Algorithmic Rural Road Planning in India: Constrained Capacities and Choices in Public Sector

Harsh Nisar[*] [†]
Ministry of Rural Development
New Delhi, India

Deepak Gupta
National Rural Infrastructure Development Agency
New Delhi, India

Pankaj Kumar
Centre for Development of Advanced Computing
Pune, India

M. Srinivasa Rao
Centre for Development of Advanced Computing
Pune, India

A. V. Rajesh
Tamil Nadu Rural Development and Panchayati Raj
Tamil Nadu, India

Alka Upadhyaya[†] [†]
Ministry of Rural Development
New Delhi, India

ABSTRACT

The use of AI and other algorithms in resource-constrained public sector settings of the developing world face unique technical and social challenges. The organizational and institutional realities of the public sector such as legacy data/IT systems, embedded work culture, bureaucratic norms, resource-constraints both explicitly and implicitly shape deliberation, design and deployment of public sector data science projects. Through the case of algorithmic rural road planning in a large-scale government program in India, our work demonstrates how algorithms can be positively utilized within the context of constrained capacities and choices. As practitioners deeply involved in the entire project life-cycle, our action-research provides an intimate and reflective account of how production of even seemingly "simple" algorithmic projects pose non-trivial complexities and challenges in the public sector. We situate the conversation around the humans in the in our setting and show how public sector characteristics impact participatory design, choice of interfaces, data inequities and algorithm design decisions. Further, we show how the preparation and production of technology by constrained capacities can be counter-productive and detrimental even before the technology is put to use. This further expanding the scope of the debate concerning the use of public sector algorithms. Understanding the nuances, practices and constraints in production of data science in the public sector will not only allow more just production of data sciences, but also help formulate realistic strategies to mitigate the risks involved in the use of algorithms in high-stakes public policy situations.

[*]Primary Author
[†]Alka Upadhayaya is a member of the IAS and currently at the Ministry of Rural Development in Government of India. However, all views and errors are solely of the primary author and this paper does not represent the official views of the Government of India

EAAMO '22, October 6–9, 2022, Arlington, VA, USA

ACM ISBN 978-1-4503-9477-2/22/10…$15.00
https://doi.org/10.1145/3551624.3555299

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in collaborative and social computing*; • **Social and professional topics** → *Socio-technical systems*; • **Applied computing** → **Computing in government**.

KEYWORDS

public sector, social good, algorithms, AI/ML

ACM Reference Format:
Harsh Nisar, Deepak Gupta, Pankaj Kumar, M. Srinivasa Rao, A. V. Rajesh, and Alka Upadhyaya. 2022. Algorithmic Rural Road Planning in India: Constrained Capacities and Choices in Public Sector. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), October 6–9, 2022, Arlington, VA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3551624.3555299

1 INTRODUCTION

AI for Good is a growing field where advances in AI/ML are being applied to public policy, United Nation's Sustainable Development Goals or societal issues at large such as climate change, wildlife conservation etc [48]. Governments across the world have also been experimenting with applications of algorithms, AI/ML or mechanism design in service delivery, resource allocation, inspection prioritisation, fraud identification etc [[4], [32], [11], [3]]. There are academic venues which focus on AI solutions for the developing nations where technological solutions to problem statements common to the global south are encouraged [[1], [2]]. While there is growing research to understand public sector algorithmic interventions, it is largely built by parsing policy documents, reverse engineering [27] [18] or simulating of public sector algorithms or by conducting interviews [54] with public sector practitioners. The positionality of academia or corporate communities may limit a holistic perspective on the challenges within the public sector especially in the Global South. Direct reflective public sector accounts are limited [1] but rising [40]. This is partly due to the constraints within which public sector operates. It is non-trivial to setup experiments and structured interviews of stakeholders you are working with given the capacity constraints, over-burdened and closed nature of many government organizations. The role of public sector in applied AI

[1]None of the authors of ML4D workshop in 2020 belonged to public sector organizations

research shouldn't be limited to the site of deployment but also be an active stakeholder in deliberation, design, development and research. Research about AI in public sector or developing countries cannot be centered only around social issues common to these settings which can be solved with the help of AI or other algorithms, but instead what it means to produce algorithmic solutions in these settings, especially by actors local to these settings. We attempt to contribute to the filling of the gap by providing a reflective account of a large scale algorithmic intervention in India. All authors of this paper worked in the government during the period of the research and were deeply involved in the policy, design, development and use of algorithmic intervention talked about in this paper.

There is growing research about the practices of data science in corporate settings and how they separate from academic and research sites of production [43]. Researchers [24] are trying further distinguish between large corporations and smaller organizations which might not be facing the same problems given resource-constraints. Ackermann et al [4] introduces a framework for thinking about deploying machine learning projects for public policy. This framework emphasizes the value of attending to the special characteristics of the *site* influence the production of data science right from deliberation, design, development and its deployment. We argue that using algorithms in the public sector especially in resource-constrained environments pose non-trivial complexities and risks. The realities of existing data systems, bureaucracy and government's weak organizational and state capacity explicitly or implicitly end up influencing the deliberation, design and deployment of algorithms. The hard part of applying algorithmic solutions in these settings is not only the complexity of the algorithms or the novelty of the engineering solutions deployed but the socio-technical externalities [47] which end up shaping the pipeline for data science product from deliberation to deployment.

There is mounting evidence of public sector algorithms adversely impacting marginalized communities [19][39][27]. Passi and Barocas [41] demonstrates the uncertain and messy process of formulating data science projects in corporate setting and highlight the role of formulation in overall fairness of data science projects. Suresh and Guttag [52] offer a framework to identify sources of biases through the machine learning pipeline. We extend that understanding the specific constraints in formulation, production and deployment of data science projects in the public sector can help us better mitigate, to an extent, the risks of harmful and counterproductive applications of algorithms increasingly used in high-stakes public policy settings across the world. We use the case-study of applying algorithms to guide rural road investments in one of the largest rural road construction programs in the world. In 2000, the Government of India announced Pradhan Mantri Gram Sadak Yojana (PMGSY) or Prime Minister's Village Road Program, an ambitious federal program to connect all unconnected rural habitations in India with all-weather roads. It started with a target of connecting nearly 200,000 habitations and over the last 19 years has constructed more than 6,00,000 kilometers of all-weather roads[2]. After a sizable target was completed in 2015, the government shifted their focus on consolidating the existing rural road network further by upgrading existing important arterial and major rural roads. The second phase of PMGSY (PMGSY-II) was launched with a target of upgrading/renewing 50,000 km of existing rural roads leading to *rural growth centers*[37]. In 2019, the Government of India announced the third phase of the programme, PMGSY-III. PMGSY-III is also about upgrading existing arterial rural roads but with a more specific focus on access to high schools, hospitals and agricultural markets. It has been allocated a budget of USD 11.2 billion and a target of upgrading 125,000 km of rural roads. Multiple studies have shown the effect of rural road connectivity on agriculture, health and education outcomes [7][8][5]. The case-study detailed in this paper is about the algorithmic and data-driven process employed for planning and identification of roads to be built under the PMGSY-III program. Roads are highly politicized and visual forms of development and selecting which roads get built is a high-stakes decision with many political, social and economical stakes. To add to that, it's an extreme resource allocation problem because it requires choosing 125,000 km for upgradation out of 4.16 million km of rural roads in India. [3] This makes it an important and hard selection and resource allocation problem. Through our reflective case study as practitioners involved in the overall project, we highlight the challenges faced in using algorithmic solutions in public sector such as integrating with legacy systems, operating with low-resources and designing for overburdened participants with low-technical literacy.

