Comments on the Indian Artificial Intelligence Stack proposed by the Department of Telecommunications

Sindhu A

Research Member, Indian Society of Artificial Intelligence and Law, India
executive@isail.in

Abstract. Here are the comments on the Indian AI Stack submitted to the Department of Telecommunications, the Government of India submitted by the Indian Society of Artificial Intelligence and Law.

1 Summary of the Draft

The proposed Indian Artificial Intelligence Stack[1] seeks to remove the impediments to AI deployment by proposing to set up a six-layered stack, each handling different functions including consent gathering, storage, and AI/Machine Learning (AI/ML) analytics. Once developed, this stack will be structured across all sectors, including data protection, data minimisation, open algorithm frameworks, defined data structures, trustworthiness and digital rights, and data federation (a single database source for front-end applications), among other things.

The Proposed Indian AI stack hinges on the five main horizontal layers;

(I) The Infrastructure Layer:

- Ensures setting up of a common Data controller including multi cloud scenarios- private and public;
- Ensures federation, encryption and minimization at the cloud end; and
- Ensures proper monitoring and data privacy of the data stored.

(II) The Storage Layer:

- Ensures that the data is properly archived and stored in a fashion for easy access when queried; and
- Ensures that the Hot Data/ Cold Data/ Warm data are stored in appropriate fashion to ensure fast or slow data access.

[1] AI Standardization Committee, Department of Telecommunications. (2020, 09 02). *Indian Artificial Intelligence Stack.* Tec.gov. Retrieved 10 01, 2020, from https://www.tec.gov.in/pdf/Whatsnew/ARTIFICIAL%20INTELLIGENCE%20-%20INDIAN%20STACK.pdf

Indian Strategy on AI and Law Programme – Recommendations.
Available at isail.in/strategy

(III) The Compute Layer:

- Ensures proper AI & ML analytics;
- Certain template of data access and processing to ensure open algorithm framework is in place;
- Process ensures Natural Language Processing and Decision tree;
- Deep learning and Neural networks;
- Predictive models and Cognitive models;
- Analytics includes;

o Data engineering and sandboxing
o Scaling and data ingestion
o Technology mapping and Rule execution

(IV) The Application Layer:

- Ensures that the Backend services are properly and legitimately programmed;
- Develop proper Service Framework;
- Ensure proper Transaction movement; and
- Ensure that proper logging and management is put in place for auditing if required at any point of time.

(V) The Data/ Information exchange layer:

- Provides for End Customer Interface;
- Has Consent Framework for data consent from/to customers; Provision for consent can be for individual data fields or for collective fields. Typically there could be different Tiers of consent be made available to accommodate different tiers of permissions.
- Provides various services through secure Gateway services; Ensures that Digital Rights are protected and the Ethical standards maintained;
- Provides for Open API access of the data and has Chatbots access; and Provides for various AI/ML Apps.

And one vertical layer;

(VI) The Security and Governance Layer:

- This is a cross cutting layer across all above layers that ensures that AI services are safe, secure, privately protected, trusted and assured. Encryption at different levels and Cryptographic supporting is an important dimension of the security layer.

Tackling Algorithmic bias in the following ways:

1. Openness in AI algorithms
2. Centrally controlled data
3. Proper storage framework for AI so the data is not incomplete or wrong etc
4. Changing the ‘culture’ of coders and developers

2 Issues with the Draft

1. **The security layer specifies cryptography and encryption as essential measures to ensure security in the system. The draft also points out that it needs to develop suitable encryption methodologies.**

- There is a lack of an adequate encryption policy in India.
- Owing to the ongoing encryption debate in India[2], would such a measure be safe from orders for the interception and decryption of information from law enforcement agencies? These agencies are also empowered to demand the same under section 69 of the Information Technology Act 2000 and search-and-seizure provisions like Section 91 of the Code of Criminal Procedure 1973. If the Personal Data Protection Bill is enacted, then it provides leeway for the authorities to demand access to data according to the exemption under the Bill. Since there is no current implementation of an Encryption policy, it aggravates these concerns.

