AI Ethics in Practice:

Implementing AI ethics in the policy, legal and regulatory, and technical arenas in Singapore and India

Ramathi Bandaranayake,* Viren Dias,+ Ashwini Natesan, × and Gayashi Jayasinghe

LIRNEasia, 12 Balcombe Place, Colombo, Sri Lanka

LIRNEasia is a pro-poor, pro-market think tank whose mission is *catalyzing policy change through research to improve people's lives in the emerging Asia Pacific by facilitating their use of hard and soft infrastructures through the use of knowledge, information and technology*.

Contact: 12 Balcombe Place, Colombo 00800, Sri Lanka. +94 11 267 1160.
info@lirneasia.net www.lirneasia.net

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Canada, and an unrestricted research gift from Facebook under the "Ethics in AI Research Initiative for the Asia Pacific."

* Project Lead. Please direct correspondence regarding this research report to ramathi@lirneasia.net
+ Primary Author: Case Studies
× Primary Author: Legal and Regulatory section

Table of Contents

Executive Summary

As the field of artificial intelligence (AI) has advanced and AI solutions are being increasingly deployed, debates around the ethical use of AI have arisen. There is no universally agreed upon definition of "AI ethics." A guide published by the Alan Turing Institute, UK, offers the following useful framing:

"AI ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies." [1]

Ethical principles often debated include whether the decisions taken by AI technologies are fair, whether such decisions can be explained, and who should be held accountable for them, among others. Such questions have become even more vital with the deployment of AI technologies in areas such as healthcare diagnostics and predicting recidivism, where incorrect or biased decisions have the potential to do great harm.

AI ethics, however, are difficult to study systematically. Firstly, there is often disagreement about the meanings of ethical principles such as "fairness", "accountability," and "explainability." Secondly, principles that would be ideal to achieve in theory may not always be practically implementable or enforceable. At a policy level, governments have outlined their goals for AI development, including ethical AI, through policy documents, and different countries are at varying stages of translating these visions into practice. While existing legal and regulatory frameworks can be brought to bear on AI ethics there are also gaps, and a dearth of case laws mean that rulings relating to many important ethical questions are yet to be made. Meanwhile, AI developers are searching for technical solutions to ethical problems, such as how to make their models fairer, and decision-making processes more explainable.

Given these challenges, we take a specific approach. In this research, we focus on the challenges of implementing AI ethics principles in two Asia Pacific Nations: Singapore and India. These nations are of two different sizes and are at different stages of development, but both are looking to develop their AI capacities in a very ambitious way. Singapore has many initiatives in AI ethics and AI governance, and was ranked 6th in the 2020 edition of Oxford Insights' Government AI Readiness Index.[2] India has many AI ambitions, although it is not at the level of Singapore yet (with a Readiness Index rank of 40).[3]

We conduct this analysis through the lens of a three-way implementation framework. The three points considered are:

1. Ethical AI principles as defined and advocated for in the relevant policy documents in India and Singapore.
2. The legal and regulatory landscape as it relates to the use of AI in both countries

[1] Leslie, D., (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

[2] Shearer, E., Stirling, R., & Pasquarelli, W. (2020). Government AI Readiness Index 2020. Oxford Insights. https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf

[3] Ibid.

3. Two case studies of specific AI applications, one deployed in each country, which analyze how the relevant principles are being implemented in the technical space: The SELENA+ Diabetic Retinopathy Screening tool in Singapore, and the deployment of Google's Early Warning System for floods in India.

We undertake research by analyzing policy documents, legal and regulatory analysis, and analyzing two case studies. We aim to illustrate the use of the three-way framework, and propose it as a method for analyzing the interactions between these three aspects more broadly, which would be of use to policymakers working in this space.

We do find implementation of AI ethics principles in the legal and regulatory and technical spaces. The AI ethics discourse is also very specific to applications, and the specific ethical questions that arise depend on the technology and its context of use. There is also a high reliance on the voluntary uptake of principles. We end by offering recommendations for policymakers, which include:

- Ethical, policy, legal and regulatory and technical applications must be considered together. Policymakers must bear in mind what is technically feasible, AI developers should bear both ethics and the law in mind, and regulators should be mindful of technological developments. In this three-way implementation framework, these three aspects need to be considered together.
- Sector-specific ethics frameworks and regulations need to be further developed. Specific ethical debates will emerge from individual case studies, and these must be considered in tandem with "top-down" principles.
- Consideration needs to be given to the roles of the public and private sector in adopting and promoting AI ethics principles, and the responsibilities the two may have. For data sharing agreements between the public and private sector, long term and short term benefits need to be borne in mind and ensured that benefits of the agreement continue to accrue in the long run.
- Alternate regulatory / policy mechanisms need to be used when ideal technical solutions to ethical problems do not yet exist or are still in development.
- Debates around the structure of regulatory bodies for AI need to be examined. While sector-specific regulatory frameworks have begun to emerge and need to be further developed, some have argued that overall regulatory frameworks for AI should be considered.
- Finally, given the context of the COVID-19 pandemic, the use of data and AI in urgent and emergency situations must be examined to ensure that privacy is safeguarded while responding to the emergency.

Introduction

The ethics of artificial intelligence (AI) concerns the appropriate and responsible use of AI technologies. However, words such as "ethics," "appropriate," and "responsible" are loaded terms that can take on a multiplicity of meanings depending on context, which includes the geography and domain of use. A recent review in 2019 found that most AI ethics frameworks are still concentrated in the Global North.[4] Nor does the presence of AI ethics principles necessarily assure ethical AI.[5] One of the key challenges surrounding the ethics of artificial intelligence is how they may be implemented in practice, including in the policy, legal and technical spaces.

Given this, we wish to explore the challenges of implementing AI ethics frameworks in India and Singapore. These nations are of two different sizes and are at different stages of development, but both are looking to develop their AI capacities in a very ambitious way. Singapore has many initiatives in AI ethics and AI governance, and was ranked 6th in the 2020 edition of Oxford Insights' Government AI Readiness Index.[6] India has many AI ambitions, although it is not at the level of Singapore yet (with a Readiness Index rank of 40).[7]

Research Question

What are the ethical, policy, regulatory, and technical challenges of implementing AI ethics frameworks in Singapore and India? Under this, we ask:

1. What ethical principles are being advanced and advocated for in policy and policy-relevant documents from the governments of Singapore and India?
2. How do the existing legal and regulatory landscapes of Singapore and India address these ethical principles, and what are the gaps?
3. Through case studies, what are the challenges of implementing AI ethics principles through technical applications in the real world?

We undertake this research by analyzing policy documents, legal and regulatory analysis, and analyzing two case studies. Our goal is to extract insights and recommendations that will be of use to policymakers in addressing the implementation of AI ethics.

We have considered laws, policies and case studies up to 31 December 2020.

[4] Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat Mach Intell* 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

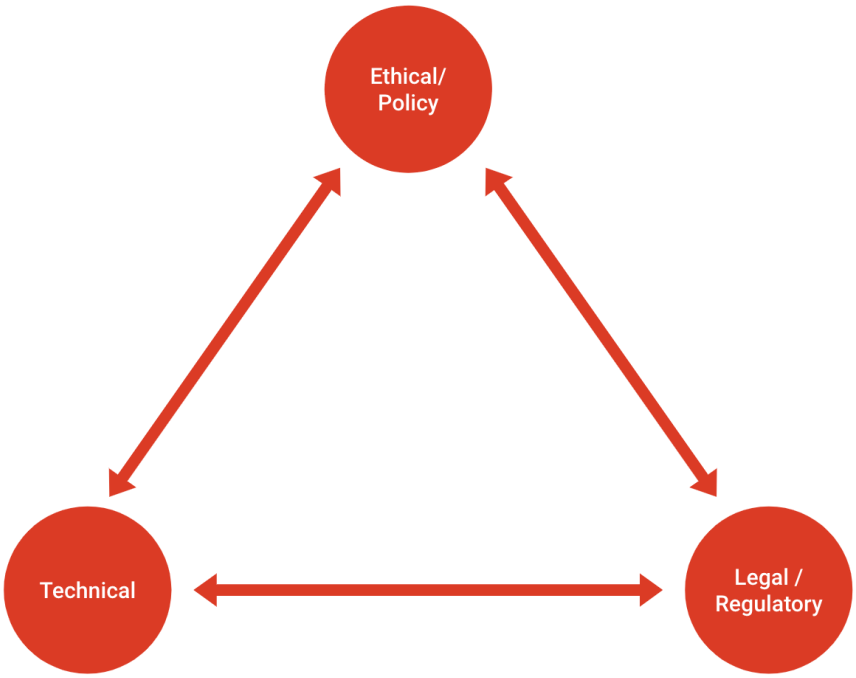
[5] Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nat Mach Intell* 1, 501–507. https://doi.org/10.1038/s42256-019-0114-4

[6] 2 Shearer, E., Stirling, R., & Pasquarelli, W. (2020). Government AI Readiness Index 2020. Oxford Insights. https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf

[7] Ibid.

The Three-Way Implementation Framework

Our goal is to study how the ethical principles that are expressed in the policy documents may be implemented in the legal and technical space. We introduce the below framework as the backbone for this analysis.



Stage 1: Policy Articulation of Ethics

This section lays out the predominant AI ethics principles articulated in the key policy documents of Singapore and India.

List of relevant policy documents

Below are the relevant policies and initiatives related to AI ethics, and the relevant institutions within Singapore and India.

Singapore

- National Artificial Intelligence Strategy (Smart Nation Singapore, Digital Government Office) [8]
- Model AI Governance Framework (in January 2020, the second edition was released) (Info-communications Media Development Authority ("IMDA") Personal Data Protection Commission ("PDPC") and SG Digital Office).[9] Henceforth "MAIGF."
- Implementation and Self-Assessment Guide for Organisations ("ISAGO") (IMDA, PDPC and SG Digital Office)[10]
- Trusted Data Sharing Framework (IMDA and PDPC)[11]
- Discussion Paper on AI and Personal Data Fostering Responsible Development and Adoption of AI (PDPC)[12]
- Principles to Promote Fairness, Ethics, Accountability and Transparency in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector (Monetary Authority of Singapore (MAS))[13]

---

[8] Smart Nation and Digital Government Office. (2019). *National Artificial Intelligence Strategy*. Government of Singapore. https://www.smartnation.gov.sg/docs/default-source/default-document-library/national-ai-strategy.pdf?sfvrsn=2c3bd8e9_4 Retrieved September 9, 2021.

[9] Infocomm Media Development Authority & Personal Data Protection Commission. (2020). *Model Artificial Intelligence Governance Framework*. Government of Singapore. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf . Retrieved September 9, 2021.

[10] IMDA, PDPC, & Digital Government Office. (2020). *Implementation and Self-Assessment Guide for Organisations ("ISAGO").* Government of Singapore. https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgisago.pdf . Retrieved September 9, 2021.

[11] IMDA, PDPC, & Digital Government Office. (2019) *Trusted Data Sharing Framework.* Government of Singapore. https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf. Retrieved September 9, 2021.
Refer also IMDA (2019, June 28) *Enabling Data-Driven Innovation Through Trusted Data Sharing In A Digital Economy.* https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2019/Enabling-Data-Driven-Innovation-Through-Trusted-Data-Sharing-In-A-Digital-Economy . Retrieved September 9, 2021.

[12] Discussion Paper On Artificial Intelligence (Ai) And Personal Data – Fostering Responsible Development and Adoption of AI (2018, June 5). Retrieved January 3, 2021, from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf

[13] Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector. (n.d.). Retrieved March 01, 2021, from

- Smart Nation: The Way Forward (Smart Nation, Digital Government Office)[14]

India

- Discussion Paper: National Strategy for Artificial Intelligence (NITI Aayog)[15]
- Report of AI Task Force (AI Task Force)[16]
- Committees on Policy Framework for AI (Ministry of Electronics and Information Technology, Government of India)
    - Report of committee – A on platforms and data on Artificial Intelligence[17]
    - Report of committee - B on leveraging Artificial Intelligence for identifying national missions in key sectors[18]
    - Report of committee - C on mapping technological capabilities, key policy enablers required across sectors, skilling, reskill[19]
    - Report of committee - D on cyber security, safety, legal and ethical issues[20]

https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf . Note: While this document is included for completion, we did not incorporate it into the analysis because we have not focused on the financial sector in particular anywhere in this research.

[14] Smart Nation and the Digital Government Office (2018). *Smart Nation: The Way Forward.* Government of Singapore. https://www.smartnation.gov.sg/docs/default-source/default-document-library/smart-nation-strategy_nov2018.pdf?sfvrsn=3f5c2af8_2 . Retrieved September 9, 2021.

[15] The National Strategy for Artificial Intelligence #AIFORALL. (2018, June). Retrieved February 4, 2021, from https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf .

[16] Report of Artificial Intelligence Task Force. (2018, March 20). Retrieved January 3, 2021, from https://dipp.gov.in/sites/default/files/Report_of_Task_Force_on_ArtificialIntelligence_20March2018_2.pdf

[17] Ministry of Electronics and Information Technology. (2019). *Report of Committee A On Platforms and Data on Artificial Intelligence.* Ministry of Electronics and Information Technology, Government of India.

https://www.meity.gov.in/writereaddata/files/Committes_A-Report_on_Platforms.pdf Retrieved September 9, 2021.

[18] Ministry of Electronics and Information Technology. (2019). *Report of Committee B On Leveraging AI for Identifying National Missions in Key Sectors*. Ministry of Electronics and Information Technology, Government of India. https://www.meity.gov.in/writereaddata/files/Committes_B-Report-on-Key-Sector.pdf Retrieved September 9, 2021.

[19] Ministry of Electronics and Information Technology. (2019). *Report of Committee C On Mapping Technological Capabilities, Key Policy Enablers Required Across Sectors, Skilling and Re-Skilling, R & D.* Ministry of Electronics and Information Technology, Government of India. https://www.meity.gov.in/writereaddata/files/Committes_C-Report-on_RnD.pdf Retrieved September 9, 2021.

[20] Ministry of Electronics and Information Technology. (2019). *Report of Committee D On Cyber Security, Safety, Legal and Ethical Issues* . Ministry of Electronics and Information Technology, Government of India. https://www.meity.gov.in/writereaddata/files/Committes_D-Cyber-n-Legal-and-Ethical.pdf Retrieved September 9, 2021.

- AI explicit Computer Framework (AIRAWAT) (NITI Aayog)[21]
- Towards responsible #AIforAll
- (Working Document- Draft) (NITI Aayog)[22]
- Indian Artificial Intelligence Stack (AI Standardisation Committee, Department Of Telecommunications)[23]
- Report by the Committee of Experts on Non-Personal Data Governance Framework [24]

Ethical Principles Identified

Based on a reading of the above documents, we found that the key principles in common in both countries could be classed as: [25]

1. Fairness and Absence of Bias[26]
2. Transparency and Explainability
3. Accountability
4. Privacy and data protection
5. Well-being and safety
6. Inclusion

Below, we offer some examples of how the aforementioned principles are discussed in the policy documents.

