From Principles to Practice

Use Cases for Implementing Responsible AI in Financial Services
## Contents

1 Introduction 6

2 Implementing AI principles: applying Singapore’s MAS and PDPC guidance 8

3 What is AI and when is its use material? 11

4 Applying responsible AI in context: Three use cases in financial services 13
   4.1 Use Case 1: Using AI to predict which cards will be used for travel 14
   4.2 Use Case 2: Using AI to automate verification of wet ink signatures 16
   4.3 Use Case 3: Using AI for regulatory compliance relating to Know Your Customer (KYC) checks 18

5 Lessons learned 22
   5.1 The importance of context and assessing materiality for specific use cases 22
   5.2 Existing internal risk management and governance frameworks, as well as legal and regulatory compliance, are highly relevant for responsible AI 23
   5.3 Technology has a key role to play in addressing responsible AI considerations 24
   5.4 The importance of diversity and culture for responsible AI 25
   5.5 Implementing multiple sets of AI principles is challenging 26
   5.6 Distinguishing between AI developers and AI deployers in how principles will be applied 27
Contributors and Acknowledgements

This paper was prepared to capture the findings from work by a group of partners in a project to explore the implementation of responsible AI principles, including Deutsche Bank, Linklaters, Microsoft, Standard Chartered, and Visa.

The group is grateful for the engagement of leadership and wider teams within their organisations, drawing on a diverse set of expertise including data science, IT, legal and regulatory affairs, and public policy, among others. The project team and those who provided valuable input from each institution included the following:

<table>
<thead>
<tr>
<th>Deutsche Bank</th>
<th>Linklaters</th>
<th>Standard Chartered</th>
<th>Visa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boon-Hiong Chan</td>
<td>Adrian Fisher</td>
<td>Shameek Kundu</td>
<td>Nuri Purswani</td>
</tr>
<tr>
<td>Chris Bezuidenhout</td>
<td>Jakub Brecka</td>
<td>Sophie Lloret</td>
<td>Meri Rosich</td>
</tr>
<tr>
<td>Abhishek Mehta</td>
<td>Joel Cheang</td>
<td>Tanty Muliani</td>
<td>Alexandra Stuart</td>
</tr>
<tr>
<td>Jiali Liu</td>
<td></td>
<td>Yin Teng Choy</td>
<td>Ajit Patil</td>
</tr>
<tr>
<td>Subramanian Venkataraman</td>
<td></td>
<td>Boon Chye Ong</td>
<td>Hasan Al-Madfai</td>
</tr>
<tr>
<td>Eugene Yeo</td>
<td></td>
<td>Jatin Badiani</td>
<td>Anubhav Narang</td>
</tr>
<tr>
<td>Peter Hung</td>
<td></td>
<td>Siddhartha Thimmavajjalla</td>
<td>Patrick Storey</td>
</tr>
<tr>
<td>John Doyle</td>
<td></td>
<td>Vijay Jairaj</td>
<td>Clement Cher</td>
</tr>
<tr>
<td>Nadim Shehadeh</td>
<td></td>
<td></td>
<td>Grace Shin Hui Tan</td>
</tr>
<tr>
<td>Kaye Ancanan</td>
<td></td>
<td></td>
<td>Doug Campbell-Pitt</td>
</tr>
<tr>
<td>Sumitraa Srinivasan</td>
<td></td>
<td></td>
<td>Olivia Maly</td>
</tr>
</tbody>
</table>

The group expresses its thanks to David Hardoon, Special Advisor (AI) of the Monetary Authority of Singapore; Zee Kin Yong, Deputy Commissioner of the Singapore Personal Data Protection Commission, and the staff of MAS and PDPC for their ongoing and constructive dialogue with the project group.

The group would welcome further engagement with regulators, industry and others that share our interest in translating responsible AI principles into practice.
Foreword

There is great interest in financial services – as in other industries – in the promise of AI to unlock new opportunities for innovation, increased efficiency, and improved services for customers. This promise is fast becoming reality, with growing adoption of AI by financial institutions across Asia.

In Microsoft’s conversations with our financial services customers and with regulators, there is a shared focus on ensuring that AI is used responsibly. There is a clear view that in order to retain societal trust in AI, there will need to be greater efforts to identify not just the new opportunities AI presents, but also the ways in which it might present new challenges, along with practical approaches to managing these challenges.

In Singapore, the Monetary Authority of Singapore (MAS) and Personal Data Protection Commission (PDPC) have both issued guidance documents that set out various considerations for how organisations can use AI in a responsible manner. Both regulators have encouraged companies to pilot the guidance and share the lessons learned.

In that spirit, Microsoft has been privileged to collaborate with Deutsche Bank, Linklaters, Standard Chartered, and Visa in this project to explore the application of responsible AI guidance to specific use cases in financial services. This white paper is the collective output of the group. We believe that the collaborative approach taken within this group – and the close dialogue with the MAS and PDPC – has been very beneficial. It demonstrates the positive impact of cross-industry partnership and engagement with regulators to advance the implementation of responsible AI.

The diversity of perspectives and open dialogue demonstrated through this project is essential if we are to build trust in AI, while maximizing the opportunities it presents. We look forward to working together with organisations across Asia as they continue their responsible AI journeys.

Marcus Bartley Johns
Regional Director, Government Affairs and Public Policy
Microsoft Asia
1. Introduction

The rapid pace of evolution in artificial intelligence (AI) applications, coupled with improvements in computing power and availability of data, have led to growing adoption of AI across all industries. The financial services industry is no exception, with financial institutions (FIs) adopting AI across a variety of contexts.

Some of the uses for AI in financial services include:

• Providing more personalized services to consumers using Machine Learning;
• More rapidly responding to basic queries in retail banking through Conversational AI;
• Empowering employees by making data-driven insights more readily accessible across functions within FIs, and;
• Streamlining regulatory compliance, from Know Your Customer / Anti-Money Laundering checks to using intelligent systems to track regulatory changes.

A wider range of use cases for AI in financial services, across three broad categories of customer experience, employee empowerment, and AI-empowered insights, is set out in Box 1.

A recent Microsoft/IDC survey of FIs in the Asia-Pacific highlighted that AI adoption is gaining pace in the industry across the region. More than half (52%) of the organisations surveyed said they are already on their AI journey – this is higher than other industries, for which the Asia-Pacific average is 41%. Those FIs that have adopted AI across different business functions have seen improvements of between 17% and 25% in areas such as better customer engagement, higher competitiveness, accelerated innovation, higher margins, and improved business intelligence.

As the range of potential use cases for AI grows in financial services, as in other industries, it important that public and private sectors work together to ensure AI is developed and deployed responsibly. Increasing confidence amongst enterprises, regulators and society more widely will encourage further adoption and innovation that maximizes the benefits of AI, while addressing any challenges.

