TECHNICAL, LEGAL AND ETHICAL OPPORTUNITIES AND CHALLENGES OF GOVERNING ARTIFICIAL INTELLIGENCE IN INDIA

Mabel V Paul, School of Law, CHRIST (Deemed to be University).

ABSTRACT

Applications that make use of AI have, up to this point, been pushed almost entirely by the private sector and have generally concentrated on consumer goods. Because of the rapidly expanding magnitude of the technology and the ramifications it could have, it is vital that officials in government pay attention. India ought to take into consideration the various public and private funding methods for AI research that are available as a result of the early successes of AI in the United States, China, and South Korea, among other places. The traditional model of schooling followed by employment is no longer relevant in the modern economy. The nature of jobs is changing at a rapid pace, and skills can quickly go from being important to being worthless in a matter of years. It is vital that India recognizes artificial intelligence as a crucial component of its national security strategy. This is because China is making rapid progress in the research of AI-based technologies.It is crucial to promote AI-based innovation and to establish AI-ready infrastructure in order to preserve India's strategic interests and to future-proof the Indian employment market and the Indian workforce. There are currently no laws, statutory rules, or government-issued recommendations in India pertaining to AI. This article argues that technological constraints of AI systems should be considered during policy development, and that the societal and ethical problems that arise as a result of these constraints should be used to guide the goals of policy making. It examines the three crucial phases—the collection of data, the development of a model, and the integration of that model into a working application—into which machine learning (the most popular subset of AI techniques) must pass before it can be put into widespread use. The analysis presented here gives a foundation upon which such thought might be deliberated. It is built with the contemporary AI policy landscape in India as its backdrop, and it applies the suggested framework to ongoing sectoral problems in India. Its goal is to influence the policy discourse taking place in India by drawing

attention to the risks associated with data-driven decisions in general and in the Indian context in particular.

Keywords: AI, develop, infrastructure, policy, hazardous.

Introduction

Today's AI-based apps have already had an impact on people's lives, though the full extent of that impact is not often recognised or understood. The private sector has been largely responsible for the recent rapid development of AI technology, with an emphasis on commercial applications and consumer goods. The scale and ramifications of AI's applications, however, make it vital for government leaders to take notice. The development and execution of the technology cannot be left primarily to a few group of Silicon Valley businesses and its distributors. To understand the entire scope of these shifts, it is useful to examine some specific examples of the knock-on effects of AI's broad adoption. A network of algorithms processes the user's online behaviour data, such as the user's browsing history and hundreds of data points, in order to generate an educated prediction as to which products might spark the interest of the visitors to the website. From voice-activated assistants in tablets and desktops to intelligent keyboards on smartphones that reduce typing effort by anticipating the next words, machines in users' immediate personal space have become substantially more sophisticated than is commonly understood. While several studies have looked at the effects of AI on the economy as a whole, few have focused on India's developing market. The Indian government is actively pursuing two goals: increasing human capital on a national scale (with a focus on the country's youth via the Skill India project) and luring global industry to India (with the Make in India programme). The third leg of this modernization trifecta is the Digital India programme, which is making a concerted effort to increase the number of people who have access to the internet across the country. India's lack of economic stability is one reason artificial intelligence (AI) faces certain additional challenges. There is a lengthy history of government support for artificial intelligence research around the world, with funding fluctuating between periods of optimism and pessimism. However, in the past two decades, there has been a considerable movement toward private-sector support, thanks in large part to the rise of the Internet economy. The resulting attrition is an intriguing facet of this change.

The emerging areas of Artificial Intelligence

The Indian government has made it clear that it is committed to advancing AI research,

commercialization, and education because it believes that doing so will improve people's quality of life and promote social justice. In 2018, the Union government increased its spending by 100% in areas related to research, education, and training for emerging technologies such as AI.

This emphasis on digital technology is not something that is particularly novel. The objective of the effort known as Digital India, which is being directed by the Union Government, is to turn India into a "digitally empowered society and knowledge economy"[1]. The concept of "Digital India" calls for the provision of digital infrastructure as a fundamental service to each and every citizen, the incorporation of such digitization into governance, and, eventually, the empowerment of citizens[2]. The Digital India programme has increased funding for AI-related study, education, and development of related skills. As you read this, this operation will be under progress. The goal of the "Make in India" programme is to position the country as a "make it anywhere" manufacturing powerhouse. The government has started making efforts to ensure that AI technology is developed in India and serves the interests of the country. This is consistent with its Make in India effort, which aims to localise the production of AI tools and systems.

