DREWACADEMY
DATA PROTECTION & CYBERSECURITY SERVICES

INCORPORATING SINGAPORE'S MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK INTO YOUR BUSINESS

LEGAL GUIDES
2023

CONTENTS

Developments as of 11 April 2023

DREWACADEMY
DATA PROTECTION & CYBERSECURITY SERVICES

INCORPORATING SINGAPORE'S MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK INTO YOUR BUSINESS

How can companies build trust in their use of AI, with both their customers and the regulators? Beyond having strong personal data protection and cybersecurity regimes, it is also important to ensure that the data used in developing AI is accurate and representative, and decisions made using AI are correct and follow a fair process. We will refer to these collectively as "AI governance" measures.

The primary guidance for the deployment of artificial intelligence in Singapore is the Model Artificial Intelligence Governance Framework ("**Model Framework**")[1], first released in January 2019 by Singapore's Info-communications Media Development Authority ("**IMDA**") and Personal Data Protection Commission ("**PDPC**") and updated in January 2020. The Model Framework has 4 key governance areas, which we will explore, together with methods to verify compliance with those areas. Where appropriate, we will also look at cybersecurity, intellectual property protection, and personal data protection issues, which are outside the scope of the Model Framework but no less relevant. In essence, this article will take you through how you can implement the Model Framework in your business.

What does the Model AI Governance Framework apply to?

The Model Framework is a voluntary document, setting out ethical and governance principles for the use of AI and translating them into practical recommendations for organisations to adopt. It is read in tandem with the Implementation and Self-Assessment Guide for Organisations ("**ISAGO**"), a self-assessment guide for organisations developed by IMDA/PDPC to assess how their AI governance practices align with the Model Framework. The ISAGO does not set pass/fail standards for organisations, but aims to build awareness of good AI governance practices.

The Model Framework is mainly applicable to machine learning models (which have the ability to self-learn) and not pure decision tree driven AI models[2], and it is not intended for organisations deploying updated commercial off-the-shelf software packages that happen to now incorporate AI in their feature set.[3]

The Model Framework is useful guide for you if you are a business developing your own AI solution, or if you are working with a third party to create one for your business. While it is sector-agnostic, not all elements in the Model Framework will be relevant to you so you should adopt and adapt only the elements that are relevant to your needs.[4]

The Model Framework does not mention liability for when the AI does not perform as expected – this will be addressed by existing legal principles. To find out more, please see our articles "*Liability Arising from the Use of Artificial Intelligence*[5]" and "*Autonomous vehicles in Singapore*[6]".

What are Singapore's key guiding principles when using AI?

Singapore's Model Framework sets out the 2 high-level guiding principles in an organisation's use of AI to promote trust and understanding of its use. First, organisations using AI in decision-making should ensure that the decision-making process is –

[1] The Model Framework is available at: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf.
[2] See [2.10] of the Model Framework.
[3] See [2.2] of the Model Framework.
[4] See [3.2] of the Model Framework.
[5] See "Liability arising from the use of artificial intelligence" by Cheryl Seah published May 2023, The Singapore Law Gazette, available at https://lawgazette.com.sg/feature/liability-arising-from-the-use-of-artificial-intelligence/.
[6] Please see our article on "Autonomous Vehicles in Singapore", available at: https://www.drewnapier.com/DrewNapier/media/DrewNapier/Autonomous-vehicles-in-Singapore-laws-and-liability.pdf

(a) **explainable** – explaining how the AI model functions and how it arrives at a particular prediction in plain language that a non-technical user can understand;
(b) **transparent** – informing persons that AI is being used and how it affects them; and

(c) **fair** – ensuring algorithmic decisions do not create discriminatory or unjust impacts across different demographic lines (e.g., race, sex) and regularly monitoring the system and its output.

Second, AI solutions are to be "human-centric", meaning the protection of the interests of humans, including their "well-being" and "safety", should be the primary considerations in the design, development and deployment of AI. Humans should not be worse-off for having used AI.