Whether this would in any way affect the encryption measures proposed by the committee in this draft?

- Has the draft considered the provisions of the Information Technology (Procedure and Safeguards for Interception, Monitoring and Decryption of Information) Rules 2009?

It is advisable for the draft to mention what policies would apply to them and whether certain exemptions will be applicable to the data under the AI Stack.

2. **The draft specifies that in the absence of a clear data protection law in the country, EU's General Data Protection Regulation (GDPR) or any of the laws can be applied. This will serve as interim measure until Indian laws are formalised.**

- However, the draft seriously overlooks certain dissimilarities between the Personal Data Protection Bill (PDPB) of India and the General Data Protection Regulation (GDPR) of the European Union. If the current draft focusses on being in accordance with the requirements of GDPR, then once the PDPB is enacted, it will have to go through several changes. The following reasons dictate why being compliant with GDPR doesn’t necessarily mean being compliant with PDPB:

(i) While the GDPR doesn’t govern anonymised data at all, the PDPB allows governments to compel the disclosure of non-personal data and anonymised data.

[2] MOHANTY, B. (2019, May 30). *The Encryption Debate in India*. Carnegie Endowment. Retrieved October 01, 2020, from https://carnegieendowment.org/2019/05/30/encryption-debate-in-india-pub-79213

(ii) The definition of ‘personal data’ under the PDPB itself is broader than the definition under GDPR. Under the PDPB, the definition of ‘sensitive personal data’ is also broader as compared to its definition under the GDPR. Therefore, PDPB imposes a higher degree of standard while processing sensitive personal data as opposed to the standards under GDPR.

(iii) The lawful bases for processing personal data under the GDPR and PDPB are different. Under the GDPR there are six lawful bases for processing personal data namely- Consent, Performance of a contract, Legal obligation, Legitimate interests, Life protection and vital interests and Public interest. However, under the PDPB there are seven lawful bases for processing personal data, namely- Consent, Legal obligation, Medical emergency involving a threat to life or severe threat to health, Providing medical treatment or health services, Protecting the safety of individuals during a disaster, Employment purposes and “Reasonable purposes” (to be defined by the Data Protection Authority).

Under the GDPR all the grounds are placed on an equal footing, however, the PDPB classifies consent as the primary grounds and treats the other six grounds as an exception.

(iv) Under the GDPR, data localization is not essential unless international data transfer requirements are not met. While the PDPB mandates that ‘critical personal data’ (to be defined by the government) must be stored and processed in India, except under emergency circumstances or where the government has approved the transfer. ‘Sensitive personal data’ must be stored in India, but a copy of such data may be transferred outside of India in accordance with explicit consent.

(v) The PDPB definition of consent is considerably more flexible than the definition under the GDPR. The PDPB also proposes a new type of entity to help manage the consent of data principals i.e. 'consent managers' which is not available under the GDPR.

(vi) In terms of security compliance although the PDPB and GDPR are broadly based on the same principles, In the GDPR, while all Data Controllers have to undertake Data Protection Impact Assessments prior to processing some kinds of personal data (subject to limited prescribed exemptions), under the PDPB, only 'significant Data Fiduciaries' are required to do so where their processing involves (a) new technologies; (b) large-scale profiling or use of sensitive data; or (c) any other activities that carry a significant risk of harm as may be specified by regulations.[3]

[3] Wimmer, K., Maldoff, G., & Lee, D. (n.d.). *Indian Personal Data Protection Bill 2019 vs. GDPR.* IAPP. https://iapp.org/media/pdf/resource_center/india_pdpb2019_vs_gdpr_iapp_chart.pdf

(vii) The GDPR mandates data be kept in an identifiable form and certain exceptions such as public interest have been clearly laid down for increasing the storage period under the GDPR. However, the PDPB mandates that data shall not be retained beyond the period necessary to satisfy the purpose for which it is collected and has to be deleted once the purpose is fulfilled. Even if such data needs to be retained beyond the necessary period, the PDPB demands explicit consent from the data principal.

