To,
Department for Promotion of Industry and Internal Trade
ecom-dipp[at]gov[dot]in
goonjan[dot]kumar[at]gov[dot]in
[Submitted Electronically]

29 March 2019

Subject: Comments on 'DPIIT Draft National E-commerce Policy'

Dear Ma'am/Sir,

I am research fellow with the Center for Long Term Cybersecurity at UC, Berkeley. As a fellow based in India I study the applications and implications of incorporating Machine Learning and AI in essential services in the country. Many of the companies and use cases that I study fall under the ambit of the e-commerce policy. My comments are derived from my research.

While the draft touches a number of themes, my comments will be restricted to policy concerns around the future of Artificial Intelligence in the country; and balancing both interests of both consumers and domestic enterprises in this future. In specific, I have responded to Section 4 of the draft e-commerce policy. I hope my comments will be useful in the final policy formulation.

Please feel free to contact me for any clarifications or further information.

Sincerely,
Tarunima Prabhakar
Research Fellow,
Center for Long Term Cybersecurity, UC, Berkeley
tarunima@berkeley.edu
+919971199366

Table Of Contents

Executive Summary

The draft e-commerce policy (henceforth "the draft policy") highlights important concerns around protecting Indian consumers as well as domestic e-commerce platforms in a rapidly evolving digital economy. It is motivated by concerns of domestic economic growth; security concerns around privacy, counterfeit goods, law and order violations; and fairness of digital platforms. The draft policy rightly recognizes the importance of data in the digital economy but by focusing disproportionately on data misses critical strategies for enabling a competitive digital ecosystem; and for ensuring fairness of service to consumers through these platforms. The goals of data localization, one of the draft policy's primary recommendations are better served by alternate strategies such as right to data portability.

The draft policy's suggestion of an electronic grievance redressal mechanism; clamping down on inaccurate reviews on e-commerce platforms; and unsolicited commercial messages are great steps for consumer protection in the digital economy. Opening of source code however does not inherently ensure explainability or equality in decisions of AI platforms. Instead, methods of auditing algorithms with private sector participation should be explored. The liability of accumulating excessive data should be considered seriously, and data lifecycles that set clear timelines for data deletion should be defined. Such guidelines should be sector specific.

It is also recommended that an e-commerce policy look at different aspects of an e-commerce platform more holistically, and institute strategies for innovation at the intersection of data and algorithms. The draft policy highlights important concerns on pricing strategies by global monopolies that can under out-price local competition. This is a complex issue that can be handled by multiple police levers, such as taxation and competition law. This should not be discussed in isolation in an e-commerce policy.

Contrary to seeking source code disclosure, the e-commerce policy should consider building local capacity to innovate at the cutting edge of a global competitive market. To this end, it should explore recommendations of the Report of the AI Task Force released by the Department

in 2018. Investment in research and development, education, re-skilling and international bilateral cooperations are more sustainable strategies for technology transfer.

Finally, unbridled optimism on application of AI in governance can be dangerous. Not all policy challenges lend themselves to be resolved by AI. AI systems come with additional concerns around opacity of decisions through algorithmic systems and negotiating values of fairness encoded in them; and might not necessarily improve the efficiency of government processes. AI as a strategy for governance needs considerable thought and oversight; and should be covered by a dedicated policy. This could possibly be undertaken by The National Artificial Intelligence Mission (N-AIM) recommended by the AI Task Force.