2 SITES AND METHODOLOGY

This work is situated in India and primarily at the National Rural Infrastructure and Development Agency (NRIDA) which is the agency in-charge of development of rural roads in India, primarily through the PMGSY program. The agency plays the role of policy-making and monitoring of the PMGSY government program. PMGSY is implemented by respective nodal agencies at the state or provincial level. These state level agencies have district level units which in-turn do the actual construction and maintenance of the roads, whereas the state headquarter does overall monitoring and management. The national MIS which helps in monitoring the government programme is developed by a team from the public sector information technology enterprise, C-DAC. The federal government sponsors 60% of the project cost whereas the rest is borne by the respective state governments.

This research work is an collaboration between multiple public sector employees involved in the PMGSY program implementation at various levels: policy making, IT management, algorithm development and actual implementation at the state level. This includes the head of NRIDA at the time of the research, the data scientist and GIS specialist at NRIDA who designed, developed and rolled out the algorithm presented in this work, a key mid-level bureaucrat situated at the state level headquarter of a large southern state in India, and the Team Lead and Product Manager of the team from C-DAC in-charge of the Management Information System (MIS). The research covers the time period from April 2018 to April 2021, which includes the finalization of the public policy and operational guidelines, design and development of the MIS and

[2] http://omms.nic.in/ accessed on 17.9.22

[3] The total road network in India was estimated to be 5.89 million kilometers in 2017 which is the third largest in the world. About 70% ie 4.16 million km is classified as rural roads and eligible for PMGSY.

algorithms, training of various stakeholders, roll-out of the scheme, on-boarding of all Indian states to the government program and sanction of road-works under the program for more than 11 states by the end of the research period. The positionality of the authors at various levels of bureaucracy and from different disciplines allows them to provide an intimate, end-to-end and detailed insights into the production of algorithmic intervention within the public sector. The methodology employed closest matches action-research [22] and many of the concerns regarding the algorithmic support system which are raised in the work were also actively tried to be resolved during the research period. The critical examination of the impact of the intervention was carried in parallel and not at the end of the roll-out and as part of the authors' professional duties. Primary data includes the official data present in the program's MIS, official meetings, reviews, conversations and training programs held over a period of three years.

Admittedly, views of district engineers, road construction contractors and rural Indians are not directly represented in the paper. Nonetheless, each of the authors in this paper had almost daily conversations with the district engineers across the country as part of their professional duties and the same was reflected upon to represent their views. The professional duties of the authors and the nature of public sector work prevented us from formally conducting interviews or structured experiments. Each of the authors were personally involved in the roll-out of the public sector program, with different roles and responsibilities, and hence biases arising from this position may be present but have been actively tried to be avoided.

3 EXISTING PLANNING PROCESS

To understand the use of algorithms in PMGSY-III, our case, we need to understand the process adopted in PMGSY-II which preceded PMGSY-III and was similar enough in its objective to be taken as the comparative baseline. PMGSY-II's program objective was to consolidate/upgrade existing rural roads which were serving connectivity to generic *growth centers* [37] and it had a target of 50,000 km whereas PMGSY-III, our case, has a budget for 1,25,000 km and specifically for serving roads leading to agricultural markets, health and educational facilities. This target is distributed by the central government across states, which in turn distribute it across districts and blocks. [4]. The selection procedure for roads under PMGSY-II, our baseline, are explained in Table 1.

One of the characteristics of the above planning process would be that the final result would depend on the quality of the list of 10 routes (in the example) which were picked by the Block Engineer [5]. While, the calculation of utility values in Stage 4 would ensure a local maxima within the 10 routes, the competition is limited to the roads considered important by the Block Engineer in the first place. The Block Engineer could pick a specific combination of roads as candidates to ensure that a pre-decided road comes on top when the priority lists are generated. While a block may have more than a 100 rural roads, generally it would get enough target to upgrade 3-5 rural roads under the program. Within that context, there are competing demands from different interest groups on which road should be picked under the program. This in itself isn't problematic and this is traditionally how development is demanded but not every group has the same voice or leverage within the system. Some may demand a road to service trucks on-route to recently found local mining blocks, or roads leading to sites of religious importance or a new road which would drastically cut the distance of a cluster of hilly villages to the nearby town. While each road may serve a purpose for someone, it may not be consistent with our specific policy objectives. Further, Block Engineers routinely get transferred from one district/block to another and might not always know first-hand the ground reality. The challenges are exacerbated by the absence of trusted and updated traffic or road use data across India's entire rural road network. Conducting a traffic survey on every road in a country like India will be a resource intensive exercise, and may still not generate the necessary data as the process of collecting such data may inevitably add various data biases given the incentives among various stakeholders to influence road planning activities. [45] All of this made the case for an independent, data-driven and cost-effective solution to identify important routes in each block while respecting the democratic checks and balances already existing within the system.

4 TECHNICAL INTERVENTION

Among limitations identified in the above section, a key issue was the possibility of cherry-picking candidate roads leading to unfair competition within roads and the lack of independent road use or traffic data specific to our objectives for India's entire rural road network. If information on which roads could possibly be good candidates meeting the policy objectives were independently made available, we could atleast ensure that these were considered as candidates thereby leading to a "fair competition" [6]. In this manner, we would not change the existing system drastically but just add another source of recommendation at the stage of candidate road selection and keep the final scoring mechanism as it is.