(viii) The GDPR requires Data Controllers to notify the Data Protection Authority of a breach within 72 hours, only if it is likely to result in a "high risk" to individuals. However, the PDPB requires Data fiduciaries to notify the Data Protection Authority of a breach as soon as possible (regulations to decide the exact time period) even if it is "likely to cause harm to any data principal."[4]

Apart from this, there are significant differences in other requirements such as Audit requirements, collection of personal data and processing of personal data belonging to children, registration of 'Significant Data Fiduciaries', additional provisions for social media intermediaries.

Therefore, if the current Indian AI Stack is developed in consonance with the GDPR regulations, it will risk being non-compliant with the PDPB. Once the PDPB is enacted, then the draft will have to undergo significant changes to ensure compliance with the Indian Data Protection Laws.

3. **The Infrastructure layer provides for the setting up of a 'common data controller' (an entity that determines the purpose and means of processing personal data) including both public and private clouds.**

- It is not clear from the draft whether this entity- 'common data controller' is similar to the Data Controllers under the GDPR. Data Controllers as under the GDPR do not exist under the PDPB, instead the PDPB established 'Data Fiduciaries'. The functions carried out by the Data Controllers under the GDPR are similar to the functions carried out by Data Fiduciaries under the PDPB; however, the deliberate use of the word 'fiduciary' under the PDPB as opposed to 'Controller' indicates that Fiduciaries have a higher level of duty and care.
- Therefore, in the proposed AI Stack, the 'common data controllers' might not be the same as 'data fiduciaries' under the Indian Data Protection law once enacted, even though they are carrying out similar functions. The draft also fails to specify the obligations and powers of this authority, it merely specifies

[4] Roshan, R., & Srinivasan, S. (2020, March 12). *Comparative Analysis: General Data Protection Regulation, 2016 And The Personal Data Protection Bill, 2019*. Mondaq. Retrieved October 01, 2020, from https://www.mondaq.com/india/privacy/903076/comparative-analysis-general-data-protection-regulation-2016-and-the-personal-data-protection-bill-2019

that the entity will be responsible for determining the purpose and means of processing personal data. It is also not clear whether this duty of determining means of processing personal data will overlap with the duties of Data controllers and Data fiduciaries under respective Data protection laws.

4. **The Storage layer in the proposed draft ensures that the data is properly archived and stored in a fashion for easy access when queried. The draft also proposes a classification of Hot Data/ Cold Data/ Warm data according to the relevance of data and its usability.**

- Although GDPR doesn’t deman deletion of data once the purpose for which it was collected is exhausted, however, this could be inconsistent with PDPB provisions that demand that data shall not be retained beyond the period necessary to satisfy the purpose for which it is collected and has to be deleted once the purpose is fulfilled.

5. **The draft calls for all the government and private sector players, including manufacturers, service integrators, cloud service providers etc, to come together and coordinate in the development of the India AI stack in order to seamlessly cater to all sectors.**

- This warrants a closer look at whether all government employees have the required skill sets to make this AI project successful and whether the digital literacy rates among the government employees is high. The Compute Layer ensures proper AI & ML analytics and embedding AI or ML in national systems is a piece that has to come from the government, not merely private tech companies to make it successful and Digital literacy holds immense significance.[5]
- Government officials and Policymakers need to be data literate to make data-driven decisions. Therefore, there is aneed for AI literacy/education and skill rejuvenation.

6. **Under measures to tackle AI bias, the paper proposes the need to centrally control data using a single or multiple cloud controllers because the data from which the AI learns can itself be flawed or biased, leading to flawed automated AI decisions.**

[5] Chawla, V. (2020, September 10). *What Does India Need In Place To Implement Nationwide AI Systems Across Sectors?* Analytics India Magazine. Retrieved October 01, 2020, from https://analyticsindiamag.com/what-does-the-government-of-india-need-in-place-to-implement-nationwide-ai/

However, the draft is silent on how the data will be controlled centrally and does not prescribe any procedural guidelines. It is also not clear whether a separate entity will be created for this sole purpose.

