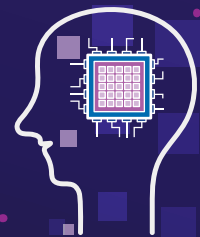
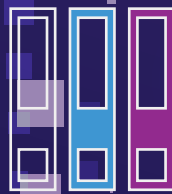


# COMPENDIUM OF USE CASES:

PRACTICAL ILLUSTRATIONS OF THE  
MODEL AI GOVERNANCE FRAMEWORK



# TABLE OF CONTENTS

<b>03</b>	Summary of the Model AI Governance Framework	<b>18</b>	MSD: Keeping Employees at the Heart of AI Use
<b>05</b>	Introduction	<b>22</b>	Ngee Ann Polytechnic: Admissions Selection Made Easier with Responsible AI
<b>06</b>	Callsign: Adopting AI Governance Practices for Identity Authentication	<b>26</b>	Omada Health: Accountable AI in Digital Healthcare
<b>10</b>	DBS Bank: Leveraging AI to Fight Money Laundering	<b>30</b>	UCARE.AI: Accountable AI for Accurate Healthcare Costs
<b>14</b>	HSBC: AI Governance in All Facets of Loan Applications	<b>34</b>	Visa Asia Pacific: Forecasting Cards Used for Travel with Ethical AI Governance

# RESPONSIBLE AI MADE EASY

## FOR ORGANISATIONS

Using Artificial Intelligence (AI) in your organisation?

Help your stakeholders understand and build their confidence in your AI solutions.

### PRINCIPLES FOR RESPONSIBLE AI



**DECISIONS MADE BY AI SHOULD BE EXPLAINABLE, TRANSPARENT AND FAIR**



**AI SOLUTIONS SHOULD BE HUMAN-CENTRIC**

## 4 AREAS TO CONSIDER



#### INTERNAL GOVERNANCE STRUCTURES & MEASURES

- Clear roles and responsibilities in your organisation
- SOPs to monitor and manage risks
- Staff training



#### DETERMINING THE LEVEL OF HUMAN INVOLVEMENT IN AI-AUGMENTED DECISION-MAKING

- Appropriate degree of human involvement
- Minimise the risk of harm to individuals



#### OPERATIONS MANAGEMENT

- Minimise bias in data and model
- Risk-based approach to measures such as explainability, robustness and regular tuning



#### STAKEHOLDER INTERACTION AND COMMUNICATION

- Make AI policies known to users
- Allow users to provide feedback, if possible
- Make communications easy to understand



FIND OUT MORE ABOUT THE PDPC'S SECOND EDITION OF THE MODEL AI GOVERNANCE FRAMEWORK AT [GO.GOV.SG/AI-GOV-MF-2](https://go.gov.sg/ai-gov-mf-2)

# LEVEL OF HUMAN INVOLVEMENT

A design framework to help determine the degree of human involvement in your AI solution to minimise the risk of adverse impact on individuals.

## SEVERITY AND PROBABILITY OF HARM

LOW

HIGH

### Human-out-of-the-loop

AI makes the final decision without human involvement, e.g. recommendation engines.

### Human-over-the-loop

User plays a supervisory role, with the ability to take over when the AI encounters unexpected scenarios, e.g. GPS map navigations.

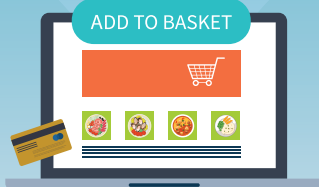
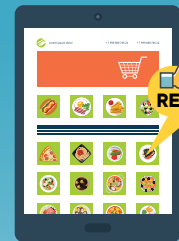
### Human-in-the-loop

User makes the final decision with recommendations or input from AI, e.g. medical diagnosis solutions.

## HUMAN INVOLVEMENT: HOW MUCH IS JUST RIGHT?

EXAMPLE

An online retail store wishes to use AI to fully automate the recommendation of food products to individuals based on their browsing behaviours and purchase history.



### What should be assessed?

#### What is the harm?

One possible harm could be recommending products that the customer does not need or want.

#### Is it a serious problem?

Wrong product recommendations would not be a serious problem since the customer can still decide whether or not to accept the recommendations.

#### Recommendation:

Given the low severity of harm, the human-out-of-the loop approach could be considered for adoption.

# INTRODUCTION

AI will transform businesses and power the next bound of economic growth. Businesses and society can enjoy the full benefits of AI if the deployment of AI products and services is founded upon trustworthy AI governance practices.

As part of advancing Singapore's thought leadership in AI governance, Singapore has released the Model AI Governance Framework (Model Framework) to guide organisations on how to deploy AI in a responsible manner. This Compendium of Use Cases demonstrates how various organisations across different sectors – big and small, local and international – have either implemented or aligned their AI governance practices with all sections of the Model Framework. The Compendium also illustrates how the organisations have effectively put in place accountable AI governance practices and benefit from the use of AI in their line of business.

By implementing responsible AI governance practices, organisations can distinguish themselves from others and show that they care about building trust with consumers and other stakeholders. This will create a virtuous cycle of trust, allowing organisations to continue to innovate for their stakeholders. We thank the World Economic Forum Centre for the Fourth Industrial Revolution for partnering us on this journey. We hope that this Compendium will inspire more organisations to embark on a similar journey.

Here are the use cases.

# CALLSIGN: ADOPTING AI GOVERNANCE PRACTICES FOR IDENTITY AUTHENTICATION

Callsign is a London-based company that leverages deep learning techniques and combines biometrics, geo-location and behavioural analytics with multi-factor authentication to help clients authenticate user identities. Providing services to companies from all over the globe, Callsign helps clients from various sectors like finance, healthcare and e-commerce flag out potential risks in user authentication.

As a company that puts priority on solutions that are transparent, and at the same time, produce reliable and accurate results, Callsign understands the importance of building and maintaining trust with its clients to enable such solutions. With this in mind, they put in place processes to govern the development and deployment of their AI models, adapting and implementing practices as recommended in the Model AI Governance Framework.



## ROBUST OVERSIGHT IN AI MODEL DEVELOPMENT

In overseeing the development of its AI models, Callsign included three parts in its implementation of **internal governance structures and measures**. The first part involved creating a multi-level assurance framework, where each department head formulates and oversees certain controls and policies under his or her purview. This framework is managed by the Chief Security Officer and Data Protection Officer.

The second part comprised a three-stage process – concept, consult, and approve. The process coordinates the engineering, product, sales and research teams to ensure consistency in the data management and development of Callsign’s AI models.



CONCEPT

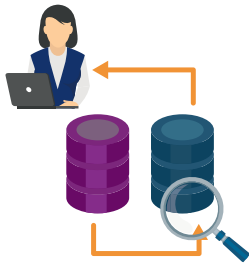
CONSULT

APPROVE

An example of this process would be when the Product Owner conceptualises a feature – either a perceived business or product value – and considers views from various teams such as the architects and security specialists. These views will then be used to enhance the security of the feature before it is presented to the Design Authority for approval. The Design Authority, comprising representatives from other teams as well as the leads of Callsign’s Chief Technology Office, Chief Security Office and Chief Information Office, approves all the AI models that have been developed within the organisation. With the different teams’ inputs and expertise, this second part helps build upon the robustness of Callsign’s governance process for its AI models.

A data provenance process is essential in ensuring accountability for AI model development. For the third part, Callsign established a specialised Data Change Governance Committee to oversee the company's data provenance process. The Committee's responsibilities include:

**Reviewing** inclusion of data points to ensure the AI solutions meet its clients' business purpose;



**Assessing** the types of data collected, including conducting reviews to check the validity and relevance of data; and

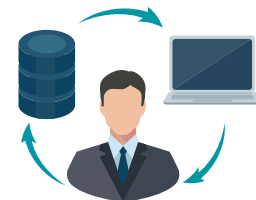


**Ensuring** that the end-to-end information lifecycle has considered controls addressing access, confidentiality, integrity, retention and movement of data.



## HUMAN INTERVENTION ACCORDING TO CLIENTS' NEEDS

While Callsign adopts a **human-over-the-loop** approach in developing the AI model, the company works closely with its clients to **determine the level of human involvement appropriate for the specific application and context**. For instance, if Callsign's client is a bank using AI to authenticate user identities for fraud detection, various security considerations will come into play in Callsign and its client's assessment on the level of human involvement:



### *Client's risk appetite*

A bank may have a larger risk appetite for corporate transactions as compared to transactions made by a retail customer. Flagging out potential risks and disrupting a corporate customer's transaction could result in serious consequences. Hence, in such cases the bank may opt for a lower degree of human intervention for corporate customer transactions;



### *User experience of its clients' customers*

If Callsign's client received poor customer satisfaction scores, the client may consider improving their user experience journey and reduce levels of human intervention in the AI model deployment; and



### *Operational cost to the client*

The cost of supporting customer feedback on the user authentication process may also urge the bank to lower the level of human involvement.



## ACCOUNTABILITY THROUGH DATA GOVERNANCE PRACTICES

CallSign has in place good accountability practices to ensure the responsible use of data for its AI model development. These include measures to avoid over-collection of data and governance frameworks to ensure data protection.