Which roads make good candidates for PMGSY-III?

The objective of the policy is to upgrade existing rural roads on which large population depend on to access agricultural markets, educational institutions and health facilities. In pursuit of this, an algorithm named "Trace Maps" was developed. The core idea behind the algorithm is to simulate traffic from every rural habitation to its nearest points of interests or facilities. While a road may be used for different purposes, the policy exhaustively lists 23 kinds of destinations under its objectives such as notified agriculture markets, degree colleges, bedded hospitals, agro-industries, banks and more [38]. Trace Map identifies the shortest path from each habitation to the nearest facilities of each of the 23 kinds. Ideally, there will be 23 routes identified for each habitation. Once all the shortest routes have been identified, the algorithm iterates through each of the roads identified as part of various habitation-facility shortest routes and sums population relying on the road weighed by

[4]As per India's Local Government Directory (https://lgdirectory.gov.in/), India's 36 States/Union Territories are administratively divided into 736 districts, 7199 blocks and 6,62,543 villages.

[5]These are government civil engineers at the block level who are tasked with managing the design, construction and upkeep of rural roads under PMGSY. In our case, they form the layer interfacing directly with private contractors and the closest to the actual site of construction.

[6]Fair competition in this setting would mean that the roads considered as candidates are in fact representative of the best options in the block meeting our policy objectives

Stage	Description	Example
Inventory	Listing of all the roads, habitations and facilities in the block.	block A has 100 roads, 30 habitations and 45 rural facilities.
Identification of Candidate Roads	Identifying a list of routes which are important to meet the objectives of the program	Block Engineer identifies 10 important routes based on these 100 roads. A route can be a combination of 1 or more roads. Block Engineer also solicits recommendations from elected political representatives of the region and other local bodies
Prioritization	Sorting the roads based on their "Utility Values" and eliminating the roads with good pavement condition or those still under "Design Life" (warranty) .	These 10 routes are scored based on a formula which gives weightage to the population and facilities in the direct vicinity of the route. Two routes are eliminated because they currently are still in good pavement condition.
DPR Preparation	Preparing Detailed Project Report for the high priority roads including cost estimates and engineering plan.	The top scoring eligible routes of the block are picked till the target allocated to the block is saturated. Target allocated may be lesser than the total length of eligible roads.

Table 1: Existing Planning Process

the policy prescribed importance of the facility. The total weighted population benefiting from each road is then used to give *Trace Map Ranks* to each road in the block.

Algorithm 1: Trace Map | Calculating population benefiting for every road

Input : H is list of habitations in the Block, *FacilityListsByCatogory* is collection of facilities by category (Eg. Schools, Hospitals etc), *RoadNetwork* is the road network of the block
Output: *RoadMaster* is mapping of all roads in the block and the population depending on it

initialization;
foreach *habitation h in H* **do**
  population $p = h.population$;
  foreach *FacilityList in FacilityListsByCategory* **do**
    path = `FindShortestRoute`(*h, FacilityList, RoadNetwork*);
    foreach *road in path* **do**
      $RoadMaster[road] = RoadMaster[road] + p$
    end
  end
end

The algorithm operates at a block level ie considering roads, habitations and facilities from a single block at a time. Once the ranks have been identified, the tool generates an A1 sized PDF map which displays the entire road network, habitations and the facilities (Figure 1). The roads are colour coded from green to red and have varying thickness based on the population depending on them. The road network contains National Highways, State Highways and Major District Roads, but only the rural roads are ranked. The rest are formatted in shades of grey. There is a tabular representation of the top-15 roads with their name, length etc on the lower right side of the map. Basic elements of traditional maps such as legend, north arrow and scale are also included. Instructions on interpreting Trace Maps are re-iterated on the map itself. The tool was developed as a Python based QGIS [44] user script which has been tested to work on QGIS 2.18 with GRASS.

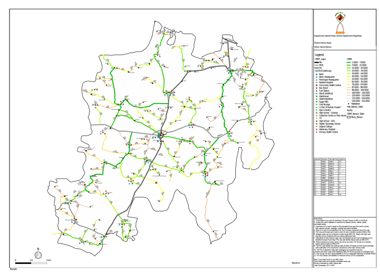

Figure 1: Trace Map for a block in Alwar District, Rajasthan

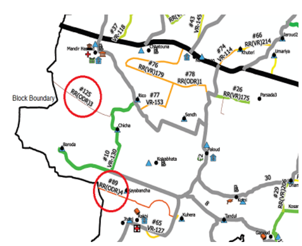

Figure 2: Example of an inter-block ODR which was not classified in a the Top 15 Trace Map ranks of the block

4.1 Usage

The Trace Map is designed to help the field engineers identify important routes during the candidate road selection process and also steer discussions with various stakeholders while soliciting their recommendations (Stage 2 in Table 1). Further, at least Top-15

Trace Map Rank roads need to be part of one or the other routes considered in the candidate road pool. The field engineers were free to choose any other road which they think is important based on feedback from elected representatives, citizens, administrators or based on the field knowledge. The scoring and prioritization mechanism is the same as PMGSY-II (Stage 3 in Table 1), but in this case the competition is assumed to be more *fair* as it includes the Top 15 Trace Map identified routes among other sources of recommendations such as from the elected representatives, field knowledge etc.

As of 20.5.2021, 6000+ Trace Maps had been prepared, uploaded and utilized for planning. This covers approximately 80 % blocks in rural India. As other states get onboarded to the programme, all the eligible blocks in rural India will be covered. Further as on 20.5.2021, more than 60,000 km of roads identified through the above mentioned planning process have been approved by the Indian government for upgradation.

An exhaustive evaluation of the entire process re-engineering is outside the scope of this paper. The Trace Map algorithm itself could be evaluated by conducting actual traffic surveys etc. But if we take the limited goal of Trace Maps to suggest competitive candidate road which have a high likelihood of being prioritized based on *Utility Values*, we can analyze the percentage of the final approved road proposals which have top-15 trace map rank roads as part of them. It should be noted that a candidate road may have a high utility value but still may not be proposed because of good existing surface condition, contractual warranty etc.

As on writing this paper, out of 8037 roads of length 62,658 km approved by the central government across 16 states, 66.2 % of approved road proposals contain one of the top-15 recommended trace map roads of their respective blocks and 24.4 % of road proposals contain at least one trace map road with rank between 15-50. The remaining 9.4 % proposals have roads with trace map ranks 50 or above.