[21] AIRAWAT- Establishing an AI specific Cloud Computing Infrastructure for India- An Approach Paper. (2020, January). Retrieved December 30, 2020, from https://niti.gov.in/sites/default/files/2020-01/AIRAWAT_Approach_Paper.pdf

[22] Working Document: Towards Responsible #AIforAll. (2020). Retrieved March 01, 2021, from https://niti.gov.in/sites/default/files/2020-07/Responsible-AI.pdf

[23] AI Standardisation Committee: Department of Telecommunications. (2020). *Indian Artificial Intelligence Stack.* Department of Telecommunications, Government of India. https://www.medianama.com/wp-content/uploads/ARTIFICIAL-INTELLIGENCE-INDIAN-STACK.pdf Retrieved September 9, 2021.

[24] Ministry of Electronics and Information Technology. (2020). *Report by the Committee of Experts on Non-Personal Data Governance Framework*. Ministry of Electronics and Information Technology, Government of India. https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf Retrieved September 9, 2021.

[25] Other principles are mentioned as well in the documents. For instance, the Singapore Model AI Governance Framework notes repeatability, robustness, traceability, and reproducibility. The Working Document: Towards Responsible #AiforAll from India also deals with security. However, we will focus primarily on the given six for the sake of comparison, and to define the scope of this report.

[26] While these two principles are not the same, they are often discussed together, so they will be treated in the same category in this report. The same applies to points 2, and 4.

Fairness and Absence of Bias

Position in Singapore

Many of the policy documents give suggestions to organizations and developers about how to avoid bias, such as the PDPC discussion paper, ISAGO and the MAIGF. For instance, ISAGO mentions that "inherent bias" should be checked for. i.e. whether the organisation took steps to mitigate unintended inherent biases in the data itself, or as a result of the method of data collection. Additionally, whether the data used to produce the model is reflective of the environment in which the model operates. The PDPC discussion paper contains the following quote on fairness:

"Fair: AI algorithms and models embedded in decision-making systems should incorporate fairness at their core. This could include the training dataset, AI engine and selection of model(s) for deployment in the intelligent system. What practices will avoid unintentional discrimination in automated algorithmic decisions? Examples include monitoring decisions to detect unintentional discrimination and accounting for how they were made."

Position in India

Several references to the problem of bias in the datasets in NITI Aayog National AI Strategy document, Towards Responsible #AIforAll, and MEITY Committee reports A and C. MEITY Committee Report D notes the following criteria for "fairness without bias or prejudice":

"a. Analysis technique should be completely neutral.

b. Training data must come from unbiased sampling.

c. System should automatically detect any bias present in it."

The report also provides the following suggestion: "In order to promote the responsible uses of AI, government should invest in the development of bias-free datasets and techniques/tools for building fairness, transparency and accountability features in the systems."

Transparency and Explainability

Position in Singapore

The MAIGF offers the following conception for explainability:

"Explainability is achieved by explaining how deployed AI models' algorithms function and/or how the decision-making process incorporates model predictions. The purpose of being able to explain predictions made by AI is to build understanding and trust. An algorithm deployed in an AI solution is said to be explainable if how it functions and how it arrives at a particular prediction can be explained. When an algorithm cannot be explained, understanding and trust can still be built by explaining how predictions play a role in the decision-making process."

The goal of explainability, as stated above, is to be able to explain how a certain prediction was arrived at, and thereby build trust. Explainability is also addressed in the National AI Strategy, ISAGO, and the PDPC discussion paper.

The MAIGF notes that transparency refers to the openness of all parties involved in data sharing to make available all information that is necessary for the successful delivery of the data sharing partnership.

The Trusted Data Sharing Framework defines transparency as "the openness of all parties involved in data sharing to make available all information that is necessary for the successful delivery of the data sharing partnership."

The PDPC Discussion Paper offers the following explanation for "transparency":

"Transparent: AI developers, data scientists, application builders and user companies should be accountable for the AI algorithms, systems, applications and resultant decisions respectively in order to build trust in the entire AI ecosystem. What are the measures and processes that stakeholders in the different parts of the value chain can incorporate in order to be able to inform consumers or customers about how and when AI technology is applied in decisions affecting them?"

Transparency has several different conceptions as shown above, but they revolve around keeping stakeholders informed of how AI and data are being used in a given system.

Position in India

The NITI Aayog strategy refers to "Transparency / opening the 'black box,'" in which the "black box" is defined as knowing the inputs and outputs of a system, but not the processes that happen inside. The document encourages explainability as a solution to this problem.

The Report of the AI Task Force conceives of transparency as follows:

> *"AI systems must be transparent i.e. they must be known to humans as machines and their performance, including their learning, must be verifiable / auditable. All relevant test & evaluation data must bs shared with the users."*

In the document proposing the development of an Indian Artificial Intelligence Stack, it is noted that some of the most powerful AI tools are a result of deep-learning algorithms, which can tackle complex problems at the cost of transparency. This lack of transparency might lead to "covert biases." Consequently, standards must be developed to ensure a minimum level explainability for decisions made by such algorithms.

Accountability

Position in Singapore

In Singapore, some of the recommendations consider how firms may build a culture of accountability and trust regarding the use of AI and data. For example, the MAIGF notes that accountability refers to demonstrating compliance with data protection laws and other rules specific to the data sharing partnership, and that each party has robust governance structures in place, and a corporate culture that encourages employees to take responsibility for the handling of data.

Furthermore, the Data Protection Certification Trustmark is "a voluntary enterprise-wide certification for organisations to demonstrate accountable data protection practices. The DPTM will help businesses increase their competitive advantage and build trust with their customers and stakeholders."

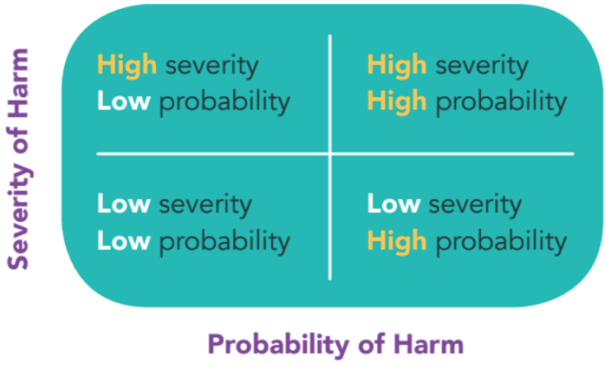
In terms of the agency of AI, the MAIGF defines three "loop" categories which specify the responsibility of an AI in relation to a human.

1. Human-in-the-loop: "suggests that human oversight is active and involved, with the human retaining full control and the AI only providing recommendations or input. Decisions cannot be exercised without affirmative actions by the human, such as a human command to proceed with a given decision."
2. Human-out-of-the-loop: "suggests that there is no human oversight over the execution of decisions. The AI system has full control without the option of human override."
3. Human-over-the-loop (also called human-on-the-loop): "suggests that human oversight is involved to the extent that the human is in a monitoring or supervisory role, with the ability to take over control when the AI model encounters unexpected or undesirable events (such as model failure)."

The probability / severity harm matrix (pictured below, from the MAIGF), is suggested as one consideration to help determine the degree of human oversight.

Figure 1

Probability / Severity Harm Matrix (Image Credit – Model AI Governance Framework)



Position in India

The Working Document Towards #ResponsibleAIforAll notes that AI decisions are influenced by many factors, making it difficult to hold any entity / cause to account for a given decision. This poses a challenge to redressing grievances. The NITI Aayog Strategy suggests negligence tests instead of strict liability: "This involves self-regulation by the stakeholders by conducting damage impact assessment at every stage of development of an AI model."

The MEITY Committee Report D includes the following factors under "accountability:"

- "a. Knowing 'things can go wrong'.
- b. Designers and deployers share responsibility for the consequences or impact an algorithmic system has on stakeholders and society.
- c. Analyze if an AI application does exactly what it is designed to do.
- d. All possible failure modes of an algorithm should be thought of
- e. Active mitigation of probable high risk failures"

Privacy and Data Protection

Position in Singapore

Singapore's Personal Data Protection Act (PDPA) will be dealt with in the section on law and regulations.

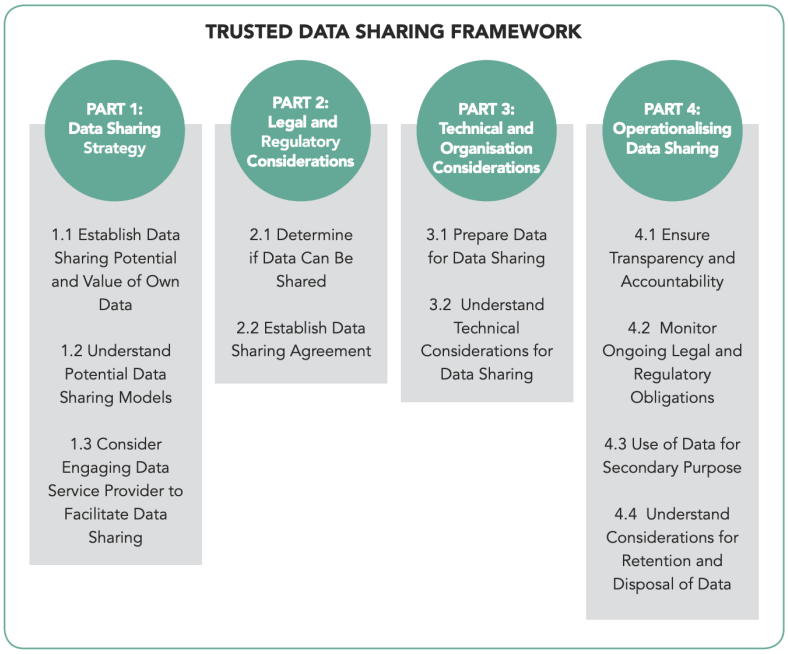
ISAGO specifies the following considerations regarding data protection:

- Personal data protection: Whether the organization implemented accountability-based practices to ensure compliance with data protection laws, regulations, and principles.
- Data lineage. Whether organizations traced and maintained a record of the lineage of data, and, if the data were obtained from a third party, whether that third party adhered to data protection practices.
- Data quality. Whether the data is an accurate representation of what it is describing—i.e. it has a complete set of attributes, it is credible and from a reliable source, it is up-to-date, it is relevant, and in the case of personal data, it was collected for the intended purpose. The guide also questions whether the organization made additional checks to ensure the quality of human-labelled data.

Furthermore, the Trusted Data Sharing Framework provides the below framework on sharing data as shown in an illustration from the document.

Figure 2

Trusted Data Sharing Framework (Image Credit - Trusted Data Sharing Framework)



Further guidelines are provided regarding the sharing of personal data, in line with the PDPA. It should be noted that "This Framework is intended for use in the commercial and non-governmental sectors but excludes data sharing in or with the public sector."

The aforementioned Data Protection Certification Trustmark is also of note here. The checklist for the Trustmark is divided into Governance & Transparency, Management of Personal Data, Care of Personal Data, and Individuals' Rights.

Position in India

There is currently a draft Personal Data Protection Bill in India, although it has not yet been passed into law. The NITI Aayog National AI Strategy proposes the establishment of a legally binding data protection framework and sector-specific regulatory frameworks, as well as other measures including investing in privacy-preserving research in AI. In India, there have also been debates and policy proposals around the regulation of non-personal data, which will be discussed further in the legal and regulatory section.

Well-Being and Safety

Position in Singapore

Several mentions of safety are made in the Singapore policy documents. For instance, the MAIGF notes that "As AI is used to amplify human capabilities, the protection of the interests of human beings, including their **well-being** and **safety**, should be the primary considerations in the design, development and deployment of AI." ISAGO states that with regards to safety-critical systems,[27] whether the organization ensured that personnel were able to take control when required, and whether the AI system provided enough information to personnel to help them make the decision to take control. Furthermore, the PDPC Discussion Paper states "AI systems and robots should be designed to avoid causing bodily harm or affecting the safety of individuals."

Position in India

Safety is mentioned in several of the policy documents in India. For instance, the Report of the AI Task Force notes that AI should be "engineered for safety and security." Safety is dealt with in depth in MEITY Committee Report D, which recommends the creation of safety guidelines, setting safety thresholds and creating safety certifications.

Inclusion

Position in Singapore

The Smart Nation Singapore strategy notes:

> *"Technology also has the power to be a social leveller. Hence, we need to dedicate resources to ensure that all Singaporeans, including the vulnerable, such as the elderly, low-income and persons with disabilities, are able to seize the opportunities offered by digital technologies."*

Position in India

It has been noted that one of the unique and noticeable features of India's AI National Strategy is social inclusion.[28] The hashtag of the Strategy is #AIforAll, and one of its targets is "AI for

[27] For ease of reference for the reader, a technical definition of "Safety-critical system" - "Safety-critical systems are those systems whose failure could result in loss of life, significant property damage or damage to the environment." As defined in Knight. J. C. (2002, May 25). Safety critical systems: challenges and directions. *Proceedings of the 24th International Conference on Software Engineering. ICSE 2002.* 24th International Conference on Software Engineering. Orlando, FL, USA. https://ieeexplore.ieee.org/document/1007998

[28] Dutton, T. (2018, June 29). An Overview of National AI Strategies. *Politics and AI.* Medium. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

Greater Good: social development and inclusive growth." The target is explained as quoted below:

> *"Beyond just the headline numbers of economic impact, a disruptive technology such as AI needs to be seen from the perspective of the transformative impact it could have on the greater good – improving the quality of life and access of choice to a large section of the country. In that sense, the recent advancements in AI seem to be custom-made for the unique opportunities and challenges that India faces. Increased access to quality health facilities (including addressing the locational access barriers), inclusive financial growth for large sections of population that have hitherto been excluded from formal financial products, providing real-time advisory to farmers and help address unforeseen factors towards increasing productivity, building smart and efficient cities and infrastructure to meet the demands of rapidly urbanising population are some of the examples that can be most effectively solved through the non-incremental advantages that a technology such as AI can provide."*

Comparison and Discussion of Principles

The ethical principles expressed in the policy documents of India and Singapore appear to be generally in line with principles discussed in other parts of the world. The synthesis of Global AI ethics guidelines by Jobin et al. (2019) cited at the beginning of this paper identified transparency, justice and fairness, non-maleficence, responsibility, and privacy among the key principles, although synthesis did note that there were differences in how the principles were interpreted.