To avoid over-collecting data, and at the same time, still deliver its services effectively, CallSign conducted extensive research and developed new, intelligent ways of gathering valuable results from a minimal amount of data. In addition, CallSign tokenised all personally identifiable information.



For username data, they are hashed by both CallSign and its clients for protection against rainbow attacks.<sup>1</sup> The hashing also allows CallSign to identify individuals while maintaining their anonymity.



For device data, CallSign adopts persistent device tagging methods to allow for device identification whilst maintaining the obfuscation of the user in its AI models.



For behavioural data, CallSign is mindful not to collect specific keys or pin codes that were entered on mobile phones or websites.



1.3XXX° N  
10X.XXXX° E

For location data modelling, CallSign adopts a data protection by design approach by masking the longitude and latitude data collected.

Developing data governance frameworks has also helped CallSign in its accountability approach to AI model development. Using the frameworks and international standards like the ISO guidelines as references, CallSign applied internal data classification principles to classify data sensitivity and created risk-based outlooks to avoid the misuse of data in situations such as data breaches.

As part of its efforts to support point-in-time explanations for its data collection and analysis, CallSign also developed data inventories<sup>2</sup>, data dictionaries<sup>3</sup>, data change processes, control mechanisms, forums and collaterals. Having a **clear understanding of the lineage of data** and being able to provide point-in-time explanations to various stakeholders has enabled CallSign to improve operational efficiency and offer better services to its clients.

<sup>1</sup> Rainbow attack is a type of attack that attempts to uncover the password from the hash

<sup>2</sup> Data inventory is a dataset containing metadata on contents of data, its sources, and other pieces of useful information.

<sup>3</sup> Data dictionary is a dataset describing the relationship between the data, where and how the data is used.



To ensure the performance of their AI models, Callsign collects and carefully creates **distinct datasets to train, test and validate its AI models**. The company conducts intensive testing on the performance of its AI models with the use of proof of concepts, prototypes and through peer reviews from its network of research communities. On top of that, Callsign puts its AI models through behavioural biometric models and tools, such as the ISO/IEC 19795 Biometric performance testing and reporting, to build the model's accuracy. These public tests not only verify the model prototypes, but also provide much needed assurance to Callsign's clients.

Once the performance of the AI models has been tested and baselined, Callsign integrates the performance evaluation into Continuous Integration<sup>4</sup> and Continuous Delivery<sup>5</sup> practices. Through this, Callsign is able to enhance **the reliability of its AI models and ensure that they serve the intended purpose** of providing well-tested services to its clients.

Explainability of AI model outcomes can help tremendously in building understanding and trust with its clients. Cognisant of this, Callsign **documents the development process of its AI models** and extracts insights of its key contributing factors. Such **documentation facilitates explainability**, and with this, Callsign is able to explain outcomes such as model and database description, evaluation parameters and error rate metrics. When technical explanations of the AI model may not be as comprehensible, Callsign provides a non-technical explanation to its clients. This boosts the clients' understanding of the AI solution and encourage buy-in from their internal stakeholders.

## AN OPEN AND TRANSPARENT APPROACH

When it comes to communicating to its clients and their users, Callsign adopts an open and transparent approach, notifying them on the types of data collected, the reasons for collection and the use of such data.

The company's Data Processing Notice is an exemplary example of this openness. In this Notice, Callsign not only details their device, location and behavioural data collection and use, but also explains how clients may use Callsign's product to manage authentication, authorisation and transaction risks. These policies can be configured or used to make automated decisions by its clients. Callsign further **provides an email address in its Notice** for clients to provide feedback on its data processing policies.



## CONCLUSION

As businesses look towards AI to solve challenges, Callsign's early adoption of governance best practices provides the necessary client confidence in using its services. These practices have also enabled Callsign to engage a diverse range of internal stakeholders that help shape an open discussion and contributed to an accountable development of its AI models.

<sup>4</sup> CI is a development practice that requires developers to integrate code changes to a shared repository several times a day; each code change triggers an automated build-and-test sequence.

<sup>5</sup> CD is an extension of CI; teams ensure changes in the code are releasable, and the integration process is fully automated.

# DBS BANK: LEVERAGING AI TO FIGHT MONEY LAUNDERING

DBS Bank (DBS) is a multinational banking and financial services corporation headquartered in Singapore. With more than 100 branches in Singapore and a presence in 18 markets globally, it is the largest bank in Southeast Asia by assets.

To improve the operational efficiency and effectiveness of its current anti-money laundering surveillance, DBS developed and successfully implemented an AI model – the Anti-Money Laundering (AML) Filter Model. The AML Filter Model identifies predictive indicators of suspicious transactions to reduce the number of false positives generated by the non-AI system, thereby reducing the number of alerts that require manual review.

While recognising the vital role AI plays in addressing the limitations of non-AI systems, the bank had an even more pertinent priority – putting in place measures to ensure responsible deployment of the AML Filter Model. After all, responsibility translates to accuracy and reliability. For this, the bank took steps to implement several AI governance processes and practices.



## ESTABLISHING A ROBUST GOVERNANCE STRUCTURE AND FRAMEWORK

To ensure robust oversight of AI deployment, DBS **introduced certain internal governance structures and measures**. These included setting up a Responsible Data Use (RDU) framework, for which a RDU Committee was appointed to oversee and govern it.

DBS made sure the RDU Committee included senior leaders from different DBS units to ensure appropriate levels of checks and balances as well as a good diversity of views.

The RDU framework involved a three-stage process that evaluated and managed the risks of all data used by DBS.



The first stage ensured that the AI model complied with fundamental legal, compliance, security and data quality issues.

## PURE

The second stage ensured responsible data use with the PURE<sup>1</sup> principles.



Finally, the framework also made sure that the AI and machine learning models conformed to the technical requirements of DBS' model governance policy.

For a smooth running of its programmes, a Global Rules and Models Committee (GRMC) within the Group Legal, Compliance and Secretariat was set up. Responsible for assessing all rules, models and score setting changes used for financial crime surveillance, GRMC reviewed the exploratory data analysis, evaluation and deployment of the AML Filter Model. The deployment of the AML Filter Model was then given the green light by the head of the Group Legal, Compliance and Secretariat.

## STRIKING THE RIGHT BALANCE

When it came to determining the level of human involvement in AI decision-making process, DBS aligned its practices to the Model AI Governance Framework and considered, among other things, the probability-severity of harm matrix. Allowing AI to sieve through voluminous transactions to identify the suspicious ones will no doubt increase operational efficiency, but DBS recognised the need to balance this against the impact of false positives being missed out. Adopting the **human-over-the-loop** approach gave the AI model freedom to decide which transactions were suspicious, while also allowing humans to intervene when the situation calls for it.

To achieve this, DBS used statistical confidence levels to determine when its staff are required to review a particular alert. If the AML Filter generated a high confidence level that a surveillance alert is a false positive, DBS did not review the alert because this meant a low likelihood of the alert representing a set of suspicious transactions.

Staff will only take action when the AML Filter Model generated a low confidence level that a surveillance alert is a false positive, because this translated to a higher probability of error in identifying suspicious transactions. In cases where alerts are tagged with a high risk rating, DBS' staff will conduct a comprehensive and detailed review.

This approach saved considerable time as it allowed DBS' staff to be efficiently deployed for the high-risk rating alert reviews.