5 DISCUSSION

The evaluation of the algorithm isn't center-piece to our research. The network analysis algorithm powering Trace Maps could be replaced with a more sophisticated algorithm than our "simplistic" customization of Dijkstra's algorithm but the challenges will more or less remain the same. An algorithmic intervention will go through multiple stages: Deliberation, Design, Development and Deployment. We examine how Trace Maps as an intervention progressed through various stages of production by adopting a sociotechnical frame[47] by focusing not only on aspects of data and algorithm but centered around the humans in the system and the social context in which they operate. We believe that challenges similar to the ones faced in our case will arise and need to be addressed whenever any algorithmic or data science projects are produced in or for governments.

5.1 Who are the humans in the loop?

Deliberation on humans in the loop is not limited to agency during the final output of an algorithmic decision making system or its design but also in its production. Questions of agency, power, hierarchy, status, and accessibility need to be accounted for in the full life cycle of a data science project from the origins of data collection to deployment and beyond. A common distinction in public sector settings in developing nations is limited state capacity and resource-constraints [29] [12]. Kshirsagar and Robinson et al [30] suggests examining the resource constraints ie compute and storage available with the partner organization during scoping of projects. We add to this and argue that resource constraints shouldn't be limited to the technical but in fact be centered around the human resources available within public sector settings. Government departments are likely to be understaffed, overburdened, under-paid [45][49][23] and at the same time function with limited in-house technical capacity or low tech-literacy. This is non-trivial to unpack and not easily visible both to external partners which are interested in working on AI for Social Good problem statements with the government or even government officials at higher levels of the hierarchy distant from the sites of implementation. Weak state capacity in our context has multiple ramifications while scoping algorithmic interventions for data scientists or policy makers. We explore themes arising from this in the following sections:

5.1.1 Counterproductive Computing.

An earlier generation of researchers have discussed how technological interventions such as e-governance solutions amplify existing inequalities [53]. The cost of deployed AI solutions is also getting [39][19] [36] its due attention. We further argue that algorithmic solutions can carry significant costs even before they are put to use, if at all. This isn't limited to direct harms of the algorithms or other technical intervention in question or limited to the bias of the algorithms, but the secondary and ripple effects [47] of burdening the existing state capacity further with interventions which adversely impact their ability to conduct their core responsibilities therefore impacting the proposed targeted outcomes and others. This we term as counterproductive computing, where the very act of introducing technology, irrespective of its eventual deployment, ends up indirectly and adversely impacting the very outcomes it proposes to improve and other allied outcomes unrelated to the intervention itself because of the added costs of using the technology itself. Dasgupta and Kapur [14] show how bureaucratic overload and resource scarcities force rural development officials to excessively multi-task and how inability to focus on managerial activities impact implementation of development schemes. We argue that algorithmic interventions carry high preparatory and deployment costs which can further overburden and distract government capacities and therefore need to be approached with caution. For example and hypothetically, a work-in-progress algorithmic solution to predict road surface deterioration may overburden the field engineers with its data collection requirements, continuous training workshops and erratic software bugs, so much so that the field engineer is no longer able to carry out his routine maintenance activities satisfactorily, thereby impacting the very outcomes the original solution intended to solve. By its nature, most data driven projects require structured and quality data for a larger set than just the target population before it can make an informed decision. Put simply, for a prediction model to identify at risk properties for fire hazards, you need information on a large number of properties both with and without fire hazards to be able to train the model. In our case, to be able to identify "important" roads, GIS data is required for all the roads in the block, every facility needs to be surveyed and only thereafter Trace Maps can be generated. The facility survey was conducted with the help of an android application by field engineers over months. This would mean covering all habitations in a block. Depending on the state, field engineers might be understaffed with many positions vacant or salaries unpaid. This exercise is in addition to their routine responsibilities of inspecting roads, monitoring construction, filing administrative reports etc. Even on the top, for the months preceding and following the launch of the program, the limited IT resources of the central team were singularly focused on building the IT infrastructure for this program which was admittedly more demanding than previous projects. This meant that at times we had to de-prioritize maintenance of existing IT systems pertaining to previous phases of the program. Even post-deployment, algorithm projects are costly to maintain [46]. A large share of IT resources continued to be deployed towards maintaining the newly developed IT modules. Burden of preparation might be higher in public sector projects in settings where the existing data collected over time is of lower quality and technical resources are limited. A common reaction to limited state capacity is to outsource key components of software design and delivery to private technical vendors. This approach has been followed globally [17] with mixed results. Capacity constraints of governments can also limit management of contracts with private vendors and leading to delayed projects or badly managed projects. Further, emerging research shows privatization of algorithmic projects can also shift accountability of key public policy decision out the purview of public sector accountability into the private where details can be shrouded under confidentiality and IP agreements [10]. Nonetheless, apart from limited capacity within the existing development team, the chocices are constrained by overburdened capacity across the bureaucracy, from upper-levels to the street. Further, many constraints are also dictated by legacy choices. Nonetheless, we suggest capacity constraints of the state should be a key consideration while thinking about such projects and is not limited to just the development of the algorithm.

Any data science project should begin with these questions: Will the preparation to the intervention and the intervention itself increase the burden on the state capacity? What's the cost on the system dis-aggregated by rungs of hierarchy? [7] Is the trade-off worthwhile and for whom? What policy/resource-planning [8] can be undertaken to mitigate or alleviate the added burden on the functionaries?

5.1.2 Data Inequities as proxy and perpetuating larger inequities. The public sector isn't a single monolith and within itself capacities vary across political and administrative boundaries, departments etc. Data science projects or algorithmic interventions may require the data to be structured and of certain quality and quantity to be useful. In our case, for the algorithm to function properly, the road network data needed to be digitised on GIS with reasonable quality and the facility survey completed on the ground had to be exhaustive. Each state independently prepared their GIS data by contracting it out to private vendors or publicly owned State Remote Sensing Centres. There is an apparent variance in the quality of the GIS data across the states which may be attributable to either their contract management, local remote sensing capacities etc. Functioning data systems are often a consequence of functioning operations, bureaucracy and availability of resources. If delivery of welfare is centered around algorithmic solutions which require resource intensive preparation of good quality data, it can delay delivery of welfare in these states. While, 13 states were on-boarded to PMGSY-III at roughly the same time, they all took varying times (4-12+ months) to complete prerequisite processes needed to get road projects approved. This included digitization of road network, ground survey of facilities, preparation of Trace Maps to preparing detailed design and cost reports for the selected roads. The variance in time can be in partially attributed to the technical capacity and data-readiness of the states among other things such administrative leadership, local elections etc. Just because the data is structured, it needn't mean its of good quality which meant even if certain states were able to eventually run the algorithm, the results were not always satisfactory [9]. Lower quality data systems also means that data from these states may not be used at the time of designing or training of algorithms thereby leading to models not functioning accurately in these sites when deployed eventually.