While AI has become a major consideration for many kinds of digital technology, a number of projects with a singular focus on AI have also emerged in recent years. This section will present an overview of the initiatives covered by this article, with a focus on a few salient points from each. It will not, however, pretend to be an exhaustive analysis of the initiatives.

Artificial intelligence task force

To "embed AI in our Economical, Political, and Legal thought processes so that there is the systemic capability to support the goal of India becoming one of the leaders of AI-rich economies," the Union Ministry of Commerce and Industry formed an AI Task Force in August 2017[3]. The report centered on the concept that AI has the potential to be applied to societal and economic issues on a massive scale. Other areas of focus include production, financial technology (also known as FinTech), agriculture, health, technology for individuals with various abilities, national security, the environment, public utility services, retail, and consumer

[1] Digital India – A programme to transform India into digital empowered society and knowledge economy. *Press Information Bureau, Government of India*
[2] Digital India – Vision and Vision Areas. *Digital India*.
[3] Artificial Intelligence Task Force.

interactions, and education. The goal of this article was to examine the role of government and the potential role of AI in resolving societal issues more closely. It proposes, for instance, the establishment of a single body, the National Artificial Intelligence Mission, to coordinate AI-related initiatives across India.

The paper does not seriously address the ethical, social, or technical restrictions that underlay the use of AI technology, although identifying the specific government bodies and ministries that could facilitate such expansion and presenting a list of enabling criteria for the wider adoption of AI. In addition, the several government agencies that could help facilitate such growth are outlined in the report. Even in the exceptional cases where privacy and data protection are discussed, the study does not go far enough to address challenges surrounding data that are specific to AI. Data sharing and third-party access to data are obstacles that the Task Force is aware of as it investigates the ethics and social safety issue. It's done as a natural part of the conversation. The potential for data-driven decision-making to reinforce and exacerbate existing bias and discrimination isn't addressed, despite the fact that data privacy concerns are raised. Algorithmic systems, even when created with the greatest of intentions, may have unintended negative consequences for those who are already at a disadvantage.

The Task Force's analysis of the industry landscape reveals the same gap. Predicting market demand and finding a happy medium between scale and innovation are two challenges that have been identified in the FinTech sector. Concerns about how the expansion of FinTech may affect those on the margins of data collection and technological inclusion are barely scratched. Perhaps more worrisome is the fact that the use of AI for "autonomous surveillance and warfare systems" was stated without qualification of the catastrophic repercussions such technologies have on privacy and freedom of expression. This is the most problematic part of the scenario.

The work being done by the Task Force has the goal of shedding light on the appropriate path that India's AI policy should evolve. Its primary emphasis is placed on the development of accessible technology, which is perhaps its greatest asset. The ethical and social analysis of India's AI environment is, at best, rudimentary; yet, this research makes it clear that there is a lack of legal and regulatory engagement within this process, as well as participation from civil society. AI for All is the National Strategy for Artificial Intelligence developed by NITI Aayog. The National Institution for Transforming India (commonly known as 'NITI Aayog'), a think tank that the government runs, has been entrusted with establishing a national AI policy in

order to direct the AI activities of the government [4]. NITI Aayog and Google formed a partnership at the beginning of May 2018 intending to increase economic productivity in India [5]. The partnership will provide training and incubation services for start-up companies that aim to develop and integrate AI-based solutions into their business models. Late in the month of May 2018, NITI Aayog also entered into a statement of intent with ABB India with the objective of "making important sectors of the Indian economy ready for a digitalized future and realising the promise of AI, big data, and connectivity." [6]

NITI Aayog outlined the overarching objective for a national AI strategy in a discussion paper that was published in June of 2018 [7]. The paper stated that the strategy should "leverage AI for economic growth, social development, and inclusive growth, and finally as a "Garage" for emerging and developing economies." NITI Aayog outlined the overarching objective for a national AI strategy in a discussion paper that was published in June of 2018. The job of NITI Aayog goes beyond simply advocating an approach to policy; it also includes involvement in the process of putting the policy into effect and deploying it. The National Strategy surpasses all other AI policy processes in two distinct respects, making it the most comprehensive. First, it notes that commercial interests have mostly driven the adoption of AI technology to this point, and it admits the 'need to find a balance between narrow notions of financial effect and the broader good.' Second, it acknowledges that applications of AI should be welcomed for the the incremental value that they provide to particular industries rather than for the value that they are rumored to bring in terms of change.