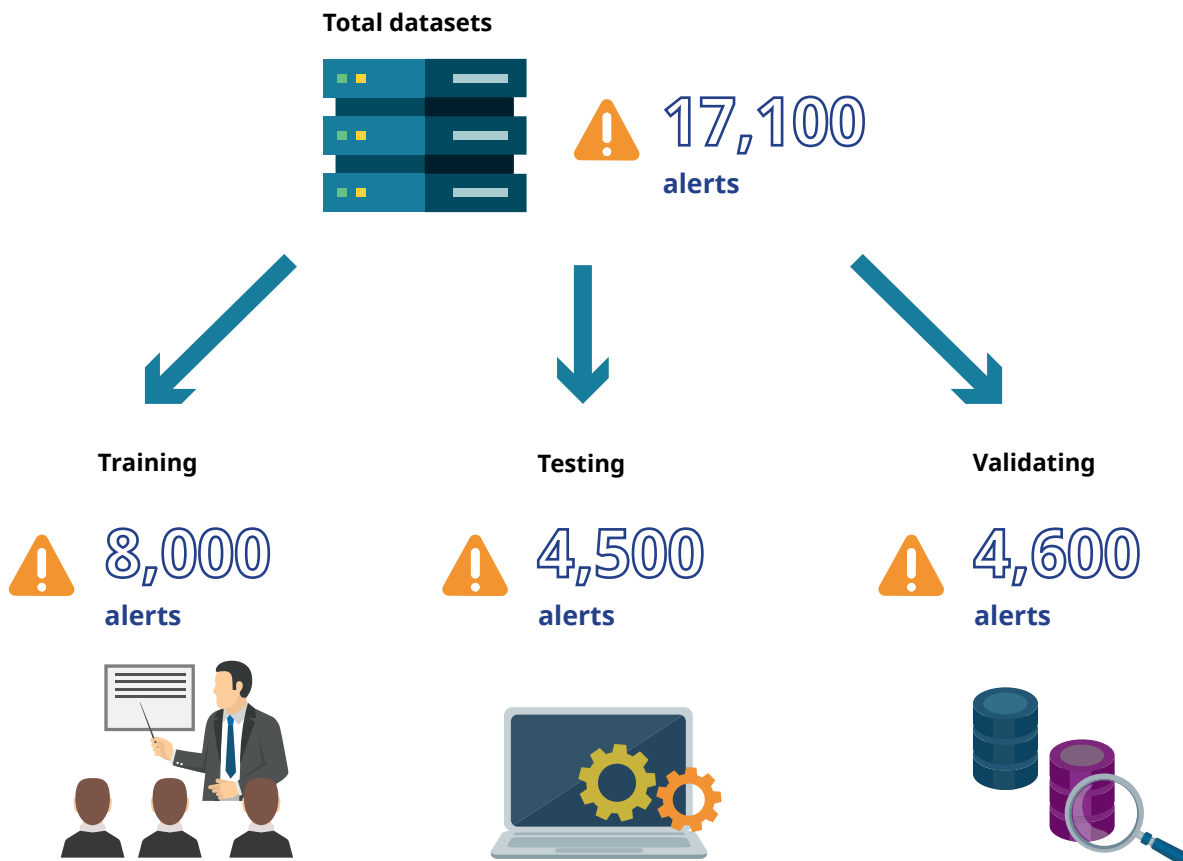
<sup>1</sup> PURE stands for Purposeful, Unsurprising, Respectful and Explainable. In essence, all data use in DBS must conform with its PURE principles.

## ENSURING GOOD DATA ACCOUNTABILITY PRACTICES AT EVERY STAGE

DBS put in place a rigorous process to ensure the responsible use of data in developing the AML Filter Model. This aligned to the data accountability practices suggested in the Model AI Governance Framework, which focused on **understanding the data lineage**, **minimising inherent bias** and **using different datasets for training, testing, and validation**.

DBS recognised the importance of **understanding data lineage** in the development of any AI model. With this in mind, DBS used data from its banking systems for the AML Filter Model. These systems have a direct link to ongoing customer transaction activity, making its source easily identifiable. DBS also obtained data from surveillance systems and historical suspicious activity reports, as a diverse form of data was another essential component for AI model development. On top of that, DBS maintained a data-mapping document that allowed **tracing of all data** used to their respective source system fields. This helped DBS identify the data source, even after it is transformed<sup>2</sup> and aggregated.

To **mitigate the risks of inherent bias**, the bank used full datasets instead of sample datasets to train, test and validate the AML Filter Model. These full datasets were then separated into **training, testing and validating data** for the AML Filter Model.



DBS built its training data from approximately 8,000 alerts triggered by about 4,000 customers over a period of one year to train the AML Filter Model. To mitigate model bias, the bank excluded the data used in training from the testing data. This back-testing stage involved approximately 4,500 alerts generated by about 3,000 customers over a period of six months. Finally, to validate the AML Filter Model, DBS conducted a parallel run<sup>3</sup>, with approximately 4,600 alerts generated by approximately 2,500 customers over a separate period of four months.

<sup>2</sup> This is the technical process where data is changed and or structured from one format/state to another.

<sup>3</sup> Parallel run refers to a practice of concurrently running the existing system after a new system is launched for a period of time until there is confidence that the new system is performing according to expectation.

## REAPING THE BENEFITS

In all, it took DBS almost two years to develop and test the system, giving the team an intimate understanding of how the AML Filter Model functions and arrives at its results.



### *Explainable*

A good understanding of the data lineage and transaction alert triggers, coupled with a transparent computation of the results generated by the AML Filter Model, provided the bank with the ability to explain how the AI model functioned and arrived at its particular risk rating prediction.



### *Responsible*

DBS tracks the model metrics every month to ensure stability of the AML Filter Model. The results from the training, back-testing and validation stages were used as a benchmark for the model metrics. These model metrics could then be fine-tuned post-deployment. To ensure good track record and consistency over time, the model is monitored monthly and reviewed once every six months by the bank's Machine Learning team. This added precaution ensured that any deviation from the pre-defined thresholds will be flagged out for the team's review. Any fine-tuning recommendations by the Machine Learning team will then be reviewed and approved by the GRMC before deployment.

In addition, DBS implemented internal controls to address the risk involved in the deployment of the AML Filter Model. For example, DBS **documented** the results of the AML Filter Model in its meeting minutes to ensure proper knowledge transfer during the various development stages and decision-making processes.



### *Accountable*

In the Model AI Governance Framework, organisations are encouraged to develop appropriate communication strategies to inspire trust through stakeholder relationship management. For DBS, the objective of developing and implementing the AML Filter Model was to increase the internal efficiency and effectiveness in identifying suspicious activities.

In establishing stakeholder confidence, DBS created a simple dashboard to document and track the AML Filter Model performance. Updated on a monthly basis, the dashboard helped the bank explain its model and results to internal stakeholders, such as its senior management and board, as well as to regulators such as the Monetary Authority of Singapore (MAS).

Given the sensitive nature of AML surveillance, the bank understood that in cases where information on the exact surveillance mechanism falls into wrong hands, there will be the inadvertent risk of bad actors avoiding detection. With this in mind, DBS does not publicly communicate detailed information about its AML Filter Model.

## CONCLUSION

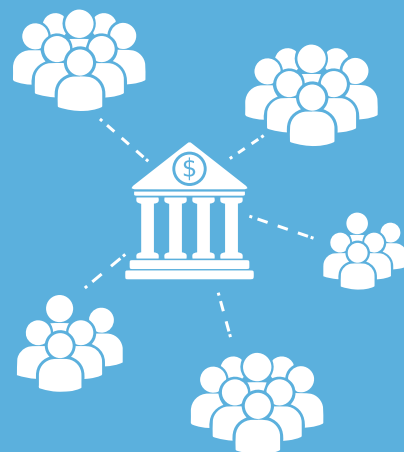
In its journey to use AI, DBS recognises the importance of leadership, governance and AI use frameworks. Only in ensuring the appropriate data and infrastructure environment, and that the governance and AI use frameworks are in place, can organisations fully optimise the benefits of AI technology and solutions.



# HSBC: AI GOVERNANCE IN ALL FACETS OF LOAN APPLICATIONS

HSBC is one of the largest banking and financial service organisations, with operations in 65 countries. Serving more than 40 million customers worldwide, HSBC primarily deploys AI in its retail banking and wealth management businesses. Amongst other applications, HSBC uses AI to promptly and effectively assess loan applications to meet the personal lending needs of its customers.

To enable optimal benefits to its customers, while keeping them protected from potential harm that could come with the unmanned use of emerging technologies, HSBC put in place measures to govern the use of AI in its businesses. From establishing committees and frameworks, to training its staff and taking careful considerations in the level of human involvement for its loan application decision-making, HSBC established practices in all aspects of its AI model development and deployment, while adhering to appropriate global regulatory standards to stay relevant and applicable.



## CHECKS AND BALANCES IN AI DEVELOPMENT

### *Establishing Clear Roles and Responsibilities*

From the start, HSBC understood the importance of keeping a close eye on the various stages of its AI models' development to ensure reliability and accuracy. For this, the bank established an internal Global Model Oversight Committee (**GMOC**). Chaired by its Chief Risk Officer, the committee comprises representatives from relevant departments with defined roles in the development of accountable AI processes for HSBC including:



The Chief Data Office, which is responsible for the firm-wide policy on data quality and governance. This office works with various businesses and functions at the global and regional levels to ensure that the data being used in models meets the intended business purpose and requirements of the data governance policy. If the data does not meet the requirements, the office will assist in putting in place appropriate remediation plans;



Heads of various sub-committees that represent different regions (e.g. US and Asia), businesses (e.g. Retail Banking and Wealth Management) and functions (e.g. Risk Management and Finance); and



### **MRM TEAM**

The Model Risk Management (**MRM**) team, which consists of a Model Risk Governance sub-team that sets the policy and standards and an Independent Model Review sub-team that is responsible for validating the AI models before deployment. The MRM team plays a significant role in reviewing and approving HSBC's AI processes prior to implementation. Besides helping the GMOC understand the risks of each AI model, this team ensures that there are sufficient controls in place to mitigate material risks during model development.

The committee also provides updates on the insights to HSBC's Advanced Analytics<sup>1</sup> to its senior management and board that help the leadership team make informed decisions in managing the risks of its AI governance processes.

To complement efforts of the GMOC, the HSBC's regional Chief Operating Officers provide direct risk and operational management support to the various businesses and functions across all regions. In particular, they ensure that AI/Machine Learning (ML) technologies being brought into the bank are within the risk appetite set by the bank by ensuring efficient and effective risk and control management.