Globally, an important step to strengthen trust in AI has been the development of principles for responsible development and deployment of AI. Multi-stakeholder processes have resulted in a number of countries and groups publishing principles for the responsible use of AI, including the OECD, G20, European Union, and Singapore, among others. Singapore is unique in having issued two sets of guidelines – one that is intended to be cross-cutting and applied across all sectors, and one that is specific to financial services. Taken together, these factors are driving strong interest in what practical measures can support the implementation of principles for responsible AI, including the two sets of guidance issued by the Monetary Authority of Singapore (MAS) and Personal Data Protection Commission (PDPC) in Singapore.

Goals of the project and this White Paper

This White Paper is the product of a cross-industry collaboration of global institutions that are strongly focused on the technology, business and legal implications of AI adoption: Deutsche Bank, Linklaters, Microsoft, Standard Chartered, and Visa. These partners have produced this White Paper with the goal of supporting efforts for the responsible adoption of AI in financial services.

The partners that developed this White Paper believe that its findings will be directly relevant for regulators, industry, and other stakeholders that want to ensure responsible AI adoption in the financial services industry.

As FIs operate in a highly-regulated industry with robust requirements covering technology use, risk management, and governance, they are able to build on an existing foundation of governance and controls, when considering the use of AI. This provides an important case study for other industries. We aim to build understanding of how existing controls and governance (across the financial ecosystem) can be leveraged to implement and demonstrate good practices of responsible AI – coupled with improvements in technology that are helping address implementation of the principles.

The partners involved in this project took an approach that we believe is unique so far in exploring the implementation of responsible AI in financial services. We decided to apply the MAS and PDPC guidance to three real-world AI use cases in financial services, and use this as a way of generating findings relevant to wider efforts to implement responsible AI. The project partners did not aim to identify a definitive set of methods for implementing responsible AI – not least because we believe this is an ongoing effort – but instead sought to derive insights and learnings from a practical exercise to apply responsible AI principles to specific use cases.

The approach taken helped underline that greater, practical understanding of the materiality of AI systems – compared with extreme interpretations of potential consequences based on imagined or low probability scenarios – is essential in guiding proportionate application of responsible AI principles.

---

1 Microsoft/IDC (September 2019), Future Ready Business: Assessing Asia-Pacific FSI’s Growth with AI, covering 15 Asia-Pacific markets, with 168 business leaders and 94 workers from the financial services industry participating in the survey.
AI Use Cases in Financial Services (Box 1)

**Customer Experience**
- Identity
- B2C B2B Digital Agents
- Digitized Customer Service
- Targeted Offers/ Next Best Action
- New Banking Products Powered by AI

**AI Powered Employees**
- Relationship Managers
- Wealth Advisory Risk and Compliance
- Robotics Process Automation
- Financial Forecasting

**AI Powered Insights**
- Identity Protection
- KYC Fraud Prevention
- Payment/AML Fraud Prevention
- Customer Churn
- Market Surveillance
2. Implementing AI principles: applying Singapore’s MAS and PDPC guidance

Singapore – as an Asian hub for financial services and technology – has been an early promoter of AI adoption, while also encouraging this adoption be done responsibly. It is also the first jurisdiction to publish both financial sector-specific guidelines and cross-industry guidelines for the responsible use of AI.

In November 2018, the MAS published the Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of AI and Data Analytics in Singapore’s Financial Sector (the MAS FEAT Principles). In January 2019, the Singapore PDPC published its Proposed Model AI Governance Framework (the PDPC AI Framework), a sector-agnostic, cross-cutting set of principles and guidelines. Both sets of guidance were developed based on consultation with industry. Notably, both the MAS and PDPC have encouraged feedback from industry and other stakeholders on implementation of the guidance documents, with a view to seeing them as “living documents” that could be amended in future.

The starting point for the project group was our desire to respond to this encouragement from the MAS and PDPC, by exploring issues raised in implementation of the two sets of guidance. Rather than doing so in isolation, the partners decided to explore the considerations involved in implementing both sets of guidance at the same time. This was seen as especially appropriate given the design of the PDPC Framework as applying across all industries, and the MAS guidance as applying specifically to FIs. The result of having both the MAS and PDPC guidance in place is that organisations in the financial services industry are encouraged to consider the implementation of both sets of guidelines when considering adopting AI technologies for their business.

Each of the project partners is a global institution, operating in many more jurisdictions than Singapore alone. This meant that an important consideration for the group was how to implement responsible AI principles globally, especially as the volume of responsible AI principles issued around the world continues to increase (see Box 2). As explained in detail below, the group also included key principles from the European Commission High-Level Expert Group’s (EC HLEG) Ethics Guidelines for Trustworthy AI.

**Approach taken to exploring implementation of the principles**

The group based its work together around three use cases of AI in financial services, each of which was drawn from the practical experience of the partners implementing these use cases. For each use case, one financial services partner led in assessing how the principles in the MAS FEAT Principles and the PDPC AI Framework would apply to that specific adoption of AI technology, and the whole group discussed and debated the findings. As noted above, the purpose of this approach was to provide practical, real world insights into the benefits, issues and challenges identified in the implementation of AI principles. Microsoft contributed analysis from its work across industries on AI systems, and efforts to lead adoption of responsible AI, including through research on technical methods relating to many of the AI principles considered in the MAS and PDPC guidelines; and Linklaters focused on the legal and regulatory considerations involved.
The group began by mapping the recommendations of the MAS FEAT Principles and the PDPC AI Framework. The product of this mapping exercise was a set of questions, organized thematically according to the responsible AI principles to assist in the assessment of each use case.2

Mapping two sets of guidelines into a list of harmonized questions was a challenging exercise. In the course of doing so, we found that the principles set out in each of the guidelines both rested on the pillars of Fairness, Ethics, Accountability and Transparency, and so are largely complementary. However, compared to the MAS FEAT Principles, the PDPC AI Framework suggested more detailed steps and measures that organisations should consider taking to align with these principles. Each set of guidelines also introduced potential challenges for organisations seeking to apply both sets of guidelines together.

One key question is that of scope: the MAS FEAT Principles uses the term “AIDA” to refer to “AI and data analytics”. The PDPC AI Framework, on the other hand, only relates to “AI”. From a technological perspective, these are distinct concepts, and while AI relies on and includes data analytics, it is more than that – it is a set of technologies that enable computers to perceive, learn, reason and assist in decision-making to solve problems in ways that are similar to how humans behave. There is therefore a risk that the term “AIDA” could be read to include relatively simple (even manual) data processing and analysis activities. Perhaps in taking this approach, the MAS may have intended for FIs to focus on the context and consequences of the way their AIDA systems are used and the materiality of the decisions being made, rather than whether the technology simply falls within the “AIDA” bucket. Based on our learnings from this project, that would be a welcome approach.

