THE CENTRE FOR internet & society

AI in Governance

Roundtable Event Report

16 March, 2018 | New Delhi, India

By **SAMAN GOUDARZI** & **NATALLIA KHANIEJO**

The Centre for Internet and Society, India

Designed by **Saumyaa Naidu**

This report provides a summary of the proceedings of the Roundtable on Artificial Intelligence (AI) in Governance (hereinafter referred to as 'the Roundtable'). The Roundtable took place at the India Islamic Cultural Centre in New Delhi on March 16, 2018 and included participation from academia, civil society, law, finance, and government. The main purpose of the Roundtable was to discuss the deployment and implementation of AI in various aspects of governance within the Indian context.

The Roundtable began with a presentation by Amber Sinha (Centre for Internet and Society - CIS) providing an overview of the CIS's research objectives and findings thus far. During this presentation, he defined both AI and the scope of CIS's research, outlining the areas of law enforcement, defense, education, judicial decision making, and the discharging of administrative functions as the main areas of concerns for the study. The presentation then outlined the key AI deployments and implementations that have been identified by the research in each of these areas. Lastly, the presentation raised some of the ethical and legal concerns related to this phenomenon.

The presentation was followed by the Roundtable discussion that saw various topics in regards to the usages, challenges, ethical considerations and implications of AI in the sector being discussed. This report has identified a number of key themes of importance evident throughout these discussions. These themes include: (1) the meaning and scope of AI, (2) AI's sectoral applications, (3) human involvement with automated decision making, (4) social and power relations surrounding AI, (5) regulatory approaches to AI and, (6) challenges to adopting AI. These themes in relation to the Roundtable are explored further below.

Meaning & Scope of AI

One of the first tasks recommended by the group of participants was to define the meaning and scope of AI and the way those terms are used and adopted today. These concerns included the need to establish a distinction between the use of algorithms, machine learning, automation and artificial intelligence. Several participants believed that establishing consensus around these terms was essential before proceeding towards a stage of developing regulatory frameworks around them.

The general fact agreed to was that AI as we understand it does not necessarily extend to complete independence in terms of automated decision making but it refers instead to the varying levels of machine learning (ML), and the automation of certain processes that has already been achieved. Several concerns that emerged during the course of the discussion centred around the question of autonomy and transparency in the process of ML and algorithmic processing. Stakeholders recommended that over and above the debates of humans in the loop[1], on the loop[2] and out of the loop[3], there were several other gaps with respect to AI and its usage in the industry today which also need to be considered before building a roadmap for future usage. Key issues like information asymmetries, communication lags, a lack of transparency, the increased mystification of the coding process and the centralization of power all needed to be examined and analysed under the rubric of developing regulatory frameworks.

Takeaway Point

The group brought out the need for standardization of terminology as well as the establishment of globally replicable standards surrounding the usage, control and

1 Automated decision making model where final decisions are made by a human operator

2 Automated decision making model where decisions can be made without human involvement but a human can override the system

3 A completely autonomous decision making model requiring no human involvement

proliferation of AI. The discussion also brought up the problems with universal applicability of norms. One of the participants brought up an issue regarding the lack of normative frameworks around the usage and proliferation of AI. Another participant responded to the concern by alluding to the Asilomar AI principles[4]. The Asilomar AI principles are a set of 23 principles aimed at directing and shaping AI research in the future. The discussion brought out further issues regarding the enforceability as well universal applicability of the principles and their global relevance as well. Participants recommended the development of a shorter, more universally applicable regulatory framework that could address various contextual limitations as well.

AI Sectoral Applications

Participants mentioned a number of both current and potential applications of AI technologies, referencing the defence sector, the financial sector, and the agriculture sector.

There are several developments taking place on the Indian military front with the Committee on AI and National Security being established by the Ministry of Defence. Through the course of the discussion it was also stated that the Indian Armed Forces were very interested in the possibilities of using AI for their own strategic and tactical purposes. From a technological standpoint, however, there has been limited progress in India in researching and developing AI. While India does deploy some Unmanned Aerial Vehicles (UAVs), they are mostly bought from Israel, and often are not autonomous. It was also pointed out that contrary to reportage in the media, the defence establishment in India is extremely cautious about the adoption of autonomous weapons systems, and that the autonomous technology being rolled out by the CAIR is not yet considered trustworthy enough for deployment.

