



PDPC'S CLOSED CONSULTATION

ON

**A PROPOSED MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE
FRAMEWORK**

COMMENTS

FROM

BSA | THE SOFTWARE ALLIANCE

7 Nov 2018

Contact:

Jared Ragland,
Senior Director, Policy - APAC

jaredr@bsa.org

300 Beach Road, #25-08 The Concourse, Singapore 199555

**PDPC'S CLOSED CONSULTATION ON A PROPOSED MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK
– COMMENTS FROM BSA | THE SOFTWARE ALLIANCE**

BSA | The Software Alliance (**BSA**) appreciates the opportunity to comment on the Closed Consultation on A Proposed Model Artificial Intelligence Governance Framework on 7 November 2018 (the **Consultation Paper**) by the Personal Data Protection Commission (**PDPC**).

A. INTRODUCTION AND STATEMENT OF INTEREST

BSA is the leading advocate for the global software industry before governments and in the international marketplace. Our members¹ are at the forefront of the software-enabled innovation that is fueling global economic growth, including cloud-computing and cutting-edge advancements in artificial intelligence (**AI**). As leaders in AI development, BSA members have unique insight into both the tremendous potential that AI holds for societies, and the government policies that can support the responsible use of AI in order to foster trust among businesses and consumers and ensure continued innovation.

The responsible use of AI, enabled in large part by cloud computing, has the potential to spur tremendous economic growth across every sector, improve human decision-making, and enable cutting-edge breakthroughs on some of the world's most pressing challenges, such as climate change and cancer research. Conversely, AI services that are not developed or used responsibly, like other ground-breaking innovations, create a risk of unintended consequences or malicious uses. Governments are therefore rightly focused on developing policies to address both the opportunities and risks associated with AI. BSA and its members are equally attuned to these challenges and are committed to developing AI in a manner that will maximize the benefits and minimize the risks associated with the deployment of the technology. BSA has published a range of materials related to AI, including AI Policy Principles, an AI Primer, and examples of AI being applied in different sectors.²

BSA commends PDPC for its ongoing efforts to foster discussion and develop thought-leadership in the nascent area of AI governance policies. The *Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI*, published by PDPC in June 2018 (the **Discussion Paper**) provided an important articulation of an accountability-based framework for discussing ways to ensure responsible and ethical development and use of AI while also ensuring commercial viability of the creation and deployment of AI technologies and solutions. Crucially it recognized the need for a flexible and risk-based governance framework that is context and use-case specific. The Discussion Paper also recognized the need to demarcate between accountability processes that are internal to an organization, and information that should be communicated to individuals in order to meet user expectations and build public confidence and trust in AI.

B. COMMENTS ON THE CONSULTATION PAPER

The Proposed Model Artificial Intelligence Governance Framework (the **Model Framework**), discussed within the Consultation Paper, represents an important and positive step towards providing guidance that organizations can voluntarily adopt if they are considering deploying AI solutions. BSA

¹ BSA's members include: Adobe, Akamai, Amazon Web Services, ANSYS, Apple, Autodesk, AVEVA, Bentley Systems, Box, CA Technologies, Cadence, Cisco, CNC/Mastercam, DataStax, DocuSign, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, PTC, Salesforce, SAS Institute, Siemens PLM Software, Slack, Splunk, Symantec, Synopsys, Trend Micro, Trimble Solutions Corporation, and Workday.

² <https://ai.bsa.org/>

sets out below our summarized comments and recommendations on the proposed Model Framework. BSA also proposes specific line-by-line amendments to the Model Framework in the **Annex**.

B1. Definition and Scope of the Model Framework

BSA appreciates PDPC distinguishing between AI Solution Providers, Organizations, Consumers, and Customers. However, the scope of the Model Framework, and in turn the definitions in paragraph 1.9 can benefit from further clarity. In this regard, **BSA recommends that PDPC consider making clear that the Model Framework is applicable regardless of whether an organization develops or deploys AI, and whether the deployment has implications on businesses or individual citizens.**

BSA also recommends that PDPC distinguish between the responsibilities that tend to rest with the AI end-user organizations and those that rest with the AI Solution Providers. In most cases, the main responsibility for implementing and operating an AI governance framework, in accordance with the recommendations of the Model Framework, ought to lie with AI end-user organizations. AI Solution Providers, including those that offer services over the cloud, may have little to no insight into how AI end-user organizations use these solutions to make decisions and hence are unlikely to be in position to implement the Model Framework in a manner that meaningfully protects end-user individual rights. This is particularly pertinent to the recommendations related to the section on “AI Deployment Processes and Operations Management”, covered in detail in section B6 of this submission.