Despite these hopeful adjustments in viewpoint, India's national strategy on AI leaves a lot to be desired in terms of its substantive suggestions and analyses. The report identifies five primary sectors where AI could have a positive social impact that requires the government to play a leading role. These sectors are education, agriculture, healthcare, smart cities and infrastructure, and smart mobility and transportation. Each of these sectors requires the government to take the lead.

During the course of its discussion of smart cities, the paper advocates for the implementation of AI in various surveillance applications. These include sophisticated surveillance systems

[4] Sharma YS, Agarwal S. 2018Niti Aayog to come out with national policy on artificial intelligence soon. *The Economic Times*.
[5] Gupta K.2018Niti Aayog partners with Google to grow India's artificial intelligence ecosystem. *Livemint*.
[6] Hebbar P.2018Niti Aayog and ABB join hands to make India AI-Ready. *Analytics India Magazine*.
[7] NITI Aayog. 2018National strategy for artificial intelligence. *Niti Aayog*.

that could keep a check on people's movement and behavior and social media intelligence platforms that can assist with public safety. AI systems that can predict the behavior of crowds and be used for crowd management are also included in this category. In the context of smart cities, the report's suggestions did not take into account the considerable limitations of accuracy and fairness, as well as the harmful impact of surveillance on fundamental rights. This is especially concerning in light of the fact that India's surveillance regime already suffers from an inadequate lack of protections [8], which are intended to protect against the potential erosion of fundamental freedoms by law enforcement authorities. Therefore, AI-powered surveillance needs to be a specifically customized exception rather than the standard operating procedure.

The paper acknowledges that prejudice is inherent in data and that there is a chance that such bias would be reinforced over time when it is discussed as a problem regarding fairness in artificial intelligence (AI) systems. It suggests that one approach to this issue could be to "identify the in-built biases and analyze their impact, and in turn discover strategies to lessen the prejudice." This is one of the recommendations made in the report. It takes a ceteris paribus approach, which means that all other things remain equal, in order to simply identify and reduce bias in datasets one can hope to yield fairer results when the reality is that biassed data emerges from a biassed, unequal, discriminatory, and unfair world. Ceteris paribus is a Latin phrase that means "all other things remaining equal." This method has the drawback of viewing artificial intelligence solely through the lens of a mathematical model rather than as a socio-technical system. It is imperative that a proper understanding and adaptation of the social milieu in which these systems will operate in order to minimize the likelihood of bias occurring inside them. In point of fact, the goal should be to steer clear of consequences that are discriminatory. As Eubanks has pointed out, the deployment of automated decision-making tools can only serve to exacerbate existing structural injustices if these systems are not first designed to eliminate existing inequities.[9]

Access to data

At the risk of employing an overused metaphor, the data that are collected are the oil that powers AI. The construction of AI systems relies heavily on data that is readily available, easily accessible, precisely described, and reasonably priced. This is something that can be difficult

[8] Bailey R, Bhandari V, Parsheera S, Rahman F.2018Use of personal data by intelligence and law enforcement agencies. *Macro/Finance Group, National Institute of Public Finance and Policy*.
[9] Eubanks V. 2018 *Automating inequality: how high tech tools profile, police, and punish the poor*, pp. 190. New York, NY: St. Martin's Press.

to do in India. Data that are correct and pertinent to a certain setting are extremely uncommon to be easily accessible. For example, there is a significant lack of reporting of criminal activity in India [10]. The country's expanding FinTech and healthcare-related start-ups are the ones that are most affected by this limitation; relatively few of them have access to data that is both accurate and economical regarding the populations that they aim to serve. This issue, which Ryan Calo refers to as the challenge of data parity [11], is one in which only a select few well-established leaders in the area can obtain data and construct databases. However, this has only had moderate success in solving the data parity problem because the vast majority of quality data in India is restricted solely to the private sector [12]. India's National Data Sharing and Accessibility Policy contemplates sharing of non-sensitive data generated using public funds through the Open Data Platform.