Issues to consider at each stage of developing to deploying an AI model based on Singapore's Model Framework

The Model Framework focuses on 4 key areas where organisations should adopt measures to promote their responsible use of AI:

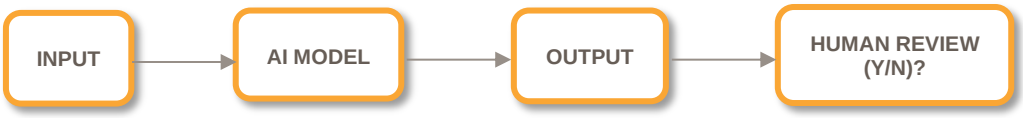
(a) Area 1: Adapting existing or setting up internal governance structures and measures to incorporate values, minimise risks and allocate responsibilities relating to algorithmic decision-making;

(b) Area 2: Determining the appropriate level of human involvement in AI-augmented decision-making based on the organisation's risk appetite for use of AI;

(c) Area 3: Operations management, such that the organisation is sensitive to potential issues when developing, selecting and maintaining AI models, including data management;

(d) Area 4: Strategies for interacting and communicating with the organisation's stakeholders.

We will discuss these 4 areas by integrating them into the lifecycle of developing and deploying AI, so that they are discussed chronologically instead of topically, as each area may apply to one or more stages of the AI lifecycle. We have identified a total of 4 distinct stages (in the diagram below), and a separate set of overarching principles that will apply across all 4 stages:



(a) 1st stage: Input – selecting data to train the model;

(b) 2nd stage: Model selection, training and calibration;

(c) 3rd stage: Output and reasons for the output; AND

(d) 4th stage: Human review of decision made prior to its implementation (if necessary).

We will also discuss some solutions and safeguards to the issues posed by the nature of AI at each stage. Notably, the issues and solutions posed at each stage would vary depending on the type of decision to be made and its impact on a person – the use of AI in giving shopping recommendations would be treated very differently from a decision that affects the rights of a person. Generally, the approach taken should be commensurate with the potential harm the AI solution deployed may cause to the affected person.

For clarity, these are the definitions of the terminology we will use[7]:

(a) "Algorithms" are a set of rules/instructions like mathematical formulas and programming commands given to a computer for it to do a task;[8]

(b) An "AI model" is created when algorithms analyse data, leading to an output/result which is examined and the algorithms iterated, until the most appropriate model emerges[9] - it is akin to what has been "learnt" by the algorithms (in the sense of them having been adjusted/calibrated) after they have analysed all the training data;

(c) An "AI system" is the model that is selected and deployed, such as by being incorporated into an application[10].

1st stage: Input – selecting data to train the model

An AI model's ability to make predictions and recommendations is due to the data it has been trained on. It is thus crucial to pay attention to the data used to train the model, to avoid a situation of "rubbish in, rubbish out". Even after the AI model is deployed, it is important to ensure that the data given to it to make a decision on is accurate, because in all cases, the wrong data will lead to the wrong decision.

One of the key things in selecting data to develop the AI model is ensuring that the data is not inaccurate (such as it being drawn from incomplete records, or that it is outdated) or biased (e.g. it was not drawn from a representative group). The IMDA/PDPC have acknowledged that it is not possible to have a dataset that is completely unbiased[11]; but when organisations are aware of this possibility and scrutinise the dataset, it helps to minimise bias.

We encourage organisations to review the questions and recommendations in section 4 of the ISAGO (accessible here), as they will guide your assessment of your existing policies around both personal and non-personal data. It is important for organisations to understand the lineage and quality of data (where it came from, how it was collected, whether it was edited, how recent it is), and keep a record for future reference. Some examples of the questions are:

(a) Did your organisation implement accountability-based practices in data management and protection (e.g. the PDPA and OECD privacy principles);

(b) If your organisation obtained datasets from a third party, did your organisation assess and manage the risks of using such datasets;

(c) Did your organisation test the AI model used on different demographic groups to mitigate systematic bias?

Where personal data is used in the training of an AI model, please see our article on "*How using artificial intelligence to process personal data may affect your compliance with the Personal Data Protection Act 2012*"[12] for more details on what to look out for. You may also follow up with our article on "*Avoiding intellectual property infringement when developing artificial intelligence systems*"[13] in relation to the datasets.

[7] These 3 terms are used but are not defined in the Model Framework, but their meaning can be inferred from the context in which they are used.
[8] As defined in the Singapore Academy of Law, Law Reform Committee's report on "Applying Ethical Principles for Artificial Intelligence in Regulatory Reform" (July 2020), available at: https://www.sal.org.sg/sites/default/files/SAL-LawReform-Pdf/2020-09/2020%20Applying%20Ethical%20Principles%20for%20AI%20in%20Regulatory%20Reform_ebook.pdf.
[9] See [3.20] and [3.21] of the Model Framework.
[10] See [3.21] of the Model Framework.
[11] See page 24 of the Model Framework.
[12] The article is available at: https://www.drewnapier.com/DrewNapier/media/DrewNapier/Using-AI-to-process-personal-data-what-does-the-Personal-Data-Protection-Act-2012-require.pdf.
[13] The article is available at: https://www.drewnapier.com/DrewNapier/media/DrewNapier/When-developing-your-AI-system-can-you-scrape-the-Internet-for-data.pdf.