In the cases of India and Singapore, the ethical principles and interpretations used appear to be more similar than different. In terms of bias and fairness, in both countries, the issue of bias in datasets is emphasized, but they also go further in encouraging the monitoring of decisions made by the AI to catch out instances of bias. While several conceptions of explainability and transparency are offered, the goal is to be clearer about when AI is used in decision making processes, who the decisions affect, and how the decisions are made. Transparency is also mentioned in relation to data sharing in Singapore as well. Although explainability is a cornerstone of many AI ethics frameworks, there is still debate around what it means and what its goals should be.[29] In the case study section of this report (Stage 3), we examine how regulators in Singapore assess the workings of medical AI devices, and the debate around explainability will be taken up again. In both countries, maintaining accountability and identifying agency is related to assigning responsibility for the AI system's decisions. An emphasis is placed on the risk of causing harm via the AI system's decisions and to what extent humans should be involved in the operations of the system. However, it is less clear who exactly should be accountable and what the details of accountability (e.g. assessment of damages) look like, and there is a lack of real-world examples in this regard. In Singapore, there has been a court case regarding a deterministic algorithm, which is discussed in Stage 2 on the law and regulation. Data protection and privacy are important considerations in both countries. Singapore is somewhat ahead in this regard since a Personal Data Protection Act has been passed and is in force. In the policy documents in Singapore,

[29] For example, see Newman, J. (2021, May 19). Explainability Won't Save AI. *TechStream, Brookings.* https://www.brookings.edu/techstream/explainability-wont-save-ai/

there are also specific proposals given to the private sector on how manage personal data. A draft bill is available in India, however. Data protection legislation will be further discussed in Stage 2, as well as the discussions in India around regulating non-personal data. In both countries, the policy documents mention safety, where the concern is not causing harm, and promote safety checks. The focus on inclusion is much stronger in India than in Singapore, as in India, inclusion is the primary theme of the AI strategy, and "AI for Greater Good" is explicitly stated as one of the targets. Furthermore, the Indian NITI Aayog AI strategy Discussion Paper "focuses on how India can leverage the transformative technologies to ensure social and inclusive growth in line with the development philosophy of the government." Unlike in Singapore, the Indian AI Strategy appears to be driven by development priorities.

It should be noted that while ethical principles can be defined at a high level, this alone may not be sufficient. For instance, Prof. Ang Peng Hwa of the Wee Kim Wee School of Communication and Information, Nanyang Technological University (NTU), Singapore contends many of the most pressing issues of AI ethics are more likely to arise from specific applications, including considering whether it is appropriate to use AI in a given context in the first place, and how the use of AI affects the ethical context of a given situation. These instances will not be covered by high-level frameworks such as the Model AI Governance Framework.[30] The role of AI ethics in specific applications will be further addressed in Stage 3.

[30] Personal Communication, 16 April 2021

Stage 2: Law and Regulation

This section lays out how the aforementioned principles are dealt with in the legal and regulatory space in Singapore and India.

The proliferation of AI across sectors warrants an analysis on the legal and regulatory challenges the ethical frameworks could pose, in addition to examining how the existing legal structure could handle those concerns. Legal regulation / intervention can be in the form of legislation, subsidiary/delegated legislation, sector specific guidance, frameworks, codes of practice, certifications etc. For AI, however, there has been an increasing acceptance across regions that – "given the profound changes widespread deployment of AI and autonomous technologies will precipitate – such deployment needs to be underpinned by an ethical framework that helps ensure those technologies improve human wellbeing."[31]

Regulating AI

In considering "AI regulation" through ethical principles it is essential to bear in mind that there are at least three connected layers that need to be subject to control / regulation -

a) The data (that is required to train the algorithm and its input into the system)

b) The algorithm

c) The application / deployment of the AI system

The research paper will attempt to look at these components that are integral to AI in implementation as a whole rather than three separate areas. However, data regulation will also be analyzed separately for better clarity.

The following ethical principles will be analyzed through the lens of the law: fairness and absence of bias; wellbeing and safety; transparency and explainability; accountability; privacy and data protection. The question of legal personhood will also be considered.

[31] Applying ethical principles for artificial intelligence in regulatory reform. (2020). Retrieved February 01, 2021, from https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles ; https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf

Interpretation and Implementation of ethical principles

Fairness and Absence of Bias

Position in Singapore

Fairness as an ethical principle finds its place in several policy documents including the Model Framework, discussion paper on AI and personal data.[32] It is also found in the Monetary Authority of Singapore's Principles to Promote Fairness, Ethics and Accountability and Transparency in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector.[33]

"Fairness" as is used in the context of AI ethical principles can be reasonably interpreted as an equitable principle that is reflected in absence of bias, and negative perceptions across communities, especially minorities, non-discrimination and inclusion of diverse demographics.[34] While it understood that "fairness" is not limited to "absence of bias", they have been clubbed together for analysis on the basis of their inclusion in the policy documents. In this context algorithmic bias and data bias have been subject of much discussion and debate.[35] The Model Framework refers to minimizing of inherent biases in data, which is commonly bias in selection data or measurement data.

The law reform report on AI ethics published by the Singapore Academy of Law's Law Reform Committee ("the LRC Report") states that, "[a]n AI system should be rational, fair, and not contain biases that are intentionally or unintentionally built into their system which may harm a community of people or an individual"[36]. As an illustration, the LRC Report suggested that where a government agency intended to deploy an AI system to assess a citizen's risk of committing certain types of offences, it should "evaluate potential impact on fairness, justice, bias and negative perceptions across affected communities, especially minorities".[37]

---

[32] Discussion Paper On Artificial Intelligence (AI) And Personal Data – Fostering Responsible Development and Adoption of AI (2018, June 5). Retrieved January 3, 2021, from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf

[33] Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector. (n.d.). Retrieved March 01, 2021, from https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

[34] Applying ethical principles for artificial intelligence in regulatory reform. (2020). Retrieved February 01, 2021, from https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles ; https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf

[35] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI SSRN Electronic Journal*. doi:10.2139/ssrn.3518482

[36] Applying ethical principles for artificial intelligence in regulatory reform. (2020). Retrieved February 01, 2021, from https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles ; https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf

[37] Applying ethical principles for artificial intelligence in regulatory reform. (2020). Retrieved February 01, 2021, from https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles ; https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf

It is incontrovertible that bias in AI should be minimized. What could the legal interpretation of "fairness" / "bias" mean? There is no reported case law on interpretation of fairness or bias in the context of AI.

The well-established legal connotation of "bias" exists in administrative law, as part of a principle of natural justice applicable to judicial and administrative agencies.[38] The rule against bias is again divided into actual, imputed and apparent bias. The applicable test for apparent bias in Singapore is "reasonable suspicion test", whether "there are circumstances which would give rise to a *reasonable suspicion or apprehension* in a fair-minded reasonable person with knowledge of the relevant facts that the [decision-maker] was biased."[39] It cannot be anyone's argument that these principles of administrative law should be imputed into all AI regulation, without limitation to judicial/administrative agencies. However, a reading of the related laws makes it clear that there can be no exhaustive or clear-cut explanation of these terms. Determinations will differ depending on the facts and circumstances of the case.

Additionally, the problem with interpretation of equitable principles such as fairness is that they cannot be straightjacketed or stringent. Singapore has been mindful of these challenges and released several additional materials such as the Implementation and Self-Assessment Guide for Organisations ("ISAGO"), Compendium of Use Cases, and one specifically for data, Trusted Data Sharing Framework ("Data Framework"). However, caution must be exercised in attempting to comply with guidance documents in making sure "fairness" is achieved for two reasons a) the guidance documents are not mandatory in nature b) judicial determination on whether they would pass muster in courts of law remain to be seen.

Position in India

Several policies in India including the National Strategy, AI Task Force Report, AIforAll[40] etc include the necessity to reduce /mitigate bias (cognitive, race, gender etc.) found in data and in the decision-making system in AI and maintain fairness.

The report of Committee D: On Cybersecurity, Safety, Legal and Ethical Issues (Draft)[41] reads as follows:

"Fairness without Bias or Prejudice:

a. Analysis technique should be completely neutral.

b. Training data must come from unbiased sampling.

c. System should automatically detect any bias present in it."

[38] Jhaveri, S. (2019, May 21). *Administrative law in Singapore: Recent developments and looking ahead*. Retrieved March 01, 2021, from https://lawgazette.com.sg/feature/administrative-law-in-singapore-recent-developments-and-looking-ahead/
[39] *Anant Kulkarni* [2007] 1 SLR(R) 85, quoted with approval in *BOI v BOJ* [2018] SGCA 61.
[40] Working Document: Towards Responsible #AIforAll. (2020). Retrieved March 01, 2021, from https://niti.gov.in/sites/default/files/2020-07/Responsible-AI.pdf
[41] Report of Committee D On Cyber Security, Safety, Legal and Ethical Issues. (n.d.). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_D-Cyber-n-Legal-and-Ethical.pdf

So as to fully appreciate the legal implications of "bias" and "fairness" it needs to be interpreted in line with the existing laws. As is seen in Singapore, in India too "bias" has been linked with administrative/judicial decision-making, as a principle of natural justice.[42] The test for bias personal or subject-matter, has been "real likelihood of bias"[43] and in some instances "reasonable suspicion test."[44] The constitutional safeguards against discrimination etc. cannot be fitted into all situations as with claims on the basis of fundamental rights.[45] The Information Technology Act 2000 does not address bias of this nature. However, the definition / explanation of "fairness" has been the subject of many judicial decisions, the question is whether these would need to be reviewed in the context of use in AI.

Unlike the General Data Protection Regulation ["GDPR"] in the EU, the proposed Personal Data Protection Bill, 2019, does not address fully automated decision making. While an objection by data subject to solely automated decision making cannot be equated with lack of bias/ fairness it requires mention here, as the provision has its roots in ensuring decisions that are subject to legal / significant effects are not fully automated. The White Paper of the data protection framework[46] mentions the hesitation in following the GDPR route (of prohibition on fully automated decision making) for two reasons. One, that only the decisions which are solely made by automated means are covered and any degree of human involvement (major or minor) will make this provision inapplicable and which consequentially, narrows the scope of the provision. Two, it is only applicable when such decision has legal or significantly similar effects, which further narrows the scope of this provision, given the fact that no criteria has been laid down that what constitutes legal or significantly similar effects.[47] In closing, the White Paper suggests that either some different forms of protection for the data subjects (which are affected by the results/decisions based on automated processing) should be adopted or a practically enforceable and effective right may be carved out.

As is the case with Singapore, several policies on ethics and use of AI have been released in India but it is essential to ensure bias and fairness are not nebulous ideals but those that can be implemented with some degree of certainty. While legislation cannot put to rest all uncertainties, it could potentially be a good starting point, provided it is drafted with care and does not make use of AI unduly cumbersome/ restrictive.

[42] *Crawford Bayley & co v Union of India* AIR 2006 SCC25; *Ramanand Prasad singh v Union of India* AIR1996 SCC64
[43] *Manak Lal v. Dr. Prem Chand* AIR 1957 SC 425.
[44] *Mineral Development Ltd v. State of Bihar* AIR 1960 SC468.
[45] Article 14 of the Constitution of India
[46] White Paper Of The Committee Of Experts On A Data Protection Framework For India. (2017). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/white_paper_on_data_protection_in_india_171127_final_v2.pdf
[47] Arora, H. (2019). *Automated decision Making: EUROPEAN (GDPR) and Indian Perspective (*INDIAN personal data Protection Bill, 2018). *SSRN Electronic Journal*. doi:10.2139/ssrn.3680409

Well-Being and Safety

Position in Singapore

The Model Framework urges that AI is used to "amplify human capabilities, the protection of the interests of human beings, including their well-being and safety, should be the primary considerations in the design, development and deployment of AI."

This can be reasonably interpreted to state that AI systems should at the very least 'do no harm' or minimize harm in unavoidable circumstances.[48] However, policy challenges arise when harms are not obvious and only insidious. The challenges arise in regulating the unintended or unknown harms. The challenges to privacy of the individual or a group of individuals are address through data protection legislation. In Singapore the Personal Data Protection Act 2012 ["PDPA"] address concerns of privacy / data protection.

The LRC Report includes some relevant examples of affective robot systems (artificial emotional intelligence or emotion AI) being used as nurses to care for some isolated elderly patients.[49] The potential psychological harms of such affective AI system tools cannot be ascertained. The Model Framework advocates human-in-the-loop / human involvement in situations of high probability of harm and where the repercussions could be severe.

By interpreting the concepts of "well-being" and "safety" to existing legal principles of civil or criminal liability could lead to results that would not only be limiting but would also mean attempting to fit a square peg into a round hole. It should be borne in mind that ethical frameworks have a broader applicability to harms that are outside the rigid confines of the law.[50] Hence what these "harms" are and how can be they prevented or mitigated needs better focus through research.

Position in India

The National Strategy in very clear terms requires "actual harms" for invocation of liability "*so that a lawsuit cannot proceed based only on a speculative damage or a fear of future damages*."[51] While this view is both practical and forward thinking, it could lead to a laxer approach being adopted by stakeholders in building safety precautions into AI systems. The

---

[48] Discussion Paper On Artificial Intelligence (Ai) And Personal Data – Fostering Responsible Development and Adoption of AI (2018, June 5). Retrieved January 3, 2021, from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf; Model Artificial Intelligence Governance Framework second edition. (n.d.). Retrieved January 20, 2021, from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

[49] Applying ethical principles for artificial intelligence in regulatory reform. (2020). Retrieved February 01, 2021, from https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles ; https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf

[50] Basu, A., Hickok, E., & Sinha, A. (n.d.). Regulatory Interventions for Emerging Economies Governing the Use of Artificial Intelligence in Public Functions. In *Artificial Intelligence for Social Good*. doi:https://issuu.com/jamfactory/docs/layout_v3_web_page

[51] The National Strategy for Artificial Intelligence #AIFORALL. (2018, June). Retrieved February 4, 2021, from https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

"potential to cause damage" is as much an important consideration as in encouraging innovation and progress. The ethics principle of "well-being" should not remain a mere ideal but a firm commitment that is required whilst developing and deploying AI systems.

In that regard, the Report of Committee D (Draft)[52] calls for safety guidelines, thresholds and certifications. The said report mentions multi-stakeholder involvement of both private and public sector in framing guidelines for safety. It also moots the idea of a safety certification in sectors where there can be threat to life such as health, transport etc. Human control /intervention is also included as a necessity in situations where there could be serious implications to life.[53]

The potential harm to privacy is an area that has been widely discussed in several Indian policies but as will be seen in the "data" section below, the data protection legislation and subsequent common law are crucial to safeguarding against the violation of privacy.

Similarities can be seen between the positions adopted in Singapore and India where both countries require human-in-the loop/ human intervention as a safety net to prevent harms or in situations where there could be greater harm.

Transparency and Explainability

It is important to also note that explainability is not equivalent to transparency. The obligation of providing an explanation does not necessarily mean that the developer should know the flow of bits through the AI system. However, since most policies include both requirements, and in several of them they have been used together, they have been clubbed together in this analysis.