7. **Under measures to tackle AI bias, the paper proposes to change the culture of coders and developers. The paper stated that there is a need to change the "culture" so that coders and developers themselves recognise the "harmful and consequential" implication of biases, the paper said, adding that this goes beyond standardisation of the type of algorithmic code and focuses on the programmers of the code. Since much coding is outsourced, this would place the onus on the company developing the software product to enforce such standards.**

The draft clearly places the onus only on the companies developing the software product to enforce standards on coders and developers to ensure that they do not impose their biases onto the algorithm.

However, the draft fails to specifically lay down these standards that the companies developing the software product need to incorporate. The draft is also silent about whether there is a separate appointed authority to check if such measures are adequately incorporated by the companies and whether these companies will be penalised/ reprimanded if they fail to comply with this requirement.

It is advisable for the draft to create an authority to ensure compliance with this requirement by periodically reviewing the company's standards and supervising if the standards are being implemented efficiently. This becomes important because tackling AI biases is one of the prominent objectives of the proposed Indian AI Stack. Another indispensable requirement is the revamp of curren training and skill development programs by governments to promote digital literacy to be more inclusive and far-reaching.

8. **The draft proposes to implement the provisions on the General Data Protection Regulation (GDPR) in order to ensure effective data protection standards. However there are certain rights granted to Data Principals under the GDPR such as the Right of rectification, right to be forgotten, right to withdraw consent and right to restriction of processing. Research has proven that these rights are largely inconsistent with AI systems. The Proposed AI Stack seriously overlooks the implementation of these rights under the GDPR which are also present under the PDP Bill.**

One right under the PDP Bill and the GDPR that is largely inconsistent with AI systems is the Right to be Forgotten (Right of Erasure); primarily because AI systems are not taught to forget the way humans are. It has been observed that when data or memory is deleted, it doesn't automatically disappear from the system however, the data is redirected onto a 'linked list' which will eventually be processed and then made part of available software memory to be re-used later. Practical implementation of the right to be forgotten in such situations may not also mean that one was complaint with the letter of the law in the traditional sense.[6] Whether the right of erasure under these data protection laws entails making the data disappear or making it unavailable is yet unclear.

Some studies have suggested methods such as the 'SISA training'- short for Sharded, Isolated, Sliced, and Aggregated training for better enforcement of these rights under the GDPR. The method involves dividing the training data into multiple disjoint shards in a manner to ensure that each training point is included in one shard only. These shards are then trained in isolation which effectively limits the influence of a point to the model that was trained on the shard containing the point. Finally, if there is a request to unlearn or delete a training point then only that shard or training point needs to be retrained.[7]

9. **The draft proposes 4 ways to tackle algorithmic bias- open algorithms, centrally control data, proper storage framework and change culture of coders and developers.**

However, there are other reasons for algorithmic bias that the draft ignores. Placing complete onus on the 'culture' of the coders and developers is not the ideal way of approaching AI bias. Biases creep into AI for several reasons including insufficient training data sets or lack of diversity in these data sets, lack of oversight in collection of data and sampling, lack of regular audits and reviews of policies etc. The Proposed draft fails to address all of these concerns and instead places the burden mostly on the coders and the developers.

[6] Green, A. (2020, March 29). *GDPR: The Right to Be Forgotten and AI*. Varonis. Retrieved October 01, 2020, from https://www.varonis.com/blog/right-forgotten-ai/

[7] Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., Lie, D., & Papernot, N. (2019, December 01). Machine Unlearning. *arXiv e-prints*, *arXiv:1912.03817*(2019). https://arxiv.org/pdf/1912.03817.pdf