Discussions further revealed that the few technologies that have a relative degree of autonomy are primarily loitering ammunitions and are used to target radar insulations for reconnaissance purposes. One participant mentioned that while most militaries are interested in deploying AI, it is primarily from an Intelligence, Surveillance and Reconnaissance (ISR) perspective. The only exception to this generalization is China where the military ethos and command structure would work better with increased reliance on independent AI systems. One major AI system rolled out by the US is Project Maven which is primarily an ISR system. The aim of using these systems is to improve decision making and enhance data analysis particularly since battlefields generate a lot of data that isn't used anywhere.

Another sector discussed was the securities market where algorithms were used from an analytical and data collection perspective. A participant referred to the fact that machine learning was being used for processes like credit and trade scoring -- all with humans on the loop. The participant further suggested that while trade scoring was increasingly automated, the overall predictive nature of such technologies remained within a self limiting capacity wherein statistical models, collected data and pattern analysis were used to predict future trends. The participant questioned whether these algorithms could be considered as AI in the truest sense of the term since they primarily performed statistical functions and data analysis.

One participant also recommended the application of AI to sectors like agriculture with the intention of gradually acclimatizing users to the technology itself. Respondents also stated that while AI technologies were being used in the agricultural space it was primarily from the standpoint of data collection and analysis as opposed to predictive methods. It was mentioned that a challenge to the broad adoption of AI in this sector is the core problem of adopting AI as a methodology – namely information asymmetries, excessive data collection, limited control/centralization and the obfuscatory nature of code – would not be addressed/

4 https://futureoflife.org/ai-principles/

modified. Lastly, participants also suggested that within the Indian framework not much was being done aside from addressing farmers' queries and analysing the data from those concerns.

Takeaway Point

The discussion drew attention to the various sectors where AI was currently being used -- such as the military space, agricultural development and the securities market -- as well as potential spaces of application -- such as healthcare and manual scavenging. The key challenges that emerged were information asymmetries with respect to the usage of these technologies as well as limited capacity in terms of technological advancement.

Human Involvement with Automated Decision Making

Large parts of discussions throughout the Roundtable event were preoccupied with automated decision making and specifically, the involvement of humans (human on and in the loop) or lack thereof (human out of the loop) in this process. These discussions often took place with considerations of AI for prescriptive and descriptive uses.

Participants expressed that human involvement was not needed when AI was being used for descriptive uses, such as determining relationships between various variables in large data sets. Many agreed to the superior ability of ML and similar AI technologies in describing large and unorganized datasets. It was the prescriptive uses of AI where participants saw the need for human involvement, with many questioning the technology making more important decisions by itself.

The need for human involvement in automated decision making was further justified by references to various instances of algorithmic bias in the American context. One participant, for example, brought up the use of algorithmic decision making by a school board in the United States for human resource practices (hirings, firing, etc.) based on the standardized test scores of students. In this instance, such practices resulted in the termination of teachers primarily from low income neighbourhoods[5]. The main challenge participants identified in regards to human on the loop automated decision making is the issue of capacity, as significant training would have to be achieved for sectors to have employees actively involved in the automated decision making workflow.

An example in the context of the healthcare field was brought up by one participant arguing for human in the loop in regards to prescriptive scenarios. The participant suggested that AI technology, when given x-ray or MRI data for example, should only be limited to pointing out the correlations of diseases with patients' scans/x-rays. Analysis of such correlations should be reserved for the medical expertise of doctors who would then determine if any instances of causality can be identified from this data and if it's appropriate for diagnosing patients.

It was emphasized that, despite a preference for human on/in the loop in regards to automated decision making, there is a need to be cognisant of techno-solutionism due to the human tendency of over reliance on technology when making decisions. A need for command and control structures and protocols was emphasized for various governance sectors in order to avoid potentially disastrous results through a checks and balances system. It was noted that the defense sector has already developed such protocols, having established a chain of command due to its long history of algorithmic decision making (e.g. the Aegis Combat System being used by the US Navy in the 1980s).

