50 persons from one family would travel to a hearing to explain that they did not know the exact



spelling of the name of their distant family members. In the initial stages of the NRC project, it was required that a person travel to the NRC Seva Kendra and provide the legacy document in return for a legacy slip or a legacy code. But the collection of legacy data and family tree details also overlooked that in many places of data collection, particularly river islands, the literacy rate is still in the single digit. Thus, it was difficult for many persons to provide exact spellings of their grandparents' names or their exact addresses from 1951.

These critical lacunae in the imagination and implementation of the NRC exercise were further exacerbated by the fact that the Seva Kendra operatives were ill-prepared to carry out a project of such large scope and consequence. Working for a salary of 4000-8000 rupees, many of these Seva Kendra employees did not even have the qualification of matriculation. They were responsible for providing legacy codes and slips; however, their work ran into profound errors and oversights, which also led to the exclusion of whole families. Let us understand this with an example. Much like spellings of distant family members, it is nearly impossible to discern the exact residence of one's ancestors from the 1950s or correctly account for all the different locations a family may have shifted to, especially in river islands. In this case, a person would provide the name of their grandfather, and the Seva Kendra operative would provide them with a legacy code in return. Another family, unknown to the first one, would also provide the same name and receive the same legacy code because it is absolutely possible and ordinary for two persons to have the same name. To the automated system, however, this appeared as strange or incorrect. Subsequently, these two families had to attend joint hearings and contend for their right to be the "original" owner of the legacy data in question. In actuality, both families would have had an ancestor with the same name. However, the machine learning software was unable to

“
On July 31, 2018, more than forty lakh people were excluded from the draft NRC. The excluded people were given a window to 'claim' their inclusion in the final NRC.




recognise this commonplace fact. Such trivial and bizarre oversights in the system led to serious results, such as the algorithm excluding entire families.

Exclusion of the already marginalised

From the experiences of those excluded, it is evident that the computerised and digitised process of data collection and verification was grossly inadequate and often led to grave results. These digital systems could not account for factors such as mass illiteracy or the living and shifting patterns of populations who live around river islands. The algorithmic patterns also could not comprehend and account for social barriers and stigma in the community, which often impacted data collection and the possession of documents. This was apparent when several stigmatised communities and marginalised groups were automatically excluded from the register. The transgender community experiences acute discrimination in Assam, and many leave their families at a young age due to the prejudice against transpersons. The NRC undertaking necessitated that they go back to their villages and families years, sometimes decades later, in order to obtain legacy documents.

The NRC exercise often made verification a harrowing process for those at the bottom of the social hierarchy, and these communities and individuals endured serious hardships in obtaining and verifying their documents. For instance, a family abandoned by their father found it very difficult to prove their legacy and provide documents. In another instance, a family paid bribes of nearly 4 Lakh rupees to officers because the family's ancestor was an illegitimate child and thus did not have a legacy document. In such cases, families get excluded from the register but also encounter intense social trauma and emotional damage. In one case, a man had two wives, and this fact got exposed during a hearing. Contestations over the ownership of



legacy data also publicly exposed private aspects of one's life and further caused distress and discomfort.

Other Domains of Exclusion Doubtful Voters and Border Police

Digital modes of collecting and storing bio-data to determine legality and citizenship have had a long history in Assam. The NRC is tethered closely to two other systems or mechanisms - the D or Doubtful voter system and the Border Police. D voters are categorised as suspected illegal immigrants and are deprived of their citizenship rights and privileges till they prove their citizenship. The exercise of discerning D-voters was initiated by the Election Commission (EC) in 1997 and was executed by doing door-to-door surveys to verify citizenship and nationality and marking 3 Lakh persons as doubtful voters or D voters ([Azad et al., 2020](#)). Both the border police and the D voter register are compiling their own bio-database and also modes through which individuals can be stripped of their citizenship.

The border police came into existence in 1962 during the PIP (Prevention of Infiltration of Pakistan) scheme of the Assam government and assisted in the deportation of nearly 2 Lakh Muslims to the then East Pakistan in the 1960s ([Hussain, 1993](#)). Today, the Assam border police is an independent organisation of over 4000 personnel and are mostly deployed in poor Muslim-majority riverlands and riverbank areas. In 2009, the border police were given the authority to question, collect the fingerprints of anyone and store them in a centralised database. Those suspected as “foreigners” by the border police were then referred to the Foreigner’s Tribunal under the Foreigner’s Tribunal Act of 1946. For my research, I have examined 7000 cases available on the website of the Assam police, and 90% of those were declared as “foreigners” ex parte. The concerned persons did not even know that there was a case against them and were declared ‘foreigners’.



The management of the database of the border police and the D voter register is arbitrary and improper. Their database runs into several hundred thousand cases, and they have marked thousands of persons as doubtful voters or as foreigners. During the last days of the NRC, the NRC officials decided to exclude all those branded as foreigners by the border police or as doubtful by the EC, even though those individuals had submitted all the required documents and also passed through all the hearings. The NRC authorities at all levels and the computer algorithm further excluded persons after this initial exclusion. If a person named “XYZ” was marked as a D voter or as a foreigner and was subsequently excluded from the NRC, all others with the name “XYZ” were also excluded. This was outlandish as there were multiple persons with the same name. One of the most prudent ways to explain the absurdity of this is through sayings that became prominent in our community. It became common to say that if you share a name with someone else, you have invited death.

NRC, Biometrics and Aadhaar

On July 31, 2018, more than forty lakh people were excluded from the draft NRC (Ranjan, 2021). The excluded people were given a window to ‘claim’ their inclusion in the final NRC (Saha, 2021). Similarly, the NRC authority allowed anyone to file an objection against any ‘doubtful’ exclusion without providing the basis of such doubts. Nearly three lakh objections were filed, which implicated an estimated 15 lakh people whose names were included in the final NRC (Saikia, 2019). In November 2018, the standard operating procedure (SOP) adopted by the NRC authority forced those excluded from the draft NRC and against whom objection was filed to mandatorily submit their biometrics during the hearings for disposal of ‘claims and objections’ (Singh, 2022). The biometric information was collected so that those excluded from the final NRC, i.e. found to be non-citizens, cannot vanish or go to other

“These digital systems could not account for factors such as mass illiteracy or the living and shifting patterns of populations who live around river islands.



states. The biometric was captured using the Aadhaar infrastructure. The false promise was given that once the NRC process was complete, these people get an Aadhaar card. In reality, more than 27 lakh people's biometric information is 'locked' by the Unique Identification Authority of India (UIDAI). These people are neither provided with an Aadhaar number nor are they allowed to apply afresh for an Aadhaar card. In the meanwhile, the government has made Aadhaar mandatory for availing social security schemes. The majority of these 27 lakh people's names have been included in the final NRC. The final NRC excluded 19 lakh people, but among them, many already have their Aadhaar card. This number is higher than the number of total excluded people from the final NRC (19 lakhs).

Conclusion

The current citizenship crisis in Assam has a long historical context. Historically marginalised communities like religious and linguistic minorities have been facing the threat of losing their citizenship in large numbers. The current citizenship regime has rendered a large number of people stateless, many of them put in detention camps, and most importantly, the fear and anxiety created by the regime have been affecting millions of people. The victims of the citizenship crisis in Assam are those who belong to the lowest strata of the society, who are mostly illiterate and living in a precarious environment facing the adverse impact of climate change. It is a shame that the benefits of scientific advancement in the field of artificial intelligence and machine learning are not being used to ease the suffering of these people. Rather, the current citizenship regime in Assam is using advanced technology and digital infrastructure like algorithms to inflict more pain and suffering on the poorest of poor people on a gigantic scale than the world has witnessed ever before. This essay is an attempt to amplify the voice for justice and accountability.

“The biometric information was collected so that those excluded from the final NRC, i.e., found to be non-citizens, cannot vanish or go to other states.

Reference

Azad, A. K. (2018a). Assam NRC: A history of violence and persecution. *The Wire*. <https://thewire.in/rights/Assam-Nrc-a-History>.

Azad, A. K. (2018b). Growing up Miya in Assam: How the NRC weaponised my identity against me. *The Caravan*.

Azad, A. K., Bhat, M., & Mander, H. (2020). Citizenship and the mass production of statelessness in Assam. *India Exclusion Report (2020)*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3718512

Dutta, U., Azad, A. K., & Hussain, S. M. (n.d.). Counter Storytelling as epistemic justice: Decolonial community-based praxis from the Global South. *American Journal of Community Psychology*. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajcp.12545>

Gill, P. (2014). *The peripheral centre: Voices from India's Northeast*. Zubaan.

Guha, A. (2013). *Planter-Raj to Swaraj freedom struggle and electoral politics in Assam 1826-1947*.

Hashmi, F. (2022). Citizenship, religion, and the politics of belonging: A case study of Shaheen Bagh. *South Asia Multidisciplinary Academic Journal*, (28). <https://doi.org/10.4000/samaj.7690>

Hussain, M. (1993). *The Assam movement: Class, ideology, and identity*. South Asia Books.

Jayal, N. G. (2019). Reconfiguring citizenship in contemporary India. *South Asia: Journal of South Asian Studies*, 42(1), 33–50.

Karmakar. (2019). Over 19 lakh were excluded from Assam's final NRC. *The Hindu*.



Ranjan, A. (2021). National register of citizen update: History and its impact. *Asian Ethnicity*, 22(3), 447–463.

Raturi, A. (2020). Analysis of the implementation of the NRC with respect to statelessness in Assam. *Christ University Law Journal*, 9(2), 77–95.

Roy, A. (2016). Ambivalence of citizenship in Assam. *Economic and Political Weekly*, 51(26/27), 45–51.

Roy, A. (2020). India: Intimations of an ending. In *The Best American Magazine Writing 2020* (pp. 271–300). Columbia University Press.

Saha, A. (2021). No land's people: The untold story of Assam's NRC crisis. HarperCollins India.

Saikia, A. (2019). In Assam, student group contests the inclusion of over 2.5 lakh people in draft NRC. *Scroll.in*.


Singh, B. (2022). Assam to approach Supreme Court for providing Aadhaar cards to NRC applicants. *Economic Times*. April 21, 2022. <https://economictimes.indiatimes.com/multimedia/politics/assam-to-approach-supreme-court-for-aadhaar-to-nrc-applicants/articleshow/90973922.cms?from=mdr>

Sufian, A. (2020). Geopolitics of the NRC-CAA in Assam: Impact on Bangladesh–India relations. *Asian Ethnicity*, 1–31. <https://doi.org/10.108>



**IDENTITY, HOMELESSNESS AND MACHINE
SYSTEMS: WORKING WITH A HOMELESS
SHELTER IN DELHI**

JATIN SHARMA



The article draws on author's work at HAUSLA's shelter near Yamuna Ghat, and focuses on the challenges faced by the homeless population of the city in accessing basic welfare services, healthcare, or even medication as part of the national scheme to eradicate tuberculosis disease.

Keywords: Homeless People, Tuberculosis, Healthcare, Welfare Services, Policy Implementation, Aadhaar, Government Schemes

Jatin Sharma

Jatin is a philosophy Post Graduate and is pursuing his PhD. He is the Director i/c of the Center For Equity Studies (CES), where he heads the National Resource Team for the Urban Homeless (HAUSLA).



Introduction

Homeless populations are one of the most precarious and vulnerable sections of the urban poor in India. The National Resource Team for the Homeless (HAUSLA) defines homeless persons as those who do not have a “home or settled place or abode” and instead live on pavements, public places, transit homes or their places of work. Though the given definition has heuristic utility, it is inadequate to capture the full spectrum of the kinds of precarious housing conditions faced by millions in the Indian metropolises. Homeless populations in Delhi live without access to basic amenities such as potable water, regular food, nutrition, and health care services and endure extreme climates of hot summers, freezing winters and monsoons. The question of homelessness is crucial to the struggle for digital justice as digitised infrastructures have further excluded homeless populations from the ambit of citizenship rights and rendered them without access to basic entitlements such as access to state welfare schemes, healthcare, and identification (ID proofs), which have become intrinsic to modern life.

As employees of the Hausla team at the Centre for Equity Studies (CES), we have been working with homeless populations in Delhi for around 12 years. We work with homeless shelters allotted by the Delhi government and have observed several issues that homeless persons encounter due to increasing digitisation and automated administrative infrastructures. In relation to datafication, the primary issue we have encountered is the relation between data and policy making. It must be scrutinised. What kind of data is collected to make policies on the homeless? What are these policies trying to address? How have these policies envisioned the ground realities of living a homeless life? This article will explain how automation and digital systems have excluded Delhi’s already marginalised homeless populations. These exclusions can be understood

through various examples and primarily through access to healthcare and the welfare schemes and services of the government.

Aadhar, Health, and Homelessness

A critical example of how automated systems have precluded homeless populations from welfare services and basic rights is the Aadhar ID issued by the Unique Identification Authority of India (UIDAI). Aadhar is a 12-digit identification number linked to biometric information that serves as proof of identity anywhere in India. The Aadhar has become very prominent since the 2010s and is now used widely as ID proof. However, this wide usage of the Aadhar, and especially the extensive linkage of Aadhar with other government IDs, has not included homeless populations in its purview. A significant issue with Aadhar is how it has become an indisputable proof of identification and has been made informally mandatory despite the 2018 Supreme Court judgment in this regard. While working with homeless populations, we have observed how it is nearly impossible to access any services or schemes of the government without the Aadhar. Simultaneously, to make an Aadhar card, other government IDs, such as proof of address etc., are required. However, homeless populations do not have address proof. Conversely, without Aadhar, obtaining other IDs, such as bank accounts or PAN cards, is difficult.

“Discontinuation of the [TB] treatment regimen makes the patient prone to developing drug-resistant tuberculosis. Since homeless populations are highly mobile, it is difficult to ensure treatment adherence.”

The pervasive reliance on automated systems creates severe problems in availing of health care services. For example, the Registrar of Births and Deaths (RBD) has been entirely automated, and this automation is now a centralised feature of the RBD record. For this reason, hospitals refuse to admit anyone who does not have an ID because if that person dies, their death cannot be registered in the system. A homeless person may get access through a homeless shelter volunteer through the latter’s ID, but if they wish to go to the hospital



alone, they will not be allowed treatment. Homeless people are excluded from healthcare and treatment even without automated rejection because of societal prejudices and biases. It is common for security guards and doctors to refuse a homeless individual entry into the hospital or a clinic, citing reasons for cleanliness and ‘hygiene’.

The impact of digitisation on the health conditions of the homeless population can be understood through the experience of treating Tuberculosis (TB) among the homeless population. TB is quite prevalent in the Indian population but largely remains latent. It takes the active form in conditions of lowered bodily immunity. TB is very high among the homeless because they live without adequate nutrition, have low immunity, and have problems with substance addiction. Moreover, homeless persons live in the open air and are continually exposed to polluted air, which further causes respiratory problems. We organise Active Case Finding (ACF) TB camps along with the Delhi TB Association (DTBA) and Medanta Hospital. We have found a significantly high rate of TB incidence (upwards of 8%) among the homeless populations compared with the national average and with data for other at-risk groups.

Anti-tuberculosis treatment (ATT) takes, on average, six to nine months to complete and, in many cases, has serious side effects owing to drug toxicity. Discontinuation of the treatment regimen makes the patient prone to developing drug-resistant tuberculosis. Since homeless populations are highly mobile, it is difficult to ensure treatment adherence. Furthermore, as patients feel symptomatic relief in the initial months of taking ATT, they will likely discontinue the treatment. If they have a history of substance abuse, there is a strong likelihood that they may resume it. Thus, merely testing and obtaining drugs is not the solution to the high occurrence of TB among the homeless.


Another crucial aspect of TB treatment is nutritious food,



without which the medicines are ineffective. The homeless mostly obtain food from *mandirs* or *gurdwaras*, which do not fulfil the requirements of adequate nutrition. Earlier, the Delhi government gave us a nominal amount (Rs. 18.30 per meal). Though insufficient and required supplementing with materials from other civil society and philanthropic sources, it allowed for flexibility in designing the menu per the patient's dietary needs. This has now been discontinued, and food is being provided, through an MoU with the Delhi government, by the Akshay Patra Foundation. However, TB patients who are our residents do not appreciate or enjoy the food provided by the Akshay Patra Foundation, nor does it fulfil their dietary requirements.

“There has been a long list of pending applications in Delhi for many years, and people have to wait for up to 6-7 years to get a ration card. More ominously, it is also possible that an application may be rejected after several years of waiting.”


The State's schemes and policies for treating TB among the homeless are unable to envision and take into account the experience of living a homeless life and the ground realities which impact the health of a homeless person. The automation of the State's health infrastructure further deprives the homeless of accessing proper treatment. The widespread requirement for the Aadhar card has made it very difficult for homeless persons to benefit from welfare schemes such as the Nikshay ID. Nikshay is a web-enabled system for TB patients and is developed by the Ministry of Health and Family Welfare and the Central TB Division. One of their schemes is to provide TB patients with Rs. 500 per month for fulfilling nutrition requirements during the treatment of TB. Homeless persons are most often unable to receive the Direct Benefit Transfer (DBT) of Nikshay because they do not have an Aadhar card or other forms of identification, such as a bank account. We have approached government officers to explain that homeless persons need ID proof to create an account in a nationalised bank where the DBT can be received. However, policymakers and officials exhibit very little enthusiasm toward ensuring that the benefits of their welfare scheme are reaching a wider population.



Digital Infrastructure and Exclusion

A huge lacuna in the imagination and implementation of digital administrative infrastructures is that the State does not consider that homeless populations cannot access these systems easily. Section 5 of the Aadhaar Act, 2016 inter-alia provides that the authority (UIDAI) shall take special measures to issue Aadhaar IDs to persons without a permanent dwelling house. Accordingly, UIDAI has provided the facility to enrol such persons without the required supporting documents through a pre-designated introducer, identified and notified by the UIDAI Registrar or the Regional offices. Thus, one can obtain an Aadhar card wherever one resides. However, a Time Password (OTP) is required to activate the ID, but most of the homeless population does not have phones for the lack of space for safekeeping and because phones require recurring expenses which they cannot afford. This gap in technological resources was most visible during the Covid-19 pandemic when populations had to be vaccinated en masse and at a fast pace. The Indian government created a web portal and mobile application called COWIN where citizens could register for covid vaccination. However, in the absence of mobile phones or any access to the internet, homeless persons could not use this elaborate digital infrastructure.

While technology is seen as a panacea, there are significant gaps in sensitivity in observation which overlook all those spaces where digitised infrastructures do not work smoothly. In the Covid-19 vaccination camps, we organised for homeless people, most came without phones or Aadhar cards. We registered several persons using our phone numbers since OTPs are required for registration. As a result of persistent advocacy by civil society groups, a separate portal was introduced on COWIN to facilitate the vaccination of homeless persons without IDs. Here registrations could be made with a single phone number of homeless shelter staff, but this portal also had multiple




glitches and errors and created many problems. One such problem was that many vaccinated persons could not obtain certificates even after they had been vaccinated. For almost 10 out of every 100 vaccinated individuals, the system would show the status as “unvaccinated”. We also experienced the opposite of this problem when people who had not taken even the first dose appeared in the system as having taken two doses.

The administration of the covid vaccine made it evident that without a phone and an Aadhar card, accessing something as essential as a vaccine during a pandemic was an arduous process. While the COWIN portal allowed corrections of glitches earlier, that facility was removed later. As a result, several persons with both jabs of the vaccine do not have certificates to prove they are vaccinated. There are also cases where a person had to retake the first dose to take the second one and get the final certificate. The automated system has not been designed keeping in mind that homeless persons cannot keep e-copies of their certificates and that as they change their place of residence often, they may have registered for different doses from different mobile numbers. Vaccine certificates are also equally common to get stolen or wet in the rain. The dependency on technology has also left no scope for recourse or redressal. For example, correcting any error or glitch in a person’s vaccination status also compulsorily requires digitally mediated OTPs.