## Framework Development

For HSBC, a key part of managing the use of AI models involved developing and rolling out a framework to align governance in its practices. The MRM team took on this task by enhancing the MRM framework that was developed based on the U.S. regulatory guidance, FRB SR 11-7, OCC 2011-12.<sup>2</sup>

The framework sets out distinct responsibilities for staff involved in the development of the bank's models including AI-based models:



- The Model Owner will ensure that all AI models comply with the MRM framework and adhere to HSBC's model development standards relating to ethics, data protection and bias.



- The Model Developer will oversee the end-to-end model development process.



- The Model Sponsor will be the approving authority, and is typically the senior executive for the business or function where the AI model was used.

In the development and deployment of the AI model for its loan applications review, HSBC's Chief Risk Officer for Retail Banking and Wealth Management business was the Model Sponsor. Before approving the deployment of the AI model, the Chief Risk Officer took into account the issues raised by the Independent Model Review sub-team and ensured that they have been addressed.

For the framework to stay relevant, the MRM team not only updated the GMOC on the existing landscape of AI and ML-based technologies frequently, but also provided essential enhancements to the MRM Framework. For one, the standards within the framework was enhanced to include key aspects relating to explainability and transparency. Biannual reviews were also done on the framework to keep it up-to-date with regulatory standards.

## Training Conducted

HSBC also took steps to keep its staff abreast on the latest changes to the framework. Besides **annual mandatory training**, the bank also provided frequent updates on the framework across various levels of the management to increase awareness of the risks relating to AI-based solutions.

Additionally, HSBC took efforts to train Model Developers and the Independent Model Review sub-team on AI/ML model methodologies to ensure that specific issues are addressed in a timely manner. This led to the creation of several cross-functional working groups, a noticeable upskilling of the model development and validation teams, and more importantly, a collective effort to improve the AI-based MRM framework.

<sup>1</sup> Advanced Analytics is a collection of techniques that examines internal and external data to yield valuable insights (e.g. identify future trends, generate predictive insights and optimise desired outcomes) that can drive business strategies.

<sup>2</sup> U.S. regulatory guidance, FRB SR 11-7, OCC 2011-12 provides guidance for banks on effective model risk management. MRM covers governance and control mechanism such as board and senior management oversight, policies and procedures as well as controls and compliance.

Following a successful pilot training of the MRM framework at the group management board level, HSBC's next step would be to roll out training of these principles to all levels of the staff, so that its internal stakeholders are well-versed with the risks that may arise from the use of AI-based solutions.

### *Keeping to Principles*

In providing a holistic approach to data management and protection, the HSBC legal team **issued a set of principles** for the ethical use of big data and AI to supplement the MRM framework and its data management policies. These principles were consistent with the core values of HSBC and its commitment to customers, shareholders, employees and regulators. With these principles, HSBC was able to increase awareness of the ethical implications of AI use, guarantee consistency in the decision making on global AI usage and encourage effective compliance with HSBC's framework and legal and ethics requirements before new uses of AI are adopted within HSBC.

### *Vendor Management Controls Implementation*

As HSBC also engaged various vendors to deploy AI-based solutions, the bank stepped up efforts in strengthening its vendor risk management practices. These included adding specific questions on AI-based solutions, such as the type of technologies to be procured to the form that each vendor had to complete when initiating a new contract.

These questions were deliberately kept open-ended so that both models and non-models (e.g. rules-based decision trees) that leverage AI and ML could be identified. Once the form has been approved, the MRM team will then reach out to the contract owner to include the tool in the model inventory system and kick-start the process. Such **internal controls** and vendor management help to manage risks and ensure that all new products and services using AI would be vetted thoroughly before deployment.

## COMPLYING WITH RIGOROUS STANDARDS

With responsibility and accuracy being the bank's top priorities in AI model development, HSBC knew it had to instil stringent standards to avoid challenges such as inherent bias for the AI model. The bank hence introduced explicit data standards in its MRM framework.

To meet the standards set out in the MRM framework, HSBC's Model Developers had to demonstrate that their AI models were compliant through detailed documentation. Data collection and management were required to be reasonable in light of the potential use of such data. Additionally, the range of data collected during the loan application process had to be standardised in consultation with the business and risk teams.

After the AI models were developed, the Independent Model Review team within HSBC then validated them by:



Conducting specific data management checks and assessing the data used by the AI model in relation to consistency, completeness, transparency and data accountability. These include requirements covering the use of internal data over external data where possible, data ownership and testing of critical data elements;



Performing several statistical checks (e.g. t-test) to ensure that the data samples are unbiased; and



Performing validation (e.g. k-fold and bootstrap validation) on the output of the model to rule out any bias in the results.



HSBC also maintained close monitoring of the AI model during its deployment to ensure that it worked within the pre-defined parameters. The bank then validated the performance of the AI models on a periodic basis to confirm the relevancy and adequacy of the models.

## Auditability

From a model management perspective, HSBC made assessments in four areas – model design and performance, implementation, governance and documentation. Specifically, documentation of the model had to enable the facilitation of a comprehensive assessment by an internal or external third party. This assessment covered the design and performance of the model prior to and during its actual implementation. To achieve this, the bank developed a Model Documentation Template that standardised the model development process into sequential steps. Model developers also had to provide inputs on the requisite information within this template to help with the auditability of the AI model.

## MANAGING RISKS FOR LOAN APPLICATIONS

Considering its large customer base and the strong regulatory scrutiny that it operates under, HSBC took a generally cautious, **human-in-the-loop approach** for all its AI-based solutions.

An example would be the way HSBC enhanced its loan application process. First, the bank developed an Artificial Neural Network (ANN) scorecard to segregate applicants into different risk tiers. After each loan application submission, the ANN scorecard would generate a score for each application. Using the score, as well as other factors like the measure of current debt to income ratio, credit scores, credit history and severe credit triggers, the HSBC staff would assess the risk of the loan being defaulted and decide whether to approve or decline a loan application. The ANN scorecard helped model owners detect non-linear relationships amongst variables like customers' past behaviour on their assets, deposit balance and credit score. These variables were not easily detectable by traditional linear models and decision trees.

## KEEPING CUSTOMERS IN THE LOOP

HSBC also prides itself on sound customer relationship management. If a customer provided feedback or appealed to HSBC regarding its decision on the loan application, the bank's customer service protocols will kick-start a review by the bank. This review will look into the rationale behind the decision made by the AI model in generating the score for the loan application in question. Relevant feedback would be used to update the AI model, and also improve on the quality of its loan services.

Moving forward, HSBC would also be putting in place notices and reaching out to inform its customers when an AI-enabled solution is used. This would give their customers the assurance that the bank is open and transparent about the way their personal information and applications are handled.

## CONCLUSION

HSBC considers AI governance as an enhancement to its existing governance processes, which have been instilled in every part of HSBC. The bank ensures that AI results can still be overseen by personnel so that their commitment to customer service will not be compromised.

By using AI in an ethical manner, HSBC will be able to improve its operational efficiency and provide a safer, more seamless customer experience while upholding its values of trust, accountability, and fairness.



# MSD: KEEPING EMPLOYEES AT THE HEART OF AI USE

As a leading global biopharmaceutical company with a mission to save and improve lives, MSD creates medicines and vaccines for many of the world's most challenging diseases. MSD views responsible AI as an exciting new frontier to apply and live out the company's commitment to its values. Founded more than a century ago, MSD established its Singapore IT Hub in 2015 to harness digital innovation for better healthcare outcomes.

The IT hub utilises AI techniques to support operations in manufacturing, human health and global services. This helps the company optimise resource allocation, achieve higher productivity and manage talent more effectively. In particular, the IT hub uses AI methods to understand employee engagement and attrition risks for one of MSD's offices.

Given the high level of sensitivity involved in attrition risk assessments, MSD implemented AI governance practices that emphasised on knowledge sharing, having clear roles and responsibilities in AI development and deployment, training and transparency, among others, to ensure that AI use is handled ethically.



## CREATING AWARENESS OF RESPONSIBLE AI THROUGH EMPLOYEE TRAINING

As AI ethics and governance is inherently cross-functional, MSD has put in effort to **raise awareness of responsible AI** amongst various departments like the data science and IT risk management teams, as well as the office of compliance and ethics. Through a series of knowledge sharing sessions, the company raised awareness on the relevance of ethics in developing and deploying AI models. It has also held separate briefings to its senior leaders and used internal platforms such as MSD's global data science symposium to strengthen efforts for awareness.

MSD ensured its HR business partners and business unit executive leader were **trained in interpreting the AI model output and its decisions**. This not only mitigated the risk of inaccurately interpreting the AI model's results, but also brought attention to how the prediction scores generated from the AI model could help list out factors that HR business partners and the business unit executive leader could use to develop programmes to improve the satisfaction of employees.