Datasets that are used to train models to run the danger of having "black spots" [30] in which groups of people and communities are ignored. It may be more difficult to gather, obtain, or verify the data coming from underrepresented minorities and underprivileged groups. The ability to generalize based on specific examples is essential to machine learning. Suppose the examples that are used to train a machine learning system do not adequately represent certain groups that exist on the margins [13]of data collection. In that case, the generalization that is provided will discriminate against either the under-represented group or the over-represented group, depending on the circumstances of the given situation. This is especially problematic in jurisdictions like as India, where the ability to have a digital footprint is a result of privilege in the first place, whether that advantage is based on gender, caste, geographical location, or social status [14]. And yet, to proactively include individuals into datasets that they would have otherwise been left out for the purpose of increasing the accuracy of learning models is, in effect, to assume that the relevance of technologies that aid in the profiling and surveillance of those individuals is a given[15]. This is because including individuals in datasets that they would have otherwise been left out of is done to increase the accuracy of learning models. This dilemma is only made worse when we consider that monitoring is never neutral; rather, it is

[10] Dey A. 2017What the National Crime Records Bureau report does not tell us about Cyber Crime in India.
[11] Calo R. 2017 Artificial intelligence policy: a primer and roadmap. *U.C. Davis L. Rev.* 51, 398–435.
[12] Open Government Data (OGD) Platform India.
[13] Lerman J. 2013Big data and its exclusions. *Stanford Law Rev. Online* 66, 55–57.
[14] Sinha A, Rakesh V, Marda V.2017Big Data in Governance in India – Case Studies. *The Centre for Internet and Society*.
[15] Murali A.2018The Big Eye: The tech is all ready for mass surveillance in India. *Factor Daily*.

fundamentally disproportionate when considered in the context of gender, caste, race, and religion.

Systemic and historical bias

The statistics reflect the biases and prejudices of their development environment. This complicates matters. Systemic bias in data is problematic because it can affect decision-making. Take FaceTagr, a facial recognition tool used by police in Chennai, Tamil Nadu. Using FaceTagr, police can collect photos of "suspect" people and query criminal databases. According to Bhatia, this is problematic since "a guy" who "looks suspect" for wandering on the road at 2 AM comes from a certain socio-economic class, which can be evaluated by his appearance. Bhatia notes. FaceTagr makes it easy to target homeless or disadvantaged people[16]. Historical and systemic bias can cause many other problems. Some harmful examples perpetuate racism, sexism, violence, and hate. Not all problems have quantifiable consequences. Consider Google's AI-powered search algorithm. Google "south Indian masala" returns results for the woman, not the spice. This isn't a flaw in the algorithm but a reflection of cultural assumptions. By embodying the environment in which it was formed, data can cement, codify, and acquire undesirable preconceptions and biases. Data reflects its creative environment. Depending on how AI is implemented, these systems may benefit or harm society, limit or assist the exercise of rights, and widen the gap between protected and non-protected classes.

Privacy

Since machine learning systems can reliably draw inferences and sort data into useful categories, they have found widespread application in areas as diverse as marketing and law enforcement. The profiling they enable has far-reaching consequences for the ways in which we view and practice privacy and anonymity in our offline and online lives. The ability of AI systems to extract information from data, recognize patterns, and predict trends allows for seemingly irrelevant information to be mined to the point of relevance and intimacy.

Consider the several AI applications that have recently emerged in India. Considering FaceTagr's projected expansion plans to additional states in South India, as detailed above [17],

[16] Bhatia G.In press. *The transformative constitution: a radical biography in nine acts*. HarperCollins.
[17] Privacy and freedom of expression in the age of artificial intelligence. *ARTICLE 19 and Privacy International*.

a more thorough analysis of the hazards involved is necessary. Law enforcement agencies in the Punjab region use the Punjab Artificial Intelligence System to digitize criminal records with the help of "proprietary, advanced hybrid AI technology"[18] and to facilitate criminal search with tools like facial recognition, which can be used to predict and identify criminal behavior. Meanwhile, researchers at the University of Cambridge have published a report titled "Eye in the sky." It describes their ambitions to train and test drones at Indian music festivals to identify hostile behavior in public [19]. The majority of these projects have as their overarching goal the reduction of crime rates, the control of crowding in public spaces, and the simplification of police procedures. While the majority of machine learning systems are well-intended, we saw in earlier chapters that this is not always the case. Surveillance and privacy violations are possible since the aforementioned devices not only function with questionable degrees of accuracy [20], but also run without restrictions to prevent their misuse.