Global momentum on responsible AI principles (Box 2)

Over the last 12 months there has been significant effort globally towards developing principles for responsible AI, including the following:

- **OECD**: based on input from a multi-stakeholder group, the 36 member countries of the OECD adopted AI guidelines in May 2019, and six other non-OECD members have also adopted the guidelines.
- **G20**: in June 2019 G20 members adopted human-centred AI principles, drawing on the OECD principles – along with the core G20 economies seven other countries participating as guests also adopted the principles.
- **EU**: A multi-stakeholder group provided inputs, along with an open consultation process, resulting in Ethics Guidelines for Trustworthy AI being issued in April 2019, along with a draft Trustworthy AI Assessment List that is intended to provide practical guidance for implementing the principles.

A number of other governments and groups, in addition to Singapore, have issued principles for the responsible adoption of AI, including in Australia, Japan, and China; and many companies (including Microsoft) have publicly issued principles for their own AI development and adoption efforts.

There is significant common ground between the various principles. At the same time, an important question for the project partners was what challenges are involved in shifting from the level of principle to practice – in other words, is it challenging for global companies and organisations to implement multiple sets of detailed responsible AI guidance simultaneously?

2 This mapping exercise is available on request for those interested in further discussions on this topic.
As noted above, in conducting this mapping exercise the group also considered the EC HLEG Ethics Guidelines for Trustworthy AI. As global organisations, each partner was already considering the relevance of the EC HLEG Guidelines for their approach to responsible AI. We agreed that we would not be applying the Singapore guidance in isolation from guidance from the EU, or other jurisdictions. Notably, the EC HLEG Guidelines explored an additional angle of looking at the technical implementation of AI solutions, specifically suggesting that organisations explore aspects of system security, business continuity/disaster recovery and data protection/data governance. We thought these were important considerations, despite not being aspects that feature in the MAS FEAT Principles or the PDPC AI Framework.

The examples above demonstrate that even though there may be increasing global convergence at the level of principles for responsible AI (see Box 2), it can be challenging for organisations to translate multiple sets of guidance on implementing these principles. The fact that this process was challenging for a partnership of several large global organisations highlights the even greater challenge that smaller businesses would face. This highlights that there could be unintended negative consequences if there is a proliferation of detailed guidance from regulators on implementing responsible AI.

The responsible use of AI is not just a country-level or sector-level objective – it must be a global one, and there is merit in having coherent and consistent approaches towards AI adoption across country lines which businesses can use to ensure an aligned approach to not only its responsible use, but also to build confidence in this new type of technology. Global standards and best practices will have an important role in promoting coherence, as discussed in Section 5.
3. What is AI and when is its use material?

AI is not a single technology, but rather a set of technologies that enable computers to perceive, learn, reason and assist in decision-making to solve problems in ways that are similar to what people do. This involves taking computers beyond the limitations of following pre-programmed routines.

AI systems are generally grouped into four categories:

- **Vision**: the ability of computers to “see” by recognizing what is in a picture or video;
- **Speech**: the ability of computers to “listen” by understanding the words that people say and to transcribe them into text;
- **Language**: the ability of computers to “comprehend” the meaning of the words, taking into account the many nuances and complexities of language (such as slang and idiomatic expressions), and;
- **Knowledge**: the ability of a computer to “reason” by understanding the relationship between people, things, places, events and the like: for instance, when you have a meeting and the last several documents that you shared with the person you’re meeting with are automatically delivered to you.

The rapid improvement in AI capabilities and adoption is largely driven by rapid increases in computing power, and the availability of data.

AI systems are not always discrete pieces of software (for example, a “chatbot” conversational AI system); and increasingly AI is being incorporated into updated versions of existing software. This deep incorporation of AI into enterprise productivity software has happened rapidly in recent years, and is likely to be less obvious a use of AI than examples like facial recognition, machine reading, or conversational AI.

The development of “intelligent features” within core functions of enterprise productivity software provided by Microsoft, and other companies, highlights this. For example, Microsoft 365 uses AI in various features, including spell checking and grammar in Word; suggesting design approaches for PowerPoint presentations; suggesting relevant emails and documents for upcoming meetings through Microsoft Outlook; or improving accessibility of PowerPoint presentations by providing live speech-to-text transcription. Microsoft Dynamics 365 uses AI in software that is used regularly by businesses, from making recommendations to sales teams engaging with customers; recommending optimal scheduling for field service teams; or providing insights on the impact of communications and marketing activities. The increasing use of AI in everyday productivity contexts like these has important implications for considering when responsible AI principles need to be applied, and when use cases of AI are material, as discussed below.

**Assessing when a use of AI is material**

A key topic explored by the group was when the use of an AI system should be considered material. Determining materiality is the first step in guiding the proportionate application of governance and controls needed to implement responsible AI principles. The group’s discussion recognized that what is material in one industry – financial services in this case – may be different from other industries. Key questions included: What does “materiality” mean in practice, and how is that determined? How does the level of “materiality” impact how an organisation applies the principles in the guidelines?

In the group’s initial discussion, we considered whether materiality was determined solely by an impact on individuals (for example through harm, denial of service, or infringement on individual rights) but agreed this would be too narrow a focus. For example, an AI system may be material if it has an impact on capital adequacy, or other impacts with systemic implications from a regulator’s perspective. At the same time, not all uses of AI will be material. The increasingly embedded nature of AI within all forms of software – as outlined above – means that this question of materiality is becoming more and more relevant in order to avoid applying detailed consideration of AI guidelines to every piece of software being used in FIs. An FI’s use of AI systems that improve spell-checking in Microsoft Word, or recommend ideal meeting times in Outlook, for example, is not likely to warrant detailed ethical consideration in terms of its responsible use.

There were two themes that emerged from these discussions. First, there is a need to understand what “material” means, and what considerations are factored into that finding. The MAS guidance provides helpful considerations to guide the materiality assessment. In the use cases discussed, our focus was primarily on the severity and probability of impact on different stakeholders, including individuals. As a tool for assessing that consideration, our partners found the PDPC AI Framework’s proposed probability-severity of harm matrix to be a good starting point. An assessment of the impact and risk of harm to data subjects by decisions made using AI was found to be a common baseline criterion that our partners considered when assessing materiality. However, the group also found that a range of other factors could feed into this determination, including:

- the existing regulatory compliance framework that an organisation faces and how the adoption of AI could impact that compliance framework;
- potential cybersecurity risks associated with using the technology;
- the data that is used for the model building process;
• what downstream impact the results of such use might have on the integrity of the market and the need for healthy competition among players, and;
• the potential reputational impressions that an adoption of such technology for a particular use case could have for an organisation.

These are a broad range of issues identified in the group’s discussions, and we noted that the MAS guidance includes additional considerations. To support the practical application of responsible AI principles within an organisation, a well-defined set of considerations to guide a materiality assessment will be needed within each organisation.

There was broad agreement in the group that a risk-based approach, starting with an assessment of the materiality of the use of AI, was needed. Such an approach would be effective in achieving the goals of implementing responsible AI, especially given that AI technologies and their applications are likely to continue to develop and expand. It will allow FIs to build on their existing governance frameworks to adapt to new considerations presented by the developing AI landscape.