Position in Singapore

The requirements of transparency and explainability are found in several guiding principles in Singapore.

The Trusted Data Sharing Framework states that "transparency" "*refers to* the *openness of all parties involved in data sharing to make available all information that is necessary for the successful delivery of the data sharing partnership*."[54] It is key to understand that explainability enhances transparency. The Model Framework rightly states as follows: "Measures such as explainability, repeatability, robustness, regular tuning, reproducibility, traceability, and auditability can enhance the transparency of algorithms found in AI models."[55]

[52]Report of Committee D On Cyber Security, Safety, Legal and Ethical Issues. (n.d.). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_D-Cyber-n-Legal-and-Ethical.pdf
[53]Report of Committee D On Cyber Security, Safety, Legal and Ethical Issues. (n.d.). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_D-Cyber-n-Legal-and-Ethical.pdf
[54] Trusted Data Sharing Framework. (n.d.). Retrieved March 01, 2021, from https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf
[55] Model Artificial Intelligence Governance Framework second edition. (n.d.). Retrieved January 20, 2021, from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

The LRC reports mentions, to ensure "transparency" "explainability" "traceability" an obligation may be imposed to "ensure as far as reasonably possible that the complex processes, actions or 'thinking' of AI systems:

(a) "are documented in a way that is understood easily, or

(b) can be explained when questioned using a standard human cognitive approach, such that the logic of those processes and decisions can be understood in non-technical terms."[56]

There are at least two critical issues with implementing transparency and explainability – a) the black box (opacity) b) autonomy of AI systems.

The black box can be explained as where developers do not really know how the algorithms used by systems operate or cannot explain how the algorithms operated.[57] An example of this can be deep learning machines that can self-reprogram to the point that even their programmers are unable to understand and / or explain the internal logic behind AI decisions.[58] In such situations, difficulty arises not only in detecting hidden biases but also in ascertaining whether they were caused by a fault in the computer algorithm or by flawed datasets.[59] For this reason, neural networks are commonly depicted as a black box: closed systems that receive an input, produce an output and offer limited explainability as to why.[60]

The need for and level of transparency/explainability is also not the same. In sectors like healthcare, judicial administration, autonomous vehicles and weapon systems the requirement can be high in comparison with AI use in retail and fashion.

The challenge of how to implement "transparency" also deserves attention. A mere "technical" transparency report explaining how a decision was arrived at would be of little use to a non-expert. To maneuver around this concern, the GDPR framework not only allows one to challenge automated decision making but also to opt-out of fully automated decision making.[61] Singapore does not have such protections in the PDPA, but the Model Framework suggests that explanations for AI decision making and opportunity to review such decisions should be given to customers.

Autonomy can be said to mean the ability or capability of the AI system to function independently without human intervention. When AI systems function independently how can they be explainable by the developer? The novel issue of autonomy arose in the case of *Quoine Pte Ltd v B2C2 Ltd.*[62] (discussed in detail in the "accountability" section). However, greater

[56] Applying ethical principles for artificial intelligence in regulatory reform. (2020). Retrieved February 01, 2021, from https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles ; https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf
[57] Yu, R., & Alì, G. S. (2019). What's inside the black box? Ai challenges for lawyers and researchers. *Legal Information Management, 19*(01), 2-13. doi:10.1017/s1472669619000021
[58] Yu, R., & Alì, G. S. (2019). What's inside the black box? Ai challenges for lawyers and researchers. *Legal Information Management, 19*(01), 2-13. doi:10.1017/s1472669619000021
[59] Yu, R., & Alì, G. S. (2019). What's inside the black box? Ai challenges for lawyers and researchers. *Legal Information Management, 19*(01), 2-13. doi:10.1017/s1472669619000021
[60]Scarlett, C. (2017, August 18). The future of law: Artificial intelligence? Retrieved March 01, 2021, from https://knowledge-leader.colliers.com/colin-scarlett/future-law-artificial-intelligence/ ;
Gershgorn, D. (2017, December 7). Ai is now so complex its creators can't trust why it makes decisions. Retrieved March 01, 2021, from https://qz.com/1146753/ai-is-now-so-complex-its-creators-cant-trust-why-it-makes-decisions/;
[61] Art 22 of the GDPR; Section 49 and 50 of the UK Data Protection Act, 2018.
[62][2019] SGHC(I) 03

clarity and guidance are needed for AI that is automated and capable of making own decisions and the explanations for such decisions are not known to the human developers.

Position in India

The AI Task force reports reads as follows:

"legal provisions that are applicable to human users of AI systems should continue to apply *mutatis mutandis* to automated machines, rights and responsibilities of autonomous entities should be examined, new standards needed for use of robots…."

It is debatable whether same provisions applicable to human users can be made applicable to AI systems. Although the report does mention of additional liability for certain types of machines, there is still a lacuna that needs to be filled in.

For example, as seen in the section under "safety and well-being" applying tests for negligence may not be entirely suitable in the context of AI. Under common law, for a person to be liable in negligence, the harm that occurred had to be "reasonably foreseeable."[63] AI systems, however, are designed to be creative and to keep learning from the data analysed. Therefore, these may act in ways that are not reasonably foreseeable by the system designers.[64]

The observations of the Report of Committee C (draft)[65] are pertinent in this regard: "transparency may be more difficult for AI than with traditional data processing. Some algorithms use hundreds of millions of adjustable parameters to function and may be continually updated based upon real-time data. In some cases, this makes it impossible to deconstruct how a particular result was produced by the algorithm to accurately trace back a cause. In other words, it may be impossible to understand how a result is achieved, consequently making AI less accountable to the user."

The National Strategy highlights the "black box phenomenon" and mentions that it would be a futile exercise to merely aim for technical disclosure or opening up of the code. The said policy aims towards explainable AI (XAI) as the goal. Sector agnostic legislation like the Right to Information Act 2005, its application and responsibilities are limited to "public authorities and do not apply to the private sector."[66]

Applying the current standards to determine transparency that is required in State authorities and other public authorities would not be correct legal interpretation.[67] The question is then what would be the accurate legal interpretation to "transparency" in the context of AI? Since

[63] *Municipal Corporation of Delhi v. Sushila Devi* AIR 1999 SC 1929.
[64] Vishwanathan, A. (2016, September 26). Indian law is yet to transition into the age of artificial intelligence. Retrieved February 10, 2021, from https://thewire.in/law/indian-law-is-yet-to-transition-into-the-age-of-artificial-intelligence
[65] Report of Committee – C On Mapping Technological Capabilities, Key Policy Enablers Required Across Sectors, Skilling And Re-Skilling, R&D. (n.d.). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_C-Report-on_RnD.pdf
[66] Right to Information Act 2005, s 3.
[67] *Manohar v. State of Maharashtra,* (2012) 13 SCC 14

both common law and statutory law do not offer clarity on that, it is certainly an important concern that needs to be addressed for the meaningful implementation of this principle.

It can be observed that the National Strategy in India and the LRC report in Singapore do not place weight on code disclosure but seek a more meaningful ideal of explainability of decision-making. Nevertheless, the more nuanced challenges posed by AI systems are gray areas in the eyes of law in both jurisdictions.

Accountability

Position in Singapore

The Model AI Governance Framework calls for compliance of its principles on grounds that it is an "accountability based" framework. For example, it states that risks associated with the use of AI can be managed within the enterprise risk management structure, while ethical considerations can be introduced as corporate values and managed through ethics review boards or similar structures.

"1. Clear roles and responsibilities for the ethical deployment of AI"

"2. Risk management and internal controls"

However, the question of "who is accountable?" deserves attention.

The LRC Report states that *"those who design and deploy AI systems should be accountable for the proper functioning of those systems…"*

The problem of accountability is not, however, as simple. A key feature of modern AI is the ability to operate without human intervention.[68] AI systems can operate "autonomously" or "independently". The problems of "autonomy" vary depending on the sphere of activity, for example the most common is autonomous vehicles (liability and punishment for harm); autonomous weapons (moral questions as to designation of life-death decisions to non-human processes) and algorithmic decision-making (the more pervasive and routine decisions).[69]

In terms of autonomous vehicles, Singapore has made provision for truly autonomous vehicles "without the active physical control of, or monitoring by, a human operator," but the provision adopted in 2017 is limited to enabling the Minister to make rules for trials of autonomous vehicles (driverless cars).[70]

Although not squarely within the ambit of autonomous decision making, a related issue arose before the Singapore International Commercial Court, *Quoine Pte Ltd v B2C2 Ltd.*[71] The parties,

[68] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540
[69] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540
[70] Road Traffic Act (Rev 2004) (regulation of autonomous vehicles- Autonomous Vehicle Rules, 2017 Amendment)
[71] [2019] SGHC(I) 03

Quoine and B2C2, used software programs that executed trades involving the cryptocurrencies Bitcoin and Ethereum, with prices set according to external market information. The defect in the software resulted in values being way above prevailing market prices. B2C2 put forth the contention that reversal of contracts would constitute breach whilst Quoine, countered by arguing that the contract was void or voidable on grounds of unilateral mistake, At common law, a unilateral mistake can void a contract if the other party knows of the mistake.[72] If it cannot be proven that the other party actually knew about the mistake, but had constructive knowledge, the contract may be voidable under equity.[73]

In this case it is crucial to note the judge's finding that the computer programs in question were incapable of "knowing" anything. The algorithmic programmes in the instant case were deterministic i.e. they do and can only do what they have been programmed to do. They do not operate independently / autonomously. They operate when called upon in a pre-ordained manner.[74]

Therefore, the only question that the court needed to decide was on the knowledge of the original programmer of B2C2's software. Quoine was made liable to pay damages to B2C2, on account of the lack of knowledge by the programmer.[75]

It would be useful in this context to site the observations (obiter) of the judge on autonomous algorithms, where he viewed it as an incremental process and went on to refer to the statement of Lord Briggs in a UK Supreme Court decision the previous year: "The court is well versed in identifying the governing mind of a corporation and, when the need arises, will no doubt be able to do the same for robots."[76]

On an appeal to the Court of Appeal[77] the following was held (amongst others):

- "Where deterministic algorithms (*i.e.*, those that always produce the same output given the same input) are concerned, it is the programmer's state of knowledge that is relevant and to be attributed to the parties"[78]
- "The relevant inquiry is whether, when programming the algorithm, the programmer was doing so with actual or constructive knowledge of the fact that the relevant offer would only ever be accepted by a party operating under a mistake and whether the programmer was acting to take advantage of such a mistake"[79]

[72] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540
[73] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540
[74] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540

[75] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540
[76] *Warner-Lambert Co. Ltd. v Generics (U.K.) Ltd.* , [2018] UKSC 56 (2018),165.
[77] [2020] SGCA(I) 02.
[78] [2020] SGCA(I) 02, [98].
[79] 2020] SGCA(I) 02, [103].

The Model Framework proposes a design framework (structured as a matrix) to help organisations determine the level of human involvement required in AI-augmented decision-making, as human in / over / out of the loop.

In today's context, algorithms are typically used to support or inform decision-making, particularly with respect to decisions that explicitly and directly involve human rights.[80] This is seen as a measure to mitigate potential harm .i.e., where a human 'in the loop' acts a safeguard.[81] It ought to be borne in mind that having a human-in- the-loop has another consequence of eliminating the confusion as to "who is accountable"[82]. However, this gives rise to issues such as the human 'in the loop's' ability to understand how the algorithm functions and therefore to assign appropriate weight to any recommendation; degree of deference granted to an automated recommendation, as there is a risk that individuals may be reluctant to go against an algorithmic recommendation.[83] This reluctance could stem from the perception that an algorithm is neutral or more accurate, or because of the difficulty in explaining why the algorithmic recommendation was overturned, rendering human-in-the-loop less effective or ineffective.[84]

Position in India

The National Strategy addresses the issue of accountability from the angle of applying existing negligence tests as opposed to strict liability. The present legal position on negligence is clear under tort law. To prove that an act was negligent, it is necessary to prove all the essentials namely duty, breach of that duty and damage as a consequence thereof. There will be no liability if the damage is not foreseeable (actual and proximate cause).[85] In general the burden of proof is on the plaintiff to prove negligence. An important maxim regarding negligence "*Res Ipsa Loquitur*" is used by the courts when the negligence "*speaks for itself*" and there is a presumption of negligence on the part of the defendant. This would be crucial in the context of AI. However, in instances where the damage was not foreseeable there would be no liability, considering the nature of AI systems this is an apparent outcome. As the National Strategy mentions, applying negligence tests would mean conducting damage impact assessments, requiring safe harbors to be formulated to insulate or limit liability "so long as *appropriate steps* to design, test, monitor, and improve the AI product have been taken". The challenge lies in defining what these "appropriate steps" are / could be. Much like the common law principle of "reasonableness", appropriate steps can only be determined having regard to facts of each case. In terms of apportionment of damages, the National

[80] McGregor, L., Murray, D., & Ng, V. (2019). *International Human Rights Law as A Framework for Algorithmic Accountability*. *International and Comparative Law Quarterly, 68*(2), 309-343. doi:10.1017/s0020589319000046
[81] McGregor, L., Murray, D., & Ng, V. (2019). *International Human Rights Law as A Framework for Algorithmic Accountability*. *International and Comparative Law Quarterly, 68*(2), 309-343. doi:10.1017/s0020589319000046
[82] McGregor, L., Murray, D., & Ng, V. (2019). *International Human Rights Law as A Framework for Algorithmic Accountability*. *International and Comparative Law Quarterly, 68*(2), 309-343. doi:10.1017/s0020589319000046
[83] McGregor, L., Murray, D., & Ng, V. (2019). *International Human Rights Law as A Framework for Algorithmic Accountability*. *International and Comparative Law Quarterly, 68*(2), 309-343. doi:10.1017/s0020589319000046
[84] McGregor, L., Murray, D., & Ng, V. (2019). *International Human Rights Law as A Framework for Algorithmic Accountability*. *International and Comparative Law Quarterly, 68*(2), 309-343. doi:10.1017/s0020589319000046
[85] *Ramesh Kumar Nayak v Union of India* AIR 1994 Ori 279; *Municipal Corporation Of Delhi v Subhagwanti & Others* 1966 AIR 1750

Strategy advocates "proportionate" payment rather than joint or several liability. Here again much is left to judicial interpretation as opposed to a more definite scheme.

Importantly, product liability in India is largely governed by the Consumer Protection Act, 1986, as opposed to the civil law. How would the imposition of liability then play out for deficient services of chatbots or other AI powered systems? These may not be pressing questions at this juncture but as AI use becomes more widespread concerns such as this would need to be dealt with.