5.1.3 Participatory Design But With Whom? There is growing research on participatory design of algorithmic projects [31] and a call to include the perspectives of those impacted by the algorithms to design the systems. Through our case we highlight that identifying the actual participants in public sector setting may require careful attention. Often, especially with technical interventions in low-resource settings, tasks designed for certain roles might actually be carried out by someone else in reality. This is in particular true for IT related tasks. Existing government employees in senior positions might already be too overburdened or without necessary IT skills. In particular, IT tasks such as interfacing with Management Information Systems (MIS) may be delegated to Data Entry Operators or Computer Operators[10] or junior/young staff members. This unofficial but common delegation of duties has multiple ramifications while designing algorithmic interventions. Not only does it draw attention to careful identification of participants for user-study and training, it also questions routine imaginaries of decision makers interacting with computers to make decisions. The use of dashboards in governance to convey information is increasing [50][9] [34], but if the staff member with the authority to decide doesn't interact with the computer, what use then is an online dashboard carefully designed to convey the output of sophisticated algorithms? In our case, we saw in many states that the planning related components (apps, QGIS, MIS) were delegated to junior engineers whereas when we conducted in person training workshops or user interviews, only the senior engineers were invited. [11]

5.1.4 Paper Realities. One of the reasons for the delegation of responsibilities by the senior staff to others is their lack of comfort

[7] Some may see an increase in convenience at the expense of others. Often the decision makers on top might see an increase in convenience but at the cost of field or downstream functionaries.

[8] For e.g. We had taken steps to expand the central IT team pre-emptively and provided funds to state departments to setup GIS units with necessary personnel and hardware.

[9] Eg. A single missing road could drastically impact the result of the eventual network analysis

[10] An entry level contractual position specially for entering offline data into spreadsheets or online information systems

[11] Interestingly with COVID cases surging in 2020, our trainings shifted online and with virtually no limit to participation, we noticed larger participation engineers that actually required training

with IT and the nature of their professional duties requiring them to be on the road frequently. Ackermann et al suggest reducing the cost of usage of data science interventions by designing interfaces which appear familiar and comfortable to the end-users [4]. We extend that familiarity may not be limited to the existing IT interfaces and that the culture and social practices of working within the government can be borrowed in designing the processes or interfaces of interaction. *Paper* as a medium holds a predominant position in bureaucracy in India [20] and governance across the world [21] [26] [25]. It continues to persist within e-governance and digital imaginaries [33] [16] [51]. We argue that affordances of paper as a medium can explicitly be utilized to negotiate space for more sophisticated forms analyses and algorithms into day to day governance. In our case, a major decision we made early on was that we wanted the Trace Map to be a printable paper by default. The alternative would have been to create interactive maps hosted online on a website which admittedly would embed significantly more information but wouldn't have had the intended adoption. Having a printed paper map which one can sign, carry around and show to different stakeholders has advantages. Specific design elements were added to the printable map such as blank space for signing and approving the map and departmental logos to officiate and build trust towards the document. It bridges the gap from the world of the algorithm to the familiar reality of paper and officialdom which allowed the algorithm to assimilate in pre-existing practices.

5.1.5 Re-alignment of Responsibilities and Power. Earlier we argued that there may be a difference in the actual and the intended users of technological interventions in our setting. Extending the argument, we warn that the technological intervention may in itself further rearrange roles, responsibilities and hierarchies within the setting it is deployed. [28] [13]. This may in some cases can be an opportunity for re-engineering entrenched forms of bureaucracies or power, but also could also end up centralizing power in pockets which can be detrimental. Prior to introduction of Trace Maps, the field engineers relied on basic maps without the additional colour-coding of roads by estimated importance. These maps were prepared using CAD software and local consultants/vendors in the districts. It makes sense for the map making activity to be decentralized and closer to the field as this would ensure more accurate maps are built adapted to local needs. In PMGSY-III, generation of Trace Maps required knowledge of QGIS and close coordination with the IT team at the national level, hence it was decided that maps will be generated at the state headquarters instead and then shared with the field engineers. This did not mean the field engineers were not involved in the preparation, instead they would travel to state headquarters and spend days there ensuring the GIS data was correct, no roads were being missed and if so would get them added to the shape-files before regenerating the Trace Maps. Certain states also outsourced large parts of the preparatory process to consultants who may not have the same local contexts or accountability as the field engineers staffed within the blocks. Conversely, shifting of power and roles can be useful as well. Preparation of Trace Maps and system-generated and tamper-proof priority lists allowed field engineers some evidence backed negotiation space against powerful but not always policy-aligned road recommendations by special interests.

5.2 Inheriting Legacy Data and Decisions

Suresh and Guttag offers a rich taxonomy to investigate various forms of biases in data [52]. In our case, we highlight how legacy data and decisions compound and impact even newly generated datasets. The programme's existing MIS system already contained a database of all the rural habitations in India. The database was created for PMGSY-I in early 2000s with the objective of connecting unconnected habitations. The original habitation data contained only rural habitations and did not include municipal towns/cities because the policy's objective was to provide road connectivity to unconnected rural habitations only. This legacy decision percolated years later into newly generated datasets (GIS and Facilities) as well. The mobile application developed for PMGSY-III in 2018 to capture socio-economic points of interest populated the legacy habitation list for the surveyor in the app. This meant that important facilities which existed in nearby urban towns and accessed by rural habitations ended up not being recorded. Consequently, rural roads leading from villages to these facilities located in nearby towns would therefore not be accurately scored by the Trace Map algorithm. We argue that the use of data collected for a different purpose than its original use will be a common characteristic of resource constrained public-sector data science projects and biases emanating from these legacy decisions need to be examined carefully.