For example, an AI Solution Provider may provide access to an AI technology that allows the end-user organization to leverage pre-trained machine learning services, or build and use their own custom machine learning models and capabilities. However, it is the end-user organization that controls and determines the way in which those models and capabilities are used, the data sets that are involved,³ and, importantly, the decisions that are made in connection with the results of such models.

In addition, it would also be useful for both AI Solution Providers and end-user organizations to consider who the end user will be – in general, end-user businesses should be considered more sophisticated users than a service offered end-users who are individuals – and this would in turn have implications on internal risk assessments or commercial viability. Furthermore, the differences between “Customer” and “Consumer” are unclear and at times seem to be used interchangeably.

Hence, BSA recommends that a consistent conceptual distinction be made in the definitions and throughout the Model Framework between:

- a. AI Solution Provider organizations and AI end-user organizations; and**
- b. End-users that are businesses and end-users that are individuals.**

This would be helpful for both AI Solution Providers and end-user organizations, as they carry out risk analysis, to assess the appropriate measures they should adopt for AI deployment and operation and customer relationship management.

B2. Government Exclusion Undermines Voluntary Nature of the Model Framework

BSA is supportive of PDPC's approach to making clear that the Model Framework is intended as a “voluntary best practice”. Paragraph 1.7(b) however states that the Model Framework “does not apply to Public Sector use of AI.” In view of the voluntary nature of the Model Framework such an exclusion is unnecessary and is potentially confusing. It suggests that its application to private sector use is

³ Except for in the case of the generalized datasets that pre-trained services and models are trained on by the technology provider

more than just voluntary and is in some way intended for compliance, hence the need to explicitly exclude public sector.

Hence, BSA recommends that the PDPC **eliminate the government exclusion and simply make clear that the Model Framework is intended for voluntary adoption by private-sector organizations.**

B3. Clarify the Relationship Between Commercial Viability and the Ethical Deployment of AI

BSA welcomes the Model Framework's recognition that organizations would deploy AI technologies or solutions to achieve a commercial objective and, as part of their commercial considerations, should take into account the potential risks that could arise if AI is not deployed responsibly. We also concur with the view that risks may vary depending on the context and societal norms in which AI is deployed.

Paragraph 1.3 of the Consultation Paper states that *"To be commercially viable, the cost of implementing AI technologies in an ethical manner should not outweigh its expected benefits."* While this is consistent with what is discussed in the subsequent section — on *"Identifying Commercial Objectives and Risks"* in paragraphs 2.6 to 2.9 — paragraph 1.3, if taken out of context, could inadvertently suggest that AI should be deployed in an unethical manner if it is not commercially viable to do so ethically. The more appropriate distinction, which is likely the one intended by PDPC, should be between the ethical deployment of AI and, if such ethical deployment is not commercially viable, the use of alternative non-AI solutions. We therefore recommend that PDPC consider revising paragraph 1.3 in line with an outcome-orientated approach, as suggested in the **Annex**.

B4. Clarify the Roles and Responsibilities for Ethical Deployment of AI

BSA welcomes the Model Framework's emphasis on internal organizational accountability and oversight as a cornerstone for accountability-based AI governance. To ensure that organizations adopt this model in a manner that is most appropriate and practical to their context, **BSA recommends revising the introductory language in paragraph 2.3 to make clear that organizations should determine the appropriate features in their internal governance structure themselves but may consider the following points (as listed in 2.4.1 and 2.4.2).** This is particularly important given the rapidly changing nature of AI technology.