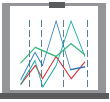
In addition, MSD **created a clear and specialised role** of Data Scientist (AI Ethics and Data Privacy) to look into the ethical considerations of scaling and deploying AI products and predictive algorithms. Besides developing and implementing business processes with ethical principles embedded, the Data Scientist (AI Ethics and Data Privacy) also oversaw the implementation of Explainable AI (XAI) techniques in projects of a more sensitive nature, such as the employee engagement project. With this role, MSD was able to weigh the benefits that the AI systems brought to the company against ethical considerations, and also uplift the maturity and awareness of AI ethics and data protection within the organisation.



Ethical considerations of scaling and deploying AI products and predictive algorithms;



Balancing organisational benefit of an AI system with ethical considerations;



Implementing Explainable AI (XAI) techniques in sensitive projects including the employee engagement project;



Developing and implementing business processes that embed ethical principles into AI projects; and



Raising the maturity and awareness of AI ethics and data protection within the organisation.

## MANAGING SENSITIVITY DELICATELY

In determining the level of human involvement in AI-augmented decision-making, MSD took into account that managing attrition risk was a sensitive subject. The consequence of allowing the algorithm to act on inaccurate predictions could result in the unfair treatment of, for instance, employee benefits.

With this, MSD knew that human oversight needed to take centre stage and adopted a **human-in-the-loop** approach in the AI-augmented decision-making process. This approach allowed the AI model to flag out employees with the highest risk of attrition, while restricting the model from making decisions or taking actions on behalf of the management or HR team.

## TAKING THE RIGHT STEPS FOR FAIR ASSESSMENT

MSD understood that to create trust in AI, quality datasets for model training and the explainability of the model's results were important building blocks. MSD used quality datasets for the employee attrition risk project and had a project team of data scientists, HR personnel and the business unit involved to explain the results of the AI model to its management and HR team. Such steps contributed to establishing a trusted ecosystem within MSD in the capabilities of AI models to solve complex problems.

## DATA QUALITY FOR MODEL DEVELOPMENT

To **ensure data quality** for the project, the project team was made to understand the meaning and origin of each feature. An internal data professional from the business unit was also brought in to work alongside the project team, to share and explain any transformation that was done to the datasets.

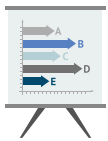
The project team also had measures to determine the appropriate amount of data used in training the model, balancing the need for sufficient training data and to take into account historical changes in business structure (e.g. change in senior leadership or government regulation may affect the relevance of datasets). The project team was also mindful to **keep datasets used for validation and testing separate from those used for training the AI model.**



The project team will understand what each feature meant and the origin of each feature in ensuring dataset relevance;



An internal data professional will act as the custodian of the datasets, sharing and explaining the human actions performed on the datasets (e.g. transforming the data) to the project team;



The project team will balance the need for sufficient training data, while taking into account possible recent changes in the company's business structure (e.g. change in senior leadership or government regulation may affect the usefulness of the datasets) when determining the appropriate amount of data to be used in AI model development; and



Datasets used for validation and testing are kept separate from those used for training during modelling practice.

After preparing the datasets for training of the AI model, the project team was also mindful to exclude features that were demographically rooted, such as gender and ethnicity, to **mitigate bias in the AI model and its results.** Prioritising human-centricity in building the AI model helped the project team to persuade internal stakeholders of the use of AI for the employee attrition risk project.

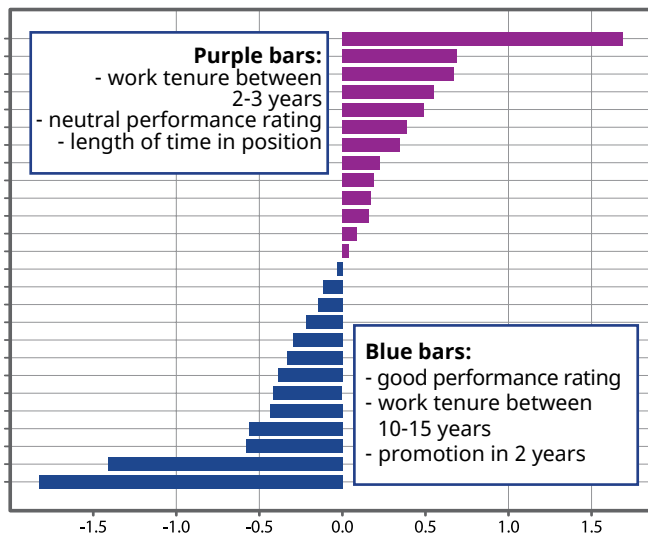
## EXPLAINABILITY

For the AI model's results to be explainable, the project team **implemented explanations for the predictive scores at both model and individual levels.** Explainability at the model level allowed MSD's management and HR team to understand the overall factors for employee attrition, which also contributed to greater trust in the model predictions (Figure 1). Explainability at the individual prediction level, on the other hand, provided insights on the different driving factors behind the same attrition risk score for two individuals, and conversely, the contrasting risk scores for individuals who have similar profiles. This gave the management and HR team a better understanding of each unique individual situation, enabling them to tailor their course of action accordingly (Figure 2).

The project team also used data visualisations to communicate to the management and HR team on how the driving factors behind the individual and model levels could be different. The model was also used to challenge the human prediction done by the business leaders who worked in the division. The differing results between the model and human prediction would spur a valuable data-driven conversation that can point the company in the right direction in terms of employee engagement and retention, as well as reinforce the use of AI within the organisation.



The features represented below are sample features used for illustration purposes and do not represent actual features selected by the AI model.



**Figure 1** Explainability at the model level. Each feature was scored based on its importance in explaining employee attrition at the model level. The purple bars represent features that are positively correlated to employee attrition; the blue bars represent features that are negatively correlated to employee attrition.



**Figure 2** Explainability at the individual level. Both models show employees with the same risk score of 0.66, but with different model features contributing to the risk scores.

## KEEPING EMPLOYEES AT THE CENTRE

A user-friendly AI interface would also provide much clarity for MSD's management and HR team. With that in mind, the project team collaborated with the User Experience (UX) team. Using observation and elicitation techniques rooted in design thinking, the UX team surfaced employee concerns to the HR and management team and kept employee interests at the centre of the solution during the process of developing the AI model.

MSD also controlled the access to the results of the employee attrition model, where only selected individuals in the company's leadership team could view the results and the list of employees with the highest risks of attrition. This helped to reinforce the confidentiality and protection of the data and its employees, while allowing the leadership team to decide on the course of action to promote employee satisfaction.

## CONCLUSION

MSD believes that operating with ethics, integrity and respect for employee rights is critical to its success as a frontrunner of digitalisation. As a global healthcare leader, MSD's work revolves around cultivating strong relationships based on trust. This meant not only listening and learning from its stakeholders, but also communicating openly about the decisions it makes and the outcomes achieved. Only with this, can MSD fulfil its commitment to its values of innovation and scientific excellence.

# NGEE ANN POLYTECHNIC: ADMISSIONS SELECTION MADE EASIER WITH RESPONSIBLE AI

Ngee Ann Polytechnic (NP), an institute of higher learning in Singapore, offers diploma courses to more than 14,000 students.

Every year, NP conducts a dedicated early admissions exercise. The Early Admissions Exercise is an aptitude-based admission exercise that allows students such as graduating Singapore-Cambridge General Certificate of Education Ordinary level (O-level) and Institute of Technical Education (ITE)<sup>1</sup> students to apply for admission prior to receiving their final grades. This exercise gives NP greater flexibility in admitting students based on their aptitudes and interests, allowing a wider range of talents to be recognised.

To automate and enhance the early admissions exercise selection process for three of the polytechnic's schools, namely the School of Business & Accountancy, School of Film & Media Studies and School of Health Sciences, Ngee Ann Polytechnic piloted an AI-powered platform with predictive analytics and a chatbot function in July 2019. Named the Early Admissions Exercise Virtual Assistant (EVA), the polytechnic implemented a framework for the responsible use of AI to carry out their admissions selection in an efficient and fair manner.

## FACILITATING THE EARLY ADMISSIONS EXERCISE

It used to take NP's staff 470 hours to manually review over 4,000 early admissions exercise applications. These applications were received annually, before candidates were shortlisted for face-to-face interviews for the three schools. Each application comprised a 600-character write-up on course-specific attributes and a separate 1,000-character write-up on talents and achievements.

With the launch of EVA, the automated review of the application write-ups was completed in two hours. EVA also conducted online "chats" where students can elaborate on their passions and aptitude for their chosen course. The chat responses were then used to curate questions for the admissions interview. All of this improved administrative efficiency and saved the three schools 135 hours on the shortlisting review, including the time taken for manual review.



135

manual review hours saved

<sup>1</sup> ITE is a public vocational education institution in Singapore that provides pre-employment training to secondary school graduates, and continuing education and training to working adults.

## OVERSEEING THE RESPONSIBLE DEPLOYMENT OF AI

The use of innovative technologies such as AI to improve on admissions selections is in line with Ngee Ann Polytechnic’s vision to be a “future-ready campus”. The polytechnic also recognised the importance of responsible AI use and made efforts to ensure that its AI governance processes were aligned with the PDPC’s Model AI Governance Framework before deployment.