5 The participant was drawing this example from Cathy O'Neil's Weapons of Math Destruction, (Penguin,2016), at 4-13.

One key reason why militaries prefer human in and on the loop systems as opposed to out of the loop systems is because of the protocol associated with human action on the battlefield. International Humanitarian Law has clear indicators of what constitutes a war crime and who is to be held responsible in the scenario but developing such a framework with AI systems would be challenging as it would be difficult to determine which party ought to be held accountable in the case of a transgression or a mistake.

Takeaway Point

It was reiterated by many participants that neither AI technology or India's regulatory framework is at a point where AI can be trusted to make significant decisions alone -- especially when such decisions are evaluating humans directly. It was recommended that human out of the loop decision making should be reserved for descriptive practices whereas human on and in the loop decision making should be used for prescriptive practices. Lastly, it was also suggested that appropriate protocols be put in place to direct those involved in the automated decision making workflow. Particularly when the process involves judgements and complex decision making in sectors such as jurisprudence and the military.

The Social & Power Relations Surrounding AI

Some participants emphasized the need to contextualize discussions of AI and governance within larger themes of poverty, global capital and power/social relations. Their concerns were that the use of AI technologies would only create and reinforce existing power structures and should instead be utilized towards ameliorating such issues. Manual scavenging, for example, was identified as an area where AI could be used to good effect if coupled with larger socio-political policy changes. There are several hierarchies that could potentially be reinforced through this process and all these failings needed to be examined thoroughly before such a system was adopted and incorporated within the real world.

Furthermore the discussion also revealed that the objectivity attributed to AI and ML tends to gloss over the fact that there are nonetheless implicit biases that exist in the minds of the creators that might work themselves into the code. Fears regarding technology recreating a more exclusionary system were not entirely unfounded as participants pointed out the fact that the knowledge base of the user would determine whether technology was used as a tool of centralization or democratization.

One participant also questioned the concept of governance itself, contrasting the Indian government's usage of the term in the 1950s (as it appears in the Directive Principle) with that of the World Bank in the 1990s.

Takeaway Point

Discussions of the implementation and deployment of AI within the governance landscape should attempt to take into consideration larger power relations and concepts of equity.

Regulatory Approaches to AI

Many recognized the need for AI-specific regulations across Indian sectors, including governance. These regulations, participants stated, should draw from notions of accountability, algorithmic transparency and efficiency. Furthermore, it was also stated that such regulations should consider the variations across the different legs of the governance sector, especially in regards to defence. One participant, pointing to the larger trends towards automation, recommended the establishment of certain fundamental guidelines

aimed at directing the applicability of AI in general. The participant drew attention to the need for a robust evaluation system for various sectors (the criminal justice system, the securities market, etc.) as a way of providing checks on algorithmic biases. Another emphasized for the need of regulations for better quality data as to ensure machine readability and processiblity for various AI systems.

Another key point that emerged was the importance of examining how specific algorithms performed processes like identification or detection. A participant recommended the need to examine the ways in which machines identify humans and what categories/biases could infiltrate machine-judgement. They reiterated that if a new element was introduced in the system, the pre-existing variables would be impacted as well. The participant further recommended that it would be useful to look at these systems in terms of the couplings that get created in order to determine what kinds of relations are fostered within that system.

The roundtable saw some debate regarding the most appropriate approach to developing such regulations. Some participants argued for a harms-based approach, particularly in regards to determining if regulations are needed all together for specific sectors (as opposed to guidelines, best practices, etc.). The need to be cognisant of both individual and structural harms was emphasized, mindful of the possibility of algorithmic biases affecting traditionally marginalized groups.

Others only saw value in a harms based approach insomuch that it could help outline the appropriate penalties in an event of regulations being violated, arguing instead for a rights-based approach as it enabled greater room for technological changes. An approach that kept in mind emerging AI technologies was reiterated by a number of participants as being crucial to any regulatory framework. The need for a regulatory space that allowed for technological experimentation without the fear of constitutional violation was also communicated.