Furthermore, while we appreciate that the Model Framework is outcome-oriented and seeks to take varying levels of risk into account, we are concerned that some of the recommendations in paragraph 2.4 may be overly prescriptive on the one hand, while failing to address some elements on the other. **We therefore recommend consolidating and simplifying the considerations outlined in paragraph 2.4, as detailed in the Annex.**

B5. Further Clarify the Considerations for Determining Risks and Decision-Making Models

The Model Framework is premised on realistic risk analysis and proposes mitigation efforts proportionate to the concrete harms that might result in specific use cases. To this end, the Model Framework's proposal in the section "Determining a Decision-Making Model", to encourage organizations to adopt a three-pronged approach to making risk-based decisions provides a useful and practical way to assist organizations in thinking about risk. These include:

- Cataloguing their commercial objectives and risks;
- Conducting a risk-assessment exercise; and
- Determining an appropriate decision-making model based on assessed risk.

It is important that the Model Framework be clear and consistent on bottom-line considerations for organizations, as this will form the basis of any cost-benefit analysis that a commercial entity will conduct when deciding whether to develop and/or deploy AI in its systems and processes. The following are suggested revisions for PDPC's consideration.

a. Attributes and underlying considerations for the choice of decision-making models should be made clear at the outset.

PDPC should consider reiterating, at the outset of this section in paragraph 2.5, the underlying considerations for decision-making. Namely the selection of an appropriate decision-making model is crucial to **ensuring public trust and confidence that AI systems are operating in a manner that is consistent with their expectations**. Selecting an appropriate decision-making model enables the identification and mitigation of the risk of unintended consequences. It will also increase public confidence that **AI is designed to and deployed in a manner consistent with pre-determined objectives and values** (including but not limited to domestic societal norms and organizations' corporate values).

b. Accounting for consumer expectations when identifying and assessing risks.

BSA welcomes the discussion that PDPC has undertaken in the section on "Identifying Commercial Objectives and Risks" as it generally provides a balanced discussion and recognizes that organizations deploy AI for commercial objectives. Paragraph 2.7 could be further enhanced by directly stating **that a key element of risk identification is understanding and accounting for consumer expectations in a given situation**. AI deployments that are consistent with consumer expectations would foster trust and confidence and enable organizations to achieve their commercial objectives for deploying AI better.

There are also elements in paragraphs 2.7 and 2.9 that could be construed as overly prescriptive compared to the rest of the section and may impact the ability of the organization to remain nimble and adapt their commercial objectives over time. BSA therefore recommends that PDPC consider amending paragraphs 2.7 and 2.9 as set out in the **Annex**.

c. Clarify the decision-making models

The Model Framework outlines three broad decision-making models, namely "human-in-the-loop", "human-on-the-loop", and "human-out-of-the-loop" and suggests that organizations could design their decision-making processes based on these approaches, depending on the probability and severity harm of a given AI deployment. However, based on the definitions and examples provided for each of the models, it remains unclear how the "human-in-the-loop" and "human-over-the-loop" models differ. Both models appear similar, involving the use of AI to provide humans with recommendations, with the human retaining the ultimate decision-making ability.

The distinction between the two models can be better articulated from the perspective of designing human oversight into the decision-making processes, rather than from the perspective of the party (whether AI or human) making the decision. In particular, we would recommend re-conceptualizing the models as follows:

- **"Human-in-the-loop"** model, 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.

- **“Human-over-the-loop”** model in contrast suggests human oversight is passive and uninvolved, with the AI being in control of executing decisions, while the human maintains the ability to override AI decisions. Decisions can therefore be exercised fully without any human intervention.
- **“Human-out-of-the-loop”** model suggests that there is no human oversight over the execution of decisions. This model has been designed such that there is no human input or interaction with the AI. Hence the AI has full control without the option for the human override.

Separately, PDPC may also want to consider whether the Model Framework should discuss a decision-making model where a human is in control of decision-making, but an AI is deployed as an oversight mechanism with the ability to override human decisions.

It is pertinent that PDPC makes the distinction between the models clear as it could have downstream implications for liability or future regulatory controls.

d. Determination of the appropriate decision-making model should be de-linked from the “severity-probability of harm” assessment.