Overarching Concerns

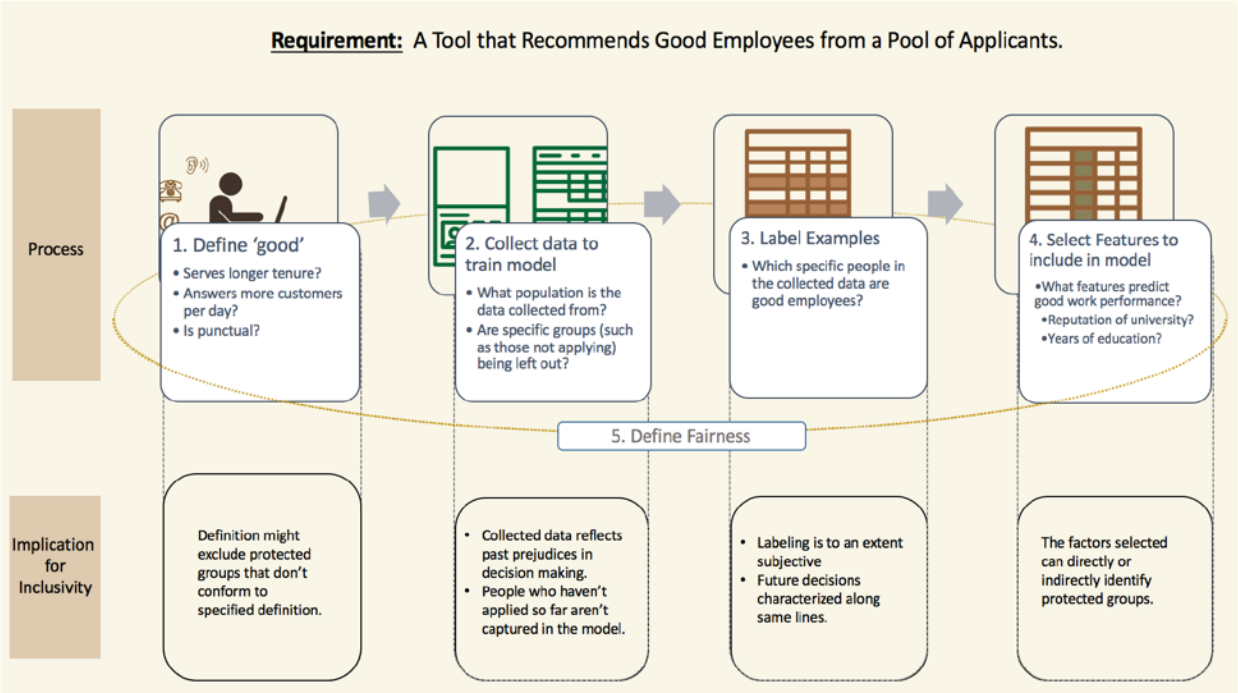
Over-emphasis on Data at the Cost of Other Factors in AI

The draft policy in equating data with oil rightfully recognizes the importance of data in powering online platforms. Access to data is a necessary condition for big data applications. The draft policy also recognizes the need for good infrastructure. However, the policy ignores other aspects that are equally important for a robust digital economy- t*echnical innovation in algorithm design* and *human talent*. Several recommendations in the draft while written from the perspective of increasing access to data locally, are antagonistic to promoting the other two.

TECHNICAL INNOVATION IN ALGORITHM DESIGN:

If data is the oil, then algorithms are the engines that process it into a valuable product. Machine Learning is a rapidly evolving field. Conceiving and adopting new approaches for data processing is as critical as getting access to data itself. Deep Learning emerged as a successful model for language translation only three years back and while popular today may not be so some years from now.[1] Businesses compete on the insights/

[1]Hao, K. (2019, Jan 25). "We Analyzed 16625 Papers to Figure Out Where AI is Headed Next". Retrieved March 25 2019: https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/



recommendations from the data they collect. *A self sustaining digital economy requires not just data, but businesses that can be on the cutting edge of research and development of new algorithmic techniques*. The policy should enable domestic businesses to leverage the best technologies that exist globally and build on them. *Data localization in particular could be prohibitive for this*. Businesses need to use services hosted on servers in other countries. These services are more important for startups that unlike big tech companies don't have the resources to reinvent the wheel. Data localization will impede newer startups from scaling.

Digital products are an interaction of both data and algorithms. The policy should encourage innovation at this intersection[2]. Source Code disclosures will be ineffective in achieving the goal of technology transfer- many of the algorithms and libraries that power online platforms are open source. There are several other factors, many of which the draft policy recognizes, that interact for a product's success. Mandating opening of source code will contribute to an environment of regulatory arbitrariness and reduced incentives for competition, without necessarily adding to the domestic knowledge base.

[2] Weber, S. (2017). Data, development, and growth. Business and Politics, 19(3), 397-423. doi:10.1017/bap.2017.3

IMPORTANCE OF HUMAN TALENT IN AI:

The draft policy refers to AI as self-learning/self-teaching[3]. It is important to recognize that even with self-learning systems, humans play an important role in design and maintenance of systems. Neural Networks, that have received a lot of attention as self-learning systems are only one kind of machine learning techniques and are not suitable for a wide range of problems.

Many decision making systems use regressions, a supervised machine learning technique which involves human judgement at different stages of the algorithm design. The above figure[4] explains different steps involved in designing a hiring system that recommends good employees from a pool of applicants. As can be seen from the figure, human decision is needed at different steps- What algorithm approach is best suited for the problem? What data should be collected? What features should be extracted from it? What kind of errors are acceptable in the system?[5] Answering these questions requires skilled human intervention. It is also the answers to these questions that sets different prediction models apart.