2nd stage: Model selection, training and calibration

After obtaining its datasets, an organisation will often consult with persons with expertise to determine what algorithms (e.g. linear regression algorithms, decision trees or neural networks) are best suited to analyse the data in order to produce the organisation's desired solution.

The selected algorithms are applied to the datasets. Based on the results that are produced, the algorithms will be iterated/adjusted until the desired level of accuracy or most appropriate results are reached – in which case the best AI model representing what the algorithms have 'learnt' from the data emerges.[14] This is the process of "training". The greater the amount of data available for training, the more accurate the results a model produces will be, as it has a more representative sample to learn from.

In selecting and training the model, the organisation should also carefully consider the following:

(a) Whether the factors to be considered in making the decision are objective or subjective – where as much as possible, clear parameters should be set – e.g. instead of requiring "good character" which is open to interpretation, the organisation might instead focus on an "absence of criminal convictions" for certain types of offences;

(b) Whether the list of factors to be considered is open or closed, as this affects whether a new scenario should be escalated to a human for review, and how the system can handle unexpected input (i.e. new variables not considered before);

(c) Weight to be attributed to each factor in making the decision (i.e. how important is each factor).

The Model Framework recommends that organisations document how the model training and selection processes are conducted, the reasons why decisions were made (e.g. why was one model selected in place of others), and measures taken to address identified risks so that the organisation can provide an account of the decisions subsequently, and also troubleshoot in the event the model does not perform as expected.

3rd stage: Output and reasons for the output

A decision, regardless of whether it is made by a human or AI, can always be subject to challenge (especially by the party who did not get their desired outcome). To build trust in the use of AI, it is important to explain how the AI made the decision – how it functions and how it arrives at a particular prediction. This allows a person to frame his/her appeal if he/she disagrees with the decision. As a general guideline, when giving reasons behind a specific AI-augmented decision, it would be sufficient to give the same level of explanation for the reasons as if the decision had been made by a human.

Some AI systems will be harder to explain than others: rule-based systems with their if-then structure are easier to understand, while machine learning where the algorithm learns its own rules and uncovers hidden relationships that humans previously could not see is more complex to explain.[15]

The Model Framework encourages organisations to provide individuals with counterfactuals (e.g. "you would have been approved if your average debt was 15% lower") or comparisons (e.g. "there are users with similar profiles to yours that received a similar decision") instead of descriptions of the model's logic.[16] This is more helpful to a non-technical audience.

However, explainability is a balancing exercise against security and commercial secrecy – for example, organisations would not reveal too much about the decision-making process if this would allow people to game the system, particularly where it comes to security or fraud detection.

[14] See [3.20] and [3.21] of the Model Framework.
[15] See "*The Law of Artificial Intelligence"* by Matt Hervey & Matthew Lavy at page 28.
[16] See [3.28] of the Model Framework.

If explainability is not possible given the level of technology available, demonstrating repeatability of the results will suffice (i.e. given the same scenario, the AI model will consistently produce the same results).

4th stage: Human review of decision made prior to its implementation (if necessary)

Whether a human should be involved in reviewing the decision made by the AI before it is implemented depends on the impact of the decision on the person affected. Broadly, there are 3 degrees of human oversight in the decision-making process:

(a) Human-in-the-loop: AI serves as a guide or recommendation only, with a human making the final decision (e.g. a doctor making the final decision on the diagnosis when using AI to diagnose a medical condition based on the symptoms);

(b) Human-out-of-the-loop: AI makes the decision with no option for human override (e.g. online retail store recommending products to individuals based on their browsing behaviour and purchase history);

(c) Human-over-the-loop: A human has a supervisory role, where he/she will intervene as appropriate, such as where there are unusual factors for consideration or the result is below a certain statistical confidence level.