It is appropriate for the Model Framework to urge organizations when adopting AI solutions, to consider whether a specific decision-making model (i.e. “human-in-the-loop, “human-on-the-loop” and “human-out-of-the-loop”) is most appropriate for their particular deployment scenario. However, the “severity-probability of harm” assessment matrix described in paragraph 2.13 is overly prescriptive and inconsistent with the context-specific and risk-based approach taken through the rest of the document.

Linking the “severity-probability of harm” assessment matrix to the decision-making models presupposes that the greater the likelihood or severity of harm, the more a human should be involved in the decision-making process. However, the “harm severity” and “harm probability” associated with any particular decision model will inherently vary based on the particular nature of a use case. There are uses, for example, where a “human-out-of-the-loop” decision model will mitigate both the likelihood and the severity of harm. For example, if an automated decision-making process has been deployed to prevent bias and discrimination, then human intervention could (re)introduce bias that may then go undetected. This would remain true regardless of the significance of the decision. Furthermore, there could be other underlying reasons (e.g., contractual or regulatory) for choosing one decision-making model over another that are not directly related to risk or severity of harm.

In addition, a “human-out-of-the-loop” decision-making process should not preclude downstream, albeit separate, appeals processes or consumer management processes that involve humans. For example, although an organization could put in place a fully automated, “human-out-of-the-loop” decision making processes, for consumer transparency, the organization could still have separate consumer management processes to broadly explain decision making approaches, or in particular instances have a human provide further analysis or review for the customer. We recommend that PDPC consider clarifying this in the section “Risk Assessment for Decision-Making Model”.

e. Other issues.

2.9 and 2.13 are repeated paragraphs. We recommend deleting paragraph 2.13.

B6. Further clarification on the scope of applicability for AI deployment processes and operations management.

BSA welcomes the discussion that PDPC has undertaken in the section “AI Deployment Processes and Operations Management.” While the list of processes provides a good starting point for both AI Solution Provider and end-user organizations to manage AI deployments, this section would be greatly enhanced by a reiteration of the conceptual distinction between these two types of organizations. Hence, **BSA recommends that PDPC outline whether and when each of the processes and operations would be applicable to the AI Solution Provider, end-user organization, or both.**

In this regard, we have identified three examples of how this conceptual distinction would enhance the usefulness of the Model Framework. This list is not intended to be exhaustive and **we recommend that PDPC continue to workshop with both AI Solution Providers and AI end-user organizations to develop a more comprehensive tool.** The three examples are as follows:

a. Clarifying how “Good Data Accountability Practices” would apply differently to technology providers and technology users.

“Good Data Accountability Practices” would apply differently to AI Solutions Providers and end-user organizations. Using the example of a cloud-based AI solution that is deployed as a “one-to-many” and “self-service” solution, it would be very difficult for the AI Solution Provider to provide data lineage information to end-user organizations, as suggested in 2.18(a). From an internal due-diligence point of view, data lineage requirements may also impact the ability of both the AI Solution Provider and the end-user organization to remain nimble and adapt their commercial objectives.

b. Balancing accountability with disclosure of proprietary information.

Paragraphs 2.18 (b) 2.20, 2.38 and 2.39 (relating to data quality and AI algorithms, respectively), do not distinguish between the AI Solution Provider and the end-user organization, and hence fail to recognize the competing objectives between intellectual property rights and accountability. For example, the AI Solution Provider’s need to protect its intellectual property rights, may be at odds with the end-user organization’s need to have access to information from the AI Solution Provider (including potentially proprietary information) in order to be accountable to its end-user individuals. Information relating to how data has been compiled or updated, and the algorithms used, are often proprietary to the AI Solution Provider and protected as intellectual property, for example, trade secrets.

We also recommend that paragraph 2.39(d) include a specific consideration for the protection of proprietary information, including the ability for organizations to choose not to disclose such information.

c. Clarifying that relying on repeatability where explainability cannot be achieved may not be appropriate in all circumstances.

Paragraph 2.24 states that in scenarios where explainability cannot be practically achieved, documenting repeatability of results could be an alternative. While repeatability assessments are generally useful and practical, there may be scenarios, particularly for AI Solution Providers, where randomness has been introduced by design and hence repeatability assessments may not be appropriate for AI end-user organizations to carry out. Hence it is important to draw the

distinction between when certain processes would be more applicable to AI Solution Providers and when it would be applicable to end-user organizations.