In a clear preference to self-regulation, the National strategy steers clear of strict liability. The English legal principles of strict liability[86] have been applied in many Indian case laws.[87] Strict liability holds a person to be liable for harm even though - they were not negligent, or had no intention to cause harm. Even though a person may have done some positive efforts to avert the harm bearing in mind the nature of the hazardous activity, liability is not excluded. It can be argued that strict liability in these early stages of AI innovation can deter progress, at the same time there is merit is imposing it especially in more critical sectors. Strict liability not only strengthens compliance with other ethical principles but would also bring in more accountability. It is suggested that "strict liability" for sectors such as health where consequences could be severe should be considered.

It is observed that the Indian position has been to stay clear of "bright-lines" as is the case in Singapore, the intention to tap the many benefits of AI systems can be seen as the reason for this approach.

Legal Personhood

In a world of more sophisticated AI usage can affording legal personality provide the answer the issues of opacity and autonomy?

Legal personality is fundamental to any system of laws. The question of who can act, who can be the subject of rights and duties, is a primary legal concern[88].

Legal personhood can be natural and juridical. Natural persons are recognised because of the simple fact of being human. Juridical persons, by contrast, are non-human entities that are granted certain rights and duties by operation of law.[89] Corporations and other forms of business associations are the most common examples. The ability to sue and be sued is one of the primary attractions of personality for AI systems, as the European Parliament has acknowledged.[90] The necessity to accord legal personhood presumes that there are accountability gaps that can and should be filled.

[86] *Rylands v. Fletcher* [1868] UKHL 1 ; *Sochaki vs. Sas* [1947] 1 ALL ER 344
[87] *Jai Laxmi Salt Works v State of Gujarat* (1994) 4 SCC 1.
[88] Chesterman, S. (2020). *Artificial intelligence and the limits of legal personality International and Comparative Law Quarterly, 69*(4), 819-844. doi:10.1017/s0020589320000366
[89] Chesterman, S. (2020). *Artificial intelligence and the limits of legal personality International and Comparative Law Quarterly, 69*(4), 819-844. doi:10.1017/s0020589320000366
[90] European Parliament Resolution with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (European Parliament, 16 February 2017), para 59(f).

Furthermore, those who argue for legal personhood to AI, point out punishment as another important factor that should be considered. The concern could be potential criminal liability, if corporations can be punished for criminal offences can AI not follow suit?

Position in Singapore

To convict a company of a criminal offense in Singapore, it is generally necessary to prove beyond reasonable doubt the following elements: [91]

- The company committed the act prohibited by the offense (the *actus reus*).
- The company had a guilty state of mind (that is, the company had the required intention when committing the act that makes it an offense) (*mens rea*).[92]

Position in India

While there is no statute in the country for making corporations criminally liable, case laws by the Supreme Court have imposed punishments on companies for criminal offences.[93]

The Committee Report on cybersecurity, safety and ethics (draft)[94] alludes to affording legal personhood for AI systems along with appropriate insurance mechanisms. It also mentions the necessity to review the existing laws in line with AI developments.

To attribute intention to artificial persons such as companies, two principles - agency and identification have been followed. But would this be desirable and effective in the context of AI? Though the question of personality is a binary, however — recognized or not — the content of that status is a spectrum.[95] As such, both jurisdictions have not yet seriously considered attribution of personhood.

[91] Martin, A., & Seng, Y. (n.d.). Corporate liability in Singapore. Retrieved March 01, 2021, from https://globalcompliancenews.com/white-collar-crime/corporate-liability-singapore/#_ftnref3
[92] It should be noted that there could be offences under "strict liability."
[93] *Standard Chartered Bank v. Directorate of Enforcement*, AIR 2005 SC 2622; *Aneeta Hada v Godfather Travels and Tours Pvt. Ltd*, [2012 5 (SCC 661)]; *Iridium India Telecom Ltd v Motorola Inc*.,(2011) 1 SCC 74.

[94]Report of Committee D On Cyber Security, Safety, Legal and Ethical Issues. (n.d.). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_D-Cyber-n-Legal-and-Ethical.pdf
[95] Chesterman, S. (2019). *Artificial intelligence and the problem of autonomy*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3450540

Privacy and Data Protection

Position in Singapore

Data is integral to use of AI. Protection and use of data in Singapore are governed by the PDPA (for private sector). The PDP Commission has from time-to-time released guidance documents on the use of data in AI.[96] Data management in the public sector is governed by the Public Sector (Governance) Act ["PGSA"] and the Government Instruction Manual on IT Management. The PSGA imposes criminal penalties on public officers who recklessly or intentionally disclose data without authorisation, misuse data for a gain or re-identify anonymized data. The recent amendments for the PDPA includes enhanced fines and penalties for de-identifying anonymized data.[97]

The Smart Nation Singapore policy mentions of a Data Innovation Programme Office to advise companies how to better harness data, and to encourage data-driven innovation projects.[98] The Info-communications Media Development Authority has also developed the Data Protection Trustmark certification to help businesses verify their conformance to personal data protection standards and best practices.[99]

The Trusted Data Sharing Framework is another source of guidance to help organisations to establish a set of baseline practices by providing a common 'data-sharing language', and suggest a systematic approach to the broad considerations for establishing trusted data sharing partnerships.[100]

Position in India

The comprehensive Personal Data Protection Bill 2019 is yet to be passed. The absence of a comprehensive data protection regime has been the subject of much criticism in India. However, the right to privacy has been recognized as a fundamental right under the Constitution. In *Justice K.S. Puttaswamy v. Union of India*[101] the Supreme Court confirmed that the right to privacy is part of Article 21 of the Constitution of India, expressly affirming its applicability to the internet.

"Informational privacy is a facet of the right to privacy. The dangers to privacy in an age of information can originate not only from the state but from non-state actors as well. We commend to the Union Government the need to examine and put into place a robust regime

[96] Discussion Paper On Artificial Intelligence (Ai) And Personal Data – Fostering Responsible Development and Adoption of AI (2018, June 5). Retrieved January 3, 2021, from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf
[97] Personal Data Protection Act 2012.
[98] Smart Nation: The Way Forward. (n.d.). Retrieved March 01, 2021, from https://www.smartnation.gov.sg/docs/default-source/default-document-library/smart-nation-strategy_nov2018.pdf?sfvrsn=3f5c2af8_2
[99] Smart Nation: The Way Forward. (n.d.). Retrieved March 01, 2021, from https://www.smartnation.gov.sg/docs/default-source/default-document-library/smart-nation-strategy_nov2018.pdf?sfvrsn=3f5c2af8_2
[100] Trusted Data Sharing Framework. (n.d.). Retrieved March 01, 2021, from https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf
[101] AIR 2017 SC 4161

for data protection. The creation of such a regime requires a careful and sensitive balance between individual interests and legitimate concerns of the state."

From its various policies it is seen that India is keen to capitalize on its large datasets made possible due to its population. The National AI Resource Platform is being envisaged as a catalyst to the development of a partnership/ collaboration/ contribution/ participation model for knowledge sharing, data sharing, meta-data structure, annotation, API framework, intellectual property creation, innovation, value added AI services, government adoption and human interactions.[102]

The observations of Committee A: On Platforms and Data on Artificial Intelligence are trite in this regard:

"It is important to define the standards and processes for data collection, data sanitization, anonymization/ pseudonymization…. Having detailed security guidelines for the use of such data (if relevant) driven by industry standards and best practices is key to building trust in the ecosystem"[103]

It is important to regulate data thorough legislation and necessary subordinate legislation so that data are not subject to misuse. It should be remembered that the fundamental right to privacy can only be claimed against the State. To ensure that the private sector is also bound by strict data protection principles, a statute is necessary. The 2018 National Strategy for AI discussion paper has high aspirations of India becoming a data marketplace and positioning India as a 'garage' for testing AI solutions applicable to the developing world. In so far as such goals are concerned it remains to be seen what the final draft of the data protection legislation would be like. The draft was subject to criticism for altering some of the clauses recommended by the Committee for Data Protection.[104] The exceptions to processing of data without consent should also be narrowly constructed so that privacy concerns are not undermined.

Unlike in Singapore, a draft regulatory framework on the protection of non-personal data has emerged in India. A draft was released in July 2020, and a revised draft released in December 2020 after stakeholder feedback. Non-personal data is defined as follows (quoting from framework):

> *"i. Non-Personal Data – When the data is not 'Personal Data' (as defined under the PDP Bill), or the data is without any Personally Identifiable Information (PII), it is considered Non-Personal Data.*
>
> *ii. A general definition of Non-Personal Data according to the data's origins can be:*

[102] Report of Committee - A On Platforms And Data On Artificial Intelligence. (2019, July). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_A-Report_on_Platforms.pdf
[103] Report of Committee - A On Platforms And Data On Artificial Intelligence. (2019, July). Retrieved March 01, 2021, from https://www.meity.gov.in/writereaddata/files/Committes_A-Report_on_Platforms.pdf
[104] Kumar, R. (2019, September 9). India needs to bring an algorithm transparency bill to combat bias. Retrieved March 01, 2021, from https://www.orfonline.org/expert-speak/india-needs-to-bring-an-algorithm-transparency-bill-to-combat-bias-55253/

o Firstly, data that never related to an identified or identifiable natural person, such as data on weather conditions, data from sensors installed on industrial machines, data from public infrastructures, and so on.

o Secondly, data which were initially personal data, but were later made anonymous. Data which are aggregated and to which certain data- transformation techniques are applied, to the extent that individual-specific events are no longer identifiable, can be qualified as anonymous data."

The draft proposes national-level regulation "to establish rights over non-personal data collected and created in India." The goals of the regulatory framework are stated as (Quoting from the framework):

"i. To create an enforcing framework that
o Establishes rights of India and its communities over its non-personal data.
o Addresses privacy, re-identification of anonymized personal data, and prevent misuse of and harms from data.

ii. To create an enabling framework that
o Ensures unlocking economic benefit from non-personal data for India and its people.
o Creates a data sharing framework.
o Provides certainty of regulations.

iii. The Committee believes that with such a regulation, India could become the first country to put in place a simple, comprehensive framework for non-personal data."

The rights over non-personal data are said to belong to communities, and a community is defined as "The Committee defines a community as any group of people that are bound by common interests and purposes, and involved in social and/or economic interactions. It could be a geographic community, a community by life, livelihood, economic interactions or other social interests and objectives, and/or an entirely virtual community."

The establishment of a Non-Personal Data Authority (NPDA) is proposed, with the function of enforcing the following (quoting from the framework):

- *"Establish rights over Indian non-personal data in a digital world.*
- *Address privacy, re-identification of anonymized personal data, prevent misuse of data.*
- *In case of data sharing for High-value Datasets, the NPDA will adjudicate only when a data custodian refuses to share data with the data trustee."*

"Data Custodian" is defined as "a Government or a Private organization that has an obligation/responsibility to share appropriate NPD when data requests are made for defined data sharing purposes."

It is of note that the framework defines a concept of "high value datasets" or ("HVDs"). HVDs are "datasets that are beneficial to the community at large, and shared as a public good." The

framework states that such datasets are useful for policymaking, job creation, creation of new businesses, research and education, alleviating poverty, financial inclusion, developing agriculture, developing skills, healthcare, urban planning, environmental planning, energy, diversity and inclusion etc. A government or NGO may request the creation of a High Value Dataset, in consultation with the Non Personal Data Authority (NPDA). The NPDA will create guidelines to determine appropriateness of the HVD and the data trustee. Specifically note that "Aggregate data level**:** Should be made available by public and private entities."

Payal Malik, Advisor and Head Economics, Competition Commission of India, notes that the aim of the proposed non-personal data regulatory framework is to democratize access to data, so that it does not remain in the hands of only a few. However, the concept is still in its early stages, and it remains to be seen what will happen with proprietary datasets that are considered "high value."[105]

An experienced Public Policy Consultant, Deepak Maheshwari, also notes that definitions need greater clarity. Firstly, it is necessary to clarify the differences in the respective mandates and scope of the Data Protection Authority proposed in the Personal Data Protection Bill, 2019 and the proposed Non-Personal Data Protection Authority (NPDA) by another committee of experts. In addition, there has to be clarity on if and how the two authorities may interact with each other, for example, in the event of disagreements on whether a particular data type is to be treated as personal or non-personal data, considering such determinations may be context specific. Another challenge arises with the concept of community data proposed within the NPD framework. Firstly, some communities may be fluid and temporary, being specific to the context of a particular place, time or occasion. For example, the members of a panel at a conference, or the passengers on a bus during a specific journey may constitute a community at a given time period, but may disperse thereafter. Secondly, it is unclear which member or members of the community will be tasked with supervising community rights and how they would be able to do so. Thirdly, an individual may be a member of many communities at any given point of time. Hence, communities need to be more clearly and narrowly defined in order to ensure that such rights are actually enforceable.[106]

In terms of ethical principles, the reasoning behind the regulatory framework appears to be in sync with the theme of inclusion, which is central to India's approach to AI. For instance, benefiting communities and the public good are aims of the framework. Poverty alleviation, and diversity and inclusion are mentioned as possible uses of high-value datasets. However, further analysis of ethical principles will require the conceptual matters discussed above to be clarified.

[105] Personal Communication, 10 April 2021.
[106] Personal Communication, 13 April 2021.

Enforceability

The foremost criticisms against ethics frameworks continue to be those on its enforceability i.e. what happens if the ethical principles are not implemented? Lack of clear and defined enforcement mechanisms leads to questionable implementation or even no implementation. While no consequences are hard coded for violation of ethical principles, it should be remembered that they stem from the seeds of "self-regulation" as opposed to enforced regulation by laws.

Until 2019, it was found that there are at least 84 'AI Ethics' initiatives have published reports describing high-level ethical principles, tenets, values, or other abstract requirements for AI development and deployment.[107] Whether these principles can be successful by themselves has been met with some skepticism. For instance, Mittelstadt (2019) argues that ethical frameworks alone are prone to fail to regulate AI solutions because unlike other fields where ethics are used as regulatory interventions such as medicine, the development of AI lacks "(1) common aims and fiduciary duties, (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms." [108] How / whether these factors will emerge over time as the field of AI develops further remains to be seen. Mittelstaedt also notes that, given the various, different technologies that are called "AI," "bottom-up" approaches to AI ethics need to be supported, and specific case studies should complement "top down" approaches.[109] Case studies will be taken up in Stage 3 of this report.

Singapore has established Advisory Council on Ethical Use of AI and Data under the Infocomm Media Development Authority to advise Singapore Government on issues arising from commercial deployment of AI that may require policy or regulatory intervention.[110] To highlight the challenges in AI enforcement, NITI Aayog had released a Working Document: Enforcement Mechanisms for Responsible #AIforAll[111] (draft for discussion) by proposing an oversight body that is tasked with broad responsibilities including managing and updating principles for responsible AI, providing clarity on responsible behavior etc. NITI Aayog has also proposed establishing an AI specific cloud infrastructure to facilitate research and solution development.[112] Regulatory authorities in individual sectors have established regulations on the use of AI in their specific sectors (an example of this in the health sector will be seen in Stage 3). In terms of data protection, there is a presence of overall regulatory bodies for enforcement of data protection regulations (the Personal Data Protection Commission (PDPC) in Singapore, and the proposed Data Protection Authority (DPA) in India's draft bill).