5.3 Constrained Choices and Hidden Parameters

Even in *simple* algorithms which don't have hyper-parameters in the machine learning sense, there exist many seemingly innocuous, implicit or explicit decisions, abstractions or assumptions which can impact the fairness of the system. These abstractions or decisions are often imperfect and can be a shaped by circumstance, accidents of opportunity, access to data, discretion or creativity [47] [42]. In our case, these decisions are a product of legacy data systems, bureaucracy or of functioning within limited technical capacity and resource constraints in the public sector. The exact parameters will be different across data science projects, but the reasoning behind why and when these decision seep in or fail may be similar and informative. In most of the cases listed below, the places where our assumptions fail are also places which are in the regional peripheries or margins and are likely to be already under-developed or under-served by the government [6]. A list of assumptions taken consciously or unconsciously while designing the algorithm are listed below:

5.3.1 People Don't Access Facilities From Different blocks. The existing policy dictated that the selection of road proposals based on competition was to be conducted for each block separately. This was done to ensure roads from every block had a chance to be picked under the program. In most cases, splitting of responsibilities for rural roads amongst engineers is done based on block boundaries. It flows, that the Trace Maps will also be generated block-wise ie the Trace Map ranks calculated were relative to roads within each block. Therefore, the QGIS plugin for Trace Map took as inputs the rural road network, habitations and facilities belonging for each block separately. This meant that for every habitation, the algorithm sought the nearest facilities inside the block itself. It could be that for certain habitations the nearest hospital or agricultural markets are actually across the block boundary. This meant that for certain inter-block roads the estimated importance in Trace Map might be lower than reality. See in Figure 2 where roads going across block boundary are ranked poorly by the Trace Map. This could potentially be rectified by additionally considering facilities situated near the block boundary from adjoining blocks.

But this would mean considerably changing the flow and modules of the online MIS which were developed such that blocks could complete processes independent of each other and not be held back because of non-performing field engineers from other blocks who had not completed digitization and other preparatory activities. Instead, the field engineers were repeatedly told to consider important inter-block roads and reminded that even the Trace Map rank of such roads is less, then they should consider these roads if they are important in reality.

5.3.2 People Prefer Accessing Their Nearest Facilities: The core algorithm assumes that the inhabitants of a habitation prefer to go to the nearest available facility whenever required. Eg. If there are more than one high schools in the block, the algorithm assumes that people prefer to go their nearest high school only. This is a very simplified notion of mobility and access. Even if the nearest destinations were preferred, shortest routes are not always the preferred routes. Preferences can change based on seasons or even time of the day. "Walking cannot be reduced to going from point a to point b. It's a social activity, embedded in cultural codes and practices" [15].

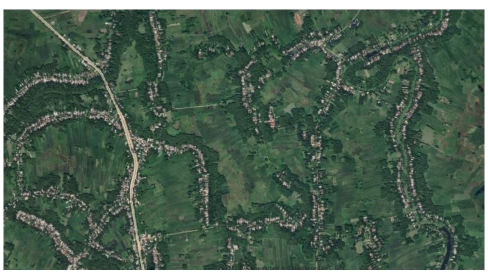

Figure 3: Linear Settlements in the Indian state of Assam

Figure 4: Point Settlements in the Indian state of Madhya Pradesh

5.3.3 Habitations can be Represented as Points. The GIS data represented habitations as points as compared to polygons. This is a reasonable decision given the scale of the maps and the relatively small population of rural habitations (Figure 4). Although, a legacy decision, this worked for Trace Maps as well because most origin-destination algorithms function point to point. The set of states that were onboarded early on to PMGSY-III were largely plain states and by the time other states in India started getting onboarded we saw our assumption breaking in specific instances. For eg. settlement patterns in Assam [12] were strikingly different. Houses in a settlement were not clustered together as point or in a circular manner, but instead were found to be thinly spread adjacent to roads and often habitations with small populations stretched across for kilometers (Figure 3). Simple patterns were also found in "plain-states" such as Rajasthan, where water was scarce and people within the same village scattered their houses across a large area. While the settlements were spread, a single point was chosen at the time of digitization of these habitations. The ramifications of the breaking of this assumption on the Trace Maps algorithms and consequently the final selection of roads needs to be examined further. But it is clear from our examples that it broke down in regions and for people that were often already marginalized, under-served and in the peripheries.

5.3.4 Facilities are located within Habitations. As mentioned in previous section, while collecting geo-tagged pictures of rural facilities, we also tagged the habitations these facilities belonged to. We decided that we'll use the lat-long of the habitations from the earlier digitization exercise as the location of the facilities instead of the actual lat-long of the facilities as captured during the mobile survey. It was assumed that the facilities would be within the habitation itself and that the rural habitations weren't that large in size to make a difference. This design decision was also influenced by the possibility that the mobile based survey would be potentially dropped because of the increased burden on field engineers and using lat-long of habitations could allow for a desk-based survey of facilities too. While this decision made sense in majority of the areas especially in plain geographies, this assumption breaks down in hilly terrains where facilities administratively belonging to certain habitations were actually constructed kilometres away because of spatial constraints. Shifting these facilities to the point mark of the habitation could adversely impact the output of the network analysis.

5.4 Naming and Terminologies - Ranks, Scores and Priorities

The policy already included terminology such as "CUCPL Ranks", "Utility Values" and "Population Benefitted". Trace Maps which were introduced in PMGSY-III, brought with itself "Trace Map Ranks" and "Trace Map Score". In our conversations with field engineers as part of our work, confusion between Trace Map ranks (ranks based on simulated traffic) and CUCPL Ranks (the final ranks based on *Utility Values*) was a common occurrence. In some cases, the field engineers would even preemptively start preparing the design specifications of high ranked trace map roads as if these roads were confirmed to

[12]A state in the north-east of India where the Brahmaputra river criss-crosses a large part of its terrain

be proposed. An alternative solution could have been that instead of ranking roads in Trace Map, we could have divided these roads into categories such high, medium, low use roads. These would mean losing a lot of variance but at the same time helpful because given all the assumptions taken in the algorithm, the individual ranks wouldn't be as precise in the first place and would only work generally. But, having ordinal ranks allows for more detailed analysis such as "Trace Map Cuts" which became routine in high level presentations or for sorting of proposals based on their trace map ranks to prioritise internal audits. While designing algorithms for public policy settings, we need to ensure that existing and entrenched definitions and terminologies prevalent in public policy are not confused with similar terminologies emanating from data science and algorithmic projects which can sometimes be very similar.