In addition, as part of PDPC's discussions in 2.24(b), **we would recommend that the rationale for exceptions, and why decisions are not repeatable, be further elaborated upon to recognize scenarios where randomness is incorporated in-system by design.**

B7. Clarifying how customer relationship processes should apply when AI is embedded in software solutions and services.

BSA appreciates the inclusion of a section on “Customer Relationship Management” in the Model Framework. Appropriate customer communication is key to ensuring public trust and confidence that AI systems are operating in a manner that is consistent with their expectations. In practice, many AI deployments tend to be embedded into other services or products, and like many other embedded technologies (e.g., data analytics, blockchain etc.), it is atypical for companies to provide detailed breakdowns to customers on all the technologies employed in every product or service. Furthermore, customers are generally indifferent to the particular technologies employed when using a service or product. What would be important to customers is the organization's ability to explain how the decisions made using AI were arrived at. The following are suggested revisions to specific customer relationship management processes for PDPC's consideration:

a. Clarifying When Organizations Should Consider Communicating to Customers

Given that the focus on customer relationship management processes covers a third of the Model Framework, **we recommend that PDPC make clear at the beginning of the section that as part of the broader risk assessment, organizations should consider whether there is a need to communicate every instance of AI embedded in services or products.** In most cases, it would be unnecessary for organizations to communicate the use of embedded AI, and over-communication could have the opposite effect – creating public concern. Rather, organizations should communicate to customers pertinently about the way a product or service is used to make decisions about them, and only draw attention to the use of AI where this is relevant to the outcome of the decision.

In addition, we have proposed some specific changes to paragraph 2.30 in the **Annex**. These suggested changes are intended to provide organizations sufficient flexibility to determine the best way to communicate with their customers. An overly prescriptive approach to communications may have the unintended consequence that, for some products or services, the communication might be unduly alarming for consumers or cause misconceptions about the safety or reliability of the product or service.

b. Clarifying When Organizations Should Consider the “Option to Opt-Out”

BSA appreciates that PDPC's discussion on “option to opt-out” in paragraph 2.35 is intended to make clear that organizations need not provide an “option to opt-out” in every instance of AI use. In particular, the Model Framework is particularly insightful in recognizing that in fully-automated decision-making processes the “option to opt-out” could create a significant challenge to using AI effectively. However, the paragraph seems to imply that there should be a default “option to opt-out” in deployment scenarios that are not fully automated. We recommend that PDPC consider further clarifying this section to make clear that decisions about whether to include an “option to opt-out” should be based on a context-specific evaluation. In this regard, we have proposed some specific changes to paragraph 2.35 in the **Annex**.

C. CONCLUSION AND NEXT STEPS

BSA once again commends PDPC's efforts to create a flexible and practical Model Framework on AI Governance to respond to the ever-evolving needs of the digital economy. BSA encourages the PDPC to consider the comments above. Making the suggested improvements to the Model Framework would better prepare organizations in Singapore to deploy AI in their business, in a manner that is considered and practical. We look forward to further engagement with the PDPC and offer our continued support for the Singapore Government's efforts towards making Singapore a global hub for AI deployment and development, through forward-looking and innovative guidelines and regulatory frameworks.

Please do not hesitate to contact Jared Ragland at jaredr@bsa.org for any further information on this submission. Thank you.