There is significant room for subjectivity in the design of AI systems, even in those that update their outputs based on newer data. The importance of human skill in designing and managing these systems should not be undermined.

---

[3] From the draft policy: "Artificial Intelligence(AI) has developed self-learning capabilities, based on analysis of data, given large enough data sets for processing."

[4] Nonecke, M. Prabhakar, T. Brown. C, Crittenden, C. (2017, May 10), Inclusive AI: Technology and Policy for a Diverse Urban Future. Retrieved March 29, 2019: https://citris-uc.org/connected-communities/project/inclusive-ai-technology-policy-diverse-urban-future/

[5] Machine learning systems can be designed to penalize misdetections (false negatives) and false alarms (false positives) differently. While it might be more important in one use case to minimize misdetections (detecting areas at high risk after a natural disasters), it might be more important to minimize false positives in another (for example in judiciary).

RECOMMENDATIONS

While the policy is rightly concerned about network effects leading to monopolies, there are multiple strategies that should be explored to ensure access to data by businesses of all sizes. The Draft Data Protection Bill lays the groundwork for data portability[6] that the e-commerce policy should build on. To this end the ministry should think about infrastructure and policies that enable data exchanges and data portability. Given the global nature of e-commerce, such deliberations will inevitably involve global negotiations that India should actively participate in and lead.

The Report of Task Force on Artificial Intelligence by Department for Promotion of Industry and Internal Trade[7] takes a more holistic view on the digital platforms and makes several recommendations that are relevant for technology transfers and building domestic talent:

- Investment in research and development of AI through public private partnerships; and a network of alliances between academia, service industry, product industry, start-ups and governments.
- Bilateral Cooperation between different countries to develop AI solutions for social and economic problems.
- Investment in re-skilling and development of an AI education strategy.

Lastly, instead of weakening the IP regime through forced disclosures, there is a need to revisit India's patent laws to see if they must be changed to align India with global trends; and to protect interests of Indian entrepreneurs.

---

[6] Committee of Experts on a Data Protection Framework for India. (2018, July 27). Section 26, Draft Personal Data Protection Bill.

[7] Government of India, Department for Promotion of Industry and Internal Trade. (2018, March 20). Report of Task Force on Artificial Intelligence

Inherent Tension Between Individual Data and Community Data

The draft policy begins by recognizing an individuals right to own her/his data. This is in line with the spirit of the draft Personal Data Protection Bill.[8] However, the draft' subsequent declaration of data as "a collective resource, a national asset that the government holds in trust" strongly contradicts an individual's right over her data.

Rights over personal data implies that the control over how that data is used lies with the individual. The government self appointing itself as a steward of users data is paternalistic and impinges on that right.

The policy would benefit from distinguishing different kinds of data and associated forms of control. Data generated through IOT devices in a city cannot be treated identically to data generated from a users' social media profile. Control and ownership of data should be dealt in more depth and with more nuance. Since concerns about data are cross-cutting across government departments and ministries, policy strategies on data control and access can be handled by a dedicated policy strategy that involves multiple government stakeholders in the drafting process.

Specific Responses

Section 4.10

In continuation, it is also important for the Government to reserve its right to seek disclosure of source code and algorithms. There will be a greater reliance on AI in decision making in future where parts of the process will become 'AI-fied'. Decisions will need to be explained. There is a need to strike a balance between commercial interests and consumer

---

[8] Committee of Experts on a Data Protection Framework for India. (2018, July 27). Draft Personal Data Protection Bill.

protection issues, as well as issues of larger public concern, like preventing racial profiling and maintaining constitutionally mandated rights, such as the right to equality.

Explainability of decision making systems is a laudable goal, especially in systems that affect access to essential goods and services. Opening access to the algorithm could increase its transparency and trustworthiness by allowing greater scrutiny among a diverse group. France for example has moved to classify source code used by government agencies as a public record subject to transparency laws.[9]

Disclosing source code and algorithms however, could be detrimental to a company's competitiveness. Disclosures could also lead to gamification, where if the decision making parameters become known, consumers change their behavior so that the algorithm gives a particular outcome. This could undermine the basic function of the platform.

Furthermore, disclosures do not inherently make the decisions of the system explainable or the systems more fair. Some machine learning algorithms—specifically those in which the application developer does not specify the predictive features—can select features that might make limited or no sense even to the engineers designing these algorithms.[10] The predictive accuracy of these algorithms however might be higher than more interpretable algorithms. In such cases, enforcing models that are explainable would reduce the performance of a platform. Building explainable AI systems is an active area of research; it is both a technical and a legal problem.