Second, once a determination of materiality has been made for a particular use case, there is a need to consider the proportionality of the responsible AI use principles and measures against that determination. We discussed that different approaches were possible: applying a subset of principles to a use case, based on an assessment of what was relevant; or applying each principle but doing so in a proportional manner for the specific use case. In certain respects, a finding that a particular principle is not relevant to a use case seemed appropriate, and organisations should have the discretion to make that determination.

Based on the group’s experience, we would encourage further discussion between regulators, industry and other stakeholders to identify common considerations in assessing the materiality of uses of AI; and to clarify that not all uses of AI are material. The end product could be the development of a concise set of questions/criteria to support best practices and help FIs assess the materiality of uses of AI; as well as clarity being provided by regulators that a materiality assessment should precede detailed consideration of a use of AI against the full set of responsible AI principles. This further discussion will inform the application of established risk management frameworks and governance mechanisms by individual FIs.

The practical ramifications of these discussions are illuminated in the three use cases that are discussed in Section 4 of this White Paper.
4. Applying responsible AI in context: Three use cases in financial services

In this section, we examine the three use cases in detail, applying the principles in the MAS FEAT Principles and the PDPC AI Framework, and capture the considerations and lessons that surfaced from each use case which we think are helpful to be shared with industry, regulators and other stakeholders. Where we have found solutions or best practice approaches to address some of these considerations, we have shared them as well.

The objective of this section is not to identify a definitive set of approaches to implementing responsible AI, and the questions raised by the MAS and PDPC guidance. In fact, there are certainly points where there are no definitive answers yet, and where this is the case, we have noted that as well, pointing to the efforts being made to obtain more clarity in that respect.
4.1 Use Case 1: Using AI to predict which cards will be used for travel

Description of this use case

An organisation is applying AI to its deep dataset of card transactional data to be able to predict which cards are likely to be used for travel in the coming months. This is a service that is provided to card issuers (i.e. the service is a business-to-business (B2B) service) and the commercial utility of this is to provide insight as to which cards are the best candidates for travel related marketing.

No protected attributes (i.e. personally identifiable information like names, gender, or age) are used in the AI model building process for this solution. Card issuers themselves match the travel propensity scores with their own data and determine who to send the campaign to.

Assessing the materiality of the application of AI

To assess the materiality of the technology application in this use case, we found the PDPC AI Framework’s proposed matrix to classify the probability and severity of harm to an individual as a result of the decision made by an organisation about that individual to be a helpful reference point. Applying the PDPC AI Framework’s proposed matrix to this use case, we determined that the probability and severity of harm to an individual as a result of the use of AI was low. The key outcome of improper decision making was low campaign success. This finding was a helpful basis for informing both the application and the relevancy of the principles to this particular use case. The group discussed how the AI developer in this use case uses a model risk management governance process which focuses on defining the materiality of the risk, how the results of the model are being used and what data points are being used to train the model. Although this was also relevant in other use cases, the group also discussed the importance of fostering an organisational culture and diversity of viewpoints in making risk assessment decisions like this.

Transparency and External Accountability – How do these principles apply to AI developers?

An important feature of this use case was that this involved the development and application of AI by an organisation as a service to another organisation to take certain decisions that ultimately affected end consumers. As a B2B service, the organisation developing and applying AI does not itself have access to, or interface with, the end consumer. Therefore, that organisation views that the accountability lies with both the AI developer (B2B service) and AI deployer (issuing bank), in meeting the explicit requirements around external accountability and transparency to data subjects directly. This includes the recommendations to inform data subjects about the application of AI, and the expectation to provide an appeal / review mechanism in relation to decisions made.

To implement shared accountability, we determined that the best approach for the AI developer to implement the Transparency principle would be to give this organisation’s issuing banks (AI deployers) model predictors and fully documented model inputs and outputs for the issuing banks to then fulfil their responsibilities under the Transparency principle by educating end users on how decisions are made using AI. As part of this process, AI deployers are then empowered to check model quality and provide feedback on model performance to the organisation for the purposes of improving and refreshing the model.

In addition to this technical information, the organisation could make available non-technical documentation explaining how the system works, so that the card issuer could use this in explaining the functioning of the AI system to consumers in plain and understandable terms. This draws on the approach used in other use cases where the technical documentation and non-technical explanation were separate documents.
Microsoft takes to providing transparency for various AI services that its customers incorporate into their own AI-infused apps. For example, the Microsoft Bot Framework, which is a set of templates and solutions for businesses to build conversational AI experiences like Q&A bots or virtual assistants, includes detailed technical documentation for developers, as well as non-technical documentation like guidelines for responsible conversational AI use that are intended for a non-technical audience. Another example is Face API, a building block for facial recognition systems available through Azure Cognitive Services, which includes a Transparency Note explaining in non-technical language how facial recognition works and the choices that facial recognition system owners to deploy it responsibly; alongside technical documentation intended for developers using Face API.

**Fairness – Context is also important in ensuring an AI model is built fairly**

In the context of this use case, travel predictions for the issuing bank’s customers may vary from time to time and are subject to seasonality. It was therefore critical to take these factors into account when building a model to ensure that individuals are not disadvantaged by decisions made in a particular season by the AI solution which is trained on data gathered from another season.

Fairness and the application of the responsible use principles must be dynamic in nature, and is not only dependent on differences in context, but also in time. In this case, the approach that is taken to ensuring fairness of this AI model is to account for seasonality in the models and refresh these models periodically to ensure new trends are captured keeping predictions accurate.

Finally, consideration was given to balancing maximising group fairness (sending offers to demographic groups that typically do not receive them) compared with individual fairness (only sending offers to specific customers that the model predicts are likely to travel). The AI developer can in this case assist the AI deployer (issuing bank) in determining the most suitable fairness scheme, though the ultimate decision is made by the AI deployer. The MAS FEAT and PDPC Framework fairness considerations appear to be determined based on the AI deployer’s needs.

**Existing regulatory and internal governance requirements align substantially with the responsible use principles**

This use case also presented helpful observations in understanding whether any existing regulatory and compliance requirements or internal governance requirements impacted the proposed adoption of this AI solution, and if so, how those aligned with the responsible use principles.

This organisation has an internal governance framework to ensure that its clients (i.e. issuers) are not given a certain quality or quantity of information that will give one client a competitive advantage over another or allow a client to understand how its competitors are faring in the market. There is also a need to ensure fairness when presenting benchmarking information. These compliance requirements naturally result in a circumspect analysis of the type, quality and quantity of the datasets being fed into the AI model, which in turn address some of the expectations in the MAS FEAT Principles and the PDPC AI Framework around the Accountability and Ethics principles. Particularly, these internal governance considerations necessitate an approach to AI development that requires alignment with the organisation’s ethical values and principles.