Yours faithfully,

BSA | The Software Alliance

Annex: Proposed Specific Line-by-Line Amendments to the Model Framework

| Section | Suggested Amendments |
|---|--|
| <p>B3. Clarify the Relationship Between Commercial Viability and the Ethical Deployment of AI</p> | <p>Paragraph 1.3</p> <p>We propose replacing the last sentence with:</p> <p>“The degree to which any specific ethical principle or implementing measure would be relevant to promoting trust in an AI system would vary depending on the circumstances. The commercial risks of AI deployment would therefore also be proportional to the impact to individuals. Where the cost of implementing AI technologies in an ethical manner outweighs any expected benefits, organizations could consider whether alternative non-AI solutions should be deployed.”</p> |
| <p>B4. Clarify the Roles and Responsibilities for Ethical Deployment of AI</p> | <p>Paragraph 2.4:</p> <p>The roles and responsibilities discussed in paragraph 2.4(1)(c) overlap with the risk measures discussed under the <i>Risk management and internal controls</i> section in paragraph 2.4(2). Paragraph 2.4(2) relates to the processes to be assigned to the roles identified in 2.4(1)(c). Hence, for clarity, PDPC may wish to consider creating a separate header for 2.4(1)(c) and 2.4(2), for instance “Examples of roles and responsibilities that should be assigned,” and map the processes in 2.4(2) to the roles in 2.4(1)(c).</p> <p>Paragraph 2.4.1(c)(iii):</p> <p>We propose re-wording this paragraph to “This includes reviewing data sets and features in those datasets used in AI model training with the aim of establishing if they are adequate for the intended purpose, accurate and unbiased”. In this way, the consideration is structured as an aim and not as a guaranteed outcome. An outcome cannot always be guaranteed.</p> <p>Paragraph 2.4.1(c):</p> <p>While the assignment of the roles and responsibilities in paragraph 2.4 is comprehensive, one notable gap is in the composition of personnel assigned such responsibilities. We propose the inclusion of the following under paragraph 2.4(1)(c):</p> <ul style="list-style-type: none"> • Identifying persons with relevant expertise who are responsible for addressing significant problems identified with operating AI systems. • Ensuring subject matter experts, especially those with knowledge of the policy landscape in which the AI system will be deployed, are available to assist computational scientists in the design and implementation phases. <p>Considering the composition of personnel in the Model Framework would help mitigate potential gaps in expertise that are crucial to ensuring that AI is deployed responsibly.</p> |

| | |
|--|--|
| <p>B5(b). Accounting for commercial objectives and consumer expectations when identifying and assessing risks.</p> | <p>Paragraph 2.7</p> <p>We propose re-wording this to “Organisations that use or intend to use AI should consider their commercial objectives prior to deploying AI solutions”. In our view this is a less prescriptive approach as the term “set out” might be interpreted as requiring formal documentation which may impact the ability of the organization to remain nimble and adapt their commercial objectives over time.</p> <p>In addition, we propose that PDPC make it clear that risk assessments should also take into account consumer expectations in a given situation, as follows: “This process lets organisations ... and to themselves. The consumers’ expectations of explainability, transparency, fairness, and human-centricity would vary depending on context. Hence any risk assessment must take into account consumer expectations in a given situation. Even within a country...”</p> <p>Paragraph 2.9</p> <p>We propose removing the overly prescriptive language around “Documenting this process through a periodically reviewed risk impact assessment... using the AI solutions” and amend it instead to “Organisations can periodically review risk impact assessments, as this would help develop clarity and confidence in using the AI solutions.”</p> <p>PDPC can also consider incorporating the concept that organizations can rely on both internal reviews or third-party certifications for this review process.</p> |
| <p>B7. Clarifying how customer relationship processes should apply when AI is embedded in software solutions and services.</p> | <p>Header for Paragraph 2.30 – 2.33</p> <p>Paragraphs 2.30 to 2.33 relate to “transparency” when organizations manage relationships with their customers. The use of the term “transparency” in the context of these paragraphs is potentially confusing as it is distinct from the concept of “transparency” referred to in paragraph 1.4(a). We propose instead distinguishing this concept in paragraphs 2.30 to 2.33 by revising the header to “Open Customer Communication”. References to “transparency” in the rest of the section should also be amended to “openness” accordingly.</p> <p>Paragraph 2.30</p> <p>We propose that this section in the Model Framework be modified as follows:</p> <p>“Organisations should provide general information to their customers on how they use AI, as the organisation deems appropriate (bearing in mind the ethical principles set out in this Framework) including the manner and form in which that information is provided. This may include disclosing (as appropriate) that AI is used in decision-making</p> |

| | |
|--|--|
| | <p>about customers, and the role and extent that AI plays in the decision-making process.”</p> <p>Paragraph 2.35</p> <p>We propose that PDPC consider moving the sentence “where decisions are fully automated ... lead to a process being commercially uncompetitive” to the start of the paragraph so that it is clear that the “option to opt-out” is not the default position that companies should take when deploying AI.</p> |
|--|--|