Identity, Verification and Homelessness


The gaps in the digitised systems and automated frameworks reflect larger shortcomings of the State’s policymaking and data collection. As per the 2011 population census of India, there are 47,076 homeless persons in Delhi, though this number comes out to be anywhere between 5 to 10 times higher in estimates by various surveys. Moreover, there are only around



200 homeless shelters in Delhi, with a capacity for approximately 10,000 individuals. Among these, shelters for homeless families are negligible in number and are in terrible conditions. Most homeless persons do not want to live in the shelters due to reasons of safety and hygiene and also because shelters are often populated by drug addicts. A conspicuous drawback of government policies on the homeless is the deprivation of homeless persons from entitlements and livelihood opportunities.

The majority of homeless persons are caught between the endless and often contradictory procedures required to obtain entitlements, make their ID proofs and access the formal banking infrastructure. For example, there are several beneficiaries in homeless shelters who are entitled to welfare schemes and pensions for persons with disabilities, old age or widows. However, it is challenging to claim these without a UIDAI. Therefore, how the State approaches policymaking and implementation for the homeless is a more fundamental problem.


We have faced multiple difficulties in opening bank accounts and getting ID proofs for homeless shelter residents. The most demanding issue is obtaining an ID proof without other ID proofs and documents to verify the former. Another difficulty is that many homeless individuals are migrant labourers who move residence often and may lose, misplace their papers or have their documents stolen. Saving your earnings in a safe place becomes impossible without a bank account. At our request, there have been some initiatives by the government, such as installing lockers in some shelters where migrant, homeless persons can store their belongings. In the last three years, the Delhi government has been running camps in homeless shelters to make voter ID cards. Through a voter ID, other ID proofs such as Aadhar and PAN cards can be obtained, and bank accounts can be opened. The documents required for voter ID creation are a voter ID form, passport-size photograph, proof of wages (wherever applicable) and verification of



residence from the homeless shelter. In such initiatives, we often seek the help of MLAs in order to get Aadhar cards or connect them to welfare schemes. However, we may not always receive their support and thus have to wait for long periods of time. Even with the help of some public representatives, structural issues of accessibility persist.

The connection of different ID proofs through automated systems has made opening bank accounts even more difficult for homeless persons. Some individuals may have an Aadhar but not a PAN card, and vice versa, but both ID proofs have become mandatory. The systems also evolve and become increasingly opaque. Earlier, it was possible to get verification from 2 or 3 police officers at the District Magistrate's office and subsequently get an Aadhar card made for a homeless individual. Now, after the National Register for Citizens becomes a priority for the government, this recourse is also discontinued. As different identity ecosystems (such as Aadhar, PAN card, and phone numbers) become linked, one ID proof does not suffice, and more are needed to give veracity to an individual. For example, we have often encountered the demand for address proofs by banks (in the form of a written statement by a shelter) even after a homeless person's Aadhar has been created. In other instances, an Aadhar card is deemed insufficient, and other verification modes are demanded. While attempting to obtain a labour card for migrant workers, we observed that they demanded ration cards even though we provided the Aadhar card. With increasing digitisation and paperless verifications, many rules also change arbitrarily, and our requests are postponed and refused often.

The arbitrary nature of ID proof requirements makes it arduous to secure the most basic rights, such as food security through NFSA. There has been a long list of pending applications in Delhi for many years, and people have to wait for up to 6-7 years to get a ration card. More ominously, it is also possible that an application may be rejected after several years of waiting. Digitization has




played a principal role in these delays and rejections as the computer software rejects applications that do not have an Aadhar card, even though Aadhar is not mandatory to get a ration card. We have filled out several forms for ration cards and have been explained by officers that the software automatically rejects the application without ID proof.

Policy and Government “in” action

The digitisation and automation of welfare policy frameworks have increased the distance between the ones who implement the policy and its beneficiaries. In the case of homeless persons, obtaining or verifying any ID proof is challenging without another person vouching for them. We have worked with the homeless in various areas in Delhi, such as Nehru Place, Kalkaji, Jama Masjid, Nizamuddin and Yamuna Bazaar and have observed that each area has 3-4 volunteers who assist in the process of ID verification and receiving entitlements. Policies also have dubious requirements which change often. For example, the Delhi Arogya Kosh scheme used to accept any of four different IDs as proof but now has made a Delhi voter ID mandatory proof. This scheme allowed persons of the Below Poverty Line (BPL) and Economically Weaker Sections (EWS) to avail of free treatments, surgeries, test implants, etc. Making the voter ID with Delhi address mandatory to avail of the benefits of this scheme has electoral connotations that arouse cynicism. This represents the broader approach the State has toward welfare policies for the homeless population.

Instead of adopting a rights-based approach toward providing essential services and capacity building for the homeless population, the government largely harbours an attitude of charity towards the homeless. This attitude is reflected at the level of both policymaking and implementation. Firstly, the government body that oversees welfare for the homeless population, the Delhi Urban Shelter Improvement Board (DUSIB), does not have the



expertise to carry out such large-scale social work since it chiefly consists of engineers and administrative officials.

The lack of political will regarding the homeless is also reflected in how the government makes and implements its decisions on related matters. For instance, we run a recovery shelter, a specialised shelter that also includes shelters for women survivors of domestic and street violence & mental illness, de-addiction shelters, etc. In 2018, the Delhi government at the highest levels decided that such specialised shelters will be transferred from DUSIB to specific relevant departments, such as health departments and women & child development, with appropriate policy frameworks and budgetary allocations in place. However, the government has not been able to implement its own orders, despite reiterations by the vice chairman, DUSIB and minister of urban development and the chief secretary, among others, and it remains a matter of disagreement among various government departments.

Another significant factor in inadequate policymaking is the lack of data on homeless populations, as it is difficult to keep track of floating and constantly migrating populations accurately. Thus, the most reliable data for us comes from outreach activities such as TB active case-finding camps, vaccination camps, or through other community engagements. A deep and rigorous understanding of the harsh realities and varied struggles of the homeless lives and experiences is crucial before designing algorithms to provide services to the homeless population in India. In the absence of extensive familiarity, machine systems shall continue to replicate existing biases of the society and establishment.


References

National Resource Team for the Homeless (HAUSLA). (2011). Delhi Homeless Shelter Plan. New Delhi, India: Supreme Court Commissioner's Office.



**THE MOVING CITY AND STATIC INEQUALITIES:
PLATFORM ECONOMY AND TRANSPORT
DATA IN INDIA**

SHAIK SALAUDDIN



In the article, author draws on his experiences from working in the union and interactions with the companies trying to negotiate getting the rights to the data the platform workers have put their effort into creating. It also explains how the algorithms are actively biased against the gig-worker, favouring the customer or the company's profits and using legal loopholes to evade responsibility.

Keywords: Platform Economy, Gig Labour, Worker Exploitation, Data Transparency, Algorithms and Profit

Shaik Salauddin


Shaik Salauddin is a union leader, the Founder and State president of Telangana Gig and Platform Workers Union and the National General Secretary of Indian Federation of App-Based Transport Workers. He is at the forefront of gig-industry organising and has spearheaded several protests.



Introduction

The app-based economy is based on digital connections between service providers and customers through phone or computer-based applications. Most of these platforms are transport, food delivery and online shopping delivery services. The platform economy began in the late 2000s in India. Today, some of the most prominent businesses in this platform-based system are Ola, Uber, Swiggy, Zomato, Flipkart, Amazon and Urban Company. These companies are regarded as signs and symbols of India entering the digital age, as they have made digital platforms and technology the primary medium between customers and service providers. Another distinguishing factor of this economy is that workers act as informal contractors or freelancers and may work with different companies simultaneously. I will share my experience working on the ground with the platform workers who work for these companies and explain how data and digitisation impact their wages, the quality of their employment, rights to their data and digital justice.

I am the National General Secretary of the Indian Federation of App-Based Transport Workers (IFAT), which was founded in 2019. IFAT is a pan-Indian union with 35,000 people-membership approximately and has contributed and participated with workers from 14-16 States. I am also the Founder and State President of the Telangana Gig and Platform Workers Union (TGPWU), founded in 2020 and limited to the State of Telangana. I formed the TGPWU with a vision for the coming ten years. Additionally, I am part of another association called the Telangana Four-Wheeler Drivers Association, which constitutes drivers hired to pick and drop off Government officials, Gazetted officers, etc., or are outsourced on a contractual basis. Lastly, I am also the founder and chairperson of the Telangana State Taxi and Drivers Joint Action Committee (JAC).



I am not a white-collar leader, and my contribution to all these organisations is achieved through working on the ground. For instance, IFAT is a union of working-class members who work at the grassroots level. The fundamental objective of all these unions and associations is to fight against the excessive exploitation of platform workers in the gig-based economy. The community of app-based workers is bound from various sides and is wedged between a private company, the customer and the government. To understand how digitisation and data operate within this system, we must first understand the structure that the worker is a part of.

“We do not possess control of such an elaborate infrastructure, advanced technology or legal expertise. Most importantly, we do not possess time. For a platform worker, to devote time to unionise and fight for their rights means borrowing time out of day-to-day earning a livelihood

Partner or Worker?


Something essential in this structure is that those who work as platform workers are designated as “partners” and not as “workers”, and this concept of partnership has been a major factor in the demolishing of our labour rights. While these companies deem us as “partners”, we are not involved in any company decision-making process. As “partners” in a company, we should be provided as much information as any other Initial Public Offering (IPO) investor. Instead, a “partner” status allows companies to withhold even the most basic labour rights that workers are entitled to, including Social Security benefits such as income security, access to health care, insurance and provident fund. More importantly, if a company deems a worker as a “partner”, then that worker should be allowed access to various forms of information and data, which these platform-based companies continuously collect, particularly about the worker. Instead, the companies do not provide the driver with any data. Essential data such as how many drives he/she completed in a month or a year, how many were cancelled, and how much money was made in a month/year is not shared with the drivers. Instead, the dashboard in the mobile app does not store data or access to any information



beyond one week, creating a sizeable statistical gulf between the driver and the private company.

The question of sharing data with the driver or the “partner” is salient because, in any transactional relationship, a detailed history is always provided and accessed by both parties. For example, if I go to a bank or pay my electricity or water bill, the government will provide me with all my data from the past. Even a car showroom will have the entire history of a car, for example, how much it has travelled and so on. While companies like Ola, Uber, Swiggy, and Zomato collect large amounts of data about their stakeholders and those they deem as “partners”, they neither share that data with the workers nor do workers know what that data is used for. So there is no security for my data even though it is my right to have it. Several companies, even those we are unaware of, and researchers have large amounts of data and details about workers. It is only the drivers and delivery boys who do not have the data that is, in fact, about them. As part of my work, I have approached Swiggy, Zomato and Ola along with delivery boys and asked for the data they collect. The companies’ responses always redirect us to the management that is either on holiday or always occupied elsewhere.

There is a direct link between the concealment of data and our earnings. The companies credit the driver’s earnings after a week, at which point it isn’t easy to account for how many rides a driver drove in the past week and for what the final amount has been credited. Simultaneously, there are other concerns regarding our payments in which the algorithm plays a crucial role. The app often directs drivers to travel to locations of surge pricing, where demand for rides is more and availability of vehicles is less. However, it is expected that drivers will drive several extra kilometres to reach the suggested location and get no rides, implying that there was no demand.



Furthermore, the driver must pay for the extra distance travelled, and the company will refuse accountability. The application will often persuade drivers with surge pricing (of 2X or 3X), and to obtain it, we have to keep our phones continually on, but surge pricing is over in minutes. (acceptance rate) These frequent experiences highlight that people sitting in offices designing the algorithm have no conception of the ground reality.

In-contingent Algorithms

The algorithm is unable to account for any contingencies, and workers have to pay any extra costs incurred due to unforeseeable circumstances such as weather changes, monsoons and traffic jams. A common thing the algorithm cannot account for is the partial knowledge of google maps, on which the app is entirely dependent. Customers want us to come right to their doorstep; however, residential complexes often have roadblocks or other forms of physical obstructions which cars cannot cross. In this case, drivers travel extra to drive through a roundabout and have to pay for this extra cost themselves. Similarly, the app will increase the ride cost, but the customer will expect to pay only the initial quoted price and demand accountability from the driver. Often the customer files a complaint against the driver, and the driver has to pay the remaining amount from their pocket. In the case of Swiggy or Zomato, if the delivery is late due to rainfall or traffic, the delivery boy has to pay for the late delivery.

Moreover, the short path algorithms are supposed to mutually benefit the drivers and the customers. The platforms, to keep the driver online, show rides in far-off places. The drivers themselves bear the cost of covering this distance. While the drivers choose to bear this cost to get a ride, the customers usually cancel the ride if the waiting period is longer. The drivers are not compensated for this either. The driver ends up paying for the cost of petrol for



this distance. The apps use another technique to keep the workers online longer. The apps would display surge prices across the map, but when one spends the petrol cost to reach the surge points, it disappears. Once again, the cost of petrol is spent by the drivers themselves. The companies, on the other hand, conveniently use this system to monopolise the market. All the taxi drivers who were independently earning a livelihood now have no means left except to keep riding for these big companies.

The struggle is different for the food delivery workers in our union. The food and grocery delivery apps promise delivery in a short time. However, the restaurants hoard orders without any realistic assessment of the time in which they can produce the orders available for the delivery staff. The responsibility of delivering food within the agreed time thus falls exclusively on the delivery workers. I was told that the help desk consistently takes stands in support of the restaurants in this case.

Customers and Data Transparency

The quest for digital justice is urgent and significant because, in its absence, there is grave and deep inequality between the customer and the worker. This is evident in the fact that there is no data transparency offered to the driver. For instance, if a customer books a ride through the app, they will receive all the information about the driver, such as name, photograph, car number, phone number and so on. The driver is not provided with any of this, and there is no way to verify whether the individual who has booked the ride is riding in the cab. The algorithm designed thus assigns the possibility of criminal activity only to the driver, whereas it is equally possible that the customer may also be a criminal. This is a concern because in incidences where drugs may be transported through Ola or Uber or




other criminal activities may be carried out, the police will hold the driver accountable even though the driver does not know what transpired. For example, if a murderer travels through Ola or Uber after committing the crime and escapes from the crime scene, but the police arrest and question the driver for this, how is this justice? What is the driver's fault? Does the driver know that it is a criminal customer? In another instance, a customer duped a driver by asking him for a note of Rs. 500 to exchange for a smaller change but disappeared instead and blocked the driver's number. For these reasons, data transparency and KYC (Know Your Customer) linkage with IDs has been long pending demands from our end.

The unequal data transparency has other critical and grievous consequences for the drivers. Customers can file complaints about anything that troubles them. For instance, if a customer cannot identify the driver properly because the driver's ID photograph is old, the customer files a complaint against the driver. What follows a complaint is a tedious process. The driver's ID gets blocked, and they lose their daily livelihood. The companies then direct the driver to get the ID unblocked by obtaining a NOC from the local police station, where no one will give such a document. The IDs of many drivers were blocked and never unblocked again. The driver then belongs nowhere as no entity, either the State or the private company, is willing to be accountable or take any responsibility. This is evident because the workers are not provided with any authorisation or ID card stating that they work for these companies. The only document identifying the driver is displayed on the car, which the police demand for security purposes.

Data transparency is also compromised through complicated legal procedures, difficult language and paperless digital agreements. Every document is digital and is on the company's mobile application. When drivers sign a "partnership" with these app-based companies, they

“The companies conveniently use this [algorithmic] system to monopolise the market. All the taxi drivers who were independently earning a livelihood now have no means left except to keep riding for these big companies.”




encounter 40-50 pages of legal documents that describe the terms of the “partnership”. A gig worker or a platform worker is only able to secure work through the app if they agree to the terms set by the company and click the “I Agree” button. These documents can only be read with competent legal expertise, which workers cannot access or afford. Moreover, this document exists nowhere after one has pressed “I Agree” and cannot be found through a website, email, phone or in a physical form. Digital platforms and the app-based economy is reinforced through convoluted legal apparatuses and automation.

Government, Laws and Absence

To understand the quality of work and livelihood of gig and platform workers, it is very important to factor in the indifference and absence exhibited by the government. The government exhibits their lack of concern in two ways. Firstly, they do not regulate the terms and policies of these private app-based companies, nor do they investigate the earnings and profits of these companies. All businesses require licenses and regulations from the government and municipalities, but in the case of companies like Ola, Uber, and Zomato, there is no inquiry or monitoring of what they are doing with such massive quantities of data and such excessive amounts of money. Secondly, the government takes no initiative or interest in designing such apps and systems for the public. On one level, there is a widespread discussion on ‘Make in India’, but on the other hand, these private, foreign companies are running extensive businesses. In one sense, the government has left us no alternative, and we are an *aam aadmi*, an ordinary person.

I assert that the government should be vigilant of how platform-based private companies are earning their profit so that there can be accountability. Let me explain this with the example of customer credit payment. Say one car gets five rides daily, and the average cost of a ride is Rs. 200. So in



1 day, a car is earning Rs. 1000, and the combined earning from 10 cars is Rs. 1000. In a broader scheme, the credit earning from 10 vehicles in a week is Rs. 70,000, and after adding interest of 8%-9%, it is approximately Rs. 76,300. In the same week, the cost of the diesel/petrol, and of any car maintenance, is borne by the driver, and they are paid only at the end of the week. There is no transparency about who earned the profit from the interest and this sizeable credited money. Simultaneously, customers will get several offers and bonuses such as 1+1 ride free while there are no such incentives for drivers. The companies often take a small percentage from our earnings and cumulatively award it to us at the end of the week as an incentive.

Conclusion

An ordinary person cannot compete with the large, complex system created by private companies as we do not possess control of such an elaborate infrastructure, advanced technology or legal expertise. Most importantly, we do not possess time. For a platform worker, to devote time to unionise and fight for their rights means borrowing time out of day-to-day earning a livelihood. Yet, we try to give 100% effort, manage both and intend to keep on fighting. It is important for us to fight from a single front as in many states, there will be 5-6 unions that will dispute each other, and this further benefits the interest of the private companies. For this reason, the founding of the Telangana State Taxi and Drivers Joint Action Committee (JAC), which is a collective of 20 organisations and is India's biggest umbrella committee so far, is a significant event.

As I keep collectivising and fighting against the digital injustice of data-driven private corporations, I also realise the importance of data and statistics, which are considered crucial for policy-making on a global scale. In 2021, I published a report on the difficulties faced by app-based cab drivers during the pandemic, and this report gained global




attention. I am working on two more reports and hope to publish them soon, but that also proves difficult due to a lack of financial resources. However, a huge success on our end has been the inclusion of gig and platform workers in the Social Security Code 2020. We achieved this after a huge fight and are now waiting for an update after giving our suggestions to the government. In September 2021, we also filed a PIL (Public Interest Litigation) in the Supreme Court with Indira Jaisingh and Gayatri Singh. The PIL is aimed at ensuring that social security benefits are awarded to delivery and transport workers who work for or with app-based companies. It also demands accountability from companies regarding the livelihood of platform workers in the context of rising fuel prices and the economic crisis of the pandemic.



**THE DECISIVE MACHINE AND ALGORITHM
EXCLUSIONS: AN OVERVIEW**

BITTU K



The article provides an overview of the several biases the data sets that Algorithms tap into may possess, be it the historical exclusion of several communities or the historical over-representation of others, or the socio-cultural conceptions interacting with the datasets- and through this, looks into the exclusions in the NRC of the transgender community, exclusion of sex-workers from benefits, and skewed results in the field of organisational psychology.

Keywords: Gender Justice, Transgender Rights, NRC, Citizenship, Medicine, Education, Organisational Psychology, Personality Research

Bittu K

Bittu K is a professor of Neuroscience at Ashoka University, Delhi and an activist who had worked with several organisations surrounding trans-rights and organising, including the Telangana Hijra, Intersex and Transgender Samiti.



Introduction

The universal and wide-ranging digitisation of the world has led to notable transformations in the storage, distribution and accessibility of data. Simultaneously, automation and digital transformations such as Artificial Intelligence (AI) have profoundly amended how data is employed and utilised. AI and automated algorithmic systems reduce human effort, often eliminate it, and thereby appear as more efficient, time-saving and accurate modes of statistical and subjective analysis. Automated knowledge systems thus stand in for humans, categorise and organise information, make decisions and also develop modes of self-improvement. AI-powered systems are used everywhere, from governance and military intelligence to drilling mineral oils, banking, healthcare and suggesting videos to millions of users on social media platforms. The algorithm is thus a ubiquitous and determining presence in everyday life and impacts human behaviour, decisions and experience.