6 LIMITATIONS

The different positionalities of the authors are useful in getting diverse perspectives and view-points among central team, the state team and the technological service provider. Nonetheless, view-points of street bureaucrats, elected representatives, beneficiaries are not directly represented. During the design and implementation of the algorithm, the authors had multiple direct conversations with different stakeholders mentioned above, but our positionality and the hierarchical nature of government may have biased the feedback we received from stakeholders lower in the bureaucratic hierarchy. Further, deployment of algorithmic projects at other public sector sites may not engender similar issues or concerns as faced in this particular project. The site i.e. the Ministry of Rural Development and within that the rural roads department has its own characteristics which may not directly map to other government verticals such as health or education in India or even globally. The Ministry of Rural Development ranks in the top three departments at the central government level under the Data and Governance Quality Index issued annually by the government's internal think-tank, NITI Aayog [35]. It has relatively better data and IT preparedness and has historically invested in MIS earlier than other departments in India. Further, the Ministry has a huge cadre of on-ground functionaries (engineers, community resource persons etc). These characteristics create space for algorithmic interventions but at the same time dictate many of our conclusions or assessments. Nonetheless, characteristics such as low-state capacity, poor data quality, algorithmic abstractions etc are common features in resource constrained public sector settings. Even within the same site, it is not necessary that all algorithmic projects will invoke similar concerns. For example, under the same government program, once the roads have been finalized, the district engineers submit detailed project reports regarding the estimated cost and design specifications of the proposed roads. The includes geo-tagged photographs of the existing road as proof of its poor condition. These pictures are then randomly audited by the central engineering team to identify road proposals which are in relatively good condition but still being proposed for re-construction. This process is now replaced with a AI model which scans through all the images and identifies a shortlist for manual vetting. Many of the concerns raised by the authors in the primary case don't necessarily apply to this intervention. The pictures were already being uploaded prior to the intervention, geo-tagged pictures of road clicked by smartphones don't have a large variance in picture quality and the fact that paved road surfaces across the country look similar. Infact, this project reduces the administrative burden of the central team while having no increase of burden downstream. Hence, further taxonomy or classification is required to understand when and in what circumstances do many of the issues faced by us in our main case can re-appear in other public sector projects.

7 CONCLUSION

The paper attempts to provide an detailed account of implementing an algorithmic intervention in a particular public sector setting. There are few detailed case-studies of public sector algorithms being utilized in a significant manner at such scale in India. The embedded culture of functioning within government, legacy IT/data systems and resource and capacity constraints shape the problem formulation, design, development and deployment of algorithms implicitly or explicitly. With the help of our case, we begin by making the case that some data science projects can be counterproductive, the cost inflicted on an overburdened capacity can outweigh any potential benefits. We bring forth the capacity constraints within which public sector staff operate and how it can impact the production of data science at various stages. If algorithmic solutions are to be employed, we highlight the importance of effectively integrating algorithmic solutions with existing sociotechnical systems and practices by careful placement within policy, integration with IT systems and using old media technologies to negotiate space for algorithms within the public sector. We argue that even in seemingly "simple" algorithms, there exist implicit and explicit assumptions and abstractions produced by imperfect circumstances which may adversely impact people already in the peripheries of the system. While some of our themes may be unique to our particular case and site, others are found in public sector sites commonly. Understanding of the site of production of data science in the public sector can help the research community better understand, regulate and mitigate the risks involved in the use of algorithms in high-stake public policy use-cases. Our future goal is to quantify the impact of our design decisions on various edge-cases and also actively understand the experiences of the various stakeholders downstream such as elected representatives, field engineers and state officials as they navigated our algorithmic intervention.

ACKNOWLEDGMENTS

Interventions of such scale are impossible without the countless many unnamed individuals involved in preparing the stage for the algorithm to do its bit. This includes many who may or may not agree with the overall objective of the intervention or remain oblivious to the lofty prose narrated in the paper but nonetheless dutifully ensured the objectives were achieved. This work would not have been possible without the field engineers who for months tirelessly surveyed every road and habitation to create the initial inventory. State IT officers who in a matter of days aced the use of QGIS, when Microsoft Excel was as far as they had ventured before. Free and open source technologies such as QGIS and GRASS allowed this intervention to be possible at this scale with our limited resources.

Lastly, Aakash Solanki, for encouragement and engagement, copy editing support, and exposing to the praxis of anthropology.