To ensure robust oversight of AI deployment, the polytechnic implemented the following **internal governance structures and measures**:



### *Final approval of deployment of all AI technologies*

by the management of NP, which is chaired by the polytechnic’s Principal and consists of its Deputy Principals and Directors from various schools.



### *Oversight and deployment of AI*

by the Academic Affairs Office. As the custodian for the early admissions exercise datasets, this Office acts as the central coordinating office that engages the AI solution provider, Impress.AI, and the three schools to ensure the relevance and effectiveness of EVA.



### *Review of non-selected applications*

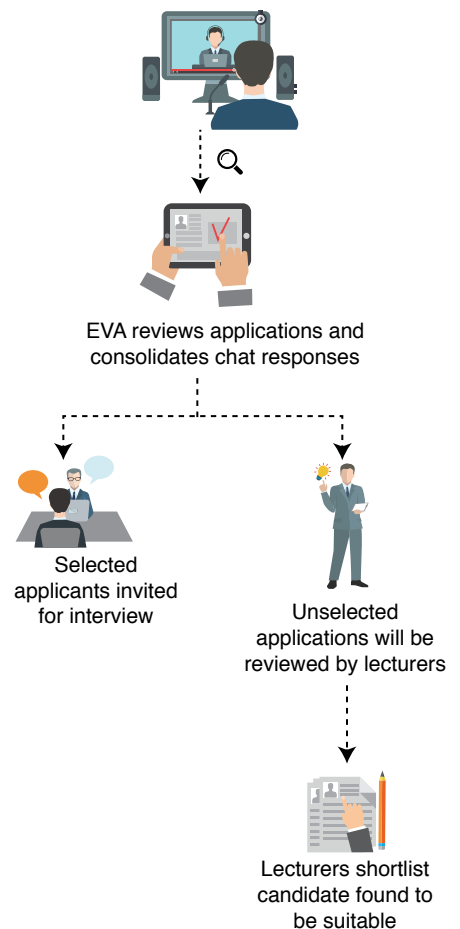
by the three respective schools to ensure that all deserving students have an opportunity to be shortlisted for the admission interview.

## ENSURING FAIR ASSESSMENT FOR ALL APPLICANTS

NP understood that the early admissions exercise had a direct impact on students, especially for those not selected. This consideration played a huge part in the polytechnic’s decision to not use EVA as a complete replacement of manual reviews. Instead, the polytechnic adopted a **human-over-the-loop approach** to determine the applicants to be invited for face-to-face interviews.

With this approach, EVA reviews the applications and selects applicants for the face-to-face interviews. The responses from the online chat will enhance NP’s engagement with the applicants during the interview.

For applicants who were not selected, lecturers from the respective schools would review their write-ups to ensure that no deserving applicants were missed out. They would also gather insights on the applicants from the online chats, such as their personalities, interests as well as their strengths in non-academic areas like leadership and teamwork, for a holistic assessment of the applicants’ suitability. If the applicants meet Ngee Ann Polytechnic’s shortlisting criteria, they would then be invited for the interview despite not being selected by EVA.



## DEVELOPING EVA FOR RESPONSIBLE DEPLOYMENT

For the **data used to develop** EVA, NP mimicked the manual review by using the same text-based, 600-character and 1,000-character write-ups. The polytechnic ensured **data quality** by using only information from the students' applications from the recent three years. This also gave the polytechnic **full clarity on the lineage of the data**.

To enhance the reliability and robustness of EVA, NP used **different datasets for training, testing and validating** EVA during model development. With the majority of the applicants for the early admissions exercise being O-level students, the polytechnic used write-ups from O-level students from 2016 to 2018 as training data for the prediction model. Separate write-ups were used as testing data to determine the model's accuracy. To validate the prediction model, the polytechnic used write-ups from ITE students from the early admissions exercise in 2018.

In using write-ups from O-level and ITE students, NP was able to minimise selection bias because they used data that was **representative of the students who participated in early admissions exercise to develop the AI model**.

Unintended bias was also a risk that the polytechnic wanted to avoid. With this in mind, personal data that was deemed irrelevant in the early admissions exercise, such as names and email addresses, were not used or analysed by the AI model.

In the early development stages of EVA, NP engaged Impress.AI to conduct a deep learning-based analysis of the write-ups gathered over the past three early admissions exercises, to gain insights on how candidates were shortlisted for various courses.

The next step was selecting an appropriate AI model. In order to do so, NP experimented with and assessed different iterations of Natural Language Processing algorithms (e.g. using bag of words for encoding with the Naive Bayes classifier<sup>2</sup> and other classification models). This is coupled with efforts from Impress.AI to explain how the AI model functions to the polytechnic. As a result, the Bidirectional Encoder Representations from Transformers (BERT)<sup>3</sup> Natural Language Processing with Neural Network classifier was adopted as the prediction model. Instead of simply checking for keywords, this prediction model had an in-depth understanding of the semantics and meanings of words, even recognising similarities between words for classification.

With its current accuracy at 76%, NP intends to **review and update the prediction model with new data** obtained from each early admissions exercise to improve on the model's accuracy and ensure relevance and quality of the datasets.

---

<sup>2</sup> A Naive Bayes classifier is a probabilistic machine learning model that is used to classify/discriminate different objects based on certain features.  
<sup>3</sup> BERT is the first fine-tuning based representation model that encodes the context as well in the representation of the word. It works by going over the data in both directions and using a "Recurrent Neural Network" and "Convolutional Neural Network" to encode the context of the word in its representation.



## BEING TRANSPARENT ABOUT THE USE OF EVA WITH APPLICANTS

NP believed that trust was critical in the successful launch and subsequent adoption of EVA, and that transparency in the use of AI was crucial to building this trust. With this, the polytechnic took steps to notify all applicants with course choices from any of the three schools that EVA would be reviewing their write-ups.

In addition, the polytechnic developed a policy for explanation and even shared the use of EVA through a media release. In its email to applicants, the polytechnic further included a video to explain how EVA would support the early admissions exercise selection process and provided instructions on how applicants could chat with EVA.

At the end of the applicants' online chat with EVA, they were also invited to provide feedback and experience ratings. The feedback showed that 92% of the applicants were satisfied with EVA. While user experience for the applicants had been positive so far, NP will be carrying out continual reviews and evaluations to improve on EVA for the next early admissions exercise.



### How NP demonstrates transparency in AI use



Policy explanation and media release on use of EVA



Notifying applicants on the use of EVA in write-up review



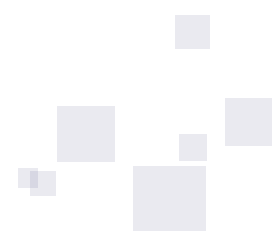
Video on how to use EVA



Feedback and ratings from students

## CONCLUSION

It is exciting times for Ngee Ann Polytechnic. Harnessing deep-technology solutions like AI has helped the polytechnic to automate manual processes and enhance student engagement. The potential of scalability in improving admissions selection and reducing administrative workload was apparent through the use of EVA. Through the pilot launch, the polytechnic also realised the benefits of adopting PDPC's Model AI Governance Framework in making EVA a reliable and effective AI model, using checks and balances to test the rigour of the technology. The polytechnic will continue to invest in such efforts to align with its vision to be a future-ready campus.



# OMADA HEALTH: ACCOUNTABLE AI IN DIGITAL HEALTHCARE

As a healthcare provider based in San Francisco, Omada Health (Omada) helps people tackle obesity-related chronic diseases, such as type 2 diabetes, prediabetes, hypertension, anxiety and even depression, with its digital care programme. Enabled through an AI platform and curriculum, Omada worked with employers, health insurance organisations and health systems to successfully roll out this programme to 300,000 participants to date.

Assigning a professional health coach to each participant, the programme analyses participants' user-provided health data like weight, blood pressure, glucose level, food intake and physical activity. It then leverages machine learning capabilities to empower the health coaches to provide personalised guidance on lifestyle changes to its participants. The guidance can take the form of meal planning or tips on integrating exercise into daily activities. Omada makes sure that the guidance is not just a one-way traffic, using the platform to further analyse the participants' responses to different tones, instructions, or other guidance before recommending appropriate coach-to-participant messages. Such interactions help participants achieve sustained weight loss while optimising the time coaches spent with each participant.

Well aware of the risks that AI model development and deployment could bring, Omada was proactive in fostering accountability within the company. It involved all employees in its governance practices, and involved its health coaches in deciding on the messages to send to participants, drawing a delicate balance between the level of human involvement and reliance on the programme's machine learning abilities. Omada's emphasis on data quality and risk management was also apparent in the governance practices implemented at every stage of the AI model development. Once the model was deployed, Omada then took steps to be transparent in the use of AI to both its health coaches and participants. These AI governance practices illustrated measures that were recommended in the Model AI Governance Framework.



## INVOLVING ALL TO MANAGE RISKS

To oversee AI model development and deployment for the digital care programme, Omada put in place internal governance structures and measures to identify and address potential risks of the AI model. For one, the company established an Omada Risk Committee (ORC) to oversee the company's risk management activities, including those within the digital care programme. This ORC, comprising the Security Officer, Privacy Officer, Compliance Officer and General Counsel, meets every quarter to review the top risks and remediation activities within the company.

