Artificial Intelligence Policy in India: A framework for engaging the limits of data-driven decision-making

Vidushi Marda[1]

Abstract

Artificial Intelligence (AI) is an emerging focus area of policy development in India. The country's regional influence, burgeoning AI industry, and ambitious governmental initiatives around AI makes it an important jurisdiction to consider, regardless of where the reader of this article lives. Even as existing policy processes intend to encourage the rapid development of AI for economic growth and social good, an overarching trend persists in India, and several other jurisdictions: the limitations and risks of data-driven decisions still feature as retrospective considerations for development and deployment of AI applications.

This article argues that the technical limitations of AI systems should be reckoned with at the time of developing policy, and the societal and ethical concerns that arise due to such limitations should be used to *inform* what policy processes aspire to achieve. It proposes a framework for such deliberation to occur, by analysing the three main stages of bringing machine learning (the most popular subset of AI techniques) to deployment - the data, model, and application stage. It is written against the backdrop of India's current AI policy landscape, and applies the proposed framework to ongoing sectoral challenges in India. With a view to influence existing policy deliberation in the country, it focuses on potential risks that arise from data-driven decisions in general, and in the Indian context in particular.

Keywords: Artificial Intelligence, Algorithms, India, Ethics, Machine Learning, Human Rights, Policy

Introduction

Over half a century after being coined, the term Artificial Intelligence (AI) is in vogue. AI technologies determine the information we consume on social media [1], determine risk

[1] Lawyer, and Digital Programme Officer at ARTICLE 19. Contact: vidushimarda@gmail.com or vidushi@article19.org. I would like to thank Corinne Cath, Mallory Knodel, Aparna Ashok, Ajit Pillai, Harsh Gupta, Srinivas Kodali, Gautam Bhatia and one anonymous reviewer for their insightful and helpful comments.

assessments of defendants in criminal sentencing [2], decide credit-worthiness of individuals [3], and even suggest the most appropriate route for one to take on their way home from work [4]. Most AI systems train on historical data, and are capable of uncovering patterns, learning from examples, and predicting future outcomes for the purposes of decision-making. These predictions and classifications are generalizations based on large datasets that humans wouldn't be able to analyse at similar speed and scale. The impact of AI is believed to be so transformative, that it has been referred to as the "new electricity"[5]. As industry and governments move towards developing and deploying these technologies, their societal and ethical implications have also come into focus.

This article proposes a framework for understanding the implications of AI, by focussing on the three main stages of bringing machine learning (the most popular subset of AI techniques) to deployment - the data, model, and application stage. I do this against the background of AI policy in India, and argue that the social, ethical and technical limitations of data-driven decision-making should form a fundamental consideration in AI policy development. To facilitate this, I also apply the framework to sectoral challenges identified in policy making processes in India. In short, I focus on potential limitations and risks that arise from data-driven decisions in general, and in the Indian context in particular. The scope and approach of this article are relevant for three reasons.

First, it offers an alternative to current policy approaches of addressing social and ethical issues which are presently grouped under the umbrella of "challenges to adoption"[6]. The existing approach is both short sighted and counter-productive. AI applications operate in societies that are chaotic, biased, unequal, and steeped in historical discrimination and disparity. To treat these important social realities as retrospective considerations, add-on features, or even as bugs to be fixed means the foundation upon which these systems are built, are inherently fragile. I demonstrate the social and ethical considerations that must be reckoned with in the *process* of building AI systems and policy, and offer a framework for the same.

Second, it invites a cross-disciplinary discussion on AI policy in India. A fundamental challenge in this space thus far has been to facilitate a conversation based on shared language and understanding between stakeholders and across sectors. By proposing this framework, I attempt to articulate policy concerns in technical terms, and vice versa.

Finally, it widens international discussions around AI, ethics, policy, and law, which are by and large currently premised on Western contexts. AI systems that we interact with today are much more than simple mathematical problems: they are socio-technical systems that depend on the contextual setting in which they function. India is an important jurisdiction to consider for a number of reasons. Its sheer size and burgeoning AI industry make it an influential power. The

Indian government's prioritisation of emerging technologies within the digital economy means that AI policy will evolve and develop rapidly in the next few years. The country is home to the world's largest biometric identity project, *Aadhaar*, which, depending on how it is used, can form a focal point for AI applications in the country. India is also at a critical point in the development of data protection regulation, which will have a profound effect on how AI technologies can and will function within it.

While the introduction discusses the problem this article intends to address, the next section will discuss its intended scope, and offer explanations to technical concepts used in the article for definitional and conceptual clarity. With an aim to provide contextual background for thinking through the proposed framework, it will then discuss the current state of India's AI policy landscape, before walking the reader through each stage of the framework. Finally, this framework is applied to sectors that are currently considered in policy making processes in India. The paper will end with conclusions and reflections.

Scope

The term 'AI' is susceptible to a plethora of interpretations. Given the definitional challenges in this area, I offer explanations around a few technical terms in this section, and define their intended scope within this article.

- An *algorithm* is a set of 'encoded procedures for transforming input data into desired output, based on specific calculations'[7]. It is a precise, step-by-step procedure that requires, in and of itself, no human intuition or guesswork [8].

- *Artificial Intelligence* refers to the ability of computers to exhibit intelligent behaviour [9]. AI is broadly of three types. The first, is artificial superintelligence, which refers to the idea that computers can surpass human intelligence, social skills, and scientific knowledge across domains. The second, is artificial general intelligence. This refers to the goal of computers exhibiting intelligence across multiple domains, to be at the very least, at par with human intelligence. Both these categories of AI are, as some experts have stated, 'logically possible and utterly implausible' [10]. The third is artificial narrow intelligence (ANI), which refers to the ability of computers to exhibit human capabilities in narrow, deliberate domains. *For the purposes of this article, the term AI will be used to denote ANI alone.*

- *Machine learning (ML)* refers to the most successful and popular subset of AI techniques today, which demonstrates the ability of a system to improve performance of a task over time [11]. Much of what is referred to as 'AI' today, is actually speaking of machine learning. Arthur Samuel referred to machine learning programs as those which have 'the ability to learn without being explicitly programmed.'[12]

- Data is key for most applications of machine learning. Developers of machine learning programmes train combinations of these algorithms on what is known as *training data*. What emerges through training is the *model*. The process of machine learning is iterative, i.e. it involves a living process of analysing these models, and adjusting training data and learning algorithms in order to optimise for success, which is defined and determined by developers.

- As mentioned above, machine learning techniques depend on directions of what to look for. A key human element in machine learning is *feature engineering*, i.e. deciding on what attributes the machine learning algorithm will train. This is in fact a much more challenging task for developers vis-à-vis conducting the actual act of learning. While there are increasing efforts to reduce human intervention in feature selection, this is unlikely [13].

State of the Field : India's AI Policy Landscape today

AI as an emerging priority area

The development, adoption and promotion of AI has been visibly high on the list of priorities of the Indian Government, an approach that rests on the premise that AI has the potential to make lives easier and make society more equal [14]. The Union government in 2018 allocated substantial funding towards research, training and skilling in emerging technologies like AI, a 100% increase from previous investment [15].