I have spent a significant portion of my life investigating neuro-functions, the concept of evolution and behavioural systems. I am a Professor of Biology and Psychology at Ashoka University and have completed my PhD in Neuroscience from Harvard University. With respect to AI-operated systems and algorithm-driven activities, I am interested in evaluating the common assumption that these expert knowledge systems produce objective and unerring information. The extensive viability of such expert systems relies on the argument that these automated knowledge systems are impartial, faultless and free from subjective human judgement. However, this pervasive reliance on digitised activity begs a deep and critical inquiry into the constitution of this seemingly unbiased data. Some of the questions we must be asking are; what is the nature of the data the algorithm employs? Who has collated and designed this data? Through what methods and for what purposes has



it been collected and classified? What are the classifications and categories that are integral to these algorithms? Are these categories indeed “objective”?

In order to do a critical evaluation of the categories that AI systems use, we must understand the paradigm in which these categories have emerged. The majority of these classifications and categories are developed by scientists and attain feasibility due to the assumption that scientific procedures and methods are objective and accurate. The Science, Technology, Engineering and Medical (STEM) disciplines operate on the premise that the knowledge they produce is a universal truth or an undisputed fact. Instead, critical social science research has established that all subjective or objective categories emerge in certain contexts with specific political and cultural histories. To put it simply, facts are a product of the social world in which they emerge and develop. The biases and constructs of the socio-political world intensely configure and structure the data and information produced therein. Thus, when we automate, the automated algorithm is as “good” as the bias involved in the data fed into it.

“...in the compiling of the National Register of Citizens in Assam, 2000 transgender persons were excluded from the NRC. As the NRC criteria did not offer any specific category for transgender persons, they were among the 40 lakhs excluded and now risk being labelled as “foreigners”

Biased Datasets, Machines and Exclusion


We must first recognise that algorithms are biased to work toward digital justice. The groups who oppose automation argue that human beings are not biased by the data they see. Instead, the data they encounter is a product of an unequal world governed by feudalism or capitalism. Therefore, data is shaped by a combination of our world and how this algorithm affects our brain and calculative capacities. Data is thus mutable and automated algorithms operate on the same principle. A sensible assessment of algorithms, human or non-human, should take into account the fact that it is humans who collect data and, thus, that data will be partial and discriminatory.



One of the severe consequences of biased data sets becoming integral to daily life is algorithmic exclusion. For example, in an educational institution, if an automated system receives an input of two mark sheets where the name and gender of two different individuals are the same, then the system declares them as fraudulent. It is common and normal for two individuals of the same gender to have the same name.

A more sinister instance of algorithmic exclusion was evident in the compiling of the National Register of Citizens in Assam, where 2000 transgender persons were excluded from the NRC. As the NRC criteria did not offer any specific category for transgender persons, they were among the 40 lakhs excluded and now risk being labelled as “foreigners”. In this regard, an application was filed by Judge Swati Baruah, who is the first transgender Judge in Assam and also the president of the All Assam Transgender Association (AATA). Furthermore, the algorithm is unable to account for the ground realities and lived experiences of the individuals it is evaluating. An exercise like the NRC demanded a large number of documents and proofs. However, this was very difficult for transpersons to provide as they had fled their homes at a young age due to abuse or stigma, and their documents were either missing or inconsistent. The fact that algorithmic categories exclude stigmatised communities further demonstrates that datasets are framed with pre-existing cultural biases and deeply subjective knowledge of the socio-political world.

How algorithms exclude and marginalise populations, cultures and ways of being is reflective of the social relations in which that algorithmic system was designed. The struggle for data justice, however, is not simply focused on including the numbers that were excluded because any stigmatised category will anyway show up in smaller



numbers. Our struggle for trans rights, for instance, is not just about numbers but about the policies that would be created and put in place even if one such person existed. Gaps in data and numbers represent larger structures of undervaluing. This is best exemplified in the case of sex work and the demand for its decriminalisation. In order to address the demands for the legalisation and/or the decriminalisation of sex work, data is required. However, sex work is considered nonexistent in a large number of places, and thus, there is no data to influence policymaking.

As long as policymakers demand numbers, they are salient, but sole reliance on statistics or numbers shall not address the combined impact of social injustice and large-scale digitisation. The more fundamental issue is how all forms of women's labour, of which sex work is one, are undervalued and underpaid in the formal economy. This would include women who work as childcare workers, Anganwadi workers, cooks or even as wives who perform sexual labour for their husbands. Women's labour is undercounted and devalued to drive down its price in the capitalist economy. Therefore, it is more urgent to re-exceptionalise sex work as work and place it on a continuum of labour that is extracted. These conceptual transformations are as, if not more, crucial than merely including categories in datasets.

A crucial question here is one of representation. In my long engagement with the struggle for trans rights and with sex workers' unions, I have observed that relatively independent Sex Workers Union (SWU) are able to better represent their demands. They also argue that through the process of NGO-isation, particular kinds of people and certain types of narratives are visibilised and promoted. There is more money to be gained if the narrative is more controlled, for example, money for HIV treatment.




Personality Research and Categorisations

“Data is shaped by a combination of our world and how this algorithm affects our brain and calculative capacities. Data is thus mutable and automated algorithms operate on the same principle. A sensible assessment of algorithms, human or non-human, should take into account the fact that it is humans who collect data and, thus, that data will be partial and discriminatory

Another example of how socio-cultural conceptions interact with datasets may be given from my own work on the science of personality research. Along with a scientist specialising in Machine Learning (ML), I am inquiring into the long classist and racist history of Personality Research. Personality research assumes that persons can be categorised into certain personality types, and this paradigm has been employed in various industries. The discipline of Organizational Psychology has been central to this widespread use of personality research. Human Resource (HR) departments often hire persons to profile employees in order to determine who they are and what kind of tasks they should be assigned. This pseudoscience was first used in the military, where applicants would be placed through personality research. More contemporarily, there is a proliferation of work where large datasets from classical psychology are fed to AI, and the results are compared to manual statistical analysis of the same data. Subsequently, it is argued that since both methods produce the same result, AI is an objective tool. This line of reasoning overlooks the fact that those personality types or categories are not objective but instead have been created by keeping a certain kind of personality in mind. Therefore, AI is fed datasets that are not objective.

In the context of ACAI (Algorithms, Computing and Artificial Intelligence), it is not as important to see whether it would be unsurprising if the data was not duplicated. The more urgent question is, how did these categories gain viability? What determines the viability of one category over another, and what is the point of cut-off which establishes that one category is more valuable than the other? Our research has shown that AI does not provide this cut-off. Furthermore, in the case of standardised data sets, there is no clear demarcation of what has an explanatory



value with five given variables and what does not. How are these variables being decided upon? In the same way, Randomised Controlled Trials (RCTs) are critiqued because if they are functioning on a flawed hypothesis or other validity issues, then the result is not so random. Therefore, automation is only as objective as the biases and the constructs that go into it.

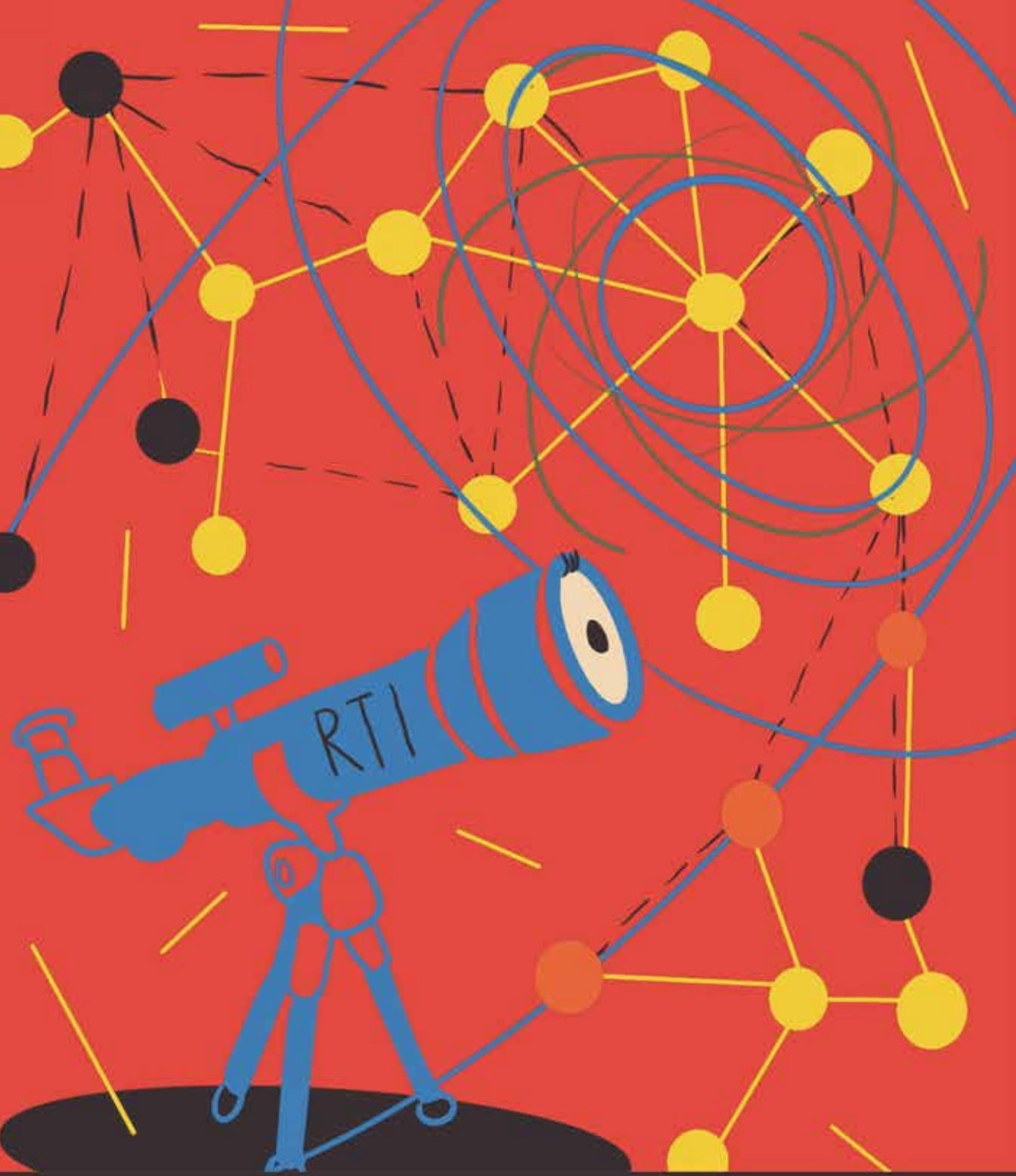
Conclusion: Possible Solutions

Algorithms and AI can also be used toward very helpful and productive ends, particularly in the field of medical research. As long as an AI is not fed with personalised information (and instead labelled as “Patient 1” or “Patient 2”) but is instead provided with centralised, unbiased records, then the promise of medical AI is very reassuring. Humans are extremely good at pattern analysis but are limited in the amount of data we can receive. For example, an algorithm can look at many cancer biopsies and learn to identify whether it is cancer or not. Moreover, the algorithm will be faster in assessment than a physician because the algorithm does not have to sleep or feed itself and can receive and process an input of millions of data sets much quicker. Conversely, an unbiased AI also enhances democratic access to information. It is becoming increasingly possible to verify medical diagnoses and other information from Wikipedia in the event that one has an incompetent doctor, and this practice keeps doctors in check.

A powerful tool that we use in our research to deconstruct biases is Machine Learning (ML). The process of deconstructing biases requires critical research wherein neuroscientists (who understand that algorithms are biased) work alongside those who use AI to duplicate human reasoning. There is a compelling need to be cognizant of the fact that we as humans see priors in how the world works and continue to believe that the world functions that way. It




is indeed possible to show that if we can feed AI with non-biased datasets, it can expose the assumptions of humans, who, with their own cognitive biases, see patterns that aren't really there. To re-construct these expert knowledge systems, we also need to urge scientists to be conscious of the categories and the algorithms that they engage with and deem to be unbiased. To develop this critical thinking and cognisance, it is important to have training in social sciences. It is equally important to diversify the workforce that deals with data because, at the moment, whether it is in social science or sciences, elite sections of society deal with data and are unaware of the disadvantages that large sections of the population face.



WHO GOVERNS “DATA GOVERNANCE”

NIKHIL DEY



The article looks at the evolution of demands from the grassroots for transparency that led to the Right to Information Act, and the renewed demands for transparency and accountability, bringing it together with the present state of data protection laws and public perceptions of privacy and how information that should be available to everyone is made inaccessible.

Keywords: Right To Information, Social Accountability, Transparency, Governance, Food Security

Nikhil Dey

Nikhil Dey is one of the founding members of the Mazdoor Kisan Shakti Sangathan (MKSS) and has been a full-time worker there since 1990. He has been involved in the struggles of the poor for justice and in the organisation's larger campaigns, most notably for the Right to Information and the Right to Work (MGNREGA).



Introduction

The use of digital technology for the collection of citizens' information and its aggregation into 'metadata' has multiple social, political and economic implications. The centralisation that this process facilitates leads to two crucial and fundamental concerns: a) who controls the data and how people can track and monitor how it is being used, and b) how to protect the data from being misused and exploited by the state or private profiteers. The issue of citizens' right to privacy arises mainly from these concerns. In other words, how can it be ensured that the government or its agencies only use the data to enable people to claim their rights and deliver public services? It has become obvious that digital technology has a multiplier effect that can have disastrous negative consequences for people, depending on who controls this meta-data.

The Apprehension and Misuse of Metadata

The basic premise of the Right to Information Act (RTI) was that the state is the custodian of people's information. To use Gandhian terms, the state was in a position of "trusteeship" regarding citizens' information, and therefore, the data so collected should be accessible for people to use within a democratic framework. Much as the term is used to deride people's suspicion of the magic of digital technology, many people asking questions are not Luddites but people who are deeply knowledgeable about the use (and misuse) of information technology. The state always claims that it only uses data for citizens' welfare. Many of these initiatives may be intended as pro-people measures regarding equity, participation, access, knowledge or issues related to non-discrimination. However, the question remains as to who controls the data. There are also many related questions about what kind of data/information



is aggregated. As the Aadhaar super project has shown, a technology dictatorship can access information from multiple data points, bring them together, and allow dictatorial control of political masters. Metadata in the hands of the state becomes a facilitator of mass control.

66
What happens within the so-called data set? Have those entitled to subsidised food grain been chosen well? Can they get Aadhar cards easily? Do their biometrics work? In these cases, those who manage the programme use digital technology to impose their “solutions”, but for those individuals with a grievance, they are no more than digital.

This kind of control over people’s data can also lead to tremendous exclusion and distress in development programmes. Let us take the case of the food security Act and its implementation in Rajasthan. Many people were unable to access their rations when the mandatory biometric authentication process was introduced. Data existed with the government, but they did not put it out in the public domain, which prevented people from figuring out how many people were on the abeyance list. Each aggrieved person had to find out what reports the government was generating for itself and where their exclusion was recorded. It was only when they found out which list they were on that they could begin to fight to get off it. There is, therefore, also a question of transparency and the kind of data the government is willing to put out proactively. This is just one aspect of the use of meta-data and the control of it.


Let’s take this a step further. The government has earmarked a certain percentage of the population under the National Food Security Act (NFSA). The central government’s target is to cover 75 per cent of the rural population and 55 per cent of the urban population. This varies from state to state. The Central Government is only concerned with the overall number. It, therefore, deals with the “metadata” and uses various tracking mechanisms such as Aadhar and biometric authentication to satisfy itself that the people are genuine. But what happens within the so-called data set? Have those entitled to subsidised food grain been chosen well? Can they get Aadhar cards easily? Do their biometrics work? In these cases, those who manage the programme use digital technology to impose their “solutions”, but for those individuals with a grievance, they are no more than



digit. It has been seen that those who control the data do not use it to ensure justice – particularly when it is politically or economically inconvenient. No matter how well the data is aggregated for the distribution of goods or services, we know by experience that in India, when lists are prepared for the distribution of entitlements or welfare, the powerful enter the data set, and edge out others more deserving. The lack of authenticity and quality of the data, once entered, can lead to a multiplication of failures and injustice. The data can be of far greater help if the process itself is made transparent by involving people at every stage as important stakeholders.

For the sake of administrative convenience, we cannot blindly go by the output produced by any kind of system – it needs constant monitoring and review by people. Interestingly, even where the data shows arbitrary exclusion, the state does not act on that information. In Rajasthan, for instance, people drawing social security benefits were to be automatically included in the NFSA beneficiary list. But social security beneficiary data shows over 4 lakh families drawing a disability pension not on the list. There are also almost 9 lakh old-age social security pensioners excluded from the NFSA list. These examples demonstrate how the use of data centralises decision-making and increases the arbitrary control of those who control the metadata.

Let's take another example from the NFSA. The Act requires the central government to provide 5 KG of food grain to roughly 70% of the population of the country. This list was prepared with the help of metadata and is now the metadata set. The central government has frozen the 2011 census data for each state to calculate its liability. It has consequently also frozen, for the purposes of this Act, the population of the country. The state insists that the "NFSA website will not allow a single additional name that exceeds the cap as per the 2011 population." If a prospective beneficiary of a scheme is not listed in the data of the



government, and the complaint wants to lodge a grievance to enlist his name, the site remains closed because the final cap has been reached. As a result, even those who are clearly entitled to food grain and who have been unjustly left out are unable to even individually “appeal” as provided under the law, as they are told, “the website is closed!”. The millions of children born in NFSA-selected households cannot be included because the number is frozen! From the perspective of the state, they brush off anomalies as being a small percentage of unavoidable errors considering the number of beneficiaries. However, for those who are left out for whatever reasons, this is a clear abrogation of their human rights. In the case of NFSA, closing off the entry of new data in the site means depriving deserving people of availing benefits of something as basic and fundamental as the right to food.

In the case of MGNREGA, its MIS (Management Information System), called “NREGA Soft”, is one of the most open websites of the government. It generates all kinds of reports for everyone to see and use and tracks the expenditure as well as the delivery of legal entitlements of the scheme. As it became clear that the law mandates compensation to the workers by automatically calculating the delay in the payment of wages, the central government, which controls the site and controls the payment of wages, decided not to show its own responsibility for delayed wage payments on the MIS. It is another case of metadata being misused by those who control it through acts of omission and commission.

We can take the control aspect of the arguments even further. The question of who controls the data is not only circumscribed to the state and its agencies. We can also bring in the role of private players and the possibility of collusion and misuse of data systems for their profits. The stealing or mining of data by the private sector for profit needs a separate discussion altogether. Even the use of big




data by the state for what seems like a legitimate purpose can result in the deprivation of the right of citizens, either willingly or unwillingly.

As one of our discussants, Jayesh, sketched out for us - the possibility of Kisan benefits under the state “Rayathu Bandhu” programme in Telangana, not being availed by the actual tillers or tenant farmers but by the “absentee landlord” who might be residing in the US. This is because land ownership is the criterion set by the government to get the Kisan benefit, and that is the data set they have. The all-India PM Kisan programme faces a similar flaw. The double whammy of landlessness and having to give a share of the produce to the owner is already a situation of exploitation for the sharecropper. However, for the sharecropper to not receive the benefit because their name is not on the data set defeats the very purpose of the scheme of giving the tiller of the land an input subsidy.

Often the argument offered by the technocrat is that avoiding “human eyeballing” takes away power from a bureaucrat who may take an arbitrary or subjective decision and therefore reduces corruption. The solution often comes from a progressive bureaucrat or policy maker, intending to take away any form of human intervention under the assumption that human decisions are subjective and biased. However, without context-specific human intervention, immense harm can be caused in decision-making. And finally, the data and how it is used are controlled by human beings, who often have a vested interest contrary to the people. We know how much data exists with the state and what kind of struggle people have put in to force the governments to publish or divulge this data in the public domain- be it mining in the protected area covered by the sixth schedule or allotment of land to the SC/ST who failed to get the possession of the land.