---

[107] Jobin, A., Ienca, M. & Vayena, E. *The global landscape of AI ethics guidelines*. *Nat. Mach. Intell.* **1**, 389–399 (2019).
[108] Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nat Mach Intell* 1, 501–507. https://doi.org/10.1038/s42256-019-0114-4
[109] Ibid.
[110] Composition of the Advisory Council on the ethical use of artificial INTELLIGENCE ("AI") and data. (2018, August 30). Retrieved March 01, 2021, from https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2018/composition-of-the-advisory-council-on-the-ethical-use-of-ai-and-data
[111] Working Document: Enforcement Mechanisms for Responsible #AIforAll. (2020). Retrieved March 01, 2021, from (2020). Retrieved March 01, 2021, from https://ourgovdotin.files.wordpress.com/2020/11/niti-working-document-enforcement-mechanisms-for-responsible-aiforall.pdf
[112] AIRAWAT- Establishing an AI specific Cloud Computing Infrastructure for India- An Approach Paper. (2020, January). Retrieved December 30, 2020, from https://niti.gov.in/sites/default/files/2020-01/AIRAWAT_Approach_Paper.pdf

Whether similar bodies would be necessary or desirable for regulating AI remains to be seen.[113]

In terms of the role of regulation, Payal Malik notes that one cannot look at the role of regulation in a binary form when considering the question of whether regulation stifles innovation. While excessive amounts of regulation can limit innovation, a fully deregulated environment can also limit innovation (e.g., through killer acquisitions targeting newer start-ups and firms, by incumbent firms). Therefore, this question cannot be conceived of as an either / or proposition. Furthermore, with regard to "soft-touch" regulation, it may work "until it doesn't," and there have been debates about to what extent soft touch regulation may work and to what extent stronger regulation may be needed.

[113] For example, Andrew Tutt argues in favor of a centralized regulator for algorithms in the style of the United States' Food and Drug Administration. Tutt, A. (2016). An FDA for Algorithms. *69 Admin. L. Rev. 83 (2017).* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994#

Stage 3: Case Studies

In this section, we explore the use of the previously analyzed ethical principles in two specific case studies – The SELENA+ Diabetic Retinopathy Screening Tool in Singapore and the deployment of Google's flood forecasting system in India.

Singapore: EyRIS's SELENA+ Diabetic Retinopathy Screening

As of 2019, the International Diabetes Federation[114] states that 463 million people worldwide have diabetes, consistently exceeding prior projections.[115] This number is expected to exceed 700 million by 2045.[116] Of this figure, approximately a third are estimated to have diabetic retinopathy (DR), which can lead to vision loss if left untreated.[117]

Vision loss can be prevented by early detection and prompt treatment, which requires regular screening (Ferris, 1993).[118] A study by Liew et al. (2014) on the causes of blindness in adults aged 16–64 years in England and Wales between 2009–2010, noted that DR was no longer the leading cause of blindness for the first time in five decades.[119] The authors attributed this to the introduction of nationwide screening programmes between 2003–2008 in England and Wales, thereby strengthening the case for regular screening.

However, current manual diabetic retinopathy screening is labor-intensive and inconsistent, requiring trained human graders who are difficult to acquire and retain.[120] As such, accurate automated DR screening tools with global scalability are immensely valuable.

---

[114] International Diabetes Federation. (2019). *IDF Diabetes Atlas 9th Edition 2019*. https://www.diabetesatlas.org/

[115] International Diabetes Federation. (2003). *IDF Diabetes Atlas 2nd Edition 2003*. https://www.diabetesatlas.org/ ; International Diabetes Federation. (2006). *IDF Diabetes Atlas 3rd Edition 2006*. https://www.diabetesatlas.org/ ; International Diabetes Federation. (2009). *IDF Diabetes Atlas 4th Edition 2009*. https://www.diabetesatlas.org/

[116] International Diabetes Federation. (2019). *IDF Diabetes Atlas 9th Edition 2019*. https://www.diabetesatlas.org/

[117] Yau, J. W. Y., Rogers, S. L., Kawasaki, R., Lamoureux, E. L., Kowalski, J. W., Bek, T., Chen, S.-J., Dekker, J. M., Fletcher, A., Grauslund, J., Haffner, S., Hamman, R. F., Ikram, M. K., Kayama, T., Klein, B. E. K., Klein, R., Krishnaiah, S., Mayurasakorn, K., … O'Hare, J. P. (2012). Global Prevalence and Major Risk Factors of Diabetic Retinopathy. *Diabetes Care, 35*(3), 556–564. https://doi.org/10.2337/dc11-1909

[118] Ferris, F. L. (1993). How effective are treatments for diabetic retinopathy? JAMA: The Journal of the American Medical Association, 269(10), 1290. https://doi.org/10.1001/jama.1993.03500100088034

[119] Liew, G., Michaelides, M., & Bunce, C. (2014). A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open, 4*(2), e004015. https://doi.org/10.1136/bmjopen-2013-004015

[120] Tufail, A., Kapetanakis, V. V., Salas-Vega, S., Egan, C., Rudisill, C., Owen, C. G., Lee, A., Louw, V., Anderson, J., Liew, G., Bolter, L., Bailey, C., Sadda, S., Taylor, P., & Rudnicka, A. R. (2016). An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technology Assessment, 20*(92), 1–72. https://doi.org/10.3310/hta20920

EyRIS's SELENA+

In October 2019, EyRIS, a Singapore-based company that specializes in artificial intelligence (AI) for healthcare, received regulatory approval from the Government of Singapore's Health Sciences Authority for the deployment of SELENA+ (EyRIS, 2019a).[121] SELENA+ is a deep learning AI that screens for DR and related eye diseases using retinal images, and was jointly developed by the Singapore Eye Research Institute (SERI) and National University of Singapore - School of Computing (NUS-SoC). SELENA+ has since been approved by regulatory bodies in: Malaysia,[122] the European Union,[123] Brazil,[124] and Indonesia[125]; and was featured in Singapore's National Artificial Intelligence Strategy.[126]

A publication in the Journal of the American Medical Association by Ting et al. (2017)[127] documents the technology behind SELENA+. The authors identified two objectives for the study:

1. **To train and validate an AI to detect DR and related eye diseases based on retinal images.** Initially, the objective was to detect only DR, however Ting et al. (2017) acknowledge an argument posited by Chew and Schachat (2015)[128] who assert that it is clinically unacceptable to not screen for glaucoma and age-related macular degeneration (AMD) when screening retinal images for DR. As such, Ting et al. (2017) expanded this objective to include related eye diseases as well.

2. **To evaluate two models of the AI system: a fully-automated model, and a semi-automated model.** The fully-automated model was intended for communities with no existing screening programmes and would not require human involvement. The semi-automated model was intended for communities with existing screening programmes,

---

[121] EyRIS. (2019a, October 1). *Regulatory Body Approves Marketing of SELENA+ That Can Detect 3 Eye Diseases*. https://www.eyris.io/latest_news.cfm?id=27

[122] EyRIS. (2019b, December 17). *NOVA Receives GDPMD Certification*. https://www.eyris.io/latest_news.cfm?id=33

[123] EyRIS. (2020a, March 9). *SELENA+ Obtains Approval for Market Access in EU*. https://www.eyris.io/latest_news.cfm?id=37

[124] EyRIS. (2020b, October 27). *SELENA+ Receives Regulatory Approval In Brazil*. https://www.eyris.io/latest_news.cfm?id=48

[125] EyRIS. (2021, January 12). *SELENA+ Receives Regulatory Approval In Indonesia*. https://www.eyris.io/latest_news.cfm?id=51

[126] EyRIS. (2019c, November 19). *National AI Strategy: The next key frontier of Singapore's Smart Nation Journey*. https://www.eyris.io/latest_news.cfm?id=31

[127] Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I. Y., Lee, S. Y., Wong, E. Y. M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N. C., Finkelstein, E. A., Lamoureux, E. L., Wong, I. Y., Bressler, N. M., … Wong, T. Y. (2017). Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA, 318*(22), 2211. https://doi.org/10.1001/jama.2017.18152

[128] Chew, E. Y., & Schachat, A. P. (2015). Should we add screening of age-related macular degeneration to current screening programs for diabetic retinopathy? *Ophthalmology, 122*(11), 2155–2156. https://doi.org/10.1016/j.ophtha.2015.08.007

where referable cases that did not meet a preset sensitivity threshold would be subject to a secondary screening by human graders.

Ting et al. (2017) described their model as a composition of 8 convolutional neural networks (CNNs), each using an adaptation of the Visual Geometry Group Neural Network (VGGNet) architecture. The model was trained to identify four diagnoses: referable DR, vision-threatening DR, referable possible Glaucoma, and referable AMD.

Datasets

Ting et al. (2017) acquired their training and primary validation datasets from patients attending the Singapore Integrated Diabetic Retinopathy Program (SiDRP). As such, only Chinese, Malay, and Indian ethnic groups were included in these datasets. Data from 2010–2013 was allocated to the training dataset, and data from 2013–2014 was allocated to the primary validation dataset. The authors acquired the external validation dataset by amalgamating multiple datasets from different sources worldwide, and included the following ethnic groups: Chinese, Malay, Indian, White, African American, and Hispanic. The sizes of the datasets used are depicted in Table 1.

Table 1: *The number of retinal images used for each dataset.*

	Dataset		
Diagnosis	**Training**	**Primary Validation**	**External Validation**
Referable / Vision-threatening DR	76,370	71,896	40,752
Referable possible Glaucoma	125,189	71,896	—
Referable AMD	72,610	35,948	—

Ting et al. (2017) stated that the labelling of each retinal image in the training and primary validation datasets was conducted by two trained senior certified nonmedical professional graders, each with over five years of experience. Conflicts were resolved by a retinal medical specialist with over five years of experience. The retinal images in the external validation dataset were labelled by similar configurations of nonmedical graders and medical specialists.

Results

With regards to their primary objective, Ting et al. (2017) reported on the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity, for each of the four diagnoses validated against the primary validation dataset. A summary of these results are shown in Table 2.

Table 2: *The results of the model for all four diagnoses, validated by the primary validation dataset.*

Diagnosis	AUC	Sensitivity (%)	Specificity (%)
Referable DR	0.936	90.5	91.6
Vision-threatening DR	0.958	100.0	91.1
Referable possible Glaucoma	0.942	96.4	87.2
Referable AMD	0.931	93.2	88.7

Additionally, Ting et al. (2017) documented the AUC for referable DR disaggregated by age, sex, and blood glucose levels, a summary of which is available in Table 3.

Table 3: *The results of the model for referable DR, validated by the primary validation dataset and disaggregated by age, sex, and blood glucose levels.*

Category	AUC
Age < 60	0.980
Age ≥ 60	0.920
Male	0.952
Female	0.948
HbA_{1c} < 8%	0.938
HbA_{1c} ≥ 8%	0.954

Furthermore, Ting et al. (2017) detailed the AUC, sensitivity, and specificity for referable DR, validated against the multitude of ethnic groups present in the external validation datasets, as shown in Table 4.

Table 4: *The results of the model for referable diabetic retinopathy, validated by the external validation dataset and disaggregated by dataset origin.*

Dataset Origin	Ethnicity	AUC	Sensitivity (%)	Specificity (%)
China	Chinese	0.949	98.7	81.6
Singapore	Malay	0.889	97.1	82.0
Singapore	Indian	0.917	99.3	73.3
Singapore	Chinese	0.919	100.0	76.3
China	Chinese	0.929	94.4	88.5
United States	African American	0.980	98.8	86.5
Australia	White	0.983	98.9	92.2
Mexico	Hispanic	0.950	91.8	84.8
Hong Kong	Chinese	0.948	99.3	83.1
Hong Kong	Chinese	0.964	100.0	81.3

In reference to their secondary objective, Ting et al. (2017) reported on the sensitivity and specificity for three sensitivity thresholds for the semi-automated model: 90%, 95%, and 99%. Table 5 documents these results alongside those of the fully-automated model.

Table 5: *The results of the fully-automated model compared with the semi-automated model at differing sensitivity thresholds.*

Model	Secondary Screening (%)	Sensitivity (%)	Specificity (%)
Fully-automated	—	93.0	77.5
Semi-automated (90%)	25.3	91.3	99.5
Semi-automated (95%)	37.0	95.1	99.5
Semi-automated (99%)	59.7	97.1	99.4

Ethical Considerations

For the most part, EyRIS has worked diligently to ensure that SELENA+ adheres to ethical practices. Some of these ethical practices and their shortcomings are documented below.

Model performance. Any medical diagnosis has 4 potential outcomes:

- True positive, i.e. a patient who has the disease and is correctly identified as having the disease. The true positive rate is represented by *sensitivity*.
- False negative, i.e. a patient who has the disease but is incorrectly identified as not having the disease. The false negative rate is represented by *1–sensitivity*.
- False positive, i.e. a patient who does not have the disease but is incorrectly identified as having the disease. The false positive rate is represented by *1–specificity*.
- True negative, i.e. a patient who does not have the disease and is correctly identified as not having the disease. The true negative rate is represented by *specificity*.

For medical diagnoses, it is imperative that patients who have the disease are correctly identified as so. As such, the pertinent outcomes are true positives and false negatives, the rates of which are represented by *sensitivity* and *1–sensitivity* respectively.

For completeness, EyRIS has reported on AUC, sensitivity, and specificity in Tables 2, 3, 4 and 5. However, as discussed above, sensitivity is the most salient metric, and EyRIS has demonstrated a sensitivity consistently greater than 90% across all models and disaggregations.

Bias and fairness. EyRIS has made a considerable effort to account for bias and fairness by utilizing large, diverse datasets. Whilst their training and primary validation datasets are biased, only including ethnicities present in Singapore, EyRIS has assembled an external

validation dataset by coalescing datasets from across the world for 6 different ethnicities. They reported consistent results for referable DR disaggregated by: age, sex, blood glucose level, and ethnicity; as depicted in Tables 3 and 4 (Ting et al., 2017).

However, it should be noted that EyRIS reported the aforementioned results only for referable DR. Recall that their primary objective included three additional diagnoses: vision-threatening DR, referable possible Glaucoma, and referable AMD (Ting et al., 2017). Consequently, their results cannot speak to the consistency of detecting Glaucoma and AMD across ages, sexes, blood glucose levels, and ethnicities excluding those residing in Singapore.