This prioritisation of digital technology is hardly new. The Union Government's Digital India initiative is aimed at transforming India into a 'digitally empowered society and knowledge economy' [16]. Digital India envisages providing digital infrastructure as a core utility to every citizen, incorporating such digitisation in governance and ultimately leading to empowerment of citizens [17]. The increase in funding towards research, training and skilling in emerging technologies such as AI is carried out under the umbrella of the Digital India programme. The Government has also begun to work towards ensuring that AI technology is made in India, and made to work for India as well, fitting squarely within its Make In India programme [18], a government initiative to promote India as a global manufacturing hub.

While AI has featured as an important consideration in digital technologies broadly, a number of initiatives focussed solely on AI have also emerged. This section will offer an analysis of certain salient features of each initiative in the scope of this article, and does not intend to be an exhaustive analysis of each.

Artificial Intelligence Task Force

The Union Ministry of Commerce and Industry set up an Artificial Intelligence Task Force in August 2017 with a view to *'embed AI in our Economic, Political and Legal thought processes so that there is systemic capability to support the goal of India becoming one of the leaders of AI-rich economies'* [19]. With the overarching view of AI being a socio-economic problem solver at scale, their March 2018 report identified ten sectors of relevance for AI in India. These include manufacturing, financial technology or FinTech, agriculture, health, technology for the differently abled, national security, environment, public utility services, retail and customer relationships, and education. The report specifically sought to understand what the role of the government should be, and how AI can solve problems at scale. It recommends, *inter alia,* the establishment of a nodal agency, the National Artificial Intelligence Mission that would coordinate AI-related activities in India.

While the report lists out enabling factors for the widespread adoption of AI and maps out specific government agencies and ministries that could promote such growth, it fails to meaningfully address the ethical, social, and technical limitations that undergird the use of AI technologies. Even in the few instances where the report considers privacy and data protection, it does not go far enough to address problems around data that are unique to AI. For instance, while discussing the issue of ethics and social safety, the Task Force acknowledges the challenges of data sharing and access to data by third parties. While it touches upon concerns of data privacy, it fails to consider the proclivity of data-driven decision-making to embed and exacerbate historical bias and discrimination. The potential for even well-intentioned algorithmic systems to have disproportionate consequences on vulnerable and marginalised communities is also left unconsidered.

This omission is seen in the Task Force's sectoral analysis as well. For example, the challenges identified in the FinTech sector include anticipation of market demand and balancing scale and innovation. Problems related to how the proliferation of FinTech will play out for people on the margins of data collection and technical inclusion are left wholly unconsidered. Perhaps most worryingly, the impact of AI technologies on the exercise of fundamental freedoms was ignored, and the use of AI for ‘autonomous surveillance and combat systems” was listed without qualification of the grave consequences such technologies have on privacy and freedom of expression.

The Task Force’s work intends to shed light on the direction in which AI policy in India should develop. Its focus on accessibility technology is perhaps its biggest strength. However, the lack of legal, policy, and civil society engagement within this process is made apparent in the cursory (at best) ethical and social analysis of India’s AI landscape.

Ministry of Electronics and Information Technology

The Union Ministry of Electronics and Information Technology has also begun to focus on AI. In February 2018, it set up four committees to prepare a roadmap for a national AI programme [20]. The four committees are presently studying AI in context of citizen centric services; data platforms; skilling, reskilling and R&D; and legal, regulatory and cybersecurity perspectives. The Committees reports remain unpublished at the time of writing this article.

NITI Aayog’s National Strategy for Artificial Intelligence: #AIFORALL

The National Institution for Transforming India (also known as ‘NITI Aayog’) , a government-run think tank, has been tasked with producing a national artificial intelligence policy to direct the government’s AI efforts [21]. In an effort to boost economic productivity in India, NITI Aayog teamed up with Google in early May 2018 to train and incubate startups that look to develop and integrate AI-based solutions in their business models [22]. NITI Aayog also entered into a statement of intent with ABB India to ‘make key sectors of Indian economy ready for a digitalized future and realize the potential of AI, big data and connectivity’ in late May 2018 [23].

In a discussion paper released in June 2018 [6], NITI Aayog states the overarching goal for a national AI strategy as one which will, *“leverage AI for economic growth, social development and inclusive growth, and finally as a “Garage” for emerging and developing economies.”* NITI Aayog’s role transcends just recommending a policy approach, its involvement includes implementation and deployment as well.

The National Strategy goes a step further than any other AI policy process in two distinct ways. First, it acknowledges that AI adoption has been largely commercially driven to date, and recognises the *“need to strike a balance between narrow definitions of financial impact and the greater good.”* Second, it recognises that AI applications should be embraced for their incremental, rather than purported transformational value in various sectors.

Despite these encouraging shifts in perspective, the substantive recommendations and analysis of India's national strategy on AI leave much to be desired. The report identifies five primary sectors where AI could have a positive social impact that require the government to play a leading role: education, agriculture, healthcare, smart cities & infrastructure, and smart mobility & transportation.

While discussing smart cities, the report promotes the use of AI for surveillance applications. These include AI systems that predict crowd behavior and can be used for crowd management, *"sophisticated surveillance systems"* that could keep a check on people's movement and behavior, and social media intelligence platforms that can aid public safety. The significant limitations, both of accuracy and fairness, and the adverse impact of surveillance on fundamental rights did not feature in the report's recommendations in context of smart cities. This is particularly worrying as India's surveillance framework already suffers from a lack of adequate safeguards to protect against potential erosion of fundamental freedoms by law enforcement authorities [24]. AI-powered surveillance should thus be the narrowly tailored exception rather than the norm.

In discussing fairness in AI systems as a concern, the report recognises that bias is embedded in data, with the possibility of such bias getting reinforced over time. It recommends that one possible way to deal with this is to, *"identify the in-built biases and assess their impact, and in turn find ways to reduce the bias"*. It takes a *ceteris-paribus* approach, i.e. all other things remaining equal, by simply identifying and reducing bias in datasets one can hope to yield fairer results, when the reality is that biased data emerges from a biased, unequal, discriminatory, and unfair world. The limitation to this approach is that it looks at AI as a mathematical model alone, and not as a socio-technical system. To reduce bias within these systems, it is crucial to understand and adapt for the social environment in which they will function. Indeed, the objective should be to avoid discriminatory outcomes. As Eubanks has pointed out, unless automated decision-making tools are build to dismantle structural inequities, their use will only intensify them [25].

Systemic Considerations

The development of AI in India thus far has been a fragmented policy process. There is no single regulatory body, ministry, or department that has been tasked with understanding the implications and opportunities arising out of AI. Instead, efforts have been largely ad-hoc, with little understanding as to the coordination or relationship between parallel efforts. While these reports and operational questions may be ironed out by the date of publication of this article, a number of systemic considerations must be noted.

First, existing initiatives are all primarily geared towards enhancing economic development and growth. While the NITI Aayog report does consider societal needs, it falls short of comprehensive analysis through this lens. By and large, AI is seen as a market opportunity. This is an integral consideration for AI policy in India, as these systems have a potentially transformational role to play in economies. However, it is not the only consideration that requires resources and depth of understanding. The ethical, social, legal implications of AI on everyday life are important to consider on an equal footing, particularly because their impact is often irreversible, while difficult to prove or detect at the same time.