This may seem like excellent news at first glance since if the contentious facial recognition technology is wrong, then the likelihood of one's privacy being compromised is low. However, this only makes things worse when it comes to applications currently used by law enforcement in India, as these technologies can result in erroneous arrests and place an undue burden on members of India's most vulnerable and oppressed communities by forcing them to prove their innocence despite legitimate privacy concerns. Furthermore, regardless of accuracy rates, public behavior and faces continue to be captured, kept, and occasionally shared or accessed without people's knowledge or agreement. The level of privacy protection or breach is also influenced by the laws that govern the use of AI technologies. The right to privacy was upheld as a basic right under the Indian Constitution in August 2017 by a majority ruling of the Supreme Court of India. The threats presented by more complex data analysis and machine learning were brought to light by this historic judgment. A "strong system for data protection" is urgently needed in the country, and the ruling emphasized the necessity of protecting people's privacy, autonomy, and identities.

In particular, the Court observed; Right to privacy includes protection from unauthorized disclosure of personal information. In this digital age, threats to privacy are not limited to governmental agencies but can also come from private entities. We urge the Union Government to look into and implement a thorough data protection regime. Establishing such a system needs

[18] Sathe G. 2018Cops in India are using artificial intelligence that can identify you in a crowd. *Huffington Post*.
[19] Vincent J.2018Drones taught to spot violent behavior in crowds using AI. *The Verge*.
[20] Staff Reporters. 2018Police facial recognition software is inaccurate. *The Hindu*.

a nuanced and thoughtful balancing act between private interests and the state's legitimate concerns. India's progress toward a comprehensive data protection framework is at a crossroads as this piece is being written. The right to privacy, on the other hand, has been resoundingly reaffirmed by the Supreme Court. However, the Union government has established a committee to study and provide recommendations on data protection regulations in the country, and that committee has just produced a draught of the Personal Data Protection Bill, 2018[21]. (Justice B. N. Srikrishna, a retired judge from India's highest court, heads the committee from which the name "Srikrishna Committee" is derived.) The bill proposal includes some commendable measures to ensure the privacy of sensitive information. This includes requiring data fiduciaries (broadly defined as all legal entities that process data, including individuals, governments, and enterprises) to follow the principles of purpose limitation, collection limitation, and data breach notification. It restates that permission may only be considered legitimate if it meets certain criteria, including being freely given, well-understood, specific, unambiguous, and revocable. It also makes it so that more specific permission is needed for the collection of sensitive personal information.

Nonetheless, the Bill's intended privacy protections are undermined by some clauses that create worryingly broad exclusions for government data processing. In its current form, the Bill allows the government to process personal data without consent if it is shown to be necessary, for "any function of Parliament or any State Legislature," necessary for the exercise of the State's "authorized by law for the provision of any service or benefit," or necessary for "the issuance of any certification, license, or permit for any action or activity." If it can be proven that processing the data is "strictly necessary," for example, in order to "perform any function of Parliament or any State Legislature" or "perform the exercise of any function of the State authorized by law for the provision of any service or benefit," then the data can be processed without the individual's consent.

This effectively means that citizens have next to no recourse with respect to State data processing.For machine learning systems, however, the distinction between the "strictly necessary" and "necessary" categories of personal data is largely academic. When trained on big datasets that include both personal and sensitive personal data, these algorithms will essentially absorb, entrench, and amplify biases and discrimination. Proxy information is still available when sensitive attributes like caste, religion, political leanings, and sexual orientation

[21] The Personal Data Protection Bill. 2018

are obscured from view[22]. Furthermore,' several slightly correlated features can be used to build high accuracy classifiers for the sensitive attribute,' they explain. The Bill is quiet on the issues of accountability and transparency inside intelligence services, and it does little to address surveillance concerns that afflict India's legal structure at the present time. As a result, there is a systemic risk that the privacy implications of AI applications will be negative.However, the State is simultaneously deploying artificial intelligence (AI) applications that engage in surveillance and profiling, effectively giving them free reign in this regard, and the legal framework already contemplates unrestricted State processing of both sensitive personal data and personal data.

Model

Although they may share some similarities, the models or decision frameworks that emerge from training data each have their own set of restrictions and raise their own set of ethical problems. The model selection involves subjective choices, as Veale and Binns[23] point out.