Level of human involvement. EyRIS has given due consideration to the appropriate level of human involvement. This is evidenced by their secondary objective to produce two models: a fully-automated model designed to be used in regions with no existing screening programmes; and a semi-automated model designed to work alongside existing screening programmes, supplementing them. The fully-automated model would correspond to the "human-out-of-the-loop" category in the "loop" categories defined in the MAIGF, as it is intended for regions that do not have a screening program. The semi-automated model corresponds to the "human-in-the-loop" category, since the AI model provides a recommendation, but a human takes the final decision. As the Ting et al. (2017) paper states: "a fully automated model for communities with no existing screening programs and a semiautomated model in which referable cases from the DLS undergo a secondary assessment by human graders."

Dataset acquisition. Retinal images fall under the category of personal data. As such, it is necessary to consider how the dataset was acquired. EyRIS acquired their training and primary validation datasets, which constitutes the majority, from the SiDRP (Ting et al., 2017). The SiDRP collects data from patients attending their programme.

Compliance with data protection laws is required in the approvals process of the Health Sciences Authority for software medical devices, which will be touched on in the next section.

Ethical AI Framework Compliance

The Model Artificial Intelligence Governance Framework (MAIGF) is government-backed. The framework claims to be algorithm-, technology-, sector-, and scale-agnostic, and covers a wide range of ethical practices concerning: internal AI governance structures and measures, determining the level of human involvement in AI systems, datasets used for model development, the model itself, and stakeholder interaction and communication. It is currently in its second iteration, having incorporated feedback from multiple stakeholders.

Having been featured in Singapore's National Artificial Intelligence Strategy,[129] we can safely assume that SELENA+ is compliant with the MAIGF.

[129] Smart Nation and Digital Government Office. (2019). *National Artificial Intelligence Strategy*. Government of Singapore. https://www.smartnation.gov.sg/why-Smart-Nation/NationalAIStrategy

Regulatory Approval

Each region in the world has its own regulatory body responsible for medical devices, such as: the Food and Drug Administration in the United States, the National Health Service in the United Kingdom, and the European Medicines Agency in Europe. However, not all regulatory bodies have a regulatory framework for AI medical devices. Those that do are members of the International Medical Device Regulators Forum (IMDRF), and as such draw heavy inspiration from a jointly produced document titled "Software as a Medical Device (SaMD): Clinical Evaluation." [130]

Regulatory procedure typically involves first classifying the AI medical device based on risk—the significance of the output of the AI medical device on making a healthcare decision, and the severity of the condition (i.e. the potential impact of the healthcare decision). The regulator then requests documentation about the AI medical device, in order to examine its lifecycle to determine whether it meets certain standards, with riskier AI medical devices having to satisfy stricter standards.[131]

As a member of the IMDRF, Singapore is no different. The Health Sciences Authority (HSA) is the regulatory body responsible for devices such as SELENA+ in Singapore. Having attained Class B approval for SELENA+, EyRIS would have had to submit documents pertaining to: the datasets used to train, test, and validate the model; methodology to ensure integrity of the datasets; the type of model used and justification for its selection; the results of the evaluation of the model; how the model would function in deployment; the continuous learning mechanism; the safety mechanism; etc.[132]

---

[130] Software as a Medical Device Working Group. (2017). *Software as a Medical Device (SaMD): Clinical Evaluation*. International Medical Device Regulators Forum. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf

[131] See Food and Drug Administration. (2019). *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*. Department of Health and Human Services, Federal Government of the United States. https://www.fda.gov/media/122535/download ; National Institute for Health and Care Excellence. (2019). *Evidence standards framework for digital health technologies*. Department of Health and Social Care, Government of the United Kingdom. https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies ; Health Canada. (2019). *Guidance Document: Software as a Medical Device (SaMD): Definition and Classification*. Government of Canada. https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/software-medical-device-guidance-document.html

[132] Health Sciences Authority. (2020). *Regulatory Guidelines for Software Medical Devices – A Life Cycle Approach*. Ministry of Health, Government of Singapore. https://www.hsa.gov.sg/docs/default-source/hprg-mdb/gudiance-documents-for-medical-devices/regulatory-guidelines-for-software-medical-devices---a-life-cycle-approach.pdf

Analysis of Ethical Frameworks and Regulatory Requirements

Ethical frameworks are voluntary, while regulatory requirements are mandatory. As such, they depict ideal and practical scenarios respectively. Table 6 documents a summary of the concerns addressed by the MAIGF and the HSA, based on information obtained from the MAIGF and the HSA[133] respectively. Table 7 details the justifications for Table 6. The list of concerns was generated by first summarizing the concerns addressed by the MAIGF and HSA separately, and then merging the lists.

Please see the table on the following page for our interpretation of how the MAIGF and HSA requirements compare.

[133] Health Sciences Authority. (2020). *Regulatory Guidelines for Software Medical Devices – A Life Cycle Approach*. Ministry of Health, Government of Singapore. https://www.hsa.gov.sg/docs/default-source/hprg-mdb/gudiance-documents-for-medical-devices/regulatory-guidelines-for-software-medical-devices---a-life-cycle-approach.pdf

Table 6: *Summary of the concerns addressed by the Model Artificial Intelligence Governance Framework (MAIGF) and the Health Sciences Authority (HSA).*

Concern	MAIGF	HSA
Internal Structures		
Existence of an internal AI governance structure	✓	✗
Existence of an internal AI risk management system	✓	✗
Dataset		
Compliance with personal data protection laws	✓	✓
Documentation of data lineage	✓	✓
Ensurance of data quality	✓	✓
Management of inherent biases	✓	✓
Utilization of testing, training, and validation datasets	✓	✓
Algorithm and Model		
Implementation of appropriate data pre-processing	✓	✓
Selection of a suitable level of human involvement	✓	✓
Selection of an appropriate machine learning model	✓	✓
Utilization of performance metrics	✗	✓
Explainability of the model	✓	✗
Repeatability of the model	✓	✗
Robustness of the model	✓	✓
Traceability of the model	✓	✓
Reproducibility of the model	✓	✓
Auditability of the model	✓	✓
Documentation of expected workflow	✗	✓

Post-Deployment		
Evaluation of real-world performance	✓	✓
Implementation of regular dataset reviews and updates	✓	✓
Implementation of regular model reviews and updates	✓	✓
Existence of the option to opt-out	✓	✗
Stakeholder Interaction		
Assessment of the impacts of the AI system	✓	✗
Existence of communication and feedback channels	✓	✗

Table 7: *Justification for the summary of the concerns addressed by the Model Artificial Intelligence Governance Framework (MAIGF) and the Health Sciences Authority (HSA).*

Concern	MAIGF	HSA
Internal Structures		
Existence of an internal AI governance structure	✓ Addressed in detail on pages 21–23	✗ Not addressed
Existence of an internal AI risk management system	✓ Addressed in detail on page 24	✗ Not addressed
Dataset		
Compliance with personal data protection laws	✓ Addressed in brief on pages 13 and 17	✓ Addressed explicitly on pages 29–30
Documentation of data lineage	✓ Addressed in detail on page 37	✓ Addressed explicitly on page 31
Ensurance of data quality	✓ Addressed in detail on page 38	✓ Addressed explicitly on page 31
Management of inherent biases	✓ Addressed in detail on pages 38–39	✓ Addressed explicitly on page 31
Utilization of testing, training, and validation datasets	✓ Addressed in detail on page 40	✓ Addressed explicitly on page 31

Algorithm and Model

Implementation of appropriate data pre-processing	✓	Addressed in brief on page 51	✓	Addressed explicitly on page 31
Selection of a suitable level of human involvement	✓	Addressed in detail on pages 28–32	✓	Addressed explicitly on page 32
Selection of an appropriate machine learning model	✓	Briefly addressed on page 44	✓	Addressed explicitly on page 31
Utilization of performance metrics	✗	Specific metrics not addressed	✓	Addressed explicitly on page 32
Explainability of the model	✓	Addressed in detail on pages 44–45	✗	Not addressed
Repeatability of the model	✓	Addressed in detail on page 46	✗	Not addressed
Robustness of the model	✓	Addressed in detail on page 47	✓	Addressed implicitly on page 33—"Safety mechanism to detect anomalies and any inconsistencies in the output result"
Traceability of the model	✓	Addressed in detail on pages 48–49	✓	Addressed explicitly on page 33
Reproducibility of the model	✓	Addressed in detail on page 50	✓	Addressed explicitly on page 32
Auditability of the model	✓	Addressed in detail on page 51	✓	Addressed implicitly—auditability is a prerequisite for HSA approval

Documentation of expected workflow	✗	Not addressed	✓	Addressed explicitly on page 32
Post-Deployment				
Evaluation of real-world performance	✓	Addressed in brief on page 24	✓	Addressed explicitly on page 34
Implementation of regular dataset reviews and updates	✓	Addressed in detail on page 40	✓	Addressed explicitly on page 32
Implementation of regular model reviews and updates	✓	Addressed in detail on page 48	✓	Addressed explicitly on page 34
Existence of the option to opt-out	✓	Addressed in detail on page 56	✗	Not addressed
Stakeholder Interaction				
Assessment of the impacts of the AI system	✓	Briefly addressed on pages 14, 22, and 29	✗	Not addressed
Existence of communication and feedback channels	✓	Addressed in detail on pages 56–57	✗	Not addressed

---

As a voluntary ethical framework, the MAIGF has a lot of leeway. Consequently, the concerns addressed include:

- Peripheral areas such as internal AI governance structures and stakeholder interactions, with a strong focus on organizations incorporating measures to internalize the effects of utilizing AI systems; and
- Concepts that are not universally applicable such as model explainability.

The MAIGF does not address specific performance metrics that could potentially be used to evaluate AI systems. In contrast, the concerns addressed by the HSA are limited by what can be practicably regulated. The HSA requirements are far more focused on the development of and the implementation of the AI model itself—whether it works as intended and whether it is safe. Hence explainability is not explicitly required, although the auditability is necessary and expected workflow needs to be documented. Fairness / Lack of bias is dealt with (management of inherent biases is required) as are privacy and data protection (adherence to data protection laws is mandatory). In terms of accountability, if the device does not comply with the standards of the HSA, liability is on the manufacturer, importer, registrant (collectively, the person who has been granted the license) of such device. This is under the Health Products (Medical Devices) Regulation 2010.[134] The HSA also takes safety into account and has a mechanism to classify risk. Devices classified as Class B are considered "Low to moderate risk" by the HSA.[135]

[134] Health Products (Medical Devices) Regulation, 2010. Singapore https://sso.agc.gov.sg/SL/HPA2007-S436-2010#top Retrieved September 9, 2021.
[135] Health Sciences Authority (n.d). *Risk classification of medical devices.* https://www.hsa.gov.sg/medical-devices/registration/risk-classification-rule . Retrieved September 9, 2021.

India: Deployment of Google's Flood Forecasting Initiative

According to the Centre for Research on the Epidemiology of Disasters (2021), floods are the most common type of natural disaster worldwide, with riverine floods being the most prevalent subtype.[136] From 2000–2020 India has experienced 108 riverine floods, affecting 258.9 million people, causing 19,908 deaths, and costing $45.5 billion in damages. Studies have shown that early warning systems (EWS) for floods have been effective in mitigating the effects of floods.[137]

India's native EWS for floods is underdeveloped with much room for improvement. A report by the Department of Administrative Reforms and Public Grievances (2013) on India's EWS for floods described lackluster results—an accuracy of 60% and a lead time of 7–18 hours—and flood warnings that were disseminated ineffectively without indicating specific areas that were expected to flood. Despite this, the department reported a reduction in the loss of life and property due to floods since the inception of the system in 2009, thus highlighting the importance of an EWS for floods.[138]

[136] Centre for Research on the Epidemiology of Disasters. (2021). *EM-DAT: The International Disasters Database* (Version 2021-01-31) [Data set]. Catholic University of Leuven. https://public.emdat.be/

[137] Rogers, D., & Tsirkunov, V. (2010). *Global Assessment Report on Disaster Risk Reduction: Costs and Benefits of Early Warning Systems* (No. 69358). World Bank Group. http://documents.worldbank.org/curated/en/609951468330279598/Global-assessment-report-on-disaster-risk-reduction-costs-and-benefits-of-early-warning-systems ; World Health Organization. (2014). *Global report on drowning: Preventing a leading killer*. https://www.who.int/water_sanitation_health/diseases-risks/risks/global-report-on-drowning/en/ ; Turner, G., Said, F., Afzal, U., & Campbell, K. (2014). The effect of early flood warnings on mitigation and recovery during the 2010 pakistan floods. In A. Singh & Z. Zommers (Eds.), *Reducing Disaster: Early Warning Systems For Climate Change* (pp. 249–264). Springer Netherlands. https://doi.org/10.1007/978-94-017-8598-3_13 ; Perera, D., Seidou, O., Agnihotri, J., Rasmy, M., Smakhtin, V., Coulibaly, P., & Mehmood, H. (2019). *Flood Early Warning Systems: A Review Of Benefits, Challenges And Prospects* (UNU-INWEH Report Series, Issue 08). United Nations University Institute for Water, Environment and Health. https://inweh.unu.edu/flood-early-warning-systems-a-review-of-benefits-challenges-and-prospects/

[138] Department of Administrative Reforms and Public Grievances. (2013). *Flood Early Warning Systems - A Warning Mechanism for Mitigating Disasters during floods*. Ministry of Personnel, Public Grievances and Pensions, Government of India. https://darpg.gov.in/financialassistance/flood-early-warning-systems-warning-mechanism-mitigating-disasters-during-floods

Google's EWS for Floods

Google first entered the realm of flood forecasting in 2018 at the NeurIPS Artificial Intelligence for Social Good Workshop. Nevo et al. (2018)[139] presented a paper which identified two major problems associated with the current flood forecasting landscape that could potentially be improved by leveraging advancements in machine learning:

- **Scarcity of data in underdeveloped countries.** In particular: high-resolution, up-to-date digital elevation models (DEMs); river discharge data; and river-specific static attributes. The authors proposed: utilizing the abundance of alternative, related data sources such as satellite imagery; and leveraging data from multiple locations for transfer learning, arguing that the underlying physical principles are the same at all locations.
- **High computational costs of traditional physics-based models.** The computational complexity of physics-based models is linear with respect to coverage area, but exponential with respect to resolution (Nevo et al., 2020).[140] The authors posited that machine learning-based models could reduce computational costs by orders of magnitude.

Google launched their EWS for floods in India in 2018—initially covering just the Patna region, but later expanding their coverage to the entirety of India by 2020. Google's methodology evolved over the years; improving accuracy, increasing lead time, and reducing computational costs. Described below is the 2020 implementation of their EWS for floods.