Omada further established an AI/Data Governance group. Headed by the company's Director of Data Science and primarily accountable for the digital care programme, this group oversees Omada's AI governance practices. The group is also responsible for managing data as a strategic asset for the company, developing and reviewing governance processes and policies to mitigate potential risks in the company's data use.

With top management support for AI governance, Omada had its technical team of data scientists and medical affairs staff work hand-in-hand with non-technical teams like legal and compliance during the development of its AI model. This was valuable in addressing ethical and governance issues as the interactions spurred open discourse, provided diverse perspectives and a well-rounded representation of expertise. More importantly, the synergy ensured alignment of best practices in the use of templates, evaluation criteria and code reviews.

Not forgetting that its people make up an essential component to a successful programme, Omada pushed out a value-based reimbursement system that tied revenue to longitudinal outcomes like participant weight loss. The system drives employees to innovate within the boundaries of AI governance practices and regulations, motivating them to deliver the best possible health outcomes for participants.

Omada also includes questions on the ethical use of AI during job interviews and conducts mandatory training on security, data protection and compliance for all employees twice yearly. This helps align and update both the potential hires and existing employees on the company's corporate values, mission and AI governance practices. Omada goes the extra mile to hold separate training sessions for its sales employees so that they are able to adequately explain to external stakeholders how Omada uses and manages customers' data in its AI and machine learning systems.

## A PERSONAL APPROACH FOR BETTER UNDERSTANDING

For the digital care programme to be successful, Omada's health coaches need to be able to relate and empathise with participants. To achieve this, Omada decided to adopt a **human-in-the-loop** approach for the programme's machine learning capabilities to recommend appropriate messages to the health coaches. The health coaches will then use what they know about their participants to review the personalised messages before sending them.



## DATA ASSURANCE AT EVERY STAGE

While seen as a healthcare provider for participants, Omada is categorised as a business associate when it comes to engaging hospitals under the Health Insurance Portability and Accountability Act (HIPAA). This means that their standards for safeguarding medical information would have to abide by HIPAA's data protection and security provisions.

To manage the risks that come with AI development, prior to the development of any AI model, Omada conducted a **risk impact assessment** to gauge the impact of the programme, probability of negative health outcomes, and at the same time, review its internal policies and controls. The relevant teams also used roadmaps and code review templates (see Figure 1) to adequately address the AI ethical and governance issues at each stage of the model development (see Figure 2). If risks were identified, the teams will make mitigation or remediation recommendations to the ORC. Tackling risks at an early stage provided clarity and helped to **minimise any potential data biases** that might arise in the subsequent stages of development.

### Code Review

### Roadmap

### Projects

Project Overview	
1	Define problem
2	Define stakeholders, resourcing plan (RACI)
3	Define evaluation plan
4	Project summary
5	Create project slack channel
6	Kickoff plan deck
7	Plan Complete - Approved by Engineering Leadership
Solution Discovery	
8	Discovery kickoff
9	Research solution
10	Research writeup
11	Requirement definition
12	Request for comment
13	Project Planning & Scope Finalization

### Notebook Templates

Figure 1: Examples of Omada's templates and notebooks

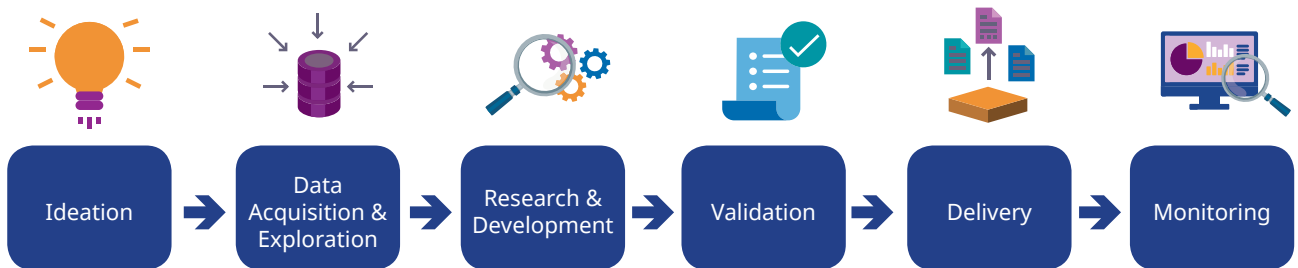


Figure 2: Omada's AI Development Lifecycle

In **ensuring data quality** for model development, Omada made sure to obtain consent before collection of any participant data. Data was also only gathered from reliable sources such as Omada's 3G cellular-enabled weighing scale, as well as recognised third-party devices such as GoogleFit, Garmin and Fitbit. For these third-party sources, the company also designed and deployed a data collection pipeline to control the quality of data.

Measurable improvements in participants' health and their continued trust in the digital care programme marked the ultimate goals for Omada. In that vein, Omada conducted rigorous tests and analyses on their AI models, using a build-and-deploy process that included automated code testing, such as unit testing and static analysis. These tests made the individual units of source code fit for purpose and highlighted possible vulnerabilities in the code before running an AI model. Omada also implemented certain measures to strengthen the rigour of their AI models:



Maintaining a logical separation between the development and production environments;



Conducting annual penetration tests with an independent third-party security firm. Any vulnerabilities found will be documented and immediately remediated; and



Performing post-mortem analyses to identify root causes and implementing future controls.

Implementing these practices not only ensured the reliability of data, but also helped Omada manage risks methodologically and implement security measures at each stage of AI model development. These practices formed Omada's approach to building the participants and health coaches' trust in its programme.

## PUTTING USERS AT THE CENTRE OF DESIGN

A smooth user experience for its health coaches was also crucial for increased programme adoption. After the deployment of the AI model, data scientists responsible for the development of the digital care programme created and enhanced the user experience of the programme as well. Alongside designers, the team conducted qualitative reviews of the programme with the health coaches, explaining to them the workings of the AI model to enable each message recommendation. This gave the health coaches confidence in using the programme, which in turn led to higher adoption rates.

On the participant front, Omada published a whitepaper, "Omada Health's Approach to Security", to help participants understand more about the digital care programme. Written in a clear and concise manner, the paper shed light on the programme's technical and security features. The company also published a blogpost sharing how machine learning was used in suggesting messages to health coaches when offering advice to their participants. It also provided a general disclosure on its website on how its participants' data was collected and used. If participants found any inaccuracies in their personal information, they can send their requests to Omada to correct it.

Data scientists, designers  
and health coaches



General  
disclosure  
on website



Qualitative  
review



Whitepaper and  
blogpost

## CONCLUSION

Recognising the value of AI governance in promoting customer trust and improved health outcomes, Omada put in place various measures, allowing their data scientists, AI managers and employees room to innovate while avoiding costly mistakes. This accelerated Omada's AI development and deployment process. The eventual result of increased efficiency and a more secure deployment of AI models boosted the healthcare provider's reputation as a responsible steward of its participants' health information.

# UCARE.AI: ACCOUNTABLE AI FOR ACCURATE HEALTHCARE COSTS

UCARE.AI is a Singapore-based start-up that specialises in providing predictive insights with its online AI and machine learning platform. Among the various solutions the platform provides, UCARE.AI's AI-powered Cost Predictor works with hospitals to deliver accurate estimations of hospital bills to patients.

One of these hospitals was Parkway Pantai (Parkway). Prior to deploying UCARE.AI's Cost Predictor, Parkway used traditional statistical methods to provide bill estimates. The statistical models generated were expensive to update and therefore not refreshed frequently, exacerbating error rates. To tackle this, Parkway deployed the Cost Predictor in all four Singapore hospitals in November 2018 and saw significant improvements. Since deployment, there have been no customer complaints and the Cost Predictor has achieved an average aggregate accuracy of 82%.

Armed with the confidence of the Cost Predictor's high accuracy, Parkway launched the Price Guarantee Programme for six hospital procedures, namely the removal of piles, breast lumps, ovarian cysts, gallbladder, thyroid and tonsils. The Programme checked and confirmed charges for these procedures against the Cost Predictor's price estimates, verifying the accuracy of the Cost Predictor. The hospital guaranteed that patients will be charged according to the initial price quoted by the Cost Predictor, regardless of whether additional treatments were included later on. Before or during the day of admission, financial counsellors also worked with patients to review their estimated medical bill. These sessions made sure that the patient had a clear estimate of the eventual medical bill, helping them make ample preparations for finances.

In its commitment to help patients make well-informed decisions with accurate cost estimations, UCARE.AI understood that trust was essential in driving adoption of its AI solutions. To achieve this, the company turned to the Model AI Governance Framework, aligning its practices in AI governance to those in the Framework to ensure reliability in its AI solutions. Besides assigning clear roles for ethical AI development and deployment, UCARE.AI concentrated efforts in good data accountability practices and treated the use of AI with openness and transparency. This provided tremendous benefits to patients in terms of seamless experiences in hospitals, greater certainty over their medical expenses and less re-financial counselling.