The irony of the whole exercise, on the part of the state,



“One should ask who is the client of the metadata. Is it the government on behalf of the people or the people themselves? Besides that, the experts in the government representing the people have to make a call based on open dialogues with citizens and citizen groups while keeping in mind the interests of the private players who would use the data eventually for their profits.”

is that- while it considers the “citizen-clients” as faceless subjects, a mere figure, it skirts around the issue of human discretion on the part of those who operationalise or control the system. However, the fact is that the governments put in the fair system when it suits their convenience and circumvents it when it is politically expedient to do so. It shows how much technology centralises power by those who are at the helm of an organisation taking these decisions and can supersede the constraints of law, policy and the constitution because digital technology gives that kind of power and platform.

This is not to assert that technology or big data is a zero-sum game. There are obviously many instances where data use has been extremely useful. It is the control over data that makes all the difference.

One important safeguard is open access to data. And open access with reports generated from the metadata that is useful to citizens can also be of immense use in increasing citizens’ participation in governance. The Jan Soochna portal in Rajasthan came from a people campaign and demanded that a “Janata information system” must be designed and used for all data held by the state. This JIS is premised on three important principles for digital data. That the data be online, that it be transaction based, and that it be open to all,

There is also the possibility of increasing efficiency and reducing corruption by using digital databases to avoid bureaucratic procedures that can be avoided.

For example, in Rajasthan, the widow pension entitlement was changed from 750 rupees to 1000 for widows over 60. Earlier, people were required to go through a gruelling process which was done away with because of the available database that enabled the enhancement to be achieved with one press of a button. However, when we tried to take the argument further, the same “omnipotent” metadata did not




automatically transfer the pension when a woman or a man attained the lawfully stipulated age-meriting pension, even when they were in an eligible beneficiary list. Since the citizens eligible for the scheme are more than the number they would like to spend, the decision-makers use discretion or convenience by not using the data because it is not expedient either in terms of financial liability or the political economy of making such a decision.

Ethics of Data Use: Citizens Vis-à-Vis State

The issue of control and discretion regarding the use of data gathered and held by the state can be classified using three broad questions: Firstly, what sort of data does the state gather and hold? Secondly, who gets access to and controls this data? And finally, how can both access and control be regulated through the law to restrain (mis)use of the data by vested interests, minimise discretion by those who hold power and are custodians of the data, and most importantly, make sure that data is collected, processed, and used as per the objectives of the law and constitution. In other words, the collection, storage, processing, and dissemination of data should meet a stated public purpose and be subject to public monitoring.

How does one ensure that the data an individual gives to a public authority is being used for the purpose for which it is intended? We have seen that once data is collected, it is deployed by governments in manifold ways. One should be wary of how decision-making vis-à-vis data unfolds in various stages: from gathering to aggregation to its final use. We need to learn from the principles of the RTI that we cannot just insist on transparency of the data or the process of gathering data but also on transparency of the decision-making itself. Given the power of meta misuse of digitised data, it cannot be left to the discretion or unmonitored trust of the custodian person or institution.



Therefore decision-making concerning the data and its use needs to be democratised as well. We have had (even before the advent of the digital age) existing laws and a Constitution that guides the ethics of the use of information/data. One should understand ‘data’ is not detached from information and knowledge. It is part of a subset - fundamentally affected by the larger universe of how information and knowledge is being shaped.

Metadata and AI as a Subset of the RTI Act

The term “Management Information System” (MIS) is so strongly established that it has insidiously established management controls over data even in a democratic framework. The assertion of building a Janta information system (JIS) comes from a need of having to re-establish the principles of the right to information act to the processes of data collection and use. This is because the bureaucracy and policymakers were quick to establish data collection and use, as a tool of management, and therefore to be naturally used exclusively by policymakers without having to apply RTI provisions and principles. Section 4 of the RTI Act, which should have found a natural facilitator in the entire process of data gathering, was somehow excluded from these MIS systems, and admin logins were used with ease and power to prevent open access to citizens’ data and the reports. Of course, similar to the exclusive control of information in the pre-RTI era, information was informally shared with the more powerful in society to further their vested interests.

When a people’s campaign for unrestricted access to MIS systems and data gathered and held by the state began in Rajasthan, the campaign was careful to connect the demand to the language of section 4(2) of the RTI Act to make the information as transparent as possible. It is a kind of relationship whereby the people’s RTI is protected and made a reality.



Section 4(2) of the Act states that:


It shall be a constant endeavour of every public authority to take steps in accordance with the requirements of clause (b) of sub-section (1) to provide as much information *suomoto* to the public at regular intervals through various means of communications, including internet so that the public have minimum resort to the use of this Act to obtain information.

The Jan Soochna portal was born as a result of this campaign, and an extraordinary amount of information was made available to the people through the principles of section 4. The fear of undermining important principles of privacy and misuse of certain categories of data were protected by exemption provisions under section 8 of the RTI Act. It was clear that the well-worked-out provisions of the RTI Act were sufficient and adequate to protect the data (like all information) from being unnecessarily put in the public domain.

However, the new Digital Personal Data Protection Bill (DPDPB) 202 threatens to undermine the RTI Act itself. It suggests an amendment in Section 8(1)(j) of the RTI act, 2005, which defines what a “private” matter is, and consequently protects the right to privacy of an individual so defined, and creates a fine and importance balance with the public interest. Section 8(1) of the RTI Act currently provides exemptions to the obligation of disclosing certain information. The proposed change to the privacy exemption under section 8(1)J appears to significantly undermine the core principles of the RTI Act. Here’s a comparison of the original text and the suggested changes:

Original Section 8(1)J:

“Information which relates to personal information, the disclosure of which has no relationship to any public activity or interest, or which would cause an unwarranted



invasion of the privacy of the individual unless the Central Public Information Officer or the State Public Information Officer or the appellate authority, as the case may be, is satisfied that the larger public interest justifies the disclosure of such information; provided that the information, which cannot be denied to the Parliament or a State Legislature shall not be denied to any person.”

Section 8(1)J if DPDPB is passed:

~~“Information which relates to personal information the disclosure of which has no relationship to any public activity or interest, or which would cause an unwarranted invasion of the privacy of the individual unless the Central Public Information Officer or the State Public Information Officer or the appellate authority, as the case may be, is satisfied that the larger public interest justifies the disclosure of such information; provided that the information, which cannot be denied to the Parliament or a State Legislature shall not be denied to any person.”~~

From the parts struck out from the exemption clause, it is easy to understand how the DPDPB, 2022, will exempt **all** personal information from disclosure, which will have larger implications for all information that in any way has any connection with any person . It will also have an impact on civil society movements that use the RTI Act to collect data on the implementation of government programs to make the government accountable. For example, details on a mining project to assess the environmental harm or details of malpractices in NREGA would, after this amendment, be inaccessible for public scrutiny.


What will the advent of AI do to metadata and its impact on people’s legal (Right to) access to data and information? While a lot of thinking is underway about the impact of AI on the use of data, very little thought has gone into how AI will affect people’s RTI and their capacity to use it to hold



those who gather, store, and use data (including government and commercial entities) to account. This is because AI will be designed (as it already is being designed by conglomerates of powerful groups who are coming together to design AI tools as per their interpretation of public good (at best) or their own interest (at worst) .

The generic use of information has always been more practical and useful for the government and the people if they can together take decisions according to their contexts. We can use data in a manager's sense to deliver goods or services, and in that, we should use AI data, but the more pertinent question is again, who is the one who is going to frame the need, who is going to use it, and for whose benefit? We cannot short-circuit democracy because it is convenient for a set of technocrats to use metadata to decide what is best on behalf of the people. This approach is not without precedent; elites, in the past as well, thought only in terms of development as they saw it, and inevitably it was their conception of development that was imposed on others. In that paradigm, information and its control played a vital role in making "informed choices". Today when a set of AI tools are going to "inform" your choices, it is obvious that the paradigm has already been fixed by those who designed the questions. In sum, data is, at best, an aid and a tool. Despite the advent of "reflective AI", which can become an even more dangerous tautology, its so-called reflection will undermine the power and complexity of democracy, and its capacity to predict is powerful but not nuanced and is often open to dangerous use.

So what do ordinary citizens do about the use of and advent of AI tools – especially after the advent of Chat GPT and its later versions? It should theoretically be designed with people so that it can be an aid in democratic decision-making and in amplifying the "voice" of the most marginalised. For people, the battle over information in the public sphere becomes even more important. It is only



an elected representative democratic government that can be forced to represent the interests of ordinary citizens and regulate the control of private commercial or political interests. In a country like India, the government is the biggest custodian of people's data. The people must retain control over the data and determine its use. They must be able to force transparency in every part of the development of an algorithm.

Even if one were to use a business term, one should ask who is the client of the metadata. Is it the government on behalf of the people or the people themselves? Besides that, the experts in the government representing the people have to make a call based on open dialogues with citizens and citizen groups while keeping in mind the interests of the private players who would use the data eventually for their profits. In this sense, whether it is data or AI, it should be seen as a smaller subset of RTI. The idea is not to criticise the technology per se but to become aware of the possibility of being used in a way unforeseen by the decision-makers themselves. In the age of AI, algorithms, and so-called reflective AI processes, this becomes even more critical.

Conclusion

In the digital age, governments are creating metadata by harnessing data provided by citizens for specific use. In the absence of a well-defined law on data, privacy, and ownership of this data, there is a high likelihood of misuse. Moreover, data or technology like AI is an aid or instrument used by governments on behalf of people. If technocrats decide that they know what the people want or need, the process of democratic decision-making is going to be severely undermined. In a democratic framework, technological aids must only be used from the vantage point of people and not where the perspective of management or decision-makers might even inadvertently upturn the priorities of the people. For that, we need to bring in a




law to include these concerns of metadata, IA or such technologies within the purview of the RTI Act. This law will have to be fundamentally different from what is under consideration in the DPDPB Bill. At a point where we sit on the cusp of yet another information revolution, we need to return to the basic democratic principle of letting the decision-making be democratic, transparent, and deliberative so that the new information systems are indeed for the greater common good, and in scenarios where there is a threat of data capture or misuse, the government actually represents the people in the battles ahead.



**DISCONNECTED BY DESIGN: UNRAVELING THE
IMPACT OF DIGITIZED ATTENDANCE ON
NREGA WORKERS**

SHANKAR SINGH



The article looks at how the hasty implementation of technological ‘solutions’ with the intention of efficiency, through the example of digital attendance for the country’s employment and poverty eradication scheme, has contributed to the loss of wages and employment.

Keywords: NREGA, Employment, Digital Divide, Labour Rights, Digitisation, Connectivity, Wage Loss

Shankar Singh

Shankar Singh is one of the founding members of the Mazdoor Kisan Shakti Sangathan (MKSS) and has been a full-time worker there since 1990. He has been involved in the struggles of the poor for justice and in the organisation’s larger campaigns, most notably for the Right to Information and the Right to Work (MGNREGA), and the pending Accountability Bill at the Assembly of Rajasthan Government. He is also a founding member of Rajasthan Asanghathit Mazdoor Union of more than 20,000 NREGA workers of Rajasthan.



66
In Rajasthan alone, I have seen approximately 8-10 crore (80 million) rupees of NREGA payments getting rejected. This massive amount is the hard-earned money of people who have been denied their due because of issues like a mismatch in Aadhaar numbers or spelling errors in identification documents.

I am Shankar Singh, an office bearer with the Majdoor Karamchari Sangh (MKSS), and I recently had a conversation with a representative from the Digital Empowerment Foundation. In our dialogue, I detailed the trials and tribulations faced by NREGA workers due to the government's overzealous digitisation drive. I would like to share those concerns with you in this report.

A glaring issue lies in the government's attendance-tracking model under the National Rural Employment Guarantee Act (NREGA). Despite our relentless hard work, the lack of a reliable internet connection for uploading our photos often labels us 'absent'. This predicament of ours remains unaddressed. In some instances, the network doesn't connect for days, rendering our efforts unaccounted for, and our pockets empty.

The existing system, as it stands, is ripe for misuse. It merely counts the number of faces in the photograph, not verifying the identity of the workers. Hence, the same workers can be counted multiple times, leading to widespread corruption. This is, sadly, the ground reality that fails to reach the policymakers ensconced in their ivory towers.

In 100 days, 15-20 days usually go by without a network connection, making it impossible for us to mark our presence. This over-reliance on technology, instead of simplifying, has further complicated our lives. It seems to me, at times, that this technology is a tool of corruption, more than an instrument of transparency and efficiency.

We see other government employees, like school teachers, using a biometric identification system. Why can't we, the NREGA labourers, have the same? It's as though we're treated differently, though we serve the same nation.

Moreover, the NREGA payments, which are directly transferred to bank accounts, often get misdirected due to single-digit errors in the account number. This simple mistake leads us into a bureaucratic maze, trying to recover money that seems forever lost.



In Rajasthan alone, I have seen approximately 8-10 crore rupees of NREGA payments getting rejected. This massive amount is the hard-earned money of people who have been denied their due because of issues like a mismatch in Aadhaar numbers or spelling errors in identification documents.

It is distressing that the process of enrolling as an NREGA worker is fraught with challenges. At times, workers are denied proof of their application to evade the provision of unemployment benefits. This also creates a false impression of limited demand for NREGA, hampering the allocation of budgets and, by extension, the workers' livelihoods.

Similarly, the pension schemes are no better. The paperwork and procedural delays are disheartening. And, Aadhaar linkage has its pitfalls too, excluding many migrant workers from welfare benefits.

Finally, the proliferation of cyber scams, ATM frauds, and biometric misuse has skyrocketed with the rise in digital dependence. Cybercrimes like those depicted in the television series 'Jamtara' are not mere fiction but reflect our current reality.

To understand the depth of the issue, one needs to get to the roots of it. The state and central government, often at odds, use us, the labourers, as a political playing field. The NREGA enrolment process, already arduous, is further complicated if the party in power at the centre differs from the one in the state. The delaying tactics include releasing payments late or refusing enrolments altogether. This creates a perception that there isn't enough demand for NREGA, directly impacting the budgetary allocation for the program and, in turn, the livelihood of the workers.

Furthermore, the system, previously managed at the Panchayat level, allowed some flexibility. If the internet was inaccessible, our attendance could still be marked offline, to be uploaded later. But the central government, in its quest for greater control, took away this power from

66
In 100 days,
15-20 days
usually go
by without
a network
connection,
making it
impossible for
us to mark our
presence.



the Panchayats and vested it with the Collectors. Now, which Collector has time to cater to the multitude of us workers? And who among us has the courage to approach the Collector? A system that was already strained has been pushed to the brink of dysfunction.

Consider the example of welfare schemes such as pensions. As a labourer, the government mandates us to fill innumerable forms, pushing us into a paper chase. In Rajasthan, the pension net was expanded to automatically cover anyone whose identity documents were with the government upon reaching pension age. But again, linking Aadhaar comes with its share of troubles - excluding migrant workers from welfare benefits.

In my opinion, the indiscriminate reliance on digital mediums has not curbed, but escalated fraud. Cyber scams, ATM frauds, and biometric misuse are more rampant than ever. I remember an instance where government officials kept our ATM cards, helping themselves to a few hundred rupees from each account. With biometric fingerprinting, uneducated labourers are often tricked. Officials take their thumbprints, claim the grain that the worker is entitled to, and then blame it on a 'thumbprint error'.

“With biometric fingerprinting, uneducated labourers are often tricked. Officials take their thumbprints, claim the grain that the worker is entitled to, and then blame it on a 'thumbprint error'.

Even in the case of electronic voting machines, we want each voter to be given a VVPAT receipt so that we can verify that our vote has gone to the party we voted for. There's an unsettling suspicion of tampering with EVM devices.

Excessive dependence on technology has not only failed to eliminate fraud but has also made us vulnerable to new forms of corruption. As a society, we need to question this trajectory and reassess our relationship with technology, especially when it involves the lives of countless labourers who form the backbone of our nation. It's not about being against digitisation; rather, it's about ensuring that digitisation doesn't leave behind those on the fringes, those without the privilege of constant connectivity. We deserve the dignity of a life that isn't hampered by the absence of a network connection or an erroneous digit in an account number.




Unpacking

AI Ethics



**HUMANISATION OF DATA AND THE DATAFICATION
OF HUMANS: GOVERNMENT, PRIVATE ENTITY, AND
THE STRUGGLE FOR DATA JUSTICE**

SRINIVAS KODALI



The chapter focuses on the instances in India surrounding the datafication of humans, like the smart city missions and the unique identification project, also examining the control and access to these datasets by the government and private entities. The chapter later argues for the humanising of data, emphasising the need to extend individual rights to privacy through community-led collection, access, and control of data.

Keywords: Humanisation of Data, Datafication of Humans, Transparency, Privacy, Community, Exclusion, Aadhaar

Srinivas Kodali

Srinivas Kodali is an independent interdisciplinary researcher working on data, internet, cybersecurity, surveillance and digital communities. He has been associated with free software and digital and internet rights movements in the country, and has written extensively on the same.



“The national exercise of data collection programme appears to be interested in claiming citizens’ bodies as the asset of the government, and it is the government which essentially decides how best these data can be used to empower people. Despite the claim of the benign intention of the government, we do not see it translate into practice.”

Introduction

Empowerment through data justice can be achieved by giving access to and control of datasets to the people. On the contrary, the government treats people’s data as government property. There is a prevalence of power asymmetry in terms of control and use of data. In addition to issues of transparency and access, there is a knowledge aspect of data use; for those who do not have at their disposal the knowledge, skill sets and infrastructure, the disclosed data is of little use. We need to extend the scope of the individual right to privacy and move in the direction of collection, access and control by the communities themselves.

Datafication of Citizens

India’s smart city’s mission was mainly about data collection.¹ The entire idea of the smart city’s mission was, in essence, a datafication at the city level.² The way urban slums are classified, they are not part of the mission. These urban slums are completely dependent upon urban funds for rehabilitation missions. In other words, if there are no funds to allocate, there exists no slum and vice versa.

Moreover, the way data is treated is dependent upon its definition. It often has social and economic implications depending upon which institution of state would be

¹The Smart City Mission is an initiative launched by the Government of India in 2015 with the aim of developing 100 cities across the country into “smart cities” that use technology and data to improve the quality of life for their residents. The mission involves a range of activities, including the development of smart transportation systems, the deployment of IoT sensors, and the creation of digital platforms for citizen engagement. See more at: Government of India. (2015). Smart Cities Mission. Retrieved from <https://smartcities.gov.in/content/>

²Datafication is the process of turning social, economic, and cultural phenomena into data that can be analyzed and used to inform decision-making. The term is used to describe the increasing use of data in a range of contexts, including business, government, and social media, and the ways in which this use of data is transforming the nature of these activities. See more at: Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

approached by the people for redressal; the response of a member of Parliament will be different from a high court judge. Let us understand, by example, how the role of AI crops up in the discussion on data.

In the city of Hyderabad, cameras with AI systems have replaced the police in controlling traffic violations. Similarly, the AI system was used to detect people without masks and was fined. Instead of distributing masks to the people, the AI system was used to exact fine. Therefore, in this case, the use of data in policing is not reformative but punitive or retributive; it is doubly unjust for those people who spend a lot of time in the street.

The government argues that we can use data to empower people, but in reality, the government has taken control of people's data. Ideally, the government should give citizens control or access to these databases to utilise for their good. The national exercise of data collection programme appears to be interested in claiming citizens' bodies as the asset of the government, and it is the government which essentially decides how best these data can be used to empower people. Despite the claim of the benign intention of the government, we do not see it translate into practice. It was the Aadhaar project that spurred the debate on surveillance, data justice and other related issues. The Aadhar project laid down a model which was unequal and shadowy to begin with. Many of the critical questions are still beyond public scrutiny. Citizens cannot inspect their data; they do not know how much data UID has kept, how data are organised and used, and so forth.

66
India's smart city's mission was mainly about data collection. The entire idea of the smart city's mission was, in essence, a datafication at the city level.