The Model Framework sets out a matrix measuring the severity of the harm (high/low) and probability of the harm occurring (high/low) to assist organisations with assessing risk:

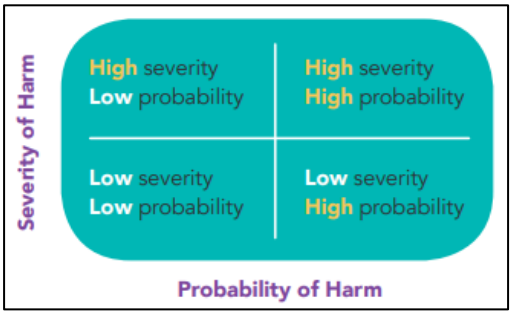


Image source: Model Governance Framework, para. 3.16

Other factors to consider include the nature of the harm (physical or intangible), reversibility of the harm, as well as operational feasibility – for example, having a human review every decision in the case of high-speed financial trading is not feasible.[17]

Across all stages: Overarching principles

First, organisations are encouraged to introduce internal governance structures such that they would have oversight over how AI technologies are used in their operations. This involves setting out clear roles and responsibilities for personnel (such as who is managing the AI model training and selection process, and who is monitoring the AI model deployed with a view of taking remediation measures where needed), and ensuring that they have sufficient expertise, resources and training to discharge their duties.

[17] See [3.17] of the Model Framework.

Second, organisations should also carry out —

(a) regular testing and tuning of the AI model, to ensure it is producing results in line with expectations, and also to cater for changes in the environment over time (e.g. a model to detect phishing emails may need to be retrained on updated datasets as new types of scams emerge);

(b) record keeping and documentation – such as ensuring that datasets and processes that lead to the AI model's decision (including those of data gathering, data labelling and the algorithms used) are documented in an easily understandable manner. In the event the AI model does not perform as expected, the organisation can look back on its records to troubleshoot (and also defend against liability). This would also facilitate any audit of the AI system. Instead of bare assertions, or having to put together documents retrospectively, comprehensive and contemporaneous documentation would be better proof that the development process was rigorous;

(c) strong cybersecurity practices, such as by limiting who has access to the AI system so that it cannot be modified without authorisation, and where the AI system continues to learn from data input into it after it is deployed, by setting acceptable use policies to ensure that users do not maliciously introduce input data that unacceptably manipulates the performance or results of the AI system.[18]

Lastly, organisations should manage their relationship with their stakeholders (and the expectations of their stakeholders) when deploying AI. This could involve disclosing the use of AI in decision-making to consumers, and providing information on what steps are taken to mitigate the risks to the consumer, and how consumers can apply to reverse the decision. Even if the organisation uses an AI system from a third party, it would be prudent to obtain enough information from the third-party provider to address the queries of consumers, as the organisation is now the one fronting (the use of) that AI system.

How do we know if the use of AI is in line with the principles in the Model Framework – what testing methods and resources are available in Singapore?

The Model Framework introduces the concept of "auditability", which is the readiness of an AI system to have its algorithms, data and design processes evaluated by internal or external auditors. These auditors can "probe, understand and review the behaviour of the algorithm through disclosure of information that enables monitoring, checking or criticism"[19]. This builds trust in the use of AI as audits can demonstrate that the design of it was well thought out, and the outcomes it produces have justification.

The Model Framework also addresses the specific concept of "algorithm audits", indicating that they are carried out at the request of a regulator (as part of a forensic investigation). Algorithmic audits can show how an AI model operates if there is any reason to doubt the veracity or completeness of the information provided about how it operates. Such algorithm audits require technical expertise, and they would be conducted when it is clear that the expected benefits of an investigation would outweigh the time and cost of conducting the audit.[20] Thus, for the purposes of providing information to customers/individuals (as opposed to regulators), providing information on how decisions are made to achieve explainability should be sufficient.

The issue is thus – what is the AI system assessed against when it is audited (not limited to algorithm audits)?

[18] See [3.57] of the Model Framework.
[19] Definition of "Auditability" in Annex A of the Model Governance Framework.
[20] In order to prepare for audits, the Model Framework recommends (at [3.43]) that organisations keep a comprehensive record of data provenance, procurement, pre-processing, lineage, storage and security. It would be ideal if the entire model creation process is documented (e.g. including reasons why a particular model was selected).

In Singapore, we have the ISAGO for organisations (accessible here), a series of questions they can run through to self-assess their compliance with the Model Framework and identify potential gaps in their governance processes.