Second, ongoing development and deployment processes of AI seem to be focussed on encouraging innovation, sometimes buying into the false dichotomy between innovation and ethics. That ethical choices can and should be built into innovative technologies as a central feature, is yet to be formally acknowledged and understood.

Third, current AI initiatives envision engagement primarily from government and industry, sometimes from academia, but not from civil society. These processes and committees have been appointed or nominated unilaterally, with no open call for participation. This risks making social and ethical considerations far more removed from the crux of discussions, whereas an approach towards AI that imbibes these considerations by design is far more sustainable. This also runs the risk of limiting the effectiveness and reach of AI policy in India, given that the AI ecosystem is complex, intertwined, and based on collaboration and partnership between different stakeholders [26].

Fourth, the limits of automated processing and technological solutions to solve systemic and societal problems are often left completely unconsidered. The potential risks and limitations of data-driven decision-making, and their ethical and social impacts needs to be reconfigured into a central consideration in India's AI policy development. The fact that a decision is "data-driven" does not mean it is fair, just, accurate, or appropriate. This will be further elaborated on in the next section.

Fifth, while State use and adoption of AI for law enforcement, defense, and national security is on the rise, policy discussions thus far have neglected to confront the contours within which such use should occur. Even in cases where national security is acknowledged, like the AI Task Force Report, the responsibilities and limitations on state power are not discussed. The relationship between individuals and the state, and the corresponding effects of AI on the exercise of rights needs more transparency and critical debate.

A framework for exploring limitations of data-driven decisions in the machine learning process

Having discussed the policy landscape in India, I now turn to the technical considerations that can help inform policy development. This section introduces a novel framework to understand the limitations of data-driven decision-making processes, with particular reference to the Indian context. It will do so by using the technical template of a machine learning system in three stages, i.e. data, models, and application. While some considerations discussed bleed into different stages, this taxonomy emphasizes the most appropriate point at which safeguards should be put in place.

1. Data

Machine learning depends on data in order to improve performance on future tasks. Existing scholarship has shown that bias in data can lead to disproportionately adverse outcomes that only entrench discrimination that exists in society [27]. The approach of identifying, assessing, and mitigating bias, as suggested in the NITI Aayog report, is one that is intuitive and simplistic at the same time. This is because bias arises not only from how data is collected, but also from the environment in which it was generated. Training data, thus, raises the following issues:

1.1. ***Access to data***: At the risk of using a be-laboured metaphor, data is the oil that fuels AI. Available, accessible, accurate, and affordable data is key to building AI systems. This is often a challenge in India. Data that is both accurate and relevant in a given context, is rarely readily available. Crime data in India, for example, is grossly underreported[28]. This constraint is felt by burgeoning FinTech and healthcare-related startups in the country; very few have access to accurate, affordable data about the populations that they hope to serve. Ryan Calo terms this the problem of data parity [29], where only a few well established leaders in the field have the ability to acquire data and build datasets. India's National Data Sharing and Accessibility Policy (NDSAP) contemplates sharing of non-sensitive data generated using public funds through the Open Data Platform, but this has had moderate success in solving the data parity problem, as a majority of quality data in India is restricted solely to the private sector [30].

1.2. ***Bias in collection*****:** Datasets used to train models run the risk of having 'dark spots'[31] where communities and classes of individuals are overlooked. Data from minority and disadvantaged groups may be more difficult to collect, access, or verify. Machine learning depends on generalizing based on examples. If the examples used to train a machine learning system under-represent certain groups that live on the margins[32] of data collection, the generalization offered will discriminate against the under or over-represented group, depending on the specific case. This is especially dangerous in jurisdictions like India, where being able to have a data footprint is a function of privilege in the first place: either due to gender, caste, geographic location, or class [33]. And yet, to proactively include individuals into datasets that they would have otherwise been left out of for the purpose of accuracy in learning models is effectively assuming that the relevance of technologies that aid in profiling and surveillance of those individuals is a given [34]. This problem is only amplified when we consider that surveillance is never neutral, it is fundamentally disproportionate in context of gender, caste, race, and religion [35].

1.3. ***Systemic & historical bias:*** Data imbibes historical discrimination that exists in the environment in which it was created. This leads to a third layer of complication. Often, systemic bias in data is considered problematic because of the outcomes at the time of decision-making. For instance, take FaceTagr, a facial recognition app currently being used by policeman in Chennai, Tamil Nadu. FaceTagr enables policeman to photograph individuals who "look suspicious"[36] and query these photos against existing criminal databases. As Bhatia observes, this is inherently problematic because, *"in the eyes of the police, 'a guy' who 'looks suspicious' for walking on the road at 2 AM will invariably come from a certain socio-economic class, measured simply by his appearance. And in this way, the FaceTagr app will invariably facilitate targeting the homeless and the destitute."*[37]. However, this is only one class of problems that arises from historical and systemic bias. Not all problematic cases lead to adverse outcomes that can be quantified; some are so because they entrench and sometimes exacerbate real world sentiments of racism, sexism, violence and hate. For instance, let us consider Google's AI-powered search algorithm. A Google search for "south indian masala" returns results not of spices, but of women [38]. This is not necessarily a flaw in the algorithm, but an uncomfortable reflection of society's stereotypes. Data can thus, not only expose problematic stereotypes and bias, but also cement, formalise, and imbibe it, simply by virtue of representing the environment in which it was created.

2. Model

The models, or decision frameworks, that emerge through training data have unique limitations and present distinct ethical questions from the training data, even though they might be closely related to one another. As Veale & Binns [39] point out, there are subjective choices embedded into even the choice of model used.

2.1. Feature selection: Models must be monitored with respect to their behavior, especially to ascertain if they demonstrate bias. Even in the case where a model inherits a perfect dataset (if such a thing existed), it could still be susceptible to exhibiting biased behavior. Design choices at the model level, such as feature engineering, determine weights assigned to attributes, which in turn determine how a particular model behaves.

Feature selection is unique to each problem, and is considered one of the more challenging parts of machine learning. Members of protected classes can be discriminated against even if adequate safeguards were built in at the time of building datasets, if feature selection involved choosing features that either serve as proxies to protected attributes [27], or because the choice of features do not place protected groups and non-protected groups at an equal footing for accurate determinations. Consider the Indian Government's plans to roll out a social media sentiment analysis tool to *"help facilitate creating a 360 degree view of the people who are creating buzz across various topics"*[40]. Momentarily pausing on the significant free expression and privacy impacts of such a tool, let us consider how these would be technically built. The tool would have to be trained on data, and specific features. The subjective decisions made at the time of feature selection - which aspects of individuals, speech and online behavior to amplify and which to downplay - would have profound influence on the classifications and outputs that such a tool would produce, and in turn, on the way in which this tool would disproportionately impact vulnerable communities.