The issue of access is also important from the perspective of the enumeration of the identity of the citizens. We do not know how the government would classify its citizens. In the case of Aadhaar, it was expected the kind of profiling Aadhaar was supposed to create. The people who understood the nature of the data economy around Aadhaar



wanted to participate to ensure, at least, a vestige of equity; however, they were not allowed to participate.

A similar modus operandi was repeated in the case of AI. The pertinent question is how to break this chain, and where can we start? The lessons we have drawn from experience can be put to use. Foundation can play a crucial role in mitigating the misuse of AI, and we do not know what kind of data sets the government or private entity is involved in generating with the aid of AI-based algorithms. Take the example of the Fintech industry. Credit Information Bureau (India) Limited (CIBIL), as a financial institution, keeps data related to financial transactions. RBI collects credit scores through a bank in the CIBIL. Thus, we have an institutional mechanism in place in the form of CIBIL, and anyone can get their CIBIL score; in this, there is a semblance of accountability, access and equity. In addition to CIBIL, there also exists a dubious private economy which is fairly structured. In the evolving economy, there are alternatives to CIBIL scores built by private companies by collecting data from mobile phones without the people's consent. People cannot inspect or access these systems and do not know what kind of algorithms the Fintech companies are running over their data to decide the terms of interest.

In this opaque system of data justice and equity, access to a financial institution is entirely contingent on data a company has collected without people's knowledge. This lack of transparency, in a sense, is disempowering. This kind of credit score makes people vulnerable to a host of loan sharks or app sharks who could coax or cajole people into backing loans. Given this kind of cagey affair, if the government is to amp a new technology up, they would have a role to ensure a boundary line. When we discuss justice as a term, it is supposed that deserving people have a claim over the welfare scheme. However, they are being exploited through the misuse of the data. A fair economy is something we all can agree upon. Still, an entire ecosystem




would have to be restructured on certain principles, and for that to happen, a diverse set of people would have to be consulted. Unless and until that happens, data justice or, for that matter, any kind of AI justice is doomed to failure.

Identity, community, and Control over Data

Let us probe into the knowledge aspect of data. Sometimes, people get access to a lot of data when the government decides to make them public. Recently, the ministry of rural development of GOI has made public a lot of data relating to *the mandi*, hospitals, rural roads etc. It is good that the government gave access to this data. One can argue that this is a case of data transparency. However, does this qualify as data justice? The moot point is that, without the knowledge and internal data in possession of the government, the disclosed data is of any use. People do not have the computational infrastructure or skill set to study and analyse them. Thus, data justice is not feasible without the availability of proper infrastructure, necessary funding and human resource.

Having said that, there are a few instances of positive stories of the use of AI-based technology-the space technology built by ISRO, especially for farms. However, similar issues of accessibility have surfaced here as well. People who have access to this information have benefited from it. In this case, information asymmetry is determined by ownership of a smartphone. Similarly, the infrastructure of Geo data with climate prediction model had great success and has ramifications across communities and everyday life. It can also be used in tackling climate change. So, in this case, we see good use of data. Of late, a new development has occurred wherein private players have been allowed to move away from the public good to a focus on private profit. So, in the end, the participation of people will determine the outcome.




Individual vis-à-vis collective rights

We also need to expand the scope of the right to privacy of an individual to include vulnerable communities. The individual right to privacy must be framed in the language of collective rights about certain communities as envisaged in the constitution. In this way, privacy is not necessarily only an individual right, but a collective right too. How a community has been affected by slight data use remained unexplored in the Indian context. Individual cases have been studied; however, a large-scale study on how it might have affected, for example, Dalits, tribes or Muslims, has yet to be examined, as we saw during the Aadhaar controversy when the government balked at the right to data privacy and refused to recognise it as a juridical right. This point of contention brought to the fore the debate about whether or not the right to privacy is a constitutional right. In the context of AI, the government would respond similarly and would raise a similar issue. The data of which community gets collected as part of data sets and the training model would have a bearing on which community would benefit and which was left out. Thus, part of collective issues versus individual issues of how AI is affecting and how individual versus AI is affecting the community as a whole is very important, more so for such a diverse country as India.

The way the Maori tribe of New Zealand is building their data trust can be an example of data justice.³ The Maori tribe has a certain kind of constitutional protection. By extending the scope of the constitutional provision to

³ The Maori tribal data protection system refers to the traditional practices and protocols that are used by Maori tribes in New Zealand to protect the privacy and confidentiality of their cultural and ancestral information. These practices are rooted in the principles of mana (authority), tapu (sacredness), and kaitiakitanga (stewardship) and involve a range of strategies, including restricting access to information, controlling the use of technology, and developing protocols for sharing information with external parties. See more at: Tawhai, V., & McCarthy, K. (2018). Maori tribal data protection in Aotearoa New Zealand. *Journal of Indigenous Social Development*, 7(1), 1-12.



include the data ecosystem, they argued that all the data relating to its tribe would be collected by themselves, which implied that they would be the custodian of their data. They will refuse access to it unless an individual has control over it. This is not an isolated case; if one were to look at the history of the internet movements across the world since the 1990s, it would become clear that several communities have built their ecosystem, and this can happen in India as well. However, communities in India cannot plunge into this stage without economic and technological aid from a host of actors. We have not attained that phase. We are still bogged down as to how to respond to the government forcing the citizens to share data or prevent data misuse.

Data Justice and Knowledge

Even labour union often talk about data and algorithms if it falls within the ambit of labour rights. We discuss work hours and health benefits on a platform of workers' unions, but there is no study on the impact of app-based work, for example, drivers sitting in a cab for long hours, on the workers' health. We should not thrust the responsibility on the state; the responsibility lies on non-governmental actors as well. When the state proposes to collect health data of the citizens, it intends to collect data from the economic point of view to build a data economy for a set of organisations. It is not interested in looking into the impact of the working environs of Uber, Ola and Swiggy on the health condition to ensure they get health benefits. This kind of logic is absent in our discussion. While we discuss economic rights accruing from data and data justice, it is the same set of data that workers would ask us to analyse and help them. They do not prevent us from collecting their data, but we are not doing justice. This is important because when we discuss economic rights, the people who are an integral part of this are ignored, and the decision of a few people get amplified. For a long time, the data justice movement has been pitching for inclusivity, but it did not appear to fructify.

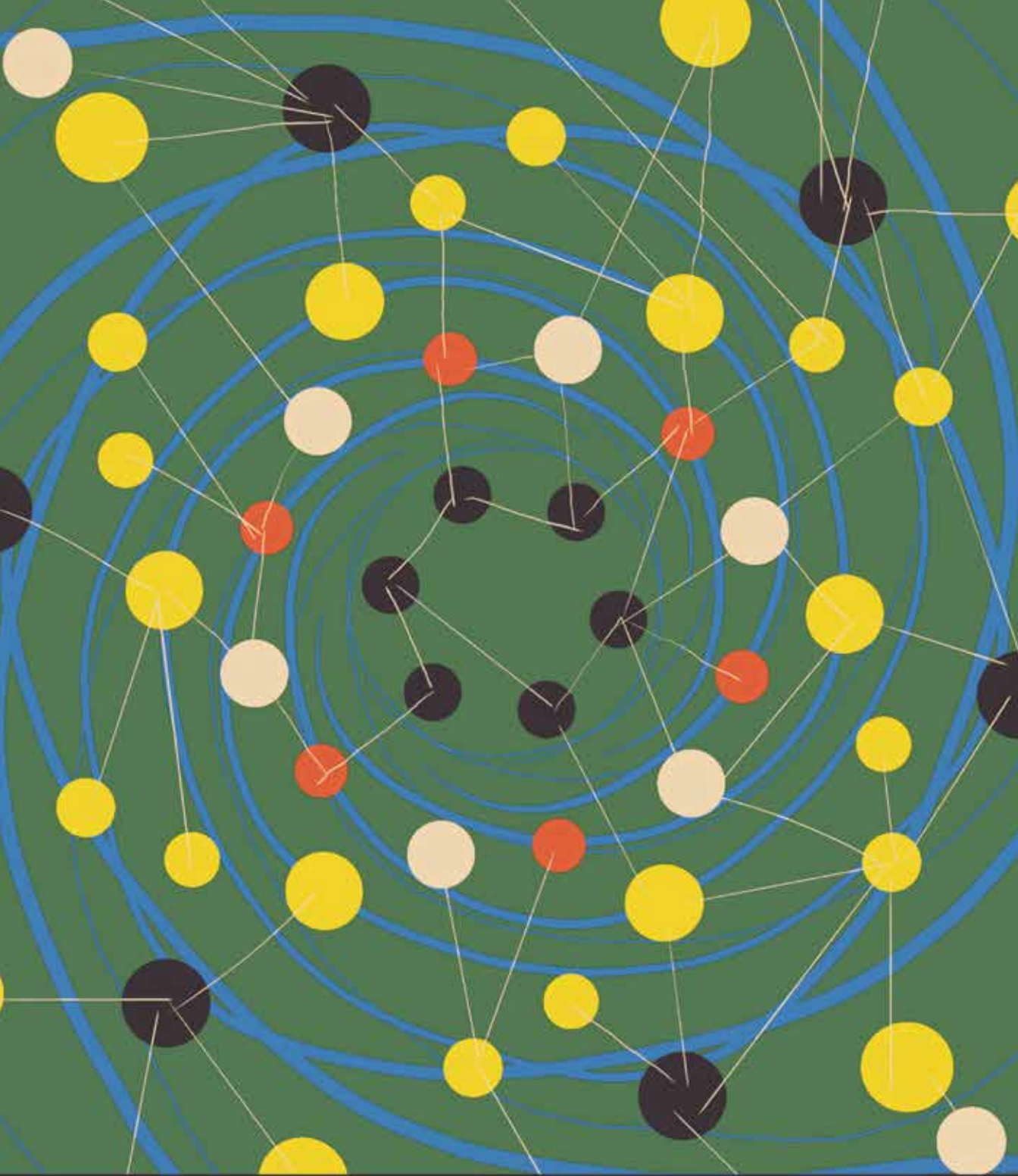


Let us hope that the AI revolution the Indian government speaks about will also take note of the workers. Some may argue that there are different intermediaries between the employees and employers and that the traditional contractual mechanism, which encompassed an employee-employer relationship, does not apply to Gig workers. However, we need to keep in mind that the state, as a custodian of law, is accountable for executing justice.

Furthermore, the issues of law and justice eventually percolate down to a political entity with the power of execution. The responsibility of upholding constitutional rights and executing justice lies on the state, and it is the responsibility of the state to look into the issues of the drivers' union. As of now, it appears that the state is abdicating its duties in favour of the big corporation.


Conclusion

The boundary of privacy rights needs to be pushed further and should be expressed in the language of collective rights as well, especially concerning certain vulnerable communities whose safeguards are laid down in the constitution. The purview of labour law needs to be examined in the face of new challenges wrought by the App-based industry and the effect of the working condition on their health. In the ultimate analysis, it is the responsibility of the government to bring in new laws and states to uphold constitutional rights and ensure justice.



**E-GOVERNANCE AND DIGITAL GOVERNANCE
IMAGINATION IN INDIA**

ABHISHEK SINGH



Abhishek Singh's chapter presents a distinct perspective on the matter of data and data governance, drawing from his association with the MyGov division of the Government of India.

Keywords: Digital Governance, Open Data, Data Policy, Transparency, Risk Management, UID, Welfare Services, Benefits, Data Monetisation

Abhishek Singh

Abhishek Singh is presently the MD & CEO of Digital India Corporation (DIC) and President & CEO of NeGD. He has formerly served as the CEO of MyGov, the Secretary of the Agricultural Scientist Recruitment Board, the Principal Secretary to the Chief Minister of Nagaland, and the Executive Director of the Food Corporation of India. He explains the principles, frameworks and infrastructure in place to enable safe and efficient use of the data that is collected through various means and the possibilities of data monetisation.



Introduction

As the President and CEO (- Chief Executive Officer) of the National e-governance division, I have seen e-Governance in India evolve and come a long way. Initially, it was the initiative of the Ministry of Electronics and Information Technology to ensure that every ministry makes e-governance a part of its long-term plan. Today, we have progressed and arrived at a point where both ministries and citizens attempt to avail more services online. At the level of government interventions and budget allocations, “digital” has become intrinsic to all aspects such as healthcare, insurance banking, employment, education or digital university. Thus, every aspect of human life has become tied to digital transformations.

At the core of this digital transformation is the question of data. A lot of data is generated through any digital application or e-governance platform. For example, if any transaction happens through UPI or on the Aadhar website, billions of data are generated, and this data has a lot of value. Data’s value increases when processed, and products are made from it. In that sense, data is very similar to crude oil. Within this dimension, where large amounts of data are collected, stored and processed for the larger public good, it is also important to track privacy and security issues. This can be done by stripping the day of personal attributes and ensuring that only aggregate data is used for policymaking, services and designing products to benefit humanity.

Personal and Non-Personal Data

The government has several approaches toward both personal and non-personal data. The Personal Data Protection Bill, 2019 (PDP) is currently in Parliament and is undergoing several consultations and deliberations so that any data the government collects is not misused, used




and used for any unintended or harmful use.¹ Hopefully, this Bill shall become a law and guide the personal data policy framework of the government. Once passed, the PDP shall be like the European Union’s GDPR (General Data Protection Regulation) that guides the regulations for managing personal data and delineating the provisions for what can or cannot be shared or collected. Non-personal data, such as mobility data, telcos and Google, help plan cities, highways and transportation networks. Thus, there is a lot of value in non-personal data. The government has set up a committee under Kris Gopalakrishnan to develop a framework for sharing non-personal data, creating data businesses, and using and monetising data.

66
At the level of government interventions and budget allocations, “digital” has become intrinsic to all aspects such as healthcare, insurance banking, employment, education or digital university. Thus, every aspect of human life has become tied to digital transformations.

There are many other initiatives and plans of the government, such as developing a new version of the National Data Sharing and Access Policy (NDSAP) and revamping the open data portal of the government of India (data.gov.in). Within these projects, the principles are clear, and the types are three-fold- Open by default, Restricted and Negative. Firstly, all data (minus the personal attribute data) about the citizens that the government holds in a fiduciary capacity should, by default, be open to all. It should be searchable, indexed and made available for anyone who wants to use it. The second type of data would be restricted, where it may not be made available to everyone as it is sensitive. However, one may source it with restricted access if one needs it for research purposes. The third type is non-open and negative data that would not be disclosed to anyone for many reasons, such as security. These would be the prime focus of every ministry.

The National e-governance division is also making attempts toward data anonymisation and data retention policies. Through a data anonymisation tool, any department, such


¹ The chapter is discussing the implications of The Personal Data Protection Bill, 2019 however, a new bill the Digital Personal Data Protection Bill 2022 has since been tabled, replacing the previous legislation.”



as education or agriculture, will publish data on open portals without revealing personal attributes. The data retention policy shall monitor the duration for which someone possesses data, whether it is for a day, an hour or months and through metadata standards, it will be ensured that everyone follows the rule. The government is also envisioning a National Data Council or India Data office that shall monitor and manage several aspects, such as the different datasets possessed by different departments, what highly valued datasets, and which of them are available through the Application Program Interface (API) on open data portals. Furthermore, the officers would check the data's geographical spread and time series. For example, are mandi prices available on the open data portal or are datasets from all states present? Therefore, the institutional framework, such as the national data council and data management units in every ministry, is as important as the legislative framework.

Artificial Intelligence and Seven Principles

Artificial Intelligence has become an essential aspect of digital transformation. Our data governance framework is focused on the responsible use of AI and is defined by seven principles. First, the principle of safety and reliability states that anything shared must result in safety and reliability in the service offered. Second, the principle of equality states that everyone should be treated equally. Third, the principle of inclusivity and non-discrimination states that data cannot be used selectively or to discriminate based on gender or religion. Non-discrimination is at the core of the AI framework. The other four principles are the principle of privacy and security, the principle of transparency, the principle of accountability, and the principle of production and reinforcement of positive human values. We expect that both government and private sector would adhere to these




values. For the larger public good, these seven principles are important.

There are also several risks, such as using data for unethical purposes or using social media platforms for their own ends. Therefore, proper regulatory and enforcement mechanisms are required for technology and innovation to prosper. It is essential to note how we use seemingly “free” social media platforms and inquire into what being free entails. When we use free platforms such as email, Google, Facebook, and Instagram, we must realise that someone is paying for them. The engineers who work for these companies will monetise and sell data about those who use these platforms, and advertisers use this data to sell their products. A Netflix film titled “The Social Dilemma” explains this phenomenon very well through a dialogue stating that if something is free, you are the product.

In this regard, the role of the government is to lay down the rules and guidelines. For instance, the Ministry of Information and Broadcasting released the “IT (Intermediary Guidelines and Digital Media Ethics Code)” in March 2021. These laid down various rules, such as how social media intermediaries function, where citizens can go if they have a grievance and if something incorrect has been posted about them. The guidelines state that if a platform has more than 50,000 users, it must appoint a nodal, grievance, and compliance officer. Furthermore, these officers must act upon a complaint within 48 hours, and if the complaint is regarding child pornography or sexual exploitation of women, they must respond within 24 hours.

However, even with these guidelines in place, there are often damaging events for many reasons. As the platforms can be used in various languages, there are many ways to circumvent the algorithm and ensure that the content remains on the platform. For this purpose, the Department of Telecommunications and Ministry of Home Affairs




“we are planning a system where one enters their basic age, gender, location, Aadhar number and other details. The system will inform them which schemes and benefits they are eligible for. Via Aadhar, authentication is also easier, and citizens do not have to provide ID proof repeatedly. Similarly, we have conducted a pilot study on the MNREGA scheme.”

have also set up handles such as cyber dost (Cyber-buddy) under their 14C scheme. These portals assist users on Facebook, Instagram, YouTube, Chingari etc., to understand what giving their consent digitally means and how to stay safe online. Part of the goal is to make sure people make informed choices and give their consent by understanding what is being asked of them in complicated legal language. Our broader goals are to develop advocacy, awareness, partnerships with non-profits and enhance school curriculum. We hope to ensure a safer state where regulations are adhered to.

Unique Id's and Interconnected Services

A crucial aspect of our agenda is also to interconnect services and create frameworks where we can predict what a citizen needs through data. For example, via a birth certificate, we know that a child of a certain age is going to school and that they belong to a particular caste group and deserve a scholarship. We endeavour to create a system that ensures the scholarship reaches the child without them having to apply. We wish to leverage data so that the government approaches citizens rather than citizens approaching the government to avail services. We evaluated thirteen key departments and discovered they have 365 citizen welfare schemes. However, citizens are not aware of these schemes.

For this reason, we are planning a system where one enters their basic age, gender, location, Aadhar number and other details. The system will inform them which schemes and benefits they are eligible for. Via Aadhar, authentication is also easier, and citizens do not have to provide ID proof repeatedly. Similarly, we have conducted a pilot study on the MNREGA scheme. We are trying to see how MNREGA beneficiaries can also avail of health insurance and Pradhan Mantri Ayushman Bharat scheme benefits.



Connecting different schemes and departments of the government digitally is one positive use of data we are planning. In this regard, we hope that a common identifier can be used to provide the benefits of multiple schemes. For example, someone who completes 100 days of work under the MNREGA scheme should also get the Ayushman Bharat card. We would not need another survey to assess the BPL (Below Poverty Line) population. In another instance, we hope to implement the ‘one nation, one ration card’ digitally and use a database to ensure that people get the benefits they are entitled to. Data can make it possible. For example, if a migrant labourer migrates from Orissa to Punjab, both he and his family should be able to get the ration in a split format. Integrating service delivery would allow citizens to avail their rights better and seamlessly.

One way in which a common identifier has been created is the Aadhar card. It functions as a magic number and allows the tracking of different documents. It is unique, de-duplicable, and a common ID linked to different federated identities. Through Aadhar, one can link their PAN card, vaccination certificate, driving license, etc. It is also a safe ID as sharing this number does not divulge any information about me, such as age, gender or area of residence.