Another aspect that merited consideration is the overlap with data protection regulation. In providing this service to issuing banks, the AI developer could assist the issuing banks in ensuring they have obtained the necessary consents from their customers before tailored marketing and advertising campaigns can be sent to them. It is ultimately the AI deployer that will need to understand the model building process so that it can satisfy relevant compliance requirements (e.g. obtaining appropriate consents to capture the processing activities contemplated, as noted above).
4.2 Use Case 2: Using AI to automate verification of wet ink signatures

Description of this use case

An organisation is looking to apply an AI solution that is trained to verify the authenticity of wet ink signatures. This solution works by comparing a signature on an instruction against the client’s specimen signature to determine authenticity.

The utility of this solution is to allow the organisation to improve consistency and accuracy in signature verification and minimize the risk of forged signatures accepted. The model is also designed to detect changes in signature patterns over time.

Assessing the materiality of the application of AI

The two-pronged discussion on materiality raised in Section 3 above (i.e. considering the materiality of the use of AI and the importance of proportionality) was particularly relevant for this use case.

This use case essentially involves the use of AI to enhance an existing manual operational process. The first question is whether the use case is material at all, warranting any consideration in terms of the responsibility issues raised. If this use case is deemed material, the next question would be the extent to which the aspects of proportionality and flexibility can feature in applying the principles set out in these guidelines.

The PDPC AI Framework’s matrix to assess the probability and severity of harm to an individual as a result of the decision made by an organisation about that individual was a helpful reference point for assessing the impact of this AI technology. In applying this matrix to this use case, we determined that the probability and severity of harm to an individual as a result of the use of AI was on the lower end of the spectrum. We reached this conclusion on the basis that the AI system was being used to support or replace an operational decision as to whether a client instruction should be processed or escalated to a human for review, and so the risk of harm to the individuals concerned seemed to be minimal.

Fairness – Considering the need for a human-in-the-loop

The consideration of the Fairness principle in this use case was an interesting one. In this particular use case, the use of AI instead of a fully manual process of having humans compare signatures provided a real benefit by reducing some of the disadvantages from aging or disability of customers. The AI model is designed to account for changes in signature patterns over time, and prompt customers when the time has come for a specimen signature on the system to be updated.

In considering which of the AI decision making models this solution should use, it was determined that a human-in-the-loop approach would be used for this application of AI, with a differential tolerance (i.e. how closely matched the signatures are) to be determined so that matches that fail this threshold are escalated appropriately. It seemed impossible to base a tolerance level off the current practice (i.e. human verification) because that is a subjective analysis, however with the use of AI, we were confronted with proposing an objective tolerance (e.g. 30, 40, 50% identicality) to the model.
Transparency – Increasing confidence in the use of AI

In applying the Transparency principle, we discussed the need to ensure that frontline staff are able to explain the use of AI to consumers in a simple yet adequate manner. However, a key concern is calibrating the level of detail provided to consumers such that consumers are not able to “game the system” from the information disclosed. That is particularly sensitive in this case as it has implications on fraud detection. This highlights a potential tension between the Transparency principle and the security of AI-enabled fraud detection or prevention systems.

In addition, we determined that given the context, it did not warrant the need for the full spectrum of accountability measures that are set out in the MAS FEAT Principles and the PDPC AI Framework (such as the need for alternative channels for redress) given the organisation already had existing established channels in place sufficient to deal with any customer concerns about this technology application. That said, we generally appreciate that the broader objective of the Transparency principle is in fact to increase public confidence and trust in the use of AI. This use case is particularly suited in that regard as it is a solution to help detect fraud better, and so proactive disclosures to customers could in fact help to increase awareness and promote acceptance.

Not all responsible AI principles are equally relevant in every context

In the course of discussing this use case, we were challenged to consider whether certain principles in the MAS FEAT Principles and the PDPC AI Framework were relevant to the use case. There were certain measures that these guidelines suggested organisations should implement which did not seem practicable in the context of this particular use case.

For instance, we discussed that the principle to provide consumers with an explanation on what data is used to make AIDA-driven decisions about the consumers is not relevant in our use case given that the only input data is their wet ink signatures. Similarly, we discussed that principles concerning the use of personal attributes of data subjects in ways that may generate bias were not relevant as there were no personal attributes being used for this particular AI application, beyond the wet ink signature itself. Like Use Case 1, the context in which AI is being implemented in this case was critical to that determination.
One of the key challenges for FIs in the KYC process is capturing quality and completeness of customer data in the shortest possible time. This is essential for building a trusted AI-based KYC model that can generate insights and decision support to subsequent KYC, Anti-Financial Crime and Compliance related process automation. The goals are to gain compliance efficiencies, improve customer experience, and do so in a cost-effective manner. AI can facilitate achieving these goals better than existing, non-AI methods.

In considering which datasets to use for these purposes, both internally sourced datasets and externally sourced datasets (from third party sources) are used to form and train models and systems. Organisations consider this appropriate and necessary to have a sufficiently complete dataset from which conclusions can be drawn about data subjects in the areas of KYC checks.

In this process, there are four steps to consider: (1) identify the additional information gaps from existing data sets and assessing the opportunities to close these gaps with other data sources; (2) assessing the quality of these other data sources, including using up-to-date data to meet accuracy obligations under privacy laws like Singapore’s Personal Data Protection Act; (3) performing profiling and segmentation computations once the super data set is determined to be sufficiently complete and of trusted quality, and; (4) performing linkage analysis.

**Data quality is an important factor in the materiality assessment**

A key risk in this use case is whether the organisation can deliver effective and adequate due diligence with an AI system for KYC. Based on the context considered by this organisation in how these AI models and systems could be used (e.g. denying customers access to bank accounts), the materiality of the decisions about data subjects to be made, assessed against the PDPC AI Framework’s probability-severity of harm matrix, was considered to be high.

This is because KYC checks (in this case AI-enabled/assisted but with “human-in-the-loop”) are determinative of whether financial products/services are offered to the data subject (i.e. whether a loan or line of credit is approved), which will have direct impacts on the individual, as well as commercial implications, and regulatory compliance implications.

It was found that the materiality was directly related to the datasets used as inputs for the AI models involved (as noted in steps 1 and 2 above in the explanation of the use case).