It is important to keep an eye on how models are behaving, especially to detect any signs of bias. Even if a model is given a flawless dataset (if such a thing existed), it can still be biased. Attribute weights, which in turn impact how a model performs, are determined by design choices made at the model level, such as feature engineering.

It is well acknowledged that feature selection is one of the most difficult aspects of machine learning since it is problem-specific. Even if proper safeguards were implemented at the time of dataset construction, discrimination against members of protected classes is still possible if feature selection involves selecting features that either serve as proxies to protected attributes[24] or because the choice of features does not put protected groups and non-protected groups on an equal footing for accurate determinations. Take, for example, the Indian government's ambitions to introduce a social media sentiment analysis tool to "assist enable developing a 360-degree perspective of the people who are producing buzz across various issues."[25] Putting aside for the moment the profound effects on individual liberties of expression and privacy that

[22] Barocas S, Hardt M, Narayanan A.2018 Fairness and machine learning, 41.
[23] Veale M, Binns R.2017Fairer machine learning in the real world: mitigating discrimination without collective sensitive data. *Big Data Soc.* 4, 1–17.
[24] Barocas S, Selbst A. 2016Big Data's Disparate Impact. 104. *California Law Rev.* 671, 671–731.
[25] Request for Proposals 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.*

such a tool would have, let us instead investigate how it might be technically constructed. The program would need training based on data and parameters. The classifications and outputs that such a tool would produce, and the way in which it would disproportionately affect vulnerable communities, would be profoundly influenced by the subjective decisions made at the time of feature selection—which aspects of individuals, speech and online behaviour to amplify and which to downplay.

While non-protected groups are prone to discrimination all across the world, in India, the risk to women, sexual minorities, and members of lower castes is particularly acute due to rigid conventional biases and societal, and cultural norms. For example, assume a case where a company is trying to hire an office manager. Instead of looking through qualitative prior references and work experience, machine learning software may be created to attach great weight to the individual's ability to spend long hours at the workplace. This is a handy perk, even if it has no bearing on a person's genuine enthusiasm and dedication to their work. Given women's relatively recent participation in the workforce and traditionally rigid social standards around an Indian woman's responsibilities at home, it is feasible that lengthy hours at the office may be represented as something that males accomplish more efficiently than women. Concerns concerning the fairness and nondiscriminatory nature of models have arisen in response to the growing reliance on machine learning systems for making important judgments. A lot of challenges arise in the context of fairness in AI systems, despite the seeming simplicity of the goal.

The first one is the difficulty associated with the definition. Computer scientists have come up with a variety of definitions in an effort to make the concepts of fairness and non-discrimination more applicable to real-world situations[26]. One common interpretation of the term "demographic parity" describes the situation in which statistical models give the same results for members of protected and non-protected groups despite the fact that the choice has no connection to the protected characteristic. This has two major flaws that need to be addressed. To begin, it may be construed as undermining desirable accuracy in situations of individual fairness in an effort to achieve group fairness[27]. This may be the case for a few reasons. Second, it disregards the situations in which discrimination is required for valid reasons, such as the

[26] Narayanan A. 2018Translation tutorial: 21 definitions of fairness and their politics. In Fairness, Accountability, and Transparency in Machine Learning Conf. 2018.
[27] Barocas S, Bradley E, Honavar V, Provost F.2017Big Data, Data Science, and Civil Liberties. *A Computing Community Consortium.*

implementation of equitable affirmative action[28]. The article "Machine Bias" by ProPublica explores accuracy equity, which recognizes that even having the same error rates for various demographic groups can preclude justice, especially if the types of errors differ[29]. One such definition of fairness is the concept of counterfactual fairness, in which a judgment is thought to be fair if it would be the same in both the actual world and a world in which the individual concerned did not have the protected attribute[30].

The issue of making compromises is the second layer of complexity. Although the trade-offs between the many interpretations of fairness in a particular scenario[31] are already complex enough, it is also necessary to take into account the trade-offs between fairness and other competing values. It has been demonstrated that getting rid of discrimination in a model makes it less accurate overall[32] . On the other hand, striving for accuracy in models might be inherently unfair because it necessitates adopting discriminating and unfair social norms and practices that already exist in society.