While non-protected groups are vulnerable to discrimination all over the world, in India, the risk to women, sexual minorities, and members of lower castes, is particularly pronounced due to rigid traditional biases and societal, cultural norms. For example, consider a situation where an employer is looking to hire an office manager. Instead of going through qualitative past references and work experience, a machine learning program could be designed to attribute considerable weight to the individual's ability to spend long hours at the office. This feature is a convenient one, even if it has no bearing on the individual's actual commitment and interest in a job. It is possible that long hours at the office could be reflected as something that men do more efficiently than women, given the comparatively modern entry of women in the workforce, and historically strict social norms about an Indian woman's responsibilities at home.

2.2. Fairness*:* The increasing use of machine learning systems to make crucial decisions gives rise to concerns about the ability of models to be fair and non-discriminatory. This seems intuitive and straightforward as a goal, but a number of complications arise in context of fairness in AI systems.

The first is a definitional complication. A number of definitions have been developed by computer scientists in an attempt to operationalise the ideals of fairness and non-discrimination [41]. One popular definition is that of demographic parity, when models produce equal outcomes for people of protected and non-protected groups with the decision having no correlation to the protected attribute. This suffers from two main problems. The first, it can be perceived to undermine desirable accuracy in instances of individual fairness, in an attempt to secure group fairness [42]. Second, it ignores the instances where discrimination is needed for legitimate purposes, like fair affirmative action [43]. ProPublica's *Machine Bias* discusses accuracy equity which recognises that even having same error rates for different demographic groups can preclude fairness, if the *type* of errors differ [2]. Another definition of fairness is that of counterfactual fairness, where a decision is considered fair if it is the same in the actual world and a counterfactual world where the individual concerned did not have the protected attribute [44].

The second complication is that of trade-offs. While the trade-off between different definitions of fairness in a given case [45] are complicated enough, the trade-off between fairness and other competing values need to be considered as well. Removal of discrimination has been shown to reduce overall accuracy in a model [46]. Conversely, striving for accurate models may undermine fairness, because accuracy involves imbibing unfair and discriminatory aspects that exist in society.

A critical policy challenge in this area is thus creating tools, checklists, and standards for what definitions of fairness are most appropriate in given applications. The Indian Constitution contemplates India as a welfare state and embraces affirmative action, both "explicitly and implicitly" [47]. The status of constitutionally protected populations and affirmative action policies in education, housing and employment will play an important role in determining what appropriate standards of fairness should look like in the case of AI systems.

2.3. Transparency and Accountability: As AI applications increasingly supplant decision-making in the public sector [48], transparency and accountability in these systems become particularly crucial in constitutional democracies all over the world, including India. The opaque, complex, invisible manner in which AI systems function has led to concerns about accountability and redressal mechanisms, particularly following application in areas like criminal justice and policing. The models applied to these problems are often inscrutable [49], with the potential to undermine democracy and accountability. This has made way for an active discussion around the need for accountability of learning algorithms, with transparency as a popular proposed mechanism[50].

The premise on which these transparency ideals lie is the subject of much debate[51]. On one hand, transparency mandates opening up opaque black boxes, and peering inside black boxes could in turn be used to hold systems accountable [52]. This assumes that knowing input data and models lead to interpretability and higher levels of trust in machine learning systems, with the potential to detect bias and unfairness in a process [53]. On the other hand, this claim is described as “obvious, but naive” because AI systems that change over time cannot be understood even with full transparency [54]. As Burrell points out, the opacity in machine learning systems can be considered an inherent property of the technology, or arising out of technical illiteracy, or even as intentional secrecy [55]. Regardless, the crux of the issue here is to ensure accountability of such complex, unpredictable systems. While there have been shifts towards demanding that systems are explainable, scrutable, and intelligibile, some experts like Winfield [56] insist that transparency is key to understanding the behavior of AI technologies.

Overarchingly, it is pragmatic for policy makers to think of transparency and intelligibility along a spectrum, as mechanisms to reach accountability of AI systems. The level of transparency or intelligibility needed in different circumstances changes depending on the nature of AI application and the purpose it is meant to fulfill, and the level of transparency possible depends on the type of model used [51].

3. Application

Depending on the manner in which AI is applied, these systems can benefit or harm society, weaken or strengthen the exercise of rights, and widen or bridge the gap between protected and non-protected classes.

3.1. Privacy: AI systems are capable of making meaningful inferences, classifications and categorisations, and their use is carried out across sectors, from advertising to law enforcement. The profiling enabled by their use significantly alters our expectations of privacy and anonymity[57], both online and offline. The ability of AI systems to extract information from data, spot patterns and predict trends means that innocuous information can be mined to the point of relevance and intimacy.

Consider current applications of AI technologies in India. The perils of FaceTagr have been discussed above, and require more deliberate consideration in the future, given their plans to expand to other states in South India [58]. Elsewhere, law enforcement agencies in Punjab use the Punjab Artificial Intelligence System (PAIS) which adopts a “smart policing” approach using *“proprietary, advance hybrid AI technology”*[59] to digitise criminal records, and facilitates criminal search by using technologies like facial recognition to predict and recognise criminal activity. Elsewhere, a research project at the University of Cambridge recently published a paper titled *Eye in the Sky* that describes plans of training drones to identify violent behavior in public spaces and to test these drones out at music festivals in India [60].

The stated intention of most of these programs is to reduce crime rates, manage crowded public spaces to make them safer, and bring efficiency to law enforcement. And yet, as we have seen in the pages above, even well-intentioned machine learning systems can lead to adverse impacts. Not only do systems discussed operate in the absence of safeguards to prevent their misuse, making them ripe for surveillance and privacy violations, they also operate at questionable levels of accuracy [61].

This may seem like a good thing: if invasive facial recognition technology is inaccurate, the probability of having one's privacy eroded is low. However, in the case of applications currently being used by law enforcement in India, this makes the problem worse because in addition to the privacy considerations, these technologies can lead to false arrests, and people from disproportionately vulnerable and marginalised communities being made to prove their innocence. Further, regardless of accuracy rates, people's behavior in public spaces and their faces are still collected, stored and possibly shared or accessed without their consent.

The extent to which privacy can be respected or eroded by AI technologies also depends on the legal framework within which they function. In August 2017, the Supreme Court of India unanimously upheld the right to privacy as a fundamental right under the Constitution of India [62]. This historic judgment recognized informational privacy as part of this fundamental right, and underscored the dangers that emanate from the ability of machines to infer information and analyse data in new, sophisticated ways. The judgment also highlighted the urgent need for a 'robust regime for data protection' in the country, emphasizing the close relationship between data protection, autonomy and identity.

In particular, the Court observed,
"Informational privacy is a facet of the right to privacy. The dangers to privacy in an age of information can originate not only from the state but from non-state actors as well. We commend to the Union Government the need to examine and put into place a robust regime for data protection. The creation of such a regime requires a careful and sensitive balance between individual interests and legitimate concerns of the state."[62]

At the time of writing this article, the development of India's data protection regulation is at a critical juncture. On one hand, the right to privacy has been unequivocally affirmed by the highest court in the country. On the other, a Union government-appointed committee to review and recommend data protection norms in the country just released a draft of the Personal Data Protection Bill, 2018 [63] (This Committee is headed by a former judge of the Supreme Court of India, Justice B.N Srikrishna, hereinafter referred to as the 'Srikrishna Committee').