A tension between two considerations was raised. On the one hand, organisations needed as complete a dataset as possible in order to produce accurate results, while on the other hand, using more data sets from third parties can increase the risks of poor dataset quality, the potential for false positives, and unintended bias in the results. Having more data may also go against the privacy principle that personal data collection should not be excessive. This would raise the gross materiality assessment of this proposed application of AI, as there is a significant risk of harm if appropriate due diligence procedures and internal risk assessments to track data lineage and to verify data sources are not implemented. While this risk would similarly exist in the use of mixed data sources for non-AI analytical methods, it was perceived that new considerations are introduced by the application of AI because of the increased speed of analysis, and the way in which AI models would generate predictions changing the way KYC checks would be processed. At the same time, if implemented effectively, these risks could be translated into benefits, by speeding up decision-making on whether to onboard customers, and improving accuracy of predictions.
Both the MAS and PDPC guidance were relevant in this use case. In both the MAS FEAT Principles and the PDPC AI Framework, data quality and provenance are considered at length under the principles of Fairness (justifiability of using certain datasets), Accountability (ensuring there is internal accountability for how models are built) and Transparency (provenance in data inputs and how AI outputs are used in decision making). Organisations also identified existing governance and controls that were relevant, including BIS requirements on data quality; determining the necessity of utilising third party data, and vendor risks assessments that would mitigate inherent risks. An important KYC/AML requirement on FIs is to use data and information from reliable sources.

**Fairness in model building**

Ensuring the Fairness principle was implemented effectively in this use case was a key consideration. The PDPC Framework’s guidance on due diligence in the selection of datasets for model development were particularly relevant to this use. In undertaking KYC checks using mixed data sources as an input for AI systems, bias could occur if those datasets do not have sufficient diversity or are skewed towards certain characteristics. For example, an external dataset might not have sufficient data, or the data included may be skewed in ways that can lead to bias, as well as leading to the risk of deanonymization of certain characteristics (like gender, location, or age). If a more representative dataset were used, these relationships may not feature, and the AI system could generate more accurate predictions without increasing FEAT-related risks.

In the context of this use case, it was agreed that the MAS and PDPC guidance on consistent monitoring for bias in AI systems and data was particularly important.

To address any residual risks, organisations need controls, governance frameworks and processes that enable proactively identifying inherent bias in data sets (including through clear documentation on the characteristics of datasets being used); assess impacts/outcomes when data is run by the algorithms; perform back-testing to recheck variance; monitor technology advances for potential mitigants to bias; and carry out internal stakeholder education. An important concern of the group was to enhance the prevention of unintended outcomes, as opposed to only being cognizant of them when risks that should have been foreseeable materialized.

It was also noted that the KYC process would not be fully automated, and a key objective was to use AI systems to provide more accurate recommendations and analysis that would then be reviewed by human compliance staff. The training of these staff to be able to identify whether specific risks of bias were generated in a KYC process – without expecting them to develop deep expertise in data science – was seen as important.

**The Transparency principle needs to be implemented in a practical way, and trade-offs may exist**

Another key lesson from this particular use case was the need to balance transparency and confidentiality. The more transparency provided on the workings of an AI model and related data, the greater the risk for the reverse engineering of anonymized data. Financial data captured in KYC checks is very likely to include sensitive information. As such, a blanket expectation of increased transparency may pose a security and privacy risk, if it leads to an increased ability to de-mask confidential data. Strong access restrictions/controls as well as the undertaking of risk assessments and monitoring is therefore suggested, especially for high materiality use cases such as KYC checks.

We also noted a potential concern that a comprehensive transparency expectation could introduce risks for the integrity of the KYC/AML process. For example, disclosure of the role and extent that AI plays in the process; meaningful summaries of ethical evaluations; explanation to the data subject on how the decision was made and the reasons behind that decision; and a channel for data subjects to submit appeals and request reviews of this decision could potentially counteract the effectiveness of an AI application for KYC. In this context, FIs will have to consider how to implement such transparency requirements in ways that are consistent with regulatory KYC/AML requirements in not compromising the integrity of a KYC/AML process, including leaving it open to manipulation. In such a case, the PDPC AI Framework has recognized there could be scenarios in which it might not be practical or reasonable to provide information in relation to an AI algorithm especially in the context of AML or fraud prevention (Paragraph 3.21 of the Model Framework). This type of recognition is important for helping strike a balance between transparency and confidentiality.
However, if implementation of transparency involved a more straightforward disclosure to a potential customer that an AI system was being used in a KYC process, this was seen as less problematic. It would still be important to manage expectations of how much information can be provided on the data sources and the process itself to the customer, as this might breach the confidentiality that may be required for the process, and also generate risks that the integrity of the KYC/AML process would be compromised.

**Existing regulations remain relevant for encouraging responsible use of AI**

The overlap with data protection regulation was also highlighted in this use case. By leveraging multiple sources of data, some of which is sensitive in nature, data privacy provisions such as those under the EU General Data Protection Regulation (GDPR) and other equivalent data privacy legislations are important. For example, consideration would need to be given to whether the consent provided for the data to be collected and processed (especially those sourced from third parties) would allow it to be used in their AI models and systems. Potentially, there could be a basis for lawful processing of personal data that could be relied on for AI-facilitated KYC, e.g. the basis of “legitimate interest”.

GDPR is also relevant in the context of implementing the Transparency principle. A data subject has a right to be informed when they are subjected to automated decision-making processes, and a right not to be subjected to a decision based solely on an automated process. This means that the Data Controller must provide Data Subjects with access to at least the right of human intervention with a simple process to access these, including potentially contesting a decision. This may further complicate the balancing between transparency and confidentiality noted above.

The group also considered that other regulations, including those restricting cross-border data flows, were relevant to the use case. Where restrictions exist on cross-border movements of personal data, it would introduce complexities for operation of the AI system for KYC, as well as having similar impacts on non-AI KYC processes.
5. Lessons learned

We conclude this White Paper by summarising below the key lessons that emerged from the application of the principles to the three use cases. It is hoped these lessons contribute to the adoption of AI in a principled and responsible way.

5.1 The importance of context and assessing materiality for specific use cases

It is important to recognize that there is a wide spectrum of AI use cases differentiated by not only functionality and purpose, but also the type of AI model used, model training (including the data that is used for that training), and impact on data subjects. Guidelines on responsible use of AI must therefore acknowledge the importance of context in translating high-level AI principles into practice. As the different use cases highlight, specific principles will have different implications in different contexts, and some principles will be more, or less relevant.

This underlines the importance of starting with a materiality assessment, and a recognition that such an assessment should inform the relevancy and proportionality of the principles (and the regulatory expectations around the measures that organisations need to take to support those principles). It also underlines the importance of ensuring that principles-based regulatory approaches are not overly prescriptive, and not technology-specific, in encouraging the responsible use of AI.

For example, when assessing an AI solution for alignment with the Fairness principle, we found that it was important not to focus on abstract concepts of what is fair or not, but what the principle means in specific contexts. An important area to focus on in that analysis is understanding how choices of datasets may have implications for bias in the way an AI system operates.

Hence, in Use Case 1 for example, we determined that there were two aspects of Fairness that produced different results – maximising group fairness (targeting offers to demographic groups that typically do not receive them) versus individual fairness (targeting only specific customers that the model predicts are likely to travel). In the context of the particular use case (having assessed also the materiality of the decisions and the impact on the end user) and the potential commercial benefits from the use of this technology, it seemed justifiable in this case for the client to prefer individual fairness over group fairness. However, we appreciate that this same conclusion may not have been reached with a different solution or impact.