Hydrologic Model

The first component of Google's EWS for floods is the hydrologic model. A hydrologic model receives inputs such as precipitation, solar radiation, soil moisture, upstream water level, river discharge, etc. and outputs a forecast for water level or river discharge—i.e. whether the river is expected to flood.

Google partnered with India's Central Water Commission to obtain hourly water level measurements from over 1000 stream gauges across India.[141] Nevo et al. (2020) then used

---

[139] Nevo, S., Wiesel, A., Hassidim, A., Elidan, G., Shalev, G., Schlesinger, M., Zlydenko, O., El-Yaniv, R., Gigi, Y., Moshe, Z., & Matias, Y. (2018). ML for Flood Forecasting at Scale. *Proceedings of the NeurIPS Artificial Intelligence for Social Good Workshop*. https://research.google/pubs/pub47651/

[140] Nevo, S., Elidan, G., Hassidim, A., Shalev, G., Gilon, O., Nearing, G., & Matias, Y. (2020). ML-based Flood Forecasting: Advances in Scale, Accuracy and Reach. *Proceedings of the NeurIPS Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop*. https://research.google/pubs/pub49993/

[141] Nevo, S. (2019, September 18). An Inside Look at Flood Forecasting. *Google AI Blog*. https://ai.googleblog.com/2019/09/an-inside-look-at-flood-forecasting.html

this data to train their hydrologic model to forecast water levels, boasting an R^2 of 0.99 and an average lead time of 20.7 hours.[142]

Inundation Model

The next component of Google's EWS for floods is the inundation model. An inundation model receives as input DEMs and either a forecasted water level or river discharge measurement, and outputs a map depicting which areas are expected to be inundated.

Publicly available DEMs were of low resolution, typically 30 meters. Therefore, Google leveraged the abundance of satellite imagery used in Google Maps, and machine learning to produce 1-meter resolution DEMs (Nevo, 2019). These DEMs, coupled with the forecasted water levels from the hydrologic model, were then fed into Google's version of the inundation model—named the morphological model—to produce an inundation map. Nevo et al. (2020) reported a precision of 76.2% and recall of 77.6% for these inundation maps at a 64-meter resolution. Additionally, the authors reported a several magnitude reduction in computational costs.

Dissemination of Flood Alerts

The final component of Google's EWS for floods is the dissemination of flood alerts. Flood alerts are delivered via Google Public Alerts (https://www.google.org/publicalerts), at the top of any Google Search result, and when using Google Maps. The alerts include information that would help people understand the severity of the flood, such as depth of water and approximate time of flood (Matias, 2020).[143] Those with Android smartphones are also able to receive alerts on their phones.[144]

In a collaborative effort between Google and the Yale Economic Growth Center, Berman et al. (2020) conducted a household survey of 810 households across 81 villages in the Indian state of Bihar, to better understand the impact of Google's EWS for floods. The authors reported that in 90% of villages at least one household received an alert before flood waters arrived, and 65% of households that received an alert took preventive measures.[145] As such, the EWS was generally impactful.

---

[142] Nevo, S., Elidan, G., Hassidim, A., Shalev, G., Gilon, O., Nearing, G., & Matias, Y. (2020). ML-based Flood Forecasting: Advances in Scale, Accuracy and Reach. *Proceedings of the NeurIPS Artificial Intelligence for Humanitarian Assistance and Disaster Response Workshop*. https://research.google/pubs/pub49993/

[143] Matias, Y. (2020, September 1). A big step for flood forecasts in India and Bangladesh. *The Keyword*. https://blog.google/technology/ai/flood-forecasts-india-bangladesh/

[144] Vincent, J. (2020, September 1). Google's AI flood warnings now cover all of India and have expanded to Bangladesh. *The Verge.* https://www.theverge.com/2020/9/1/21410252/google-ai-flood-warnings-india-bangladesh-coverage-prediction . Retrieved September 9, 2021

[145] Berman, M., Jagnani, M., Nevo, S., Pande, R., & Reich, O. (2020, July 2). Using technology to save lives during India's monsoon season. *Yale Economic Growth Center Blog*. https://egc.yale.edu/using-technology-save-lives-during-indias-monsoon-season

Ethical Considerations

Google's flood forecasting initiative isn't something that would typically require ethical considerations—it is a service that benefits the public, it is provided free of charge, it does not incur an opportunity cost to the Indian government, it does not use personal data, and it is a considerable improvement over the existing EWS for floods. In terms of ethical principles, it fits in well with the philosophy of inclusion that is present in India's approach to AI, as the flood forecasting system provides an important service to potentially vulnerable populations. However, it is important to consider what could be done better.

The ethical principle of accountability is also relevant here. While direct responsibility for disaster response lies with the relevant government authorities in India, The Berman et al. (2020) survey indicates that many residents of Bihar are using Google's service. According to a post on the Indian government's INDIAai website dated December 4, 2020, the technology now "cover[s] 200 million people across more than 250,000 sq km. Google technology is being used to improve the targeting of every alert the government sends; around 30mn notifications have been sent to people in flood affected areas, to date."[146] The same post notes, "Google.org has started a collaboration with the International Federation of Red Cross and Red Crescent Societies (IFRC) to build local networks that can get disaster alert information to people who wouldn't otherwise receive smartphone alerts directly. A partner notification infrastructure has been established to provide these forecasts for the CWC and other organisational partners that can use it to prepare for disaster management and relief efforts."[147] This makes Google's EWS an important actor in this environment as well.

In its current form, Google's flood forecasting initiative involves an exchange of information—the Government of India provides the required input data, and Google delivers a flood warning. As such, it does not contribute toward building flood forecasting capabilities locally. To illustrate why this could be a problem, consider a hypothetical point several years into the future when Google, for whatever reason, is no longer able to provide their EWS for floods service. If the Government of India had not invested in improving their own EWS for floods, and instead relied on Google's, then they would have no option but to rely on operation of an underdeveloped EWS. This problem is neatly encapsulated by the adage "give a man a fish and you feed him for a day; teach a man to fish and you feed him for a lifetime." In an article in *The Hindu* dated October 8, 2020, J. Harsha, Director of the Central Water Commission, India, argued that India lags behind in flood forecasting technology, and there is a need for a technically capable workforce.[148]

[146] INDIAai (2020, December 4). *Using AI to predict floods and save lives.* https://indiaai.gov.in/case-study/using-ai-to-predict-floods-and-save-lives . Retrieved September 9, 2021
[147] Ibid.
[148] J. Harsha (2020, October 8). Playing catch up in flood forecasting technology. *The Hindu.* https://www.thehindu.com/opinion/lead/playing-catch-up-in-flood-forecasting-technology/article32797281.ece . Retrieved September 9, 2021

Requirements for an effective EWS for floods

In their opening document on flood forecasting, Google stated that they were one of few organizations or communities that had all the necessary elements to build a successful EWS for floods, having:

- Knowledge and expertise in both flood forecasting and machine learning,
- Access to data at a global scale,
- Sufficient computational power to train the relevant models, and
- The means by which to disseminate alerts effectively (Nevo et al., 2018).

Google has been working to reduce the required computational power to train effective models. In particular, their 2020 implementation of the inundation model—the morphological model—incurs computational costs in the order of hours, which is a substantial improvement over classical models that incur computational costs in the order of years (Nevo et al., 2020). However, the other three elements remain unaddressed, and the Government of India could take initiatives to help address them.

While it is unreasonable to expect a private entity such as Google to address all of the barriers to building a successful EWS for floods for the Government of India, it is sensible to request for the sharing of knowledge to help build local expertise. This is congruent with India's goals to build local capacity to leverage AI for flood forecasting.[149] Hence, it is recommended that the government pursue partnerships with private entities to engage and collaborate with local institutions in an intellectual capacity as well.

[149] Ministry of Electronics and Information Technology. (2019). *Report by Committee B on Leveraging AI for Identifying National Missions in Key Sectors*. Ministry of Electronics and Information Technology, Government of India. https://www.meity.gov.in/writereaddata/files/Committes_B-Report-on-Key-Sector.pdf

Synthesis and Conclusions

What does implementation look like?

We do see some implementation of the AI ethics principles in the legal / regulatory and technical space. For example, the Singapore case law sets a precedent in terms of deterministic algorithms, and the Singapore PDPA sets out some regulations on how personal data can be used. Some of the ethical principles and ideas in the policy documents are present in the regulatory frameworks, and we see these considerations being taken into account in the case studies themselves.

The AI ethics principles and debates in Singapore and India are perhaps not that different to the principles and debates on AI ethics worldwide. The synthesis of Global AI ethics guidelines by Jobin et al. (2019) cited at the beginning of this paper identified transparency, justice and fairness, non-maleficence, responsibility, and privacy among the key principles.

We have also seen that the AI ethics discourse is very specific to applications. For instance, in the SELENA+ case, the rationale for regulatory approval is influenced by existing medical ethics principles regarding safety and accuracy. When it comes to implementing AI ethics principles in practice, regulators will sometimes diverge from certain ethics principles (as we saw in the SELENA+ case, where there was no explicit requirement for explainability, although explainability is highlighted in many of the policy documents). Regulators will work off of existing criteria, as well as making new ones for AI applications. Currently, it is not possible to achieve explainability for many AI applications, so these technical limitations will have to be accounted for. Overall, however, the legal and regulatory space is still quite sparse in terms of enforcing AI ethics principles in both countries. There is also a dearth of case laws to set precedents.

The high value datasets (HVDs) of the NPD proposed regulatory framework in India could lead to public greater sharing and democratization of data. However, it remains to be seen if the use of data could become more bureaucratic as a result.[150] Furthermore, there are considerations to explore regarding public-private partnerships, how data should be shared, and what the responsibilities of the public and private parties might be.

Moreover, there is still a high reliance on the voluntary uptake of principles. In Singapore the AI ethics discourse in fact appears to be geared largely towards giving suggestions to the creators / developers and encouraging them to adopt these (e.g. the Compendium of Use Cases, ISAGO, Data Protection Certification Trustmark). A preference for avoiding strict liability appears in India's NITI Aayog National AI Strategy as well.

[150] For criticisms on the regulation of non-personal data, see, for instance Burman, A. (2020, July 30). Regulations proposed by draft report on non-personal data need a relook. *The Indian Express.* https://indianexpress.com/article/opinion/columns/licence-raj-data-protection-bill-regulation-6529852/. Retrieved September 9, 2021

Recommendations for Policymakers

Here, we offer some learnings for policymakers from the above analysis.

More attention could be paid to drafting principles and frameworks for specific sectors. This is being seen in Singapore to some extent, for instance in the financial sector and medical sector. While overarching frameworks are useful, specific ethical questions and ways of working through them will emerge through case studies, and these findings need to be accounted for. Furthermore, algorithmic impact assessments have been proposed as a way for public agencies to ensure accountability and gauge the impact of the use of algorithmic technologies on the communities in which they are deployed.[151] The Government of Canada has also released an online Algorithmic Impact Assessment Tool, for instance.[152] Such assessments could also be a useful way of evaluating the effect of specific AI technologies on those who will be affected by the technologies, and minimize potential harms.

The role of the government and the role of the private sector in promoting AI ethics could be further interrogated. For instance, the government of India advocates heavily for the government's role in providing data for AI applications, including datasets without bias. While the government has a role to play in facilitating greater access to open data and championing representative datasets, there are challenges in correcting biases in datasets.[153] These include historical factors, where histories of discrimination against certain marginalized groups could have biased the way data has been collected, or the composition of the data (i.e. certain groups being over / underrepresented in certain categories). These may be difficult to correct. It has also been argued that there is no method that can be used to satisfy all types of fairness.[154] Governments should also consider how data sharing partnerships can be used to build AI for social good applications and the long term implications of these partnerships, such as how to ensure those benefits continue to accrue.

Furthermore, Prof. Ang Peng Hwa notes that the issue of surveillance in the context of AI and data should be considered. For instance, the use of data in emergency scenarios is coming to the forefront given the COVID-19 pandemic. Prof. Peng Hwa illustrates this by pointing to the debates around contact tracing applications and surveillance, including a recent controversy in which it was found that the police could be able to access data from Singapore's TraceTogether contact tracing application for criminal investigations, despite assurances that the data would only be used for COVID-19 contact tracing.[155] Hence, while emergency scenarios may require extensive collection of personally identifiable information,

---

[151] For instance, see Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability.* AI Now. https://ainowinstitute.org/aiareport2018.pdf

[152] Government of Canada. (n.d.). *Algorithmic Impact Assessment Tool.* https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html . Retrieved September 9, 2021.

[153] For more on these challenges, see Dias, V., Lokanathan, S., & Wijeratne, Y. (2020). *A Brief Primer on Bias in Machine Learning and Algorithmic Decisions.* LIRNEasia. https://lirneasia.net/2020/05/a-brief-primer-on-bias-in-machine-learning-and-algorithmic-decisions-whitepaper/

[154] Ibid.

[155] See also Illmer, A. (2021, January 5). Singapore Reveals Covid Privacy Data Available to Police. *BBC News.* https://www.bbc.com/news/world-asia-55541001 . Retrieved September 9, 2021.

firm guardrails need to be put in place to ensure that the data is protected and accessed only by certain authorities for narrowly defined purposes.

It is also useful to think about alternatives to explainability until technology advances to such a level that explainability is technically feasible. Many of the government policies in both countries stress explainability. However, there are other ways of evaluating AI devices. For instance, the regulatory priorities for approving AI in medical devices were whether the device was safe and accurate, among other principles. Other sectors could look into similar criteria, depending on the regulatory contexts of different sectors.

The legal landscape concerning AI is likely to evolve further as more case laws come into play, and policymakers will be able to take learnings from these cases. The organization of regulatory bodies also needs to be considered, including regulation by existing sector specific agencies, as well as if any kind of overarching regulatory body is needed.

Finally, we recommend considering ethical, legal, and technical matters together, in a three-way framework. As we have shown above, all three are interlinked. Regulators need to take into account what is technically feasible for AI devices, while maintaining the ethical standards of their sectors. Furthermore, while ethical principles are useful for setting benchmarks and goals for the kinds of AI that we want and the role they should play in society, ethics principles must converse with the legal and technical space to ensure that the principles are implementable in practice. Finally, those working in the technical space should be responsive to ethical and legal concerns in order to ensure that AI is deployed in an ethical manner that maximizes benefit to society while minimizing harm. We hope we have succeeded in illustrating in this research what this three-way consideration may look like.

Acknowledgments

We wish to thank:

- The following experts for offering their insights, quoted in this report:
    - Prof. Ang Peng Hwa – Wee Kim Wee School of Communication and Information, Nanyang Technological University (NTU), Singapore
    - Payal Malik - Advisor and Head Economics, Competition Commission of India
    - Deepak Maheshwari – Public Policy Consultant, India

- Helani Galpaya and Rohan Samarajiva for their feedback on draft versions of this report.
- Udesh Habaraduwa for his assistance in background research for the case studies.