The draft bill takes some positive steps towards data protection. For example, it places responsibility on data fiduciaries (broadly defined as all legal entities that process data, including individuals, governments, and companies) to adhere to principles of purpose limitation, collection limitation, and data breach notification. It reaffirms consent to be valid only if it is free, informed, specific, clear, and capable of being withdrawn. It also adds an elevated requirement of explicit consent in the case of sensitive personal data.

However, certain provisions also undermine privacy protections envisaged in the Bill by carving out worryingly broad exceptions for government data processing. In its current form, the Bill allows government processing of personal data without consent if is shown to be *necessary*, for *"any function of Parliament or any State Legislature,"*, *necessary* for the exercise of the State *"authorized by law for the provision of any service or benefit.",* or for *"the issuance of any certification, license or permit for any action or activity"* of the individual by the State.

Sensitive personal data can be processed without consent if the processing is shown to be "*strictly necessary"*, for *"any function of Parliament or any State Legislature"*, or "*the exercise of any function of the State authorised by law for the provision of any service or benefit..".*

This effectively means there is little to nothing that citizens can do in context of data processing by the State. While the standard for exemption in the case of sensitive personal data (strictly necessary), is slightly higher than that for personal data (necessary), these matter little in context of machine learning systems. When trained on vast datasets that include both personal data and sensitive personal data, these systems will essentially imbibe, embed, and amplify biases and discrimination. Even in the case where sensitive attributes such as caste, religion, political leanings, sexuality are subject to forced blindness altogether, proxies still exist, and additionally, as Barocas et al explain, *"several slightly correlated features can be used to build high accuracy classifiers for the sensitive attribute"*[64].

The Bill does not address surveillance concerns that currently plague India's legal framework, and in fact remains completely silent on the question of accountability and transparency in re: intelligence agencies. The privacy implications of AI applications are thus systemically poised to be adverse in some senses: on one hand the legal framework contemplates unrestricted State processing of both sensitive personal data and personal data, and at the same time, the State is rolling out AI applications that carry out surveillance and profiling, effectively granting them free rein in this regard.

3.2. Freedom of expression: Freedom of speech and expression is a fundamental right under Indian constitutional law[65]. The Supreme Court of India has repeatedly relied on it as an integral part of democracy, and has also found that this freedom includes the right to know [66].

Freedom of expression is profoundly impacted by AI, given the increasing reliance on these systems for moderation of content online, and increasing use of AI applications in everyday life, from smart assistants to autocorrect technology on mobile devices [57].

AI is offered by both technology companies and governments, as a panacea [67] to complex problems like hate speech, violent extremism, and misinformation online [68]. This is a dangerous trend, given the limited competence of machine learning to understand tone and context. Automated content removal risks over-broad censorship and take-down of legitimate speech, and this risk is made even more pronounced by the fact that it occurs unilaterally by private companies, sometimes acting on governmental instruction.

Surveillance driven by AI technologies, while having the obvious implications for privacy discussed above, also impacts freedom of expression. By blurring the lines between the private and the public, AI driven surveillance has a chilling effect on expression, where self-censorship is widely adopted by individuals who are unsure of the status and implications of their speech.

Sentiment analysis tools are increasingly deployed to gauge the tone and nature of speech online, and are often trained to carry out automated content removal. As discussed above, the Indian Government has expressed interest in moving towards using AI to carry out sentiment analysis, identify fake news, and boost India's image on social media platforms and even email, first in the form of Project Insight in 2016 [69], and most recently via a public tender issued by the Ministry of Information and Broadcasting [40]. The limitations of using AI for content regulation in context of private platforms in the past become amplified when similar attempts are made by states [70].

The nexus between technology and freedom of expression was most emphatically pronounced by the Supreme Court in 2015 [71], when it struck down Section 66A of India's Information Technology Act. The Section imposed criminal liability on online communications that were found to be grossly offensive, menacing, or annoying. The Court struck down this provision of law on grounds of being overbroad, vague, and having a chilling effect on free speech. It reaffirmed the importance of democracy, informed citizenry and an open culture of dialogue in India's tradition of free speech, while placing emphasis on the contours of reasonable restrictions to free speech under Constitutional law in the age of technology. The push towards AI solutions, particularly by State actors, must be studied in context of established law in India.

This taxonomy provides a framework for thinking through how bias, discrimination, surveillance, and profiling occurs through a machine learning system. The findings and considerations mentioned above transcend sector specific use, and provide a blueprint for thinking through the implications of AI technologies.

Sectoral Challenges

While there are no specific regulations in place for AI in India as a whole, sectoral laws apply to applications of AI. For example, the security markets regulator of India, namely the Security Exchange Board of India (SEBI), developed guidelines on algorithmic decision-making for investors looking to get into wealth management and algorithmic trading [72]. Studying the extent to which sectoral laws can regulate the development and deployment of AI systems is an important area of research, but beyond the scope of this current article.

One of the goals that *does* fall within the scope of this article is to inform ongoing AI policy development in India. In the sections above, I have presented societal considerations and technical limitations that arise across sectors, at various stages in the machine learning process. With a view to complement existing policy development in India specifically, I will now return to some sectors of relevance, and highlight how the proposed framework could be applied in these sectors. While an exhaustive analysis is beyond the scope of this article, I hope to present some key issues that can encourage further research and policy deliberation.

Manufacturing

Manufacturing is viewed as one of the most transformative areas for AI applications. The Task Force Report envisions using AI to enable smarter production design, using predictive tools for maintenance and planning, among others. The following considerations must also be reckoned with at the time of developing policy for this sector:

1. Data:
    a. Training: The move to autonomous cars and smarter production planning means that systems will have to be trained using existing data on safety, efficiency, and accuracy. How will it be ensured that data is representative and equally applicable in different contexts?
    b. Processing: In the case of autonomous vehicles, how will personal data be dealt with?
    c. Data parity: Given that only large incumbents have the kind of data required to train systems, is there a way to solve this problem of data parity?
2. Model:
    a. Safety: What levels of accuracy and precaution are built into systems?
    b. Testing: How are they tested before deployment?
    c. Assessment: Is there an ongoing impact assessment of models that must be adhered to?
3. Application:

    a. Transition costs: The move to AI driven manufacturing could involve a possible displacement of workers, companies, and skills. What adverse effects does the transition to AI-driven manufacturing involve? Does it leave particular communities or demographics vulnerable?
    b. Competition: Does the rise in AI applications fracture healthy competition in this sector?
    c. Human liability: To what extent is automation desirable? Does the human in the loop have a risk of simply turning into a moral crumple zone [73]?