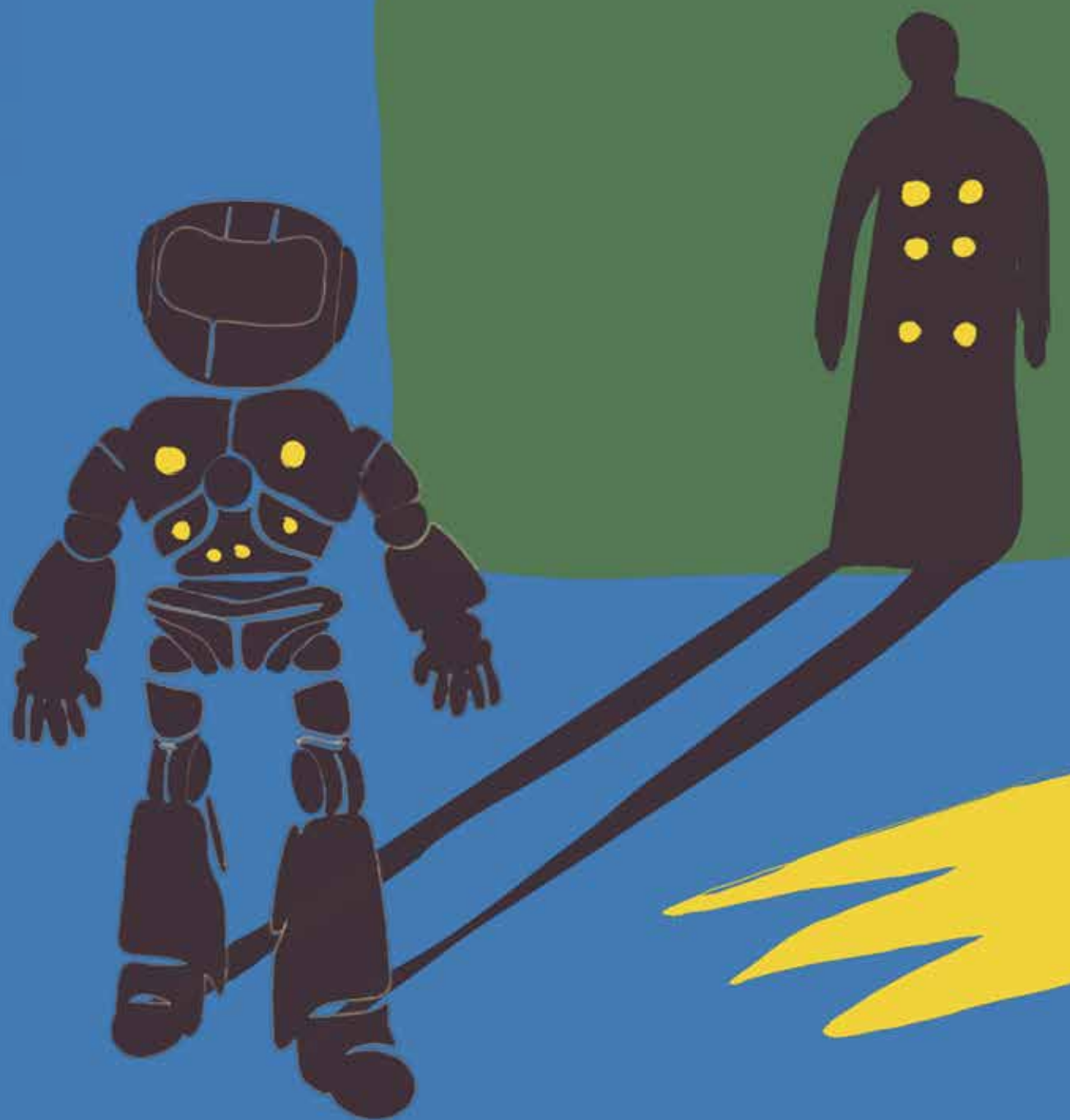
Conclusion: Managing Risks

In planning a robust and seamless system, there is also the possibility of risks and malfunctions that exist in any system. To make an analogy, accidents and mishappenings are always possible when governments build highways and roads. However, those are tackled with safer road constructions, making safer cars, spreading awareness among persons and introducing speed limits. Similarly, for a data-based IT system built for the larger public good, the aim should be to build a strong system where the benefits outperform the risks. To achieve this, a good grievance redressal system must be there and also have rigorous third-




party audits so that the system is transparent. While there may be persons or elements who would attempt to subvert the system, our efforts are to minimise their presence as much as possible.

The passing of the PDP bill would be a step toward ensuring data protection and allowing data to empower citizens. It shall also ensure the redressal of grievances and make space for a system where citizens give informed consent for data usage. Misuse of data and larger public good should be the bigger aim. This can be achieved through strong audits and imposing penalties on companies who do not follow guidelines and work towards creating a data market. Moreover, because data is sold, a good system should allow me to price my data to create an exchange ecosystem, and citizens can get better value-added services while enjoying their rights. We must view data at an aggregate level because entities (such as Google or Telcos) own data, and new databases are created as databases are bought and sold. A more legit framework for data use can be created through data monetisation.



**NON-NEUTRAL TECHNOLOGIES AND AI POLICIES:
EVALUATING BIASES AND ALGORITHMS**

URVASHI ANEJA



The article argues that even with the existing robust frameworks for the responsible use of AI, the discussion largely remains at the level of principles. The article also critically looks at the state mechanisms and their inability to keep pace with the evolving technologies and their associated challenges.

Keywords: Responsible AI, legal frameworks, Institutional Mechanisms

Urvashi Aneja

Urvashi is the Founding Director of the Digital Futures Lab, where her work examines the ethics and governance of AI in the global south; digital public infrastructure and platforms for public service delivery, and labour rights and well-being. She is also Associate Fellow at Chatham House and IIT Bombay and holds a PhD from the Department of Political Science and International Affairs, University of Oxford.



Introduction

The impacts of the complex relationship between technology and society have increasingly permeated vast areas of political, economic, and social life. I am the founder and director of Digital Futures Lab in Goa and a consultant to the Indian government on Artificial Intelligence (AI). Digital Futures Lab is part of a consortium of various think tanks and research organisations researching AI, policy and data governance in India. Through my engagement with these spaces and as a member of civil society and a user of digital networks, I have developed various insights about the myriad ways technology and society influence each other and our lives. At a broad level, this influence manifests in two ways— one, in how technology is produced, and two, in the type of technology produced. Technology is not a neutral system but is produced through the values of the society that creates it. The prevailing socio-cultural norms, interests and power relationships shape technology design and determine who benefits from it and who loses out. The impact that technology creates in the social world is not universal. It differs based on contexts, locations, social groups and individuals.

The vast digital divide in India has a definitive role to play, affecting both access to technology and the experience of using digital spaces. The socio-cultural norms of Indian society affect the usage and experience of technology, and digitisation is a varied phenomenon across social groups. For example, the prevailing norms of gender restrict women's freedom and participation in the digital space. However, simultaneously, technology production is legitimised in India as it is seen as enabling socio-economic development. The question remains: Who truly benefits from technological advancement when its development does not consider the realities of certain social groups?

Individuals, Communities and Structural Impact

Artificial Intelligence and Machine Learning systems are widely used in public sector services and governance. As a result, algorithm-backed decision-making processes are also proliferating. This can be understood at three levels: impact on individuals, communities, and structural impact. At the individual level, some concerns regarding AI usage pertain to laws of agency, laws of privacy and the undermining of individual data security. At the community level, AI systems are linked to perpetuating prevailing biases and discriminatory practices. For example, several communities are treated unfairly by AI-driven systems due to their social location. This is evident in facial recognition systems, credit scoring systems, welfare systems, hiring systems, and the workplace, which are primarily trained on data from majority social groups. At the structural level, several harms are caused via AI systems to labour markets, general market competition, and political institutions' accountability. Therefore, damages can be traced across social, cultural and political domains and are evident across levels of individual, community and structural impact.

“While using AI for diagnostics may benefit the health sector, the inaccuracy rate is nearly 10-20%. AI in judicial services reinforces the digital divide, as many people cannot access judicial services via technology. AI is often used to ascertain individuals' creditworthiness, which may lead to biased results as the algorithm detects worthiness based on pre-existing assumptions

To understand the relationship between digital systems and governance in India, it is important to focus on the processes of implementation of AI systems and frameworks. There is a wide gap between the use and deployment of AI and its governance in India. AI-powered systems are used by private players and for public services such as law enforcement of beneficiary identification. However, the legal systems, regulatory frameworks and the larger institutional capacity to deal with these transitions are inadequate.

Governance options are deeply connected to how a problem gets framed – the framing of the problem shapes how issues are prioritised and negotiated, rendering some policy options desirable while restricting others. In addition to technology



developing faster than policies can catch up, policymakers are reluctant to regulate too much or too soon in an attempt to maintain their competitiveness in the global AI market. In the Indian context, the broader narrative also favours AI-backed systems for socio-economic development, so weak state and institutional capacity pose a significant hindrance. India has joined a global partnership on AI and has proposed a framework for the responsible use of AI; however, several of its proposals are merely at the level of principles. For these frameworks to be implemented thoroughly, much thought and capacity building are still required.

AI and ML systems function as status quo systems and create futures based on what happened in the past. In that sense, they are not transformative because they are built on existing data, which they reproduce. So, when you create systems and introduce them into a society that is already profoundly polarised or has cleavages of gender, caste, religion and class, it is likely that those will be reproduced. Moreover, it is possible that over a period of time, these systems also reduce the space for different kinds of imaginations and forms of emancipatory potential. Thus, in evaluating any AI system, it is crucial to see the political culture surrounding it. If there exists a fair and equitable democratic structure, then it is possible to imagine more fair and equitable AI systems. However, when they are introduced into political cultures that are restrictive or that are inequitable, then they are likely to reproduce existing inequalities.

Across the spectrum, there are several examples of misrepresentation of data and the misuse of Artificial Intelligence. In the law enforcement sector, the use of facial recognition technology and productive policing is concerning. While using AI for diagnostics may benefit the health sector, the inaccuracy rate is nearly 10-20%. Therefore, it needs to be approached with caution. AI in judicial services reinforces the digital divide, as many people cannot access judicial services via technology. Lastly, AI is



often used to ascertain individuals' creditworthiness, which may lead to biased results as the algorithm detects worthiness based on pre-existing assumptions.

Solutions And Possibilities

One way the bias of AI systems may be tackled at the policy level would be to include more stakeholders in discussions, consulting them sustainably. Decision-making about AI happens very far away from people and their lived realities. Therefore, the role of civil society organisations and public institutions needs to be strengthened to facilitate better community participation. Through greater community engagement, people will understand fair use and data exchange. Furthermore, there is a need to understand the ground level and consider people's needs rather than corporate interests. However, this is an ideal and challenging situation because of the widespread belief in the power of data and technology. Another effort that we must make is to build capacity and awareness among grass-root level organisations and empower them to participate more in policymaking processes and tech policy creation. Thus, the significance of participatory and representative institutions cannot be overstated.


Every technology carries with it harms and benefits. At the policymaking level, developing risk assessment frameworks for using AI and doing rights-based work is important. These frameworks must be developed with expertise and would facilitate better legislative and policy measures. They also allow us to understand whether we can address those harms. While there is a discourse around data trusts, models and co-operatives, they are currently inadequate and mostly unsuccessful. This is also because corporate interests tend to hijack several emerging models. Academic institutions can play a significant role in creating a dialogue among communities, research and industries - a historical necessity of our time.

66
AI and ML systems function as status quo systems and create futures based on what happened in the past. In that sense, they are not transformative because they are built on existing data, which they reproduce. So, when you create systems and introduce them into a society that is already profoundly polarised or has cleavages of gender, caste, religion and class, it is likely that those will be reproduced.



**CORPORATE IMAGINATIONS OF ARTIFICIAL
INTELLIGENCE AND THE NEED FOR STRUCTURAL
REFORM**

DEEPAK P



The article draws on author's observations on labour in the lower ranks of the IT sector pyramid, including questions of dignity and agency as the employers unknowingly provide data to add into the system that could eventually lead to their unemployment.

Keywords: Machine-Learning, Training Algorithms, Labour, Dignity, Unemployment, Data Collection, Good and Bad AI

Deepak P

Deepak P is an Associate Professor of Computer Science at Queen's University Belfast. Besides his work in the area of data analytics and AI, Deepak is also interested in and has written extensively on AI Fairness from a philosophical perspective, science communication, and the politics and ethics of technology.



Introduction

In this article, I will share my experience of working in the field of Information Technology and reflect upon the ways in which algorithms and systems of Artificial Intelligence are developed, implemented and relied upon in these spaces. I was previously employed in the research division of a big MNC and engaged in multifold research responsibilities, many of them pertaining to processing data from message boards and myriad forms of intelligent suggestions. It was primarily a Business to Business company, which means that it does not operate services from end to end users but instead delivers services from one business to another. Working within the company, I developed several observations on AI and its injection into a vast spectrum of departments and sectors. The most significant observation, one that would eventually change my perspective on AI and technology in general, is regarding the depreciation of human agency through the wide use of AI.

IT companies collect, collate and organise large amounts of data; with the introduction of AI-operated systems and the growing need to transform more services into AI-managed resources, I happened to be involved in projects involving what could be called a paradoxical development. Customer care workers, who are employees at the bottom of the IT pyramid, were asked to collect data that could be used to create an automated customer care system. Human workers were collecting data as raw material for a machine that was about to make them jobless. This ironical instance and system design can be evidenced across the world in different industries that are moving toward AI-backed functioning. The service provider Uber has begun envisioning a long-term plan of introducing driverless cabs. In order to achieve this goal, they would collect and utilise the data that is generated by drivers. Essentially, this means a loss of the dignity and agency of the worker and the driver. This may also be read alongside observations



“Customer care workers, who are employees at the bottom of the IT pyramid, were asked to collect data that could be used to create an automated customer care system. Human workers were collecting data as raw material for a machine that was about to make them jobless. This ironical instance and system design can be evidenced across the world in different industries that are moving toward AI-backed functioning.

in early Marxist literature about how rampant division of labour reduces the agency, self-respect and satisfaction of workers. With data-driven AI, this is being operationalised at never before seen scales and in ways unforeseeable in the pre-AI era.

Another example may be sourced from the National Register of Citizens that is proposed to be implemented in India. The NRC, which proposes to identify illegal immigrants and establish who is a citizen of India, is an exercise heavily dependent on algorithmic storage and analysis of data. Human effort and assessment becoming increasingly irrelevant have severe consequences in operations such as the NRC as well if the reported usage of data analyses within it is true. For example, one part of the NRC operation is labelling several individuals as “Doubtful” or “D” voters, and they have to subsequently bear the burden of proving their citizenship even though the problem exists with the software. This framework may be observed as rooted in solutionism, a belief that there exist technocratic means for social problems (in this case, assessing whether an individual’s citizenship is doubtful). This framework additionally paves a rich pathway for data-driven technology like AI to slowly creep into the pipeline, even in an opaque manner. Eventually, a caseworker handling multiple such cases would be more likely to trust the algorithmic decision of labelling as illegitimate rather than actively investigating the cases, especially since the latter is much more time-consuming, burdensome and inefficient.

Tech Giants and Corporations

In order to contextualise and evaluate the effects of AI, we must understand the larger aim of tech giants and corporations. Their goal is to maximise their profits and that of their shareholders while minimising the labour workforce. The design and use of AI is also a component of




the same trajectory. In the USA, there are multiple software being developed to assess people's eligibility for welfare programs. As the decisions rest with the software, not a person, one cannot contest their rejection, and thus there is no space for dispute and dialogue.

During my industrial research career, I witnessed that, unlike popular perceptions, call centre workers often had to exercise significant skill, knowledge, reasoning and diligence while attending to sophisticated issues. Experienced workers could spot issues and solutions through their knowledge. Often, through their gut feeling, a call centre worker could suggest alternative solutions to a customer's problem. At that time, there also existed the option for workers to suggest solutions to other teams. In my opinion, this awarded respect and importance to human experience and contribution. The facility and use of human input contributed to efficiency and improvement and considered the worker as a respectable person. On the contrary, AI de-skills the call centre solution pipeline, and while it may be able to address several simple issues, it would need to fall back on human expertise when the going gets tough. But then, eventually, after several years, there may not be any human expertise to fall back on since their learning process has been automated by the very same AI.

Demographics and Social Composition

Another aspect that impacts the human-AI relation in IT spaces is the social composition of those who develop these algorithms and technical systems. At a broad level, tech giants and the spaces of engineering and software are majorly populated by elite sections of society. For example, the book "Technically Wrong" cites the demographic homogeneity of software developers as a significant issue and uses the acronym "WEIRD" to refer to the composition



of Silicon Valley. WEIRD stands for White, Educated, Industrious, Rich and Democratic.

An indistinguishable demographic lacks the capability to think deeply about social issues and consider the relation to AI and technological designs. Amartya Sen has argued that persons who have experienced discrimination have a deeper vision and capacity to identify and comprehend issues such as social discrimination. In addition to a homogenous demographic among developers, there is also the significant issue of the lack of social awareness and exposure to humanities and social sciences. This is evident in the curriculum of IITs and engineering colleges in India. The lack of awareness is also evident when cases of sexual harassment are filed, and women have had to resign while the man they complied against is promoted. While tech giants claim to have zero tolerance toward gender discrimination and sexual harassment, in practice, the policies are not enforced in the same manner.

The lack of social science awareness also reflects the fact that caste is often the elephant in the room in India. The majority of the persons employed in the tech industry in India are from privileged upper castes, and there is negligible discussion on developing caste-sensitive projects. There is sensitivity around race, gender, and sexual orientation in the tech industry in the West; however, that cannot be replicated in the global south as caste has its specific structures and textures. The company I worked for did have initiatives where their employees were sent to poverty-stricken countries for a few months and conducted socially relevant projects for the underprivileged. However, such initiatives are very few in number, and the tech industry could improve with more such projects. In my experience of working among tech corporations, I did not witness stakeholders and impacted communities being consulted for their inputs, even when the large corporations aspired significantly to deliver projects for the public sector,




one that would impact large and diverse populations. Even in the projects I was engaged in, we took inputs from managers, but inputs from call centre workers were not taken. These social and political differences deeply impact the design of AI software and are inextricably linked to the policies and the manner in which digital ecosystems are deployed and implemented.

“The facility and use of human input contributed to efficiency and improvement and considered the worker as a respectable person. On the contrary, AI de-skills the call centre solution pipeline, and while it may be able to address several simple issues, it would need to fall back on human expertise when the going gets tough. But then, eventually, after several years, there may not be any human expertise to fall back on

Good and Bad Aspects of AI

While the lack of AI literacy is a problem, the good and positive aspects of AI cannot be overlooked. There are many useful and beneficial ways in which AI or ML systems can be used. AI and ML are very data-driven and prove efficient in many ways. For example, in the field of bioinformatics or in the medical domain, there are many images from a pathology lab that help in ascertaining the presence of diseases. In such images and data sets, the possibility of human bias is very low. There are many examples where AI can mitigate human biases. AI imaging is also useful in pest detection in agriculture in poultry farms. However, in criminology and justice systems, AI-based systems may not provide the best-unbiased framework. While there are asymmetries and biased data collection in the medical field as well, they tend to have more potential to mitigate biases and provide efficient frameworks.

The negative aspects of AI can be understood by evaluating the process of aggregation and the demand for accuracy. For example, if a class has three times the number of males and females, then our assessment of the average performance of that class is going to be influenced three times more by the male population than by the female population. Furthermore, if we're optimising for average performance as something that you're targeting to improve, then we would eventually privilege policies that are oriented towards males rather than females, and this could be true for any dataset.



Ordinarily, in the last 40-50 years (which is also called the neoliberal era), AI has developed to serve utilitarian measures and optimise time, effort, labour, energy and so on. Therefore, AI tries to achieve optimum accuracy. But as emphasised earlier, if we're optimising accuracy over the entire population, we might be introducing a majoritarian bias implicitly and by design. Over time, these datasets become large, and the sophisticated patterns within them get encoded in neural networks. They become just a set of numbers, and it is difficult to 'see' the learning of biased patterns or even trace them back within the operation of a neural network.