In Use Case 2, Fairness did not feature as a relevant consideration because no relevant personal attributes are used. In fact, the group identified that the application of that AI technology can have a positive impact in reducing disadvantages which might otherwise exist in the manual process of signature verification, where inherent biases may be present.

Use Case 3 presents a helpful contrast to Use Cases 1 and 2. Because of its assessed high materiality, there were more considerations in the application of the AI principles (particularly Fairness and Internal Accountability) to ensure the risks raised by the use of AI in this use case were mitigated appropriately.
5.2 Existing internal risk management and governance frameworks, as well as legal and regulatory compliance, are highly relevant for responsible AI

In the already highly-regulated sector of financial services, we found that various existing regulations already affect an FI’s approach to adopting AI, and that these are relevant in translating the responsible use principles detailed in the MAS FEAT Principles and the PDPC AI Framework into practice. For example, regulations and industry standards on outsourcing, technology risk management (including cybersecurity risk), and the use of cloud will apply whether AI is part of the technology being used or not. This was an important recognition, given the growing adoption into a wide range of software used by FIs, as discussed in Section 3.

Additionally, as AI solutions tend to rely on significant amounts of internal and/or external data, we found that sector-agnostic compliance requirements (e.g. data protection and cybersecurity) continued to be important considerations for our partners when analysing the use cases against the MAS FEAT Principles and the PDPC AI Framework, despite the topics of security and data protection being absent from these publications. For example, not all of the use cases the group considered involved personal data, but for those that did, compliance with existing privacy regulatory obligations was a clear consideration. This reinforces that many existing regulatory considerations and compliance requirements are equally applicable in the context of adopting and implementing responsible AI.

Given this, based on the use cases explored, we did not identify any current regulatory gaps that suggested a need for new AI-specific regulation to ensure that AI is used responsibly by FIs. Placing overly prescriptive rules with respect to the use of AI could cause any guidance to fall out of step with rapid technological developments in this space, undermining their effectiveness, while also having the undesired effect of adding compliance costs. To that end, we commend the MAS and the PDPC for taking principles-based approaches to responsible AI, as this will give financial services providers some discretion in how they choose to operationalize the principles in relation to their adoption of AI. It will also allow for ongoing dialogue between the industry and regulators to identify whether regulatory gaps exist, including if these relate to sensitive industry-agnostic issues like the use of facial recognition technology.

Aside from legal and regulatory compliance requirements, established internal risk assessment and governance frameworks represent solid starting points for the financial services industry to implement the MAS FEAT Principles and the PDPC AI Framework. We noted the explicit reference to this concept in the MAS guidance. Based on the group’s experience, it is clear that existing frameworks for model risk management, and wider operational risk management, will continue to apply and have relevance for AI. Despite this, the use cases assessed by the group highlighted that careful consideration needs to be applied within existing risk management frameworks to the use of AI, to determine the incremental risk posed, according to the nature of the system being used and the materiality of its use. For example, the challenges posed in ensuring transparency and intelligibility of AI models (as discussed across all use cases), or through the wider use of mixed data sets as inputs for AI systems (as in Use Case 3), warrant careful consideration through existing risk management frameworks.
For industry, this means that the adoption of AI in an accountable way (by ensuring that the risks are properly identified and managed) need not be viewed as a daunting exercise. We found the processes, risk tolerances and governance frameworks that our partners already had in place were highly relevant in the context of AI adoption. Recording the justifications for FI’s decisions on materiality, and the governance and controls applied, would assist good governance. That would continue to make humans accountable for the decisions that AI systems make, while ensuring that the role of humans is proportionate to the potential impacts of an AI system.

However, we recognize the need for detailed discussions with organisations to understand how the technology works, what risks arise from its implementation, and how those risks can be mitigated. The partners recognized that identifying whether new risks are generated through the use of AI will be an ongoing effort that will require partnership among FIs, technology firms, regulators, and stakeholders more broadly. For regulators, this suggests that a major overhaul of governance or setting out prescriptive requirements around internal accountability specific to AI adoption is unnecessary. This does not mean there is no role for regulators – far from it. The group’s own interaction with regulators in Singapore – the MAS and PDPC – underlined that open dialogue with regulators to facilitate two-way sharing of information and practices was extremely helpful. Regulators can also have an important role to play in convening dialogue and partnerships to share information on technical methods, best practices and standards, governance mechanisms, and regulatory considerations involved in implementing responsible AI. Such public-private engagement will be essential for encouraging the further deployment of beneficial AI solutions whilst ensuring potential risks are addressed in the most appropriate manner.

5.3 Technology has a key role to play in addressing responsible AI considerations

As regulators and government continue to grapple with acceptable approaches to AI adoption, there is a concurrent effort by technology developers to increase the set of tools available to help ensure the responsible use of AI by organisations. For example, tools like Azure Model Management (MLOps) have capabilities that enable management, deployment and monitoring of models, contributing towards transparency and accountability. Interpretability approaches are also increasingly available, like the model interpretability features in Azure Machine Learning, which can be used to explain how machine learning models have arrived at specific predictions; or the open source InterpretML toolkit to generate explanations of the behavior of so-called “black box” models or their predictions.

Cross-industry partnerships and research efforts are contributing towards the growing availability of technical approaches relevant to the responsible use of AI. For example, the Partnership on AI is working on a project (“ABOUT ML”) to improve transparency through the creation of standardized documentation on machine learning models and datasets. The scope of this effort includes documenting the design process and purpose, the data sources and reasons for selecting that data, the approach to training, testing and correcting a model, and the purposes for which it is not suited. This initiative builds on a proposal to develop standardized “datasheets for datasets”, drawing on the experience of the hardware industry in producing standardized datasets that describe the features and functions of specific pieces of hardware.

The experience of the group across the three use cases is that while such technical methods and approaches are relevant, the use of technology alone will not ensure the responsible use of AI. The wider range of efforts involved for assessing material uses of AI – including risk management and governance – will be essential. However, the group viewed that it will be important for regulators to recognize that technical methods are improving to ensure the responsible use of AI. Regulators can play an important role in raising awareness of these methods and encouraging adoption of these methods. It also underlines the importance of taking a technology-neutral approach, given not only AI models, but also methods to ensure they are used responsibly, are being innovated on an ongoing basis.
5.4 The importance of diversity and culture for responsible AI

A common theme that surfaced during this project was the need to involve a range of stakeholders in implementing AI responsibly, and to foster a culture within organisations that recognizes the importance of responsible AI. One of the participants suggested that greater awareness and education should be a focus at this stage of development, rather than rushing to new guidance or regulation (which could generate a wide range of interpretations on how they should be implemented).