FinTech

Financial Technology (FinTech) is a burgeoning field in India, made particularly fertile by Governmental push for projects like India Stack, which *'allows governments, businesses, startups and developers to utilise an unique digital Infrastructure to solve India's hard problems towards presence-less, paperless, and cashless service delivery'* [74]. I propose the following considerations for Fintech use of AI:

1. Data:
    a. Diverse financial footprints: Given the wide disparity in financial capability and capacity in India, how are FinTech models trained?
    b. Accounting for bias in data: Much of the disparity mentioned above is a function of systemic and historical bias. How do developers intend to correct for such bias, if at all? Is this a consideration developers are made mindful of? If so, how?
    c. Unequal collection: Given that having financial data is a function of privilege, how do developers ensure that this inevitable imbalance of *whose* data is collected and from *where,* is addressed?
    d. Data margins: What deliberation has occurred to address the fact that a large majority of Indians live at the margins of data collection, either because they are physically remote or digitally excluded?
    e. Consent: One idea to address some of the considerations above could be to collect more data or include people in datasets. At this juncture, particularly with State applications of FinTech (if any), how do we ensure that consent of individuals isn't undermined? Inclusion with agency is integral, what processes and strategies are put in place for this?
    f. Proxies: How can proxies for protected characteristics be addressed?
    g. Security: What safeguards are put in place to protect user's financial data?
2. Model:
    a. Granularity: What level of granularity is offered to individuals who are the subject of consequential decision-making in this sector? Is the provision for such granularity built into the models and frameworks being developed?

   b. Scrutability: Given the increasing reliance of automation at various levels of decision-making processes, is there a clear division of responsibility and liability between the human and automated decision-making?
3. Application:
   a. Impact assessment: Given the critical nature of FinTech systems, what ongoing impact assessments can be put in place to monitor and reflect on applications in this sector?
   b. Privacy: Do FinTech applications undermine user's privacy, and are there positive steps to ensure that privacy is protected?
   c. Appeal: What appeal mechanisms can be put in place? How can we ensure that these are both technically viable and institutionally implementable?

Agriculture

The promise of AI in agriculture is widely hailed as revolutionary, particularly given India's ongoing agrarian crisis[75]. Small and marginal farmers are often left in the dark corners of policy making, and this threat is even more pronounced in the age of AI. Some critical questions include:

1. Data
   a. Collection: Given the isolated status of most small and marginal farmers, how do we intend to build systems that are accurate and representative of a group of people, whose troubles and challenges have been only cursorily considered, at best?
   b. Historical complications: The potential and outcomes of India's agricultural sector is not an isolated relationship between efficiency and output. The problems in this sector are fundamentally political, and systemic. How will training data be representative of this, both in collection and training?
2. Model:
   a. Limits to benefits: What are the limits to promised benefits on AI-driven agriculture? What can models *not* do? Does the use of AI in this sector provide enough additional value to justify the possible risks to farmers working on small margins?
   b. Impact assessment: What makes a model successful? How is this definition of success arrived at?
3. Application:

a. Resilience: What is the most adverse potential effect of these models? What safeguards are built into AI systems to mitigate the possibility of the worst case scenario occurring?
b. Liability: What do redressal mechanisms look like?
c. Equality: Is the introduction of AI in this space creating more opportunity and bridging the socio-economic divide? If no, what measures are being taken to ensure that this is more accessible?
d. Accessibility: How can AI technologies be made accessible to farmers in India? Is there a risk of exacerbating the existing problem of access to resources, by making the standard for farming something that is far beyond most farmers' financial capability?

Healthcare

Much like the domain of FinTech, Healthcare has witnessed a significant increase in the number of AI-based startups in India. Some questions to be considered at this stage include:

1. Data:
    a. Contextual data: An overarching problem in this sector is accessing data that is reliable, accurate, and relevant. This leads to a potential solution of obtaining high quality health data from other jurisdictions, and applying learnings in India. And yet, accuracy rates drop significantly when training data and test data are different[64]. How can this be meaningfully tackled?
    b. Security: Given the vast volume of sensitive information, how is the security of healthcare data ensured?
    c. Consent: How do we ensure that consent of individuals isn't undermined?
2. Model:
    a. Applicability: While data is key to building successful models, what measures are taken to ensure that the model is relevant to the socio-economic conditions and unique populations they hope to serve?
    b. Impact assessment: Is there an ongoing impact assessment of models that must be adhered to?
3. Application:
    a. Trust: What safeguards are critical to build in order to build trust and encourage buy-in? How are patients brought into the deliberation, explanation process? Trust can also include trust in ability of AI systems to be secure. What additional safeguards does health-related data mandate?

b. Liability: Are the interactions and division of responsibility between human-based and automation-based decision-making? What should internal verification mechanisms and checks and balances look like?

National Security and Law Enforcement

The Task Force Report looks at Google's Project Maven as an example of successful application of AI in national security. It is important to note that this move has been heavily criticized, most recently by Google's own staff, who highlight the ethical and human cost of biased AI [76]. Other challenges include:

1. Data:
    a. Collection: What is the process and mechanisms used to collect data for the purpose of building AI systems in this sector?
    b. Bias: Given the particularly sensitive nature of this sector, have there been efforts to recognise and address historical biases that give rise to problematic data?
    c. Consent: How is this data collected? How is the inevitably sensitive nature of data dealt with? What standards do data collection, handling, and processing adhere to for these applications?
    d. Proxies: How is the issue of proxies for protected characteristics addressed?
2. Model:
    a. Feature engineering: At the time of training models, what are we training these models to look for? What aspects of individuals are considered problematic or worth flagging?
    b. Accuracy: What is the desirable level of accuracy in AI systems needed for issues of national security to be delegated to them? What levels of inaccuracy are we willing to tolerate, and in what forms will such inaccuracy play out?
3. Application:
    a. Scope: Given the proclivity of terms like 'national security' to assume all encompassing definitions, it is critical to explain what is meant by national security, what kinds of AI applications are intended within this domain, and what definition of 'success' applications will be broadly trained to aspire to. How are objectives and purposes for predictive policing deliberated and set out?
    b. Rights: Do AI applications proposed in this domain, including autonomous surveillance and combat systems, threaten to violate constitutionally protected rights of privacy and freedom of expression? In the event that they do, are these applications within the realm of reasonable restrictions contemplated in the Constitution and other relevant laws?

c. Security: While the Ministry of Defence has begun to look into AI for National Security and Defence, and also begun to list use cases[77], has it considered scenarios where the same AI applications can be used in an adversarial manner? What security measures have been put in place?

Conclusion

This article was overarchingly intended to influence existing AI policy deliberation in India, and invite a cross-disciplinary discussion on the issue. It analysed the current policy landscape in India, and argued that the limitations of data-driven decision-making should be a fundamental consideration in AI policy development, and *not* a retrospective one. Given the multiplicity of efforts in the space already, it focussed on highlighting those aspects of the debate that have been only peripherally examined so far.

By walking the reader through the process of machine learning, I have argued that data-driven decision-making is susceptible to inaccuracies, discriminatory outcomes, embedded and exacerbated bias, and even unintended consequences due to various limitations that occur through the process. Technical research in the field is currently looking at ways in which to meaningfully address these concerns, and policy development must confront the same. The proposed framework attempts to bridge the gap between the two, and develop a shared understanding of these issues. Importantly, it demonstrates that AI systems cannot be thought of as isolated mathematical problems, or as neutral in nature, or as only beneficial because of their efficiency. Rather, AI technologies are complex social systems that cannot, and should not, be evaluated only on the basis of efficiency and accuracy.