It should be noted that interpretability is not be equally important in all domains. Explainability of technologies in judiciary is far more important than explainability of algorithms in ad targeting or credit scoring. Explainability may come at the cost of predictive accuracy and efficiency. Hence, standards of explainability should be domain specific.

---

[9] Rieke, A. Bogen, M. Robinson D.G. (2018, Feb 27). Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods. Retrieved March 29, 2019, from https://www.omidyar.com/insights/public-scrutiny-automated-decisions-early-lessons-and-emerging-methods

[10] Berreby, D. (2015, August 06). Artificial Intelligence Is Already Weirdly Inhuman - Nautil.us. Issue 27: Dark Matter. Retrieved January 25, 2017, from http://nautil.us/issue/27/dark-matter/artificial-intelligence-is-already-weirdlyinhuman

Techniques of scrutinizing automated decisions by public and governments are still emerging. The draft policy rightly points out the need for balance between commercial interests and issues of larger public concern. *At this nascent stage different strategies should be adopted* to ensure explainability and fairness. Following are other strategies that should be considered:

EXPLORE POSSIBILITIES IN ALGORITHMIC AUDITING

An audit involves an outside entity coming in to review how a company develops its product without compromising that company's trade secrets.[11] Several new ventures are exploring services around algorithmic auditing of other companies. Third party vetting makes businesses appear more trustworthy to consumers and funders. The draft policy puts the responsibility of scrutiny entirely on the government. Not only does this increase the liability on the government to scrutinize these systems responsibly, it is also a missed opportunity in mobilizing broader interest for economic opportunities. The policy should support research on auditing and relatedly fairness of online platforms and explore private sector participation in developing these verticals. They can also create incentives for businesses to self enroll for external auditing.

DEFINE GUIDELINES FOR DATA LIFECYCLES IN DIFFERENT SECTORS

There is a need for principles to guide the flow of data- from acquisition, to storage, securing, processing, archiving and deletion- in digital platforms. These principles can be different for different kinds of e-commerce services but should be thought through especially for critical sectors like health and law enforcement.

Predictive algorithms make decisions about the future based on historical data. This is true of many e-commerce platforms such as mobile lending apps, recommendation systems and advertisement models. But these decisions influence what information, content or products are visible to consumers. User behavior changes over time, and data

[11]Hempel, J. (2018, Aug 05). Want To Prove Your Business is Fair? Audit Your Algorithm. Retrieved March 28 2019, from: https://www.wired.com/story/want-to-prove-your-business-is-fair-audit-your-algorithm/

used in decisions must be updated so that users are not locked in to decisions made on them based on historical behavior that they may have changed.

Continuous updating of underlying data also addresses concerns about group biases creeping from underlying data. If the data used to train an algorithm is skewed against one group, the algorithm trained on that will be skewed against that group. For example, if historically women were less likely to be approved for loans in a regions, algorithms built on this data would also identify fewer women as trustworthy borrowers.[12] Such biases can be corrected for through the algorithm but if such correction requires that all potential biases in the dataset are identified. A more feasible alternative is to enforce recollection of data at regular intervals. This timeline should be sector specific and take into account the nature and sensitivity of the data collected; and the importance of the service in a consumer's life.

Finally, data lifecycles enhance privacy of users and reduce liability on platforms from data breaches.

Section 4.19

Departments should aim to use AI tools and attempt predictive approach to policy making.

AI technologies work well for problems that "lend themselves to formalization" in a way that computers can understand it;[13] *not all policy challenges lend themselves to be resolved by AI*. Furthermore, AI systems come with additional concerns around opacity of decisions through algorithmic systems and negotiating values of fairness encoded in them. In cases where AI is used in policy making, the efficiency gains expected from incorporating AI in that process should be clearly delineated prior to deployment of the system and measured post deployment. In

---

[13] Barocas, S. & Selbst, A. (2016) Big Data's Disparate Impact. California Law Review. DOI: http://dx.doi.org/10.15779/Z38BG31. Retrieved from: http://www.californialawreview.org/wpcontent/uploads/2016/06/2Barocas-Selbst.pdf

absence of such metrics, it is impossible to evaluate the tradeoffs in value added from AI vis-a-vis concerns of opacity, fairness and accountability of decisions.

AI as a strategy for governance needs considerable thought and oversight; and potentially a dedicated policy. More generally, *application of AI in civic governance should not be in the ambit of an e-commerce policy*. The National Artificial Intelligence Mission (N-AIM) recommended by the AI Task Force is a possible location for work in this domain[14].

[14] Government of India, Department for Promotion of Industry and Internal Trade. (2018, March 20). Report of Task Force on Artificial Intelligence