The diversity of the project group analysing the three use cases was seen as a strength in our approach. We found the input of our partners’ data scientists particularly insightful when the AI solutions for each use case were being explained and what their capabilities (and limitations) were. Equally, regulatory and compliance perspectives were needed, as were those from areas handling government and wider stakeholder engagement, given the societal interest in responsible AI.

Additionally, a diversity of perspectives also adds a healthy challenge to ensuring the responsible use principles are appropriately applied, and that potential concerns are identified and dealt with early. While these cross-practice conversations were not necessary or natural just a few years ago, the principles-based approach to AI adoption (which requires a deep understanding of the model and its impact) necessitates this type of collaboration.

A strong culture of principled adoption within an organisation will help to ensure that AI is responsibly used. There is no “one size fits all” approach to this, and that is evidenced from the way our partners have their own unique approaches in this respect. Two examples from Microsoft and Visa highlight this.

Microsoft takes a range of approaches to build a culture where a responsible approach to AI is taken, both within the company and through its relationships with partners, customers and society. In 2018 the company issued six governing principles for its development and deployment of AI, and has encouraged public engagement and internal discussion on the implementation of these, including in sensitive areas like facial recognition technology. The AI and Ethics for Engineering and Research (AETHER) Committee was convened in March 2018, bringing senior leaders together from across the company to formulate internal policies and tools for the implementation of Microsoft’s AI principles. The company has introduced training on responsible AI, and is focusing on building awareness in engineering groups working on AI systems to understand the considerations involved in responsible AI development and deployment. Cultural change is also a question of how companies engage with their customers and partners. Responsible implementation of AI is a growing topic of discussion with executives and IT teams in Microsoft’s customers: in addition to engaging in these discussions directly, INSEAD and Microsoft have also taken the approach of providing free online courses (the “AI Business School”) that include modules on responsible AI and AI governance for specific industries, including financial services.

The experience of Visa was also seen as relevant in ensuring a diverse set of perspectives are provided internally on responsible AI development and deployment. In addition to the role of its formal governance structure reviewing the use of data and development of AI, Visa is incorporating data privacy, ethical use of data, and responsible AI into its regular trainings for data science teams.

Model development teams at Visa are being encouraged to consider how AI systems should be developed and deployed responsibly, through a set of guidelines that they have to complete, covering technical issues, data-inputs, model explanation and interpretation, methodology, fairness and quality/accuracy. This is then reviewed by the model risk management committee. Teams are also asked to apply the company’s ethical data science principles in their work:

1. Needs: Start with the problem and not with the data.
2. Law: List the pieces of legislation that apply.
3. Data: Use data proportionately.
4. Limitations: Identify the limitations of the data source(s).
6. Transparency: Make your work transparent and be accountable.
7. Deployment: Describe the steps taken.
8. Test and Learn: Embed tests to refine deployment learning.

The experience of the project partners in exploring the three use cases highlighted the importance of a team coming together with diverse perspectives, from different organisations within and across industries; from different teams within organisations; and from the industry and regulators. Within the companies involved, the teams that worked on the use cases included members of teams responsible for government and regulatory affairs, legal and compliance, data science, data policy and AI developers and users, among others.
5.5 Implementing multiple sets of AI principles is challenging

There is a growing consensus globally on a set of core principles for responsible AI, as outlined in Section 1. The development of more detailed guidance for translating these principles into practice is at an early stage. The experience of the group applying guidelines to the use cases in this White Paper have already demonstrated some practical difficulties for organisations that are subject to multiple sets of guidelines. Despite the convergence on high-level principles in these guidelines, our concern is that the issuance of multiple sets of principles (that are both sector-specific and cross-cutting) is likely to present practical difficulties in implementation across all the guidelines, especially if industry perceives the guidelines as tacit requirements by regulators for implementation. This could result in organisations with global operations being discouraged from implementing more AI technologies at scale.

This further supports the need for greater consensus on core global AI principles. What was apparent in the use cases for this project is that the existence of multiple sets of guidelines potentially presents decisions for which trade-offs need to be made. Even when applying one set of guidance, organisations have to consider various trade-offs. For example, Use Case 2 presented a trade-off between transparency (the ability to request for information about how AI works) and security (preventing bad actors from knowing how to “game” the system which auto-updates changes in signatures). The complexity increases significantly when navigating these trade-offs across multiple sets of guidance. This is especially so for global organisations, like those in our group, which were applying guidance not just from one jurisdiction like Singapore, but also from others like the EU.

The responsible use of AI is not just a country-level or sector-level objective – it must be a global one, and there is merit in having coherent and consistent approaches towards AI adoption across country lines which businesses can use to ensure an aligned approach to not only its responsible use, but also to build confidence in its use. These standards should also be accompanied by public-private discussions on what it means to put these principles into practice in specific industries, and what issues might be particular to that industry (such as the impact of existing compliance regulations). Indeed, global standards (such as International Standards Organisation standards) and global best practices (such as those fostered through global partnerships like Partnership on AI or industry-specific ones such as the Basel Committee on Banking Supervision) do exist and have proven to be helpful in this regard.
Use Case 1 was helpful in underscoring an important distinction that regulators must recognize when introducing principles for responsible use of AI: they will be interpreted differently depending on whether for each use case an organisation is developing AI, or deploying AI systems procured from a third party. In some cases, institutions will be both developers and deployers of a system, but given the increased integration of AI in all forms of software (as highlighted in Section 3 above), and the growth in development of specialized AI systems for the financial industry, this distinction between developers and deployers is likely to be increasingly relevant. This raised questions about how some of the principles and measures in the MAS FEAT Principles and the PDPC AI Framework (especially the data subject-facing measures such as providing transparent information about the AI decision making process) applied to AI developers that provide this technology as a product or service to other organisations.

Implementation of the Transparency principle highlights this issue. The group thought that as AI adoption grows in financial services, this will happen through a combination of in-house development of AI, and use of AI systems (and their outputs, as in Use Case 1) supplied by third parties. Transparency will need to be implemented in a context-specific way and depending on its purpose: for example, explaining to a consumer that an AI system is being used in a consequential way that may affect them, or from a third-party developer providing technical information on the features and limitations of an AI system to a financial services firm that will deploy it.

We appreciate that both developers and deployers of AI technologies have a responsibility to take an ethical approach, but their considerations will be different. For example, key concerns for AI developers will be the model building process, and ensuring AI solutions are robust and explainable. AI deployers that procure these technologies as products will be concerned about getting sufficient information to understand the model so that they are able to explain the model, make informed choices in their use of the system, and be accountable to their customers. AI deployers should be empowered to leverage AI developer institutions’ help in providing sufficient information to their customers. It is therefore important that regulators and governments recognize this distinction between developers and deployers when preparing guidance for implementing responsible AI principles.
Further Information

Useful links

Microsoft Approach to AI

Microsoft Financial Services
banking-and-capital-markets

Deutsche Bank
www.db.com

Linklaters
www.linklaters.com

Standard Chartered
www.sc.com

Visa
www.visa.com