Consequently, the development of tools, checklists, and standards for determining which conceptions of fairness are most suited to certain situations becomes a significant obstacle to overcome in the realm of public policy. Both "explicitly and implicitly"[33] affirmative actions is supported by the Indian Constitution, which envisions India as a welfare state and recognizes the need for it. When it comes to artificial intelligence systems, the status of populations that are guaranteed protection under the constitution, as well as policies regarding affirmative action in education, housing, and employment, will play a significant role in determining what appropriate standards of fairness should look like.

Disclosure and responsibility

Transparency and accountability in these systems become especially important in constitutional democracies like India's when AI applications replace decision-making in the

[28] Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R.2011Fairness through awareness. *Computing Complexity*.
[29] Angwin J, Larson J, Mattu S, Kirchner L.2016Machine Bias. *ProPublica*.
[30] Kusner M, Loftus J, Russel C, Silva R. 2018Counterfactual Fairness. In 31st Conf. on Neural Information Processing Systems (NIPS 2017).
[31] Kleinberg J, Mullainathan S, Raghavan M. 2016Inherent trade-offs in the fair determination of risk scores. In Proc. of Innovations in Theoretical Computer Science.
[32] 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*.
[33] Jacobsohn GJ. 2016Constitutional identity. In *The Oxford handbook of The Indian constitution*. Oxford, UK: Oxford University Press.

public sector[34] . Concerns concerning accountability and restitution methods have arisen in response to AI systems' opaque, sophisticated, and unseen nature, especially after its deployment in fields like criminal justice and policing. The approaches used to solve these issues are frequently opaque[35], which threatens public trust in government. This has sparked a lively debate on the importance of holding learning algorithms to account, with more transparency, as one proposed mechanism[36].

It is controversial where exactly these concepts of openness are based[37] . One way to ensure systems are held responsible is through transparency regulations requiring them to open up impenetrable black boxes and allow outside observers to peep inside[38]. This is based on the assumption that machine learning systems are more trustworthy when their users understand the data and models they use and so have the ability to identify instances of bias or injustice. However, this assertion has been called "evident but stupid" due to the fact that AI systems that evolve over time are incomprehensible, even if all of their workings are made public. According to Burrell, opacity in machine learning algorithms can be explained by technical ignorance, deliberate concealment, or a combination of the two[39] . Whatever the case may be, the primary problem is how to hold complex, unpredictable systems accountable. While there has been a movement toward requiring that systems be explicable, scrutable, and understandable, experts like Winfield[40] maintain that transparency is essential to comprehend the actions of AI technologies.

The accountability of AI systems can be achieved through several means, and policymakers would be wise to consider a spectrum of transparency and intelligibility as they consider these options. However, the degree of transparency that is actually achievable relies on the sort of model utilized, and the degree to which it is required varies depending on the nature of the AI application and the function it is designed to fulfill.

---

[34] Veale M, Kleek MV, Binns R. 2018Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems, Paper No. 440.
[35] Mittelstadt B, Patrick A, Taddeo M, Wachter S, Floridi L. 2016 The ethics of algorithms: mapping the debate. *Big Data Soc.* 3, 1–21.
[36] Diakopoulos N. 2014Algorithmic accountability reporting: on the investigation of black boxes. *Tow Centre for Digital Journalism*.
[37] Marda V. 2017 Machine learning and transparency: a scoping exercise. *Working Paper Series*.
[38] Pasquale F. 2015 *The black box society*, p. 106. Cambridge, MA: Harvard University Press.
[39] Burrell J.2016How the machine ‘thinks’: understanding opacity in machine learning algorithms. *Big Data & Society*, January – June 2016, pp. 1–12.
[40] House of lords select committee on artificial intelligence. 2018AI in the UK: ready, willing and able?*House of Lords* 36.

Issues with data collecting bias

It is possible that datasets used to train models have 'black spots' in which groups of people or entire categories of people are ignored. It could be more challenging to gather, obtain, or validate data from marginalised communities. Generalization from instances is essential for machine learning. If a machine learning system is trained on examples that do not adequately represent particular groups, especially those on the periphery of data gathering, the resulting generalisation may be biassed towards those groups. This is especially concerning in countries like India, where access to a digital footprint depends on factors such as gender, caste, region, and socioeconomic status [41]. To assume the relevance of technologies that aid in profiling and monitoring of persons is a given is, however, to deliberately incorporate individuals into datasets that they would have otherwise been left out of for the purpose of accuracy in learning models. This issue is exacerbated by the fact that monitoring is inherently biassed based on factors such as a person's gender, social status, colour, or religion [42].