Given the complex terrain of navigating challenges posed by AI systems, it is integral that future deliberation, policy making, and regulation of AI is informed by multiple disciplines on an equal footing. These must be ethically, legally, technically, and philosophically informed throughout the process. The pace of development is quick, the nature of development is opaque, and the effects of development are profound and often irreversible. Traditions of building processes and deploying technology first, and deliberating on their effects next will not work with AI. I hope the proposed framework will help other researchers, policy makers, lawyers and technologists to explain, deliberate, and understand the challenges and opportunities of AI in their unique contexts.

References

1. Balkin, J. 2017 Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. *U.C. Davis L. Review,* Forthcoming 2018, available from: https://papers.ssrn.com/sol3/papers. cfm?abstract_id=3038939.
2. Angwin, J., Larson, J., Mattu, S., Kirchner, L. 2016 Machine Bias. *ProPublica.* Available from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
3. O'Neil, C. 2016 *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* pp. 278 - 315. New York: Crown Publishing Group.
4. Liping, Z. 2018 China is using AI to tackle traffic jams. *World Economic Forum.* Available from https://www.weforum.org/agenda/2018/01/china-is-using-ai-to-tackle-traffic-jams.
5. Lynch, S. 2017. Andrew NG: Why AI is the new electricity. *Stanford Graduate School of Business,* Available from https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity.
6. NITI Aayog. 2018 National Strategy for Artificial Intelligence. *Niti Aayog,* 46. Available from http://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf.
7. Gillespie, T. 2014 The relevance of algorithms, *Media technologies: essays on communication, materiality and society,* 167.
8. MacCormick, J. 2012 *Nine Algorithms That Changed the Future,* pp. 3. Princeton University Press.
9. Elish,M.C. & Hwang, T. 2016 *An AI Pattern Language*, Intelligence and Autonomy initiative (I&A) at Data & Society. Available from https://www.datasociety.net/pubs/ia/AI_Pattern_Language.pdf (2016).
10. L. Floridi. 2016 Should we be afraid of AI?, *Aeon*. Available from https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible.
11. Surden, S. 2014 Machine Learning and the Law. *Washington Law Review,* Vol 89, No 1.
12. Munoz, A. 2014 Machine Learning and Optimisation. New York University, retrieved from https://cims.nyu.edu/~munoz/files/ml_optimization.pdf.
13. Domingos, P. 2012 A few useful things to know about machine learning. *Communications of the ACM,* Vol. 55 Issue 10, pp 78-87.
14. 2018 Make 'Artificial Intelligence' work for India: PM. *The Economic Times.* Available from https://cio.economictimes.indiatimes.com/news/government-policy/make-artificial-intelligence-work-for-india-pm/62980259.

15. Speech of Arun Jaitley, Minister of Finance. 2018 Budget 2018 - 2019. Available from https://www.indiabudget.gov.in/ub2018-19/bs/bs.pdf.
16. 2014 Digital India - A programme to transform India into digital empowered society and knowledge economy. *Press Information Bureau, Government of India.* Available from http://pib.nic.in/newsite/PrintRelease.aspx?relid=108926.
17. 2018 Digital India - Vision and Vision Areas. *Digital India.* Available from http://digitalindia.gov.in/content/vision-and-vision-areas.
18. Make in India. Available from http://www.makeinindia.com/about.
19. 2017 Artificial Intelligence Task Force. Available from https://www.aitf.org.in/.
20. Agarwal, S. 2018 IT Ministry has formed four committees for Artificial Intelligence: Ravi Shankar Prasad. *The Economic Times.* Available from https://economictimes.indiatimes.com/news/economy/policy/it-ministry-forms-four-committees-for-artificial-intelligence-ravi-shankar-prasad/articleshow/62853767.cms.
21. Sharma, Y. S. & Agarwal, S. 2018 Niti Aayog to come out with national policy on artificial intelligence soon. *The Economic Times.* Available from https://economictimes.indiatimes.com/news/economy/policy/niti-aayog-to-come-out-with-national-policy-on-artificial-intelligence-soon/articleshow/63387764.cms.
22. Gupta, K. 2018 Niti Aayog partners with Google to grow India's artificial intelligence ecosystem. *Livemint.* Available from https://www.livemint.com/Industry/fpnGnNQ8duTCRZOEpk2P6M/Niti-Aayog-partners-with-Google-to-grow-Indias-artificial-i.html.
23. Hebbar, P. 2018 Niti Aayog and ABB join hands to make India AI-Ready. *Analytics India Magazine.* Available at https://analyticsindiamag.com/niti-aayog-abb-india-partner-to-make-india-ai-ready/.
24. Bailey, R., Bhandari, V., Parsheera, S., Rahman, F. 2018 Use of Personal Data by Intelligence and Law Enforcement Agencies. *Macro/Finance Group, National Institute of Public Finance and Policy.* Available from http://macrofinance.nipfp.org.in/PDF/BBPR2018-Use-of-personal-data.pdf.
25. Eubanks, V. 2018 *Automating Inequality: How high tech tools profile, police, and punish the poor, pp.* 190. St. Martin's Press, New York.
26. Cowls, J. & Floridi, L. 2018 Prolegomena to a White Paper on an Ethical Framework for a Good AI Society. *Working Paper Series.* Available from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3198732.
27. Barocas, S. & Selbst, A. 2016 Big Data's Disparate Impact. 104 *California Law Review* 671. (DOI 10.15779/Z38BG31).
28. Dey, A. 2017 What the National Crime Records Bureau report does not tell us about cyber crime in India. *The Scroll.* Available from https://scroll.in/article/860254/what-the-national-crime-records-bureau-report-does-not-tell-us-about-cyber-crime-in-india.

29. Calo, R. 2017 Artificial Intelligence Policy: A Primer and Roadmap. *U.C. Davis L. Review,* Vol. 51, pp. 398 - 435.
30. Open Government Data (OGD) Platform India. Available from https://data.gov.in/.
31. Crawford, K. 2013 Think Again: Big Data. *Foreign Policy*. Available from http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data.
32. Lerman, J. 2013 Big Data and Its Exclusions. *Stanford Law Review Online* Vol. 66, pp. 55, 57.
33. Sinha, A., Rakesh, V., Marda, V. 2017 Big Data in Governance in India - Case Studies. *The Centre for Internet and Society.* Available from https://cis-india.org/internet-governance/blog/big-data-in-governance-in-india-case-studies.
34. Pasquale, F. 2018 Odd Numbers. *Real Life Mag.* Available from http://reallifemag.com/odd-numbers/.
35. Shepard, N. 2016 Big Data and Sexual Surveillance. *Association for Progressive Communications.* Available from https://www.apc.org/sites/default/files/BigDataSexualSurveillance_0.pdf.
36. Murali, A. 2018 The Big Eye: The tech is all ready for mass surveillance in India. *Factor Daily.* Available from https://factordaily.com/face-recognition-mass-surveillance-in-india/.
37. Bhatia, G. 2019 The Transformative Constitution: A Radical Biography in Nine Acts. *HarperCollins* (forthcoming).
38. Retrieved from https://www.google.co.in/search?q=south+indian+masala&source=lnms&tbm=isch&sa=X&ved=0ahUKEwi26NP2tYPdAhUO148KHUjFC-QQ_AUICigB&biw=1440&bih=700.540
39. Veale, M., & Binns, R. 2017 Fairer machine learning in the real world: Mitigating discrimination without collective sensitive data. *Big Data & Society* July - December 2017, pp. 1- 17.
40. Request for Proposals (RFP) invited for Selection of Agency for SITC of Software and Service and Support for function, operation and maintenance of Social Media Communication Hub. *Broadcasting Engineering Consultants, Ministry of Information and Broadcasting.* Available from http://www.becil.com/uploads/tender/TendernoticeBECIL01pdf-04836224e38fdb96422221c4e057f6c5.pdf.
41. Narayanan, A. 2018 Translation tutorial: 21 definitions of fairness and their politics. *Fairness, Accountability and Transparency in Machine Learning Conference 2018.* Available from https://fatconference.org/static/tutorials/narayanan-21defs18.pdf.
42. Barocas, S., Bradley, E., Honavar, V., Provost, F. 2017 Big Data, Data Science, and Civil Liberties. *A Computing Community Consortium.* Available at https://arxiv.org/abs/1706.03102.
43. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R. 2011 Fairness Through Awareness. *Computing Complexity.* Available at arXiv:1104.3913.