Takeaway Point

The need for a AI-specific regulatory framework cognisant of differentiations across sectors in India was emphasized. There is some debate about the most appropriate approach for such a framework, a harms-based approach being identified by many as providing the best perspective on regulatory need and penalties. Some identified the rights-based approach as providing the most flexibility for an rapidly evolving technological landscape.

Challenges to the Adopting AI

Out of all the concerns regarding the adoption of algorithms, ML and AI, the two key points of resistance that emerged, centred around issues of accountability and transparency. Participants suggested that within an AI system, predictability would be a key concern, and in the absence of predictable outcomes, establishing redressal mechanisms would pose key challenges as well.

A discussion was also initiated regarding the problems involved in attributing responsibility within the AI chain as well as the need to demystify the process of using AI in daily life. While reiterating the current landscape, participants spoke about how the usage of AI is currently limited to the automation of certain tasks and processes in certain sectors where algorithmic processing is primarily used as a tool of data collection and analysis as opposed to an independent decision making tool.

One of the suggestions and thought points that emerged during the discussion was whether a gradual adoption of AI on a sectoral basis might be more beneficial as it would provide breathing room in the middle to test the system and establish trust between the developers, providers, and consumers. This prompted a debate about the controllers and the consumers of AI and how the gap between the two would need to be negotiated. The debate also

brought up larger concerns regarding the mystification of AI as a process itself and the complications of translating the code into communicable points of intervention.

Another major issue that emerged was the question of attribution of responsibility in the case of mistakes. In the legal process as it currently exists, human imperfections notwithstanding, it would be possible to attribute the blame for decisions taken to certain actants undertaking the action. Similarly in the defence sector, it would be possible to trace the chain of command and identify key points of failure, but in the case of AI based judgements, it would be difficult to place responsibility or blame. This observation led to a debate regarding accountability in the AI chain. It was inconclusive whether the error should be attributed to the developer, the distributor or the consumer.

A suggestion that was offered in order to counter the information asymmetry as well as reduce the mystification of computational method was to make the algorithm and its processes transparent. This sparked a debate, however, as participants stated that while such a state of transparency ought to be sought after and aspired towards, it would be accompanied by certain threats to the system. A key challenge that was pointed out was the fact that if the algorithm was made transparent, and its details were shared, there would be several ways to manipulate it, translate it and misuse it.

Another question that emerged was the distribution of AI technologies and the centralization of the proliferation process particularly in terms of service provision. One participant suggested that given the limited nature of research being undertaken and the paucity of resources, a limited number of companies would end up holding the best tech, the best resources and the best people. They further suggested that these technologies might end up being rolled out as a service on a contractual basis. In which case it would be important to track how the service was being controlled and delivered. Models of transference would become central points of negotiation with alternations between procurement based, lease based, and ownership based models of service delivery. Participants suggested that this was going to be a key factor in determining how to approach these issues from a legal and policy standpoint.

Takeaway Point

The two key points of resistance that emerged during the course of discussion were accountability and transparency. Participants pointed out the various challenges involved in attributing blame within the AI chain and they also spoke about the complexities of opening up AI code, thereby leaving it vulnerable to manipulation. Certain other challenges that were briefly touched upon were the information asymmetry, excessive data collection, centralization of power in the hands of the controllers and complicated service distribution models.

Conclusion

The Roundtable provided some insight into larger debates regarding the deployment and applications of AI in the governance sector of India. The need for a regulatory framework as well as globally replicable standards surrounding AI was emphasized, particularly one mindful of the particular needs of differing fields of the governance sector (especially defence). Furthermore, a need for human on/in the loop practices with regards to automated decision making was highlighted for prescriptive instances, particularly when such decisions are responsible for directly evaluating humans. Contextualising AI within its sociopolitical parameters was another key recommendation as it would help filter out the biases that might work themselves into the code and affect the performance of the algorithm. Further, it is necessary to see the involvement and influence of the private sector in the deployment of AI for governance, it often translating into the delivery of technological services from private

actors to public bodies towards discharge of public functions. This has clear implications for requirements of transparency and procedural fairness even in private sector delivery of these services. Defining the meaning and scope of AI while working to demystify algorithms themselves would serve to strengthen regulatory frameworks as well as make AI more accessible for the user/consumer.