Conclusion: Motivations and Data Collection

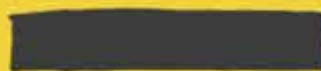
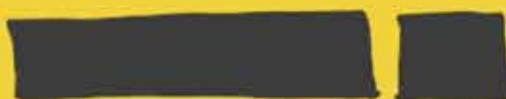
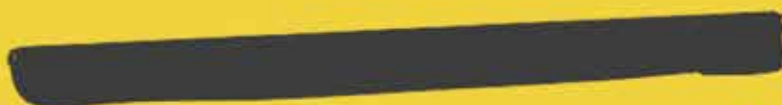
In evaluating the role of AI in the tech world, we cannot exclude a deep acknowledgement of the fact that all these systems are immersed in a neoliberal corporate culture, whether profit is the fundamental motive. Companies such as Uber won't automatically become more participatory and include workers and drivers in their decision-making because they are focused on profit maximisation. Therefore, the incentive to collect and use data in a capitalist world must be scrutinised and foregrounded. Simultaneously, the monopoly that corporations such as Ola, Uber, food delivery companies or mobile service providers have created must also be challenged using a worker-focused perspective. Another significant issue that emerges is how to challenge the decisions made by AI; what would an appeals process look like across various sectors?

I would argue that we need to challenge the assumption that collecting all and any kind of data is a good exercise. Collecting data is definitely an exertion of power of some kind and can very easily be misused. The popular idea that AI is a fair domain of judgment needs to be challenged. It is important to recontextualise and understand how




we understand the concept of fairness. Different kinds of fairness are in contest with other kinds of fairness. The predominant stream of understanding fairness in AI equates accuracy for all demographic rules and equates accuracy, facial recognition accuracy, which is created for males and females etc. This is a very easy technical way of defining fairness. However, fairness is much more than that and theories of justice are much more nuanced. I assert that for a particular domain, we need to look at what is fairness for that domain and scrutinise that question. We need to get people from different domains in the discussion process even before the algorithm conceptualisation stage. However, this is not happening at all because AI algorithm development is spearheaded by the corporate world, and there are also conflicts within AI ethics. In certain cases, some ongoing AI ethics projects have even been rightly criticised as an effort to produce a veneer of fairness and to let the underlying process go ahead without restraint. A sustained and rigorous collective effort toward altering the reforming of the AI ecosystem and addressing the structural issues within it will be the first step in developing more just and fair AI and ML systems.

DATA STATEMENT



WHY EXPLORING THE LEGAL STRUCTURE OF DATA USAGE, ENGAGEMENT WITH IT SYSTEMS AND DIGITAL ARCHITECTURE IS NECESSARY FOR AN ETHICAL AI

RAKSHITA SWAMY



The article looks at available legal frameworks that govern systems run on AI, and the status of ownership of the datasets the system uses. In it, the need for transparency in understanding how our data is being used by the state and the right to access public, developmental data is argued for.

Keywords: Developmental Data, Data Statement, Access to Information, Transparency, Legal Framework, Utility Audit, Data Literacy

Rakshita Swamy

Rakshita Swamy is a public policy practitioner, a visiting faculty at National Law School of India University, Bangalore, and the founding Head of Social Accountability Forum for Action and Research (SAFAR). She has collaborated with state governments and civil society organisations across the country on building accountability frameworks, and is also a member of the Social Audit Advisory Body at the Ministry of Social Justice and Empowerment.



Introduction

The service delivery system has become contingent on the issues of data justice, which involves two essential elements- social audit and transparency. We need to parse these concepts of digitisation having distinct paradigms – digitisation of records, automation of business processes involved in the delivery of services and functioning of institutions, the collection of information, its organisation and disclosure; in addition to that, we have Artificial Intelligence (AI) and Machine Learning (ML) using the paradigm mentioned above for decision-making. Beyond the general ethical framework of AI, we also need to explore the legal structure of data usage, its engagement with IT systems and digital architecture to ensure that the principles of social justice are followed in its deployment.

Need for Utility Audit and Data Literacy.

In common parlance, technology has become a byword of digital technology; however, many technologies are not yet digital. When campaigns like ours engage with state governments, we tend to stress only certain aspects of digitisation, automation, information collection and disclosure because we tend to have concerns with other kinds of things. The IT system is created for NREGA or another kind of welfare scheme to enable service delivery and track it openly and publicly through a digital platform. When we got access to many of these Information systems internally, it has shown that there have been cases of the administration having privileged access to information that was not made accessible to the public and certain controls that were administered that were violative of the law.

There is a real need to audit these Information Systems from the point of the law; for example, if a certain NREGA law has stipulated that citizens have the right to work when they



demand work within fifteen days, and if they fail to get it, then, the citizen should know as to what is in the IT System which is limiting that right? Only a checklist of broad ethics is not enough to ensure ethical compliance; instead, an IT system created for the delivery of a particular law or scheme should be appraised as per the guideline or legal provisions of that very scheme. The pertinent question is whether the IT system has enabled or stifled the right. We know that only a few stakeholders, for example, programmers, government officials or social activists working on digital justice, know about it.


“

People should also have the right to make decisions about the government regarding planning and participatory budgeting. There should not be a scenario where the government has access to all the data to make decisions, and people are denied access to make decisions on everyday governance.

Can We Attain a Balance?

Many state governments have metadata like Samagra, Kutumba (Karnataka), and Jan Aadhar (Rajasthan). Given that governments spend huge amounts of public money to collect this information to build a database, it is natural that a question might arise as to whether the government would use it to provide entitlements to people or still wait for people to apply. There should be a fine balance between data as an aid or supplement in the decision-making vis-à-vis data deciding by itself. In my opinion, there should be a delicate balance between them. To make sure that ethical utterances are not reduced to mere rhetorical flourishes, there should be clear safeguards and processes that have to be put in place to see when data are being used to facilitate decision-making or when it assumes the role of a decision-maker and takes the role for itself and make a decision.

Not merely the decision-making of government people should also have the right to make decisions about the government regarding planning and participatory budgeting. It should not be a scenario where the government has access to all the data to make decisions, and people are denied access to make decisions on everyday governance. To ensure the impact of data justice on development, democracy, and so forth, we need to expand the



There is a real need to audit these Information Systems from the point of the law; for example, if a certain NREGA law has stipulated that citizens have the right to work when they demand work within fifteen days, and if they fail to get it, then, the citizen should know as to what is in the IT System which is limiting that right? Only a checklist of broad ethics is not enough to ensure ethical compliance

conversation beyond the domain experts, government representatives and lawyers and engage with civil society groups, grassroots movements, trade unions and the people through a public education campaign. Governments can also campaign with a creative eye so that these conversations can happen at the grassroots level. We do not want to wait till people will face the brunt of it and realise it.

There is another issue of transparency in ensuring data justice. Since state governments wield their authority to use the databases that may or may not have been given by the people voluntarily; therefore, people should have a digital statement like on a bank passbook to keep track of the purpose of the usage, particularly in the case of state governments that have the 360-degree profile of an individual. If the digital statement is generated, it will make the data available to the people so that they can track how their data, once submitted, is being used by other departments for making a different kind of decision. It will inform and educate people about the process and purpose for which the government has used data.

Conclusion

The domain experts, policy writers, and government representatives are still learning the entire data governance system, even though we are familiar with parts of the whole. There are attempts from the state to spread awareness about data literacy. Civil society groups involved in training should understand that a digital system and data collection organisation directly impact every issue and sector we work in. Therefore, we must make this a part of our work vis-a-vis training and litigation to transmit it to the mainstream consciousness.




Perspectives on AI Innovations:

**Exploring Visions
and Realities**



**AI IN THE GLOBAL SOUTH- GAPS IN DATA AND
POSSIBLE FRAMEWORKS**

RAHUL PANICKER



This article is based on author's experiences working for the AI-for-Social-Good project in rural India, where digital solutions were innovatively conceived to solve existing social issues and inequalities.

Through some examples, the article argues for the importance of working across the stack and incorporating human workflows into innovative design, while warning us of the possible historical biases that may be built into the available data.

Keywords: AI for Social Good, Planning, Rural, Detection, Human Workflows, Multidisciplinary Approaches.

Rahul Panicker

Rahul Panicker was the founding Chief Research and Innovation Officer at Wadhvani AI, leading a team that looked into using AI for Social Good in India. He has also worked with NITI Aayog in helping write India's national AI Policy. He is presently the Co-founder and CEO of Tendril AI, based in San Francisco.




Introduction

In this article, I will elaborate upon my experiences working on algorithms, Artificial Intelligence (AI) and innovation to detail the ways in which existing machine-learning-based systems can be improved in serving their functions. I have always kept an active interest and participated in policy affairs. However, I write this article in a personal capacity and wish to reflect upon the knowledge I have gathered over the years by working in various organisations and sectors. I worked at the Wadhvani Institute for AI, an AI-for-social-good research lab in Mumbai, India, where I was the founding Chief and Research and Innovation officer. I will draw my arguments from my history of working in the social sector and my background in technology and share a few examples of how existing social inequalities impact innovative solutions designed to solve specific social issues.

Section: Designing AI Governance for Social Good

In 2008, I started a company that developed technology for social purposes. We developed an incubator that could work without electricity for premature babies, which is suitable for village settings and portable. We scaled that to over 25 developing countries, and to date, we have probably helped half a million babies and have received much recognition for that work. In fact, our work was also procured by WHO back in the day to help Syrian refugees. I helped co-write India's national AI policy that Niti Aayog put out in 2018. I am also part of the partnership on the AI steering committee for shared prosperity. Lastly, I was part of the steering committee and was part of the judging committee of the Canadian government's See Far initiative.

At Wadhvani Institute, we were basically trying to create large-scale social impact through AI-based technologies. One significant example of our work was our project on




Our endeavours differ from the Western approach to Artificial Intelligence. While they attempt to prevent the bad uses of AI, we try to make AI for social good, and these approaches are significantly different. There is a danger in following the West too much. In our case, we try to fill the gaps in the system because we do not have enough doctors and agricultural scientists and so on.

neonatal health. Too many low birth weight babies are born, and they are not weighed correctly at birth; therefore, they are missed from receiving appropriate care. United Nations International Children's Emergency Fund (UNICEF) data shows that less than 50 per cent of babies are weighed correctly at birth, and similar numbers also hold in India. This is concerning because low birth weight babies contribute disproportionately to infant mortality. However, if the low weight is discovered in time, they are relatively easy to help. This is why public health systems are trying to solve this and create a balance with ASHA (Accredited Social Health Activist) workers.

We have been deeply invested in this project for 15 or 20 years; however, it has not helped as we are still missing a large number of these babies. We have identified multiple reasons why the spring balance, the hand-held device which measures the baby's weight, has not worked. There are many reasons for this. Often, the supply chains don't work, and the spring balances don't reach healthcare workers but, of course, sit in some storage facility in the state headquarters or district headquarters. There are also maintenance issues and there are cultural issues. For example, many families don't like outsiders touching the baby for the first few days. Finally, there are also data quality issues. For instance, in an Auxiliary Nurse Midwife (ANM) register form in Rajasthan in 2019, all babies born in the region magically and precisely weighed 2500 gms. They were registered as such because 2500 gms is the minimum cut-off. It is the mandated level. These are some of the issues we face.

The system was designed with several digital components as a crucial part. ASHA workers, Anganwadi workers and ANM were given smartphones, and digital pipelines were established. Each of the workers has various apps on their phones, and various states also have digital pipelines. Multiple countries have digital pipelines, and




South Africa has MomConnect, which is a mobile-based initiative which provides information about pregnancy and labour to women. The system we designed worked in a way that, without contact, through a video, we could estimate the weight of the baby. This system worked offline as well and would work even if one did not have internet connectivity. It worked as long as the phone had inference, was geotagged, had a timestamp, and no one was punching in any numbers anymore.

The weight estimate was done automatically, and therefore, the numbers were much harder to fudge. We tried to make the system accommodate different constructions, weight sizes and complex backgrounds, which is typical in village homes.

Section: Agriculture and Artificial Intelligence Technologies

Another project I have worked on is in the domain of agriculture and worked as a system to prevent the destruction of cotton crops. Our system tried to detect pink bollworms, which is a form of pest or infestation which wipes out the cotton crop. Over a hundred million farmers depend on cotton cultivation globally, and the devastation of the cotton crop is a huge concern. There are wide varieties of bollworms, and around 50% of pesticides used in India are used to treat cotton crops. Our system worked with the farmers and the frontline workers who took pictures of the pest trap. There are different forms of pest traps, and we basically identify which of the bugs on the pest trap are actually harmful ones. There are many general insects that are not harmful, and it is important to distinguish between the two. We detect and recognise the harmful ones, count them and check whether their number is more than the acceptable amount and also how much time they are progressing. And we also discern information about



“The product itself is only one part of the solution, and there are multiple aspects to be cognizant of. For example, who is going to ensure the authenticity of data, who will ensure that the interventions are being done correctly, and how will it be certain that actions will be taken when an alert sounds in the system? Therefore, human workflows, ASHA workers, ANMs, and doctors are all a part of the larger flow of things

what is going on in the larger area to address problems of infestation at the larger village level as well. We attempt to detect the pests when they initially come as flies and try to prevent them from reaching the cotton balls. Once they reach inside the ball, not much can be done. Our system thus works as an early warning system which works along with advisory programmes at the village level and advises how much pesticide to apply and when.

Working on these projects, I have observed that our endeavours differ from the Western approach to Artificial Intelligence. While they attempt to prevent the bad uses of AI, we try to make AI for social good, and these approaches are significantly different. There is a danger in following the West too much. In our case, we try to fill the gaps in the system because we do not have enough doctors and agricultural scientists and so on. However, we also try to prevent bad usage. Technology is merely an amplifier of human intent. Therefore, my reflections in this article are geared toward explaining how systems need to be strengthened across the spectrum to ensure things such as data justice and good, judicious use of AI.

We must first ask ourselves, how are good products designed? Essentially, we work with users, identify the market gaps and iterate them. This is the classical approach toward product innovation. To use technology for development, we must work with a large spectrum. It is not just enough to work with the community. For example, in the case of weighing a newborn baby, it is not just the mother who deals with the visual weighing machine or whether the anthropometry solution will be used by the public health system or not.

Similarly, the ASHA worker alone does not make decisions; however, each of these persons is involved in the system. The mother or the ASHA worker can also deny the usage of the machine. In order for the information to benefit, the local primary health doctor has to use the data in some




meaningful way, and the state health care system has to adopt the system. These systems also have to be part of a national initiative because many of these apps and data systems are established at the national level and funded by the national health mission. And these also continually get updated; for example, we have a lot of digital health initiatives now.

Therefore, if we think of it like a stack, we must work across it. We must incorporate the social sector organisations that have the trust of the community and who know the problems and can help one pilot and iterate solutions. Simultaneously, one must also design systems to fit large-scale programmes, and at such a massive scale, there is no single user profile. Multiple persons are involved in decision-making, and many people have to take decisions and actions for these systems to function. One good example of understanding this is that public health systems don't work in a Business-to-Consumer fashion. Instead, they work closer to the Business-to-Business model. Hence, we work across the user stack to ensure that our systems are well-designed.

Section: Working Across the Stack of Technological Systems

It's important to understand that artificial intelligence is only one part of the whole technology stack. For any product or any system to work efficiently, it has to actually fit into existing workflows, and there have to be featured around the core technology to enable optimum usage. Moreover, the product itself is only one part of the solution, and there are multiple aspects to be cognizant of. For example, who is going to ensure the authenticity of data, who will ensure that the interventions are being done correctly, and how will it be certain that actions will be taken when an alert sounds in the system? Therefore, human workflows, ASHA workers, ANMs, and doctors are



all a part of the larger flow of things. And this is true for agriculture extension workers as well.

The solutions we design have many components, such as the product, the technology and the AI. Thus, a multidisciplinary approach is necessary. For example, we work with agricultural experts and people with a social sector background in deploying programmes, doctors, product designers, engineers and so on. Thus, in providing solutions, one cannot only focus on AI. The real problem is not with AI but elsewhere in the system; solving AI will not make any difference. We have realised this quickly. This can be further explained with an example from another project I worked on. We worked on sputum microscopy for Tuberculosis detection. Diagnosing TB involves taking a sputum sample, staining it and looking through the microscope. At the point when the TB vacillates over, we count it through a computer vision solution to check whether the sample is TB-positive or not. And this works similarly to the pest detection model.

We discovered that the problem was not in computer vision but elsewhere in the chain. The real issue lies in collecting the sample and in preparing the slides. Looking through and then observing the sample is not the hardest part. Therefore, while we could have thought of the AI aspect as central to the problem, it would make no difference in the overall scheme of things. The challenges in the system are not in object recognition but somewhere else in the chain. There are many such examples where an engineer will find an AI component to fix but fixing them does not make any difference.

A certain level of expertise, experience and spending time with many levels of the social sector stack. One must spend time with the user stack to identify where one can actually make a difference. In one way, engaging with the impacted communities may be helpful. There is a discourse




around consulting village communities and beneficiary communities. However, these are not enough to provide solutions. They are merely necessary but not sufficient. In the context of India, the relationship between technology and power is important to consider. We must have mechanisms to mitigate and prevent excessive concentration of power, which often happens with digital technologies. Checks and balances are thus crucial.

In the Indian context, institutions of democracy are still weak, and in many parts of the country, they are also subverted. There are significant disparities in power, and these apply across communities, religions, castes, social and economic strata, gender, age, educational level and so on. Any public solution will have to account for all of these factors. We must make mechanisms so that the effectively marginalised and unpowered communities are not left out. We must also ensure that we do not take the Business-to-Consumer approach that is followed in the West. The B-to-C approach does not help with development. In the B-to-C approach, there is no one between the company and the consumer. Instead, we must develop intermediaries, for example, ASHA workers or agricultural extension workers. Solutions must take a mediated form with human layers in between.

“Any public solution will have to account for all of these factors. We must make mechanisms so that the effectively marginalised and unpowered communities are not left out.”

This can be elaborated further with an example. If an ASHA worker does not visit certain parts of the village, which happens very often, then none of this applies. In many Indian villages, a road divides neighbourhoods based on caste. Upper castes live on one side and lower castes on the other; data gaps can be observed from this distinction. Therefore, if there are gaps in the human system, then there is only so much that technology can do. For example, the baby weighing system we developed works well in lighting conditions in homes like mine. But the home of a poor person in a village is dark and does not have too many windows. Solutions would usually be based on data from



better-off communities, and these factors won't get caught. This technological challenge is typically not addressed, and one would not get enough images or videos from darker environments. Similarly, one can observe that a baby's shape or skin colour differs due to nutrition levels or other factors, which also affect data distribution.

The role of field tests comes much later in the development and deployment of AI technologies. One must create multiple test sets for testing various conditions. For example, if one wants sufficient gender representation, then one must ensure that their performance is set across different genders and create multiple data sets to be informed by different conditions. However, finding enough samples for all phases is hard, and many conditions to be tested are rare. Therefore, field testing is necessary but not always sufficient.

Conclusion: Possible Solutions and Frameworks

In relation to data justice, we must bear in mind that whenever newer technologies are introduced, there are always people that are left behind. In human history, this happened in the case of electricity as well. There is always a focus on what is easy to implement and what is easier to achieve success. The channels through which any technology is deployed are always limited. Thus, technology is one part of a larger subset of societal problems and political problems. The real question is how do we ensure justice while knowing the reality of things? We must face this fact upfront instead of merely stating that technology leaves people out. There are also many examples to show that AI can be used for social good. Thus, we need to focus on maximising the benefit and minimising the harm.

There are many examples that show how historical biases




are built into data. For example, for similar levels of illness in cardiology, data has shown that men receive more interventions than women. Similarly, credit worthiness of persons is based on the number of loans that have been rejected in the past. For identical income or wealth levels, people who belong to communities that have wealth and therefore have greater informal access to credit will have had fewer loans rejected historically than people who come from communities that don't have wealth and therefore are likely to have greater loans. One can observe this to be true even in data on law and order. If we look at past convictions, people who can afford lawyers will have fewer convictions. So if the conviction rate is the proxy used, or if past loan rejections are a proxy used for creditworthiness, proxies can be very misleading. Therefore, the consequences and the impact of biased datasets have to be looked at very carefully.