44. Kusner, M., Loftus, J., Russel, C., Silva, R. 2018 Counterfactual Fairness. *31st Conference on Neural Information Processing Systems (NIPS 2017).* Available from https://arxiv.org/pdf/1703.06856.pdf.
45. Kleinberg, J., Mullainathan, S., Raghavan, M. 2016 Inherent Trade-offs in the Fair Determination of Risk Scores. *Proceedings of Innovations in Theoretical Computer Science.* Available at https://arxiv.org/abs/1609.05807.
46. Zliobaite, I. 2015 On the relation between accuracy and fairness in binary classification. *The 2nd workshop on Fairness, Accountability, and Transparency in Machine Learning (FATML) at ICML'15.* Available from https://arxiv.org/pdf/1505.05723.pdf.
47. Jacobsohn, G. J. 2016 Constitutional Identity. In *The Oxford Handbook of The Indian Constitution*. Oxford University Press. (DOI: 10.1093/law/9780198704898.001.0001).
48. Veale, M., Kleek, M.V., Binns, R. 2018 Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* Paper No 440.
49. Mittelstadt,B., Patrick, A., Taddeo, M., Wachter, S & Floridi, L. 2016 The Ethics of Algorithms: Mapping the Debate. Big Data & Society Vol. 3. (DOI: 10.1177/2053951716679679).
50. Diakopoulos, N. 2014 Algorithmic Accountability Reporting: On the Investigation of Black Boxes. *Tow Centre for Digital Journalism.* Available from http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf.
51. Marda, V. 2017 Machine Learning and Transparency: A Scoping Exercise. *Working Paper Series.* Available from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3236837.
52. Pasquale, F. 2015 *The Black Box Society, pp.* 106. Harvard University Press, Cambridge Massachusetts.
53. 2017 Machine Learning: The Power and Promise of Computers that Learn by Example. *Royal Society. Available from* https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf.
54. Kroll, J. et al. 2017 Accountable Algorithms. *University of Pennsylvania Law Review* Vol. 165, pp. 633 - 705.
55. Burrell, J. 2016 How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society* January - June 2016, pp. 1-12.
56. House of Lords Select Committee on Artificial Intelligence. 2018 AI in the UK: ready, willing and able?, *House of Lords,* 36. Available from https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf..
57. 2018 Privacy and Freedom of Expression in the Age of Artificial Intelligence. *ARTICLE 19 and Privacy International.* Available from https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf.

58. Narayanan, V. 2018 FaceTagr database to get wider. *The Hindu.* Available from https://www.thehindu.com/news/national/tamil-nadu/facetagr-database-to-get-wider/article23722444.ece.
59. Sathe, G. 2018 Cops in India are Using Artificial Intelligence That Can Identify You in a Crowd. *Huffington Post.* Available from https://www.huffingtonpost.in/2018/08/15/facial-recognition-ai-is-shaking-up-criminals-in-punjab-but-should-you-worry-too_a_23502796/.
60. Vincent, J. 2018 Drones Taught to Spot Violent Behavior in Crowds Using AI. *The Verge.* Available from https://www.theverge.com/2018/6/6/17433482/ai-automated-surveillance-drones-spot-violent-behavior-crowds.
61. Staff Reporters. 2018 Police Facial Recognition Software Inaccurate. *The Hindu.* Available from https://www.thehindu.com/news/cities/Delhi/police-facial-recognition-software-inaccurate/article24764781.ece.
62. *K.S. Puttaswamy v. Union of India,* (2014) 6 SCC 433.
63. The Personal Data Protection Bill, 2018. Available from http://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf.
64. Barocas, S., Hardt, M., Narayanan, A. 2018 Fairness and Machine Learning, 41. Available from fairmlbook.org.
65. Article 19(1)(a), Constitution of India.
66. *State of Uttar Pradesh v. Raj Narain.* (1975) 3 SCR 333.
67. Marda, V. 2018 Facebook Congressional Testimony: "AI tools" are not the panacea. *Article 19.* Available from https://www.article19.org/resources/facebook-congressional-testimony-ai-tools-not-panacea/.
68. Li, S., Williams, J. 2018 Despite what Zuckerberg's Testimony May Imply, AI Cannot Save Us. *Electronic Frontier Foundation.* Available from https://www.eff.org/deeplinks/2018/04/despite-what-zuckerbergs-testimony-may-imply-ai-cannot-save-us.
69. Seth, S. 2017 Machine Learning and Artificial Intelligence Interactions with the Right to Privacy. *Economic and Political Weekly,* Vol LII 51, pp. 66-70.
70. Marda, V. 2018 Regulating Social Media Content: Why AI Alone Cannot Solve the Problem. *ARTICLE 19.* Available from https://www.article19.org/resources/regulating-social-media-content-why-ai-alone-cannot-solve-the-problem/.
71. *Shreya Singhal v. Union of India.* AIR 2015 SC 1523.
72. Goudarzi, S., Hickok, E., Sinha, A. 2018 AI in Banking and Finance. *The Centre for Internet and Society*. Available from https://cis-india.org/internet-governance/files/ai-in-banking-and-finance.

73. Elish, M.C. 2016 Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *We Robot 2016 Working Paper*. (DOI: 10.2139/ssrn.2757236)
74. India Stack. Available from http://indiastack.org/about/.
75. Sainath, P. 2018 In India, Farmer's Face a Terrifying Crisis. *The New York Times.* Available from https://www.nytimes.com/2018/04/13/opinion/india-farmers-crisis.html.
76. Eckerslet, P., & Cohn, C. 2018 Google Should Not Help the U.S. Military Build Unaccountable AI Systems. *Electronic Frontier Foundation.* Available from https://www.eff.org/deeplinks/2018/04/should-google-really-be-helping-us-military-build-ai-systems.
77. 2018 Raksha Mantri Inaugurates Workshop on AI in National Security and Defence. *Press Information Bureau, Government of India.* Available from http://pib.nic.in/newsite/PrintRelease.aspx?relid=179445.