Freedom of expression

Under Indian law, the right to freedom of expression is guaranteed as a basic freedom.[43] The highest court in India has often cited it as a cornerstone of democracy, and it has also ruled that the right to information is an intrinsic aspect of that freedom[44]. The rising usage of AI applications in daily life, ranging from smart assistants to autocorrect technology on mobile devices, significantly impacts the right to freedom of speech[45]. Artificial intelligence (AI) is being promoted by internet corporations and governments as a silver bullet for thorny issues like online bigotry, terrorist recruitment, and false news. Given the limitations of machine learning's ability to recognize tone and context, this is a worrying trend. The unilateral nature of automated content removal by private corporations, often in response to government instruction, increases the potential of overly wide censorship and take-down of valid expression. Consequences for privacy and free speech are both affected by surveillance driven by AI technologies. Artificial intelligence (AI) powered monitoring has a chilling impact on

[41] Sinha A, Rakesh V, Marda V.2017Big Data in Governance in India – Case Studies. *The Centre for Internet and Society*.
[42] Shepard N.2016Big Data and Sexual Surveillance. *Association for Progressive Communications*.
[43] Article 19(1)(a), Constitution of India.
[44] *State of Uttar Pradesh v. Raj Narain.* (1975) 3 SCR 333.
[45] Privacy and freedom of expression in the age of artificial intelligence. *ARTICLE 19 and Privacy International*.

expression since it blurs the distinctions between the private and the public, leading many people to resort to self-censorship out of fear of the consequences of their words.

Tools that use sentiment analysis to determine the flavor of online discourse are becoming more common, with the latter often being programmed to do automated content removal. In 2016, with Project Insight[46] and most recently with a public tender issued by the Ministry of Information and Broadcasting, the Indian government has expressed interest in moving toward using AI to carry out sentiment analysis, identify fake news, and boost India's image across social media platforms and even e-mail. When governments try to regulate content using AI, the same problems plagued private companies when they tried it [47].

The Supreme Court of India made a strong statement about the connection between technology and free speech in 2015, [48]when it overturned Section 66A of the country's Information Technology Act. Offensive, threatening, or unpleasant messages sent via the internet were considered criminal offenses under this section. For being overbroad, ambiguous, and restrictive of free expression, the court ruled that this part of the statute must be nullified. The judgment emphasized the bounds of appropriate constraints to free speech under Constitutional law in the age of technology while reaffirming the value of democracy, educated citizenry, and an open culture of discourse in India's tradition of free speech. It is important to examine the push toward AI solutions, especially by State actors, in the context of current law in India.

Using this classification method, you may analyze the potential for bias, discrimination, surveillance, and profiling in an automated learning programme. The aforementioned research and concerns can be applied to any industry and serve as a guide for weighing the potential consequences of AI.

CONCLUSION

This study aimed to impact AI policy deliberations in India and stimulate cross-disciplinary conversation. It analyzed India's current policy landscape and argued that data-driven decision-making restrictions should be a fundamental, not retrospective, factor in AI policy development. Given the many initiatives in the sector, it concentrated on minor parts of the

[46] Seth S.2017Machine Learning and Artificial Intelligence Interactions with the Right to Privacy. *Econ. Political Weekly*, vol. LII 51, 66–70.
[47] Marda V.2018Regulating social media content: why AI alone cannot solve the problem. *ARTICLE 19*.
[48] *Shreya Singhal v. Union of India.* AIR 2015 SC 1523.

discussion. By describing machine learning, I've shown that data-driven decision-making is prone to mistakes, discriminating outcomes, inherent and amplified prejudice, and unforeseen effects. Technical research is looking for methods to resolve these concerns, and policy must do the same. The proposed framework aims to bridge the gap and build a shared understanding. It shows that AI systems can't be seen as separate mathematical problems, impartial, or just efficient. AI technologies are complicated social systems that can't be evaluated on efficiency and accuracy alone. Given the complexity of AI concerns, future deliberation, policy-making, and regulation must involve numerous disciplines. Ethics, law, technology, and philosophy must be considered throughout. Development is fast, opaque, and frequently irreversible. AI will change the way we create procedures and deploy technologies. I hope the proposed paradigm will help researchers, policymakers, attorneys, and technologists clarify AI's challenges and prospects in various contexts.