In May 2022, Singapore launched A.I. Verify[21], an AI Governance Testing Framework and Toolkit for AI system owners and developers who wish to verify their AI systems against 8 internationally accepted ethics principles – (1) transparency, (2) explainability, (3) repeatability/reproducibility, (4) safety, (5) robustness, (6) fairness, (7) accountability and (8) human agency and oversight. AI Verify consists of:

(a) **The Testing Framework** — specifies the testable criteria relevant to each AI ethics principle, testing process (i.e. actionable steps to take to ascertain if the testable criteria is satisfied – such as technical tests, statistical tests, or producing documentary evidence) and metrics for each testable criterion. The testable criteria could be a combination of non-technical factors (e.g. processes and organisational structure) and technical factors.

(b) **The Toolkit** — used to execute technical tests in the Testing Framework, and contains widely used open-source libraries (software) for testing.

"Process checks" involve looking at documentary evidence and are recorded in the form of a checklist, while "technical tests" involve using software. In terms of what principles can be assessed through technical tests or process checks or both, the AI Verify toolkit states that:

(a) Explainability, Robustness and Fairness can be assessed using both technical tests and process checks; and

(b) Transparency, Repeatability/Reproducibility, Safety, Accountability and Human agency and oversight are assessed through process checks only.

Similar to the ISAGO, AI Verify does not define standards (i.e. it is not pass/fail). Organisations can use it to see how they measure up to the ethics principles and if there are any improvements they wish to undertake. It is important to note that completing the tests under AI Verify does not guarantee that the AI system is free from risks or bias, or that it is completely safe. AI Verify is presently in its pilot testing stage and IMDA/PDPC is inviting companies to come on board.

Conclusion

As the use of AI continues to grow in Singapore, we anticipate that regulators will issue further guidance on how organisations may adopt AI in a way that is aligned to internationally accepted principles. We will continue to watch this space and provide regular updates on the latest regulatory developments. For more information, please explore the rest of our articles on our AI resource page.

---

[21] For more information, please refer to https://file.go.gov.sg/aiverify.pdf.

DREW DATA PROTECTION & CYBERSECURITY ACADEMY

Drew Data Protection & Cybersecurity Academy (Drew Academy) was established in 2020 by Drew & Napier to help our clients build their capabilities and develop and implement organisational strategies, structures, policies and processes to meet their legal, regulatory and compliance obligations. Drew Academy offers a range of courses in areas such as data protection, cybersecurity, data governance and in-house commercial practice. A particular focus for us is the delivery of workplace learning solutions and development of customised training courses. We also offer outsourced DPO services and data protection consulting services through our experienced team of practitioners.

Drew Academy is helmed by Lim Chong Kin and David N. Alfred. Our course leaders are experienced in various aspects of data and cyber governance, data protection, cybersecurity engineering and in-house commercial practice.

ARTIFICIAL INTELLIGENCE AND DIGITAL TRUST

Drew & Napier's Artificial Intelligence (AI) and Digital Trust practice brings together its expertise across several technology-related domains and in fields as diverse as data protection, cybersecurity, healthcare, Fintech, intellectual property and competition law (to name a few) to advise clients on the full range of legal issues relating to AI and Digital Trust. In addition to advising on commercial, regulatory and international / cross-border issues, our advice extends into areas such as governance and ethics as we seek to enable our clients to navigate areas where laws and legal principles are still emerging.

Working together with the Drew Academy, we provide solutions that reflect our deep understanding of underlying technologies, the risks and uncertainties involved and practical business considerations. Internationally, there is a growing consensus on AI governance.

For more information on our experience, please contact:

Lim Chong Kin
Managing Director, Corporate & Finance;
Co-Head, Data Protection,
Privacy & Cybersecurity Practice;
Co-Head, Drew Data Protection &
Cybersecurity Academy

T: +65 6531 4110
E: chongkin.lim@drewnapier.com

David N. Alfred
Director, Corporate & Finance;
Co-Head, Data Protection,
Privacy & Cybersecurity Practice;
Co-Head and Programme Director,
Drew Data Protection &
Cybersecurity Academy

T: +65 6531 2342
E: david.alfred@drewnapier.com

Cheryl Seah
Director, Corporate & Finance

T: +65 6531 4167
E: cheryl.seah@drewnapier.com

DREWACADEMY
DATA PROTECTION & CYBERSECURITY SERVICES

10 Collyer Quay
10th Floor Ocean Financial Centre
Singapore 049315

www.drewnapier.com/Academy

T: +65 6531 4152
F: +65 6535 4864
E: academy@drewnapier.com

In association with

DREW & NAPIER