REFERENCES

[1] 2020. *Machine Learning for Developing World (ML4D) Workshop, NuerIPS 2020.*
[2] Rediet Abebe and Kira Goldner. 2016. *Mechanism Design for Social Good.* http://md4sg.com
[3] Rediet Abebe and Kira Goldner. 2018. Mechanism design for social good. *AI Matters* 4, 3 (2018), 27–34.
[4] Klaus Ackermann, Joe Walsh, Adolfo De Unánue, Hareem Naveed, Andrea Navarrete Rivera, Sun-Joo Lee, Jason Bennett, Michael Defoe, Crystal Cody, Lauren Haynes, et al. 2018. Deploying machine learning models for public policy: A framework. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 15–22.
[5] Shilpa Aggarwal. 2018. Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics* 133 (2018), 375–395.
[6] Sam Asher, Karan Nagpal, and Paul Novosad. 2018. The Cost of Distance: Geography and Governance in Rural India. *World Bank Working Paper* (2018).
[7] Sam Asher and Paul Novosad. 2016. Market access and structural transformation: Evidence from rural roads in India. *Manuscript: Department of Economics, University of Oxford* (2016).
[8] Sam Asher and Paul Novosad. 2018. *Rural roads and local economic development.* The World Bank.
[9] Jamie Bartlett and Nathaniel Tkacz. 2017. Governance by dashboard: A policy paper. (2017).
[10] Hannah Bloch-Wehba. 2021. *Transparency's AI Problem.* https://knightcolumbia.org/content/transparencys-ai-problem
[11] Samuel Carton, Jennifer Helsby, Kenneth Joseph, Ayesha Mahmud, Youngsoo Park, Joe Walsh, Crystal Cody, CPT Estella Patterson, Lauren Haynes, and Rayid Ghani. 2016. Identifying police officers at risk of adverse events. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.* 67–76.
[12] The Hindu Business Line CP Chandrasekhar Jayati Ghosh. 2019. *The withering trend of public employment in India.* https://www.thehindubusinessline.com/opinion/columns/the-withering-tren-of-public-employment-in-india/article28750003.ece
[13] Kate Crawford. 2021. *The Atlas of AI.* Yale University Press.
[14] Aditya Dasgupta and Devesh Kapur. 2020. The political economy of bureaucratic overload: Evidence from rural development officials in India. *American Political Science Review* 114, 4 (2020), 1316–1334.
[15] Jonathan P Demenge. 2013. The road to Lingshed: Manufactured isolation and experienced mobility in Ladakh. *HIMALAYA, the Journal of the Association for Nepal and Himalayan Studies* 32, 1 (2013), 14.
[16] Richard H Dykstra, Joan S Ash, Emily Campbell, Dean F Sittig, Ken Guappone, James Carpenter, Joshua Richardson, Adam Wright, and Carmit McMullen. 2009. Persistent paper: the myth of "going paperless". In *AMIA Annual Symposium Proceedings*, Vol. 2009. American Medical Informatics Association, 158.
[17] David Freeman Engstrom, Daniel E Ho, Catherine M Sharkey, and Mariano-Florentino Cuéllar. 2020. Government by algorithm: Artificial intelligence in federal administrative agencies. *NYU School of Law, Public Law Research Paper* 20-54 (2020).
[18] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency.* PMLR, 160–171.
[19] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.
[20] Megan Finn, Janaki Srinivasan, and Rajesh Veeraraghavan. 2014. Seeing with Paper: Government Documents and Material Participation. In *2014 47th Hawaii International Conference on System Sciences.* 1515–1524. https://doi.org/10.1109/HICSS.2014.195
[21] Ishita Ghosh, Jay Chen, Joy Ming, and Azza Abouzied. 2015. The Persistence of Paper: A Case Study in Microfinance from Ghana. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development* (Singapore, Singapore) *(ICTD '15).* Association for Computing Machinery, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/2737856.2738029
[22] Greg Hearn and Marcus Foth. 2005. Action research in the design of new media and ICT systems. *Topical issues in communications and media research* (2005), 79–94.
[23] The Hindu. 2020. *Over 6.83 lakh vacant posts in central govt departments: Personnel Ministry.* https://www.thehindu.com/news/national/over-683-lakh-vacant-posts-in-central-govt-departments-personnel-ministry/article30741372.ece
[24] Aspen Hopkins and Serena Booth. 2021. Machine Learning Practices Outside Big Tech: How Resource Constraints Challenge Responsible Development. (2021).
[25] Matthew S Hull. 2012. Documents and bureaucracy. *Annual review of anthropology* 41 (2012), 251–267.
[26] Matthew S Hull. 2012. *Government of paper: The materiality of bureaucracy in urban Pakistan.* Univ of California Press.
[27] Surya Mattu Julia Angwin, Jeff Larson and ProPublica Lauren Kirchner. 2016. *Machine Bias.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[28] Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature* 583 (07 2020), 169–169. https://doi.org/10.1038/d41586-020-02003-2
[29] Devesh Kapur. 2020. Why Does the Indian State Both Fail and Succeed? *Journal of Economic Perspectives* 34, 1 (2020), 31–54.
[30] Meghana Kshirsagar, Caleb Robinson, Siyu Yang, Shahrzad Gholami, Ivan Klyuzhin, Sumit Mukherjee, Md Nasir, Anthony Ortiz, Felipe Oviedo, Darren Tanner, et al. 2021. Becoming Good at AI for Good. *arXiv preprint arXiv:2104.11757* (2021).
[31] Bogdan Kulynych, David Madras, Smitha Milli, Inioluwa Deborah Raji, Angela Zhou, and Richard Zemel. 2020. Participatory Approaches to Machine Learning. International Conference on Machine Learning Workshop.
[32] Avishek Kumar, Syed Ali Asad Rizvi, Benjamin Brooks, R Ali Vanderveld, Kevin H Wilson, Chad Kenney, Sam Edelstein, Adria Finch, Andrew Maxwell, Joe Zuckerbraun, et al. 2018. Using machine learning to assess the risk of and prevent water main breaks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 472–480.
[33] Megh Marathe and Priyank Chandra. 2020. Officers Never Type: Examining the Persistence of Paper in e-Governance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20).* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376216
[34] Shannon Mattern. 2015. Mission control: A history of the urban dashboard. *Places Journal* (2015).
[35] Development Monitoring and Government of India Evaluation Office, NITI Aayog. 2020. *DATA GOVERNANCE QUALITY INDEX (DGQI).* https://dmeo.gov.in/sites/default/files/2021-03/DGQI_Toolkit.pdf
[36] Safiya Umoja Noble. 2018. *Algorithms of oppression: How search engines reinforce racism.* nyu Press.
[37] Ministry of Rural Development. 2015. *PMGSY-II Programme Guidelines.* http://pmgsy.nic.in/PMGSY_E_J_2015.pdf
[38] Ministry of Rural Development. 2019. *PMGSY-III Programme Guidelines.* http://pmgsy.nic.in/PMGSY3gl.pdf
[39] Cathy O'Neil. 2008. Weapons of Math Destruction: How Big Data Increases Inequality and. (2008).
[40] Marion Oswald, Jamie Grace, Sheena Urwin, and Geoffrey C Barnes. 2018. Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental'proportionality. *Information & Communications Technology Law* 27, 2 (2018), 223–250.
[41] Samir Passi and Solon Barocas. 2019. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* 39–48.
[42] Samir Passi and Steven Jackson. 2017. Data vision: Learning to see through algorithmic abstraction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.* 2436–2447.
[43] Samir Passi and Steven J Jackson. 2018. Trust in data science: Collaboration, translation, and accountability in corporate data science projects. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.
[44] QGIS Development Team. 2021. *QGIS Geographic Information System.* QGIS Association. https://www.qgis.org
[45] Naresh Chandra Saxena. 2019. *What Ails the IAS and Why It Fails to Deliver: An Insider's View.* SAGE Publications India.
[46] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. 2015. Hidden technical debt in machine learning systems. *Advances in neural information processing systems* 28 (2015), 2503–2511.
[47] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency.* 59–68.
[48] Zheyuan Ryan Shi, Claire Wang, and Fei Fang. 2020. Artificial intelligence for social good: A survey. *arXiv preprint arXiv:2001.01818* (2020).
[49] The Times of India Siddharth Tiwari. 2021. *Asha workers to go on strike today over salary.* http://timesofindia.indiatimes.com/articleshow/82891204.cms
[50] Aakash Solanki. 2019. Management of Performance and Performance of Management: Getting to Work on Time in the Indian Bureaucracy. *South Asia: Journal of South Asian Studies* 42, 3 (2019), 588–605.
[51] Aakash Solanki. 2020. *13. UNTIDY DATA: SPREADSHEET PRACTICES IN THE INDIAN BUREAUCRACY.* Lives of Data, Institute of Network Cultures.
[52] Harini Suresh and John V Guttag. 2019. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002* (2019).
[53] Kentaro Toyama. 2015. *Geek heresy: Rescuing social change from the cult of technology.*

[54] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.