Far too often, algorithms are seen as monolithic or linear processes where a developed algorithm will work exactly as imagined when deployed. However, deployment is a different domain, and several things, such as law, society, people's lives and politics, are not captured in this picture. In India and the global south, we must also not replicate the values and systems of the global north. A possible solution is that there have to be legal frameworks around these different circumstances, and they should exist on both papers and in practice. The gap between written and enforced laws must be bridged for such frameworks. A comprehensive AI vision for data justice will take into account the aspect of people, knowledge, capabilities of the public systems, and the power imbalances that exist in the system.



**BETWEEN EMERGING TECHNOLOGIES AND THE
COMMON GOOD: EXPLORING A DELICATE
BALANCE**

JAYESH RANJAN



Jayesh Ranjan writes on Telangana's success story in employing digitisation and AI technologies, including the state AI Mission's comprehensive framework and plans to integrate AI into education and innovation. He argues against the criticisms of the Samagra project, citing the improved accuracy of data retrieval and the fixed welfare leaks.

Keywords: Innovation, Government, AI for Social Good, Exclusions, Samagra

Jayesh Ranjan

Jayesh Ranjan is an IAS officer and the Principal Secretary, Industries & Commerce (I&C) and Information Technology (IT) Departments, Telangana, where he works in developing policy frameworks, attracting new investments, identifying opportunities of utilising IT in various government processes, and promoting digital empowerment of the citizens.




Introduction

The aggregation of big data through emerging technologies like AI and issues relating to the ethics of its use, transparency and right to privacy have been raised by citizens and various stakeholders. Ploughing metadata through new technologies for government programmes is not an unmixed blessing. However, one should see the brighter aspect of any new technology, not an unmitigated disaster. The concerns of transparency in the operation of big data, issues of privacy without sacrificing the integrity and confidentiality of the processes and institutional mechanisms to oversee its ethical use should be considered a challenge. They can be addressed without falling into the reductive trap of manager versus worker discourse. There is a scope to explore a win-win situation for the government and the people.

A Balanced Approach towards Emerging Technologies

Data has multiple connotations, but we should focus on the optimistic aspect of it, and according to my understanding, it has the potential to do a lot of social good. Like any new technology, data can have many pros and cons. As digitisation is moving rapidly, adverse effects are also coming to the fore; however, it has two sides to it. There are umpteen instances where data have been used for social good, and we should focus on identifying those areas where data can be used for such purposes. Furthermore, the data themselves are meaningless unless there is intelligent use of these data enabled by emerging technology such as Artificial Intelligence (AI). However, it needs to be reiterated AI has a limitation, particularly in the domain of accuracy. AI is being used to feed humongous data into certain algorithms, but there are no foolproof measures to attain cent per cent accuracy.



66
A large portion of our IT initiative focuses on agriculture because it is the rural economy's mainstay and supports many people's livelihoods. Successive governments have promised to double the farmers' income, but no one has a roadmap for achieving the goal. We, in Telangana, are using the technologies to accomplish that goal.

In Telangana, we run lots of algorithms for agriculture; I have no compunction about admitting that none of my algorithms can claim to be hundred per cent accurate. There is a scope for that margin of errors that reflected, inter alia, in either the exclusion or inclusion errors. Consequently, it may end up excluding some sections of a vulnerable population. It is an open question whether such an anomaly of the output can be attributed to the bias of those who run and manage the data or write the algorithms. As we know, such a bias is demonstrated in the case of the historically vulnerable Black minority in the US—a society in which the colour of white skin is inextricably linked to social power informed by historical prejudice against Black people. With such historical antecedents between White and Black people, there is a possibility of such bias creeping into the system, which might be prejudicial to black skin. And if there is a possibility for such a thing, the algorithm may do more harm than good. In India, however, the moot question is to ascertain whether those identifiable markers can militate against the historically marginalised section of society because, in India, it would be difficult to differentiate people on the basis of skin colour.

Most of us, in our approach to appear holier-than-thou, speak of ethics in the usage of big data, AI or such emerging technologies, but Telangana is the first state to put in parameters which cover the knotty intersection of governance, ethics and concerns of privacy. At times, sweeping statements are made to cast aspersion on government initiatives. However, how to translate these concerns in terms of action is of paramount importance.

To address these concerns of civil society and other stakeholders, we have constituted an artificial intelligence ethics group. Typically such groups are formed by government representatives; however, in the case of Telangana, there is no government representative in this group from Hyderabad or Telangana. We conducted three



roundtables to take inputs from industry, consumers and researchers to form a guideline on ethical usage. We also made sure that AI procurement guidelines were commensurate with the standard of ethics and governance, as we mentioned before. Data ethics courses will be introduced in our institution, mainly where AI is not an independent subject. Besides that, we are also very conscious about sensitising and training the government agencies that run the database. In short, as a kind of government agency, we are cognizant of the fact that there could be ethical issues more virulent than the discrimination aspect.

A large portion of our IT initiative focuses on agriculture because it is the rural economy's mainstay and supports many people's livelihoods. Successive governments have promised to double the farmers' income, but no one has a roadmap for achieving the goal. We, in Telangana, are using the technologies to accomplish that goal.


We came across an inconsistency in the data while running a pilot project. We wanted to create a database to predict the pest attack so that we could control them—what is known as precision agriculture. It is only through prediction and prevention that we can use either natural or organic or even chemical pesticides. To accomplish the project, we began to collect individual farmers' data to create a database to analyse the increased productivity. While collecting the data, it became apparent, though not a surprise, that the names of actual or tenant farmers were not in the land record portal- what is also called dharani identification. When we tried to verify the anomaly with the village, we came to know that the *Dharani* data did not capture the people who were actual cultivators. In other words, the database we had failed to tally with the actual state of affairs. So we realised that farmers were not receiving the advisories we used to send as alerts to prepare them to take precautionary measures.

What is the *Samagra Vedika*?

Samagra is a Telugu word which means holistic or comprehensive. It is a mistake to assume it is the name of a new database. We have not created new metadata; instead, we have collated existing data from thirty-odd government agencies. These data continue to reside with each department. People have a digital footprint; in some cases, explicit consent is not required. For example, the details of the particular taxpayer are put in the database of taxpayers of that concerned place; in this case, consent is not mandatory. In this sense, the database replicates a manual recording keeping system into an electronic system. It is enabled by *Samagra* using Artificial Intelligence (AI). Despite spelling inaccuracies, *Samagra*'s algorithm can remarkably accurately retrieve records from databases. Earlier, the accuracy rate was 72 per cent; now, it is improved to 87 per cent. At this level of accuracy, the *Samagra* algorithm can retrieve records from various databases for the government programme. *Samagra* database ensures that the government programme reaches the intended beneficiary by separating the names of those who are not entitled as per the rules.

66 *Samagra* is a Telugu word which means holistic or comprehensive. It is a mistake to assume it is the name of a new database. We have not created new metadata; instead, we have collated existing data from thirty-odd government agencies. These data continue to reside with each department.

Earlier, there used to be a procedure of enquiry for verification of beneficiaries-driven programmes to fill the gap concerning the data available to the government; now, that gap or lacunae is bridged by using data from various departments which are authentic and available. In this, we do not use the identifiers sanctioned by the Supreme Court, for example, Aadhaar. Without relying on Aadhaar and phone numbers, which are privacy concerns, we use generic information like name, father's name and address from multiple databases to provide input to a department to target the deliveries more efficiently. Many people express their apprehension and accuse the government of robbing the right to privacy. Just as one does not object to




voluntarily submitting personal information and abide by the verification rule to get a passport, so does one need to subject oneself to the rule of enquiry to become eligible for the government programme. The only difference is that we supplement our verification electronically instead of relying entirely on manual processes. Moreover, it must be mentioned that, in Telangana, the *Samagra* result is just an input. The department concerned has all right to rely solely on desk verification.

Conclusion

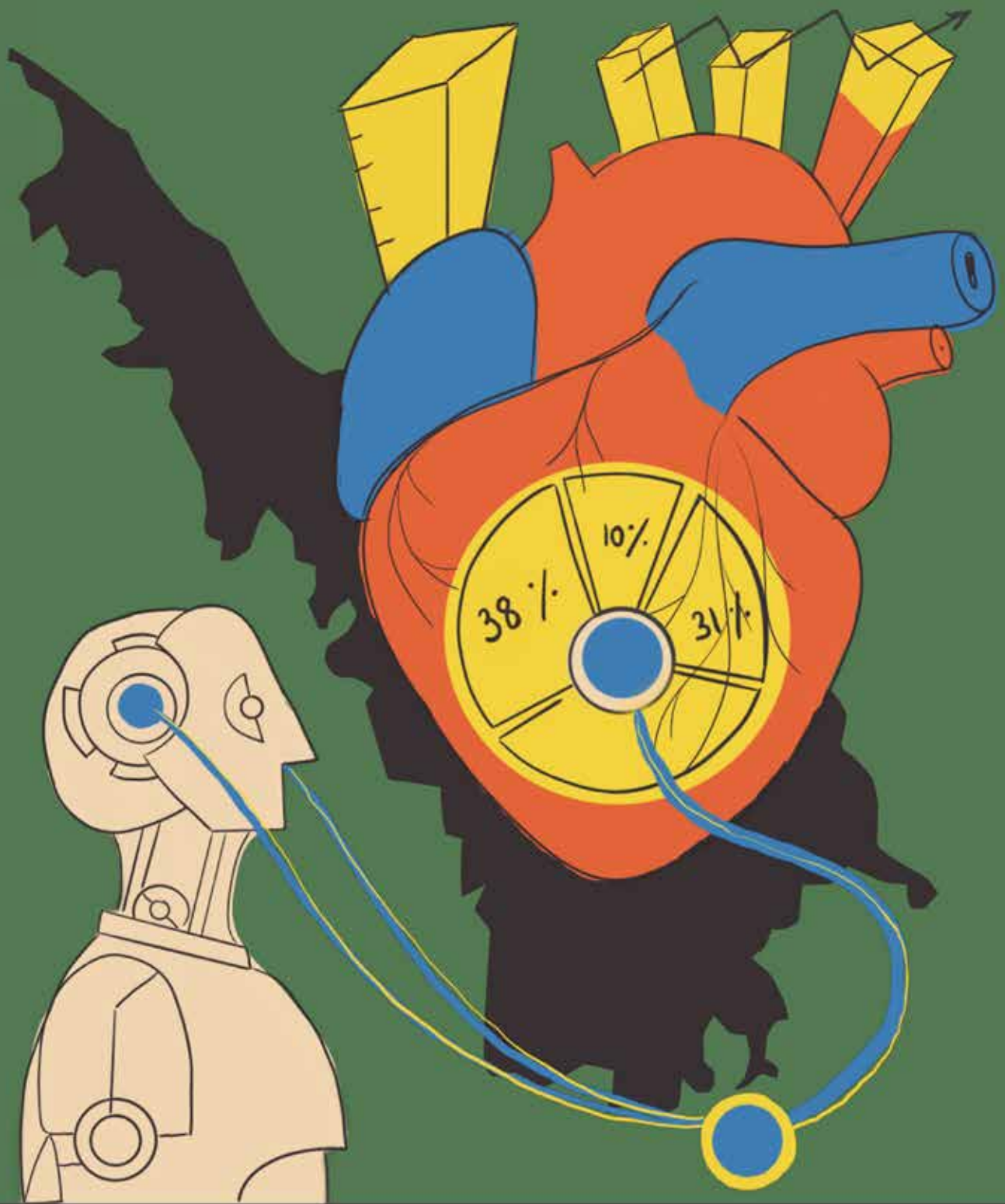
Data Justice is about achieving the six ideal paradigms—power, equity, identity, participation, knowledge and access—to make citizens empowered vis-à-vis digital justice. We are also aware that the misuse of data by a private company is closely linked to the infringement of the right to privacy. We aim to ensure privacy and transparency by designing an inbuilt safeguard in the algorithm. Besides, by involving people from across sections with proven credibility in an oversight mechanism, we can redress these concerns. Regular audits of the data and putting it in the public domain can also rectify the control and transparency aspect of the use of data. As for me, I entirely concur with these viewpoints.

However, I still have reservations about the manager versus the workers' approach. We need to arrive at a win-win situation wherein the manager benefits so does the worker. In a way, ensuring the targeted people of a government programme can be an efficient way of optimising public money. Only because government functionaries have been able to match a certain number of people in the database does not necessarily imply that these decisions are anti-people. And it is essential to point out that these are just automation of manual processes that used to be conducted with many flaws. These governmental inquiries used to be quite perfunctory in the past. Many enquiries during the




verification of the eligibility of candidates were conducted in complete confidentiality. People had the right to redressal if certain services were rejected despite having a faultless record. However, we still have to explore ways and means to become more transparent without sabotaging the integrity and confidentiality of individual data.

In Telangana, in the proposed bill on data protection, we intend to add a provision for data protection authority. The debate about who will head the authority and its various modalities is still underway. There is consensus in the government on the need for an oversight mechanism. A person with judicial experience will probably be tasked with heading a committee or group.



**IMAGINING DIGITAL GOVERNANCE AND HEALTH IN
ANDHRA PRADESH**

G S NAVEEN KUMAR



The article overviews the initiatives taken by Andhra Pradesh to digitise healthcare records and access, showcasing the state's strides in equipping primary health centres with digital records, and in making over 1.5 Cr health IDs. It also elaborates on the plans of integrating Artificial Intelligence to estimate needs and make efficient use of medical personnel and resources, and outlines the potential challenges that the project could encounter.

Keywords: Digitising Healthcare, Health Data, Ayushman Bharat, Digital Health Records, Health ID, Resource Efficiency

Naveen Kumar

Naveen Kumar, IAS, is serving as Special Secretary of Health, overseeing the complete digitization of all operations in government health facilities. He previously worked as the founding Managing Director of Uttar Pradesh Medical Supplies Corporation, and as Special Secretary in the IT and Electronics Department.



Introduction

In Andhra Pradesh (A.P.), the government has executed several initiatives toward digitising health and developed plans for future implementation. A.P. comprises 13 districts, and the state government has introduced e-hospitals in 54 hospitals in the entire state. This initiative has covered teaching hospitals, area hospitals and district hospitals. A.P. is also part of the Education, Health and Care Plan (EHCP) plan and receives funding from the central government. All the urban primary health care centres are equipped with Electronic Health Record systems (EHR), and all doctors enter their data through the EHR app dashboard. The EHR will also be extended to rural Primary Healthcare Centres (PHC), and soon, 1600 PHCs will be connected via the EHR application. Andhra Pradesh has also successfully made 1.5 crore health I.D.s and is the fifth state in the country to do so.

In the administrative management of Covid-19, the healthcare system in Andhra Pradesh tried to reduce panic and administratively manage the pandemic through telemedicine and tele-consultancy networks. A.P. had one of the lowest Covid-19 mortality in the country. During the Covid crisis, we managed by conducting triaging at multiple stages and locations. When the traffic of patients went to tertiary care centres, we did a few localised surveys. We conducted covid tests, and based on the results, we referred the patient to a hospital, the Covid Care Centres (CCCs), or the tele-consultancy network. Through the phone number (104) we set up, doctors could also check up on the patients every alternate day, giving psychological support to the patients.

Regarding the digital health policy in the state of Andhra Pradesh, we currently follow the Ayushman Bharat Digital Mission (ABDM) program of the Central government. The program has credibility because the Government of India




Artificial Intelligence has not been introduced to healthcare yet, but it could be instrumental. For example, if an IoT (Internet of Things) is installed within the medical oxygen system, it can automatically alert nearby concerned doctors whenever levels are low. These systematic changes will take time. Data is like currency or diamonds, and it must be mined.

has designed it, and we hope to build upon this model and borrow from models in Estonia and Israel. Based on the existing models, we are constructing the digital system slowly, and we anticipate that there will be several gaps in the system. Thus, we are proceeding with rudimentary and basic models. One issue with the digital governance of health is the question of data security and data protection. For example, with the HPR (Healthcare Professionals Registry), doctors must register themselves through the application. However, several doctors work in multiple hospitals, which may become a problem for their income tax records. If individuals would feel that someone wants to procure data about them, that could cause the digital health system not to function properly. In terms of digital policy, more generally, the A.P. state government has introduced the Disha app to help women and citizens in emergencies.

Gaps In Digitising Of Healthcare

There are several gaps in a digitised healthcare system. These became evident during the Covid-19 vaccination process when the vaccination process was registered and verified through the Cowin application. In the administrative position, the app only showed me the number of vaccinations pending but did not show who the unvaccinated persons in my area were. If we procure that knowledge, we could persuade those persons to vaccinate. Furthermore, many freefall systems were operating during the vaccination process. Due to many other factors, such as digital illiteracy, many doses expired and could not be used. For efficient and optimal policy implementation, having a Chief Security Officer or an Ombudsman specifically working for data and information is very important. These authorities should be judiciary-based or independent agencies so that people can own their data.

In other digitised initiatives for health, Andhra Pradesh has made several efforts. For example, all ASHA (Accredited




Social Health Activist) and Anganwadi workers have been given smartphones. Furthermore, I have a volunteer network under which 2.5 Lakh smartphones are distributed in Andhra Pradesh, and we pay 32 crores for telephone network connectivity.

Policy Recommendations

Data privacy and ownership should be an agenda of primary importance in any policy design and implementation. Data collection is a sensible exercise only if we appropriately share this data. The Cowin app was successful because we shared the API (Application Programming Interface) with protection with the public on a large scale so that the maximum number of persons could get vaccinated. Therefore the purpose and strategy of engaging with data matters and protecting the rights of the data owners are essential. There were many protests against Aadhar for collecting data; however, the firmware of cell phones manufactured in China also collects data such as photos and many other things. The UPSC website provides my home address, phone number and many other details, which is a data privacy risk.

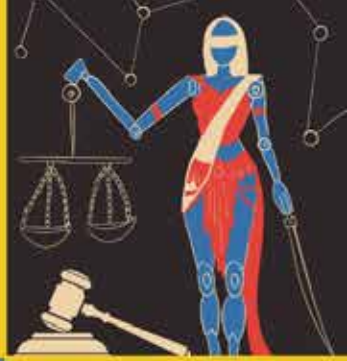
In the framing and implementing data policy, it is crucial to understand the data pyramid in which data becomes information, information becomes knowledge, and knowledge becomes wisdom. We are still in the data stage and are merely collecting the data. Artificial Intelligence has not been introduced to healthcare yet, but it could be instrumental. For example, if an IoT (Internet of Things) is installed within the medical oxygen system, it can automatically alert nearby concerned doctors whenever levels are low. These systematic changes will take time. Data is like currency or diamonds, and it must be mined. For example, to obtain data on cancer, leading cancer research institutions, academics, ICBG (International Cooperative Biodiversity Groups) and practising doctors



would have to come and study the data and institute healthcare.

Another significant aspect of policy design is the setting up of global parameters compatible with local realities. For example, tribal areas are often excluded from the digital imagination. Geographies alter the impact of technology. Parameters should consider variation and be derived after social and cultural research. This is best explained through my experience administering the purchase of neo-natal ventilators in Gorakhpur, Uttar Pradesh. I was in charge of purchasing neo-natal ventilators, and many factors must be considered in this exercise. Ventilators have tidal volume; for example, a 2 Kg baby would need a 200 ml ventilator, and a 3 Kg baby would need a 300 ml one. The price of a 200 ml ventilator was 1 lakh more than the 3 Lakh ventilator. In order to make a well-informed decision, we collected ANM (Auxiliary Nurse Midwife) data across all regions. We were surprised to learn that the average weight of a baby was neither 2 nor 3 but 1.9 or 2.2. Therefore, ensuring that parameters correspond to global and local situations is essential.

Many obstacles come in the way of efficient management. The first is an overload of work and responsibilities. District Commissioners often handle more than 90 departments that report to them, and thus, there are gaps in administration. This is where the role of NGOs and CSOs is important so that these gaps can be filled. At the administrative level, we are collecting large amounts of data and are trying to build a beneficiary management system. When doctors are often asked to empanel themselves in the HPR, they inquire about what will happen to their data. I reassure them by arguing that there has been no data theft in the last 20 years. Thus, digital systems would also function more smoothly if people put their trust in the digitised setup and governance.



Scan QR Code to Read this Book



info@defindia.org
www.defindia.org