## ASSIGNING CLEAR ROLES FOR AI OVERSIGHT

A critical part of AI governance is the need for oversight of the company's use of data and AI. For this, UCARE.AI put in place certain internal governance measures for its company and client projects. One of which involved **assigning clear roles and responsibilities for the ethical development and deployment of AI.**

The approach UCARE.AI took was to have all projects include primary and secondary data science leads to concurrently develop AI models for the same problem statement. Once completed, the data science leads would then present their results to UCARE.AI's internal team, which consists of the Chief Executive Officer, Chief Technology Officer, Chief Security Officer, project managers and the client services team for validation. During the course of the project, UCARE.AI also conducted weekly check-ins with its clients to ensure quicker and more reliable iterations of its AI models. A final step before submission of the models to the client was to have UCARE.AI's appointed medical advisors assess the models' outputs for accuracy.

After the models and its results have been submitted to the client for blind testing and approval, UCARE.AI's Quality Assurance team would then be brought in to review and ensure that the model was production-ready before deployment.



Primary and secondary data science leads develop AI model for the same problem statement

Results are presented to the internal team for validation, with weekly check-ins with clients for quicker and reliable iterations of the AI models

Appointed medical advisors assess the models' outputs for accuracy before submission of the models and results to clients for blind testing and approval

Quality Assurance team reviews and makes sure AI model is production ready

Deployment of AI model

## MINIMISING RISKS WITH ROBUST VALIDATION FRAMEWORKS AND FEEDBACK

UCARE.AI also conducted rigorous feasibility studies before developing the Cost Predictor. These studies helped address potential risks such as reduced accuracy in forecasted healthcare costs. With the studies, UCARE.AI then worked with its clients to create a validation framework to strengthen the AI model's accuracy, making sure to obtain patients' feedback on the framework for further fine-tuning. The Cost Predictor's AI model then underwent User Acceptance Testing, where the end business users from each hospital were invited to test the solution and provide feedback on various predictions.

## ENSURING SAFEGUARDS ARE IN PLACE

Accountability in its data management practices saw UCARE.AI taking proactive measures for data safeguards to ensure the Cost Predictor's functionality and effectiveness after deployment.

As a first step, when handling personal data for AI model development, UCARE.AI adhered to the requirements of various personal data protection laws and draft bills in its operating regions. Singapore's Personal Data Protection Act (2012) was one such law UCARE.AI kept in mind. Besides obtaining consent prior to any collection and use of personal data, UCARE.AI also made efforts to securely encrypt sensitive data. Its connectors- software components that can extract and transform original data sources into standardised formats - were also designed to automatically detect such sensitive data and where possible, the algorithm was trained to minimise the use of this data in developing the AI model.

To further boost efforts in data protection, UCARE.AI anonymised client data at source before using it for development, thereby minimising the risk of inappropriate access to personal data. This also ensured that in the unlikely event of a breach, personal information could not be easily used to trace back to an individual.



Obtaining consent before collection and use of personal data

Encrypting sensitive data and minimising use of this data during AI model development

Anonymising client data at source

Understanding the lineage of data was also central in the accountable use of AI. Knowing this, UCARE.AI logged data consistently across multiple components and collected data in a secure and centralised log storage. In ensuring data quality, the company was also careful to transform its data into a usable format so that the properly formatted data could be used to build AI models. The company also prioritised creating AI models that were unique to clients, obtaining reliable datasets from the client to build models instead of using third-party datasets. Such a practice provided distinctions between patients' profiles, and the eventual features selected for each AI model differed for each hospital, contributing to greater accuracy in the bill estimations for patients that visited the hospitals.

Another pertinent part of AI model development was **minimising the risk of bias**. For this, the objective and consistent machine predictions gave patients customised, data-driven predictions of their hospital bills instead of those subjected to human biases in algorithm development.

After the deployment of the Cost Predictor, UCARE.AI **continuously monitored and iterated the algorithm**, improving the data and simplifying the process for better accuracy. This continual training of the AI models ensured that the algorithms remained up-to-date and functioned with more precision after each data input. More importantly, the methodology of continuous validation of the AI models with client inputs helped to boost confidence in the accuracy of the platform's predictive insights.



## TRANSPARENT IN THE USE OF AI AND DATA

To build greater confidence and trust in the use of AI, UCARE.AI was mindful to be transparent in its use of AI with various stakeholders. UCARE.AI not only disclosed the exact parameters used in developing the AI model to its clients, but also provided detailed explanations on all algorithms that had any foreseeable impact on operations, revenue or customer base. Understanding that the accuracy of bill projection is highly regarded by hospitals and patients, UCARE.AI made a conscious decision to **declare the use of AI** in its analysis and prediction of bill amounts to Parkway's data managers and its patients.

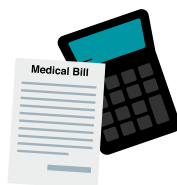
The company also actively reinforced its commitment to data protection, painstakingly cataloguing and evaluating every use of data that could be accessed by clients. Clients with concerns about bill predictions were also encouraged to highlight them through UCARE.AI's communication channels. For instance, Parkway's admission staff can easily provide feedback on bill predictions to UCARE.AI via its business owners and IT departments. The feedback would then be forwarded to UCARE.AI for review. This gave clients and external auditors the necessary assurance on UCARE.AI's policies and processes for responsible AI use.



Disclosing exact parameters to its clients



Providing detailed explanations on algorithms with foreseeable impact on operations, revenue or customer base



Declaring the use of AI in its analysis and prediction of bill amounts

## CONCLUSION

As a company that employs heavy use of personal data for AI model development, UCARE.AI is vigilant and committed to data protection. This is especially important, given that the nature of its work is in healthcare and the call for ethical and responsible use of data is paramount.

Educating clients on the importance of implementing the Model AI Governance Framework so that patients are given the assurance that their data is safe remains one of UCARE.AI's top priorities. With the company's well-tested approach in handling personal and sensitive data, UCARE.AI was able to demonstrate its experience in this field and gain the confidence of its clients. The shared professional trust and respect between UCARE.AI and its clients in turn helped to build the recognition of the company as a reliable and trusted partner in data management and developer of AI models.



# VISA ASIA PACIFIC: FORECASTING CARDS USED FOR TRAVEL WITH ETHICAL AI GOVERNANCE

For Visa Asia Pacific (Visa), responsible use of AI within its products and services is of utmost importance, given the sensitive nature of digital payments. Visa is a leading global payments technology company digitally connecting consumers, businesses, banks and governments. Visa leverages data analytics and machine learning algorithm tools to provide services to its clients.

One of these tools is Travel Predict, a recommendation engine that leverages past transactional behaviour to help Visa's issuing banks forecast the credit and debit cards more likely to be used for travel. Visa's issuing banks will take into consideration Travel Predict's forecast, together with other factors like the success of their past card promotions, to provide benefits to cardholders who are the best candidates for travel related marketing.

Mindful that trust plays a significant role in the take-up of its tools and promotions, Visa has been proactive in building governance practices into its AI processes to ensure continued client success while complying with data protection laws.

From setting up clear governance structures, to rolling out accountability measures in model development and even engaging stakeholders for better understanding of AI decision-making, Visa concentrates its efforts on developing sound policies and practices to fully realise the benefits of technologies and deliver better services to its clients.



## A STREAMLINED GOVERNANCE PROCESS

Once the company set its sights on the adoption of AI-enabled tools like Travel Predict, one of the first steps Visa took was to put in place a **comprehensive governance structure** that reviewed:



The use of data to ensure adherence to ethical data principles



Protection and management of personal data and associated risks



Model risk management processes to audit and oversee all AI solutions and evaluate the materiality of risk

Visa is continuously working to put in place comprehensive governance structures and encourages accountability in the end to end development and deployment of its AI models.

## ENGAGING EMPLOYEES TO BUILD A RESPONSIBLE AI CULTURE

To ensure a fully integrated and accountable AI programme, Visa promotes ethical data and AI practices amongst its relevant employees, holding **regular training** on compliance, legal requirements and key controls.

Visa also designed a **training curriculum** for its data scientists, covering topics such as data protection, ethical use of data and responsible AI. The relevant employees dealing with AI systems are also required to undergo assessments for specialised training and knowledge retention.

Such efforts pave the way for the relevant employees to play a part in responsible AI use within Visa. Through manpower training, Visa is able to approach AI governance more holistically.

## STEPPING IN WHEN NECESSARY

Travel Predict generates a score on the propensity to travel for each Visa card through the use of past transaction behaviour. Visa then provides the scores to issuing banks, which use these scores in determining the best strategy for engaging their cardholders.

During model development for Travel Predict, Visa tracks accuracy and quality related metrics at an aggregate level during the AI model selection, training and validation phase. The issuing bank, then performs the final filtering of cards and assessment of which cardholders would receive the offer.

With Visa as the model provider, and the bank as the model deployer, this **human-over-the-loop** approach allows Visa the flexibility to provide a certain degree of human intervention when needed in AI decision-making. In taking this approach, Visa had three considerations:



Materiality of the AI solution on the issuing bank’s cardholders



Impact on the bank’s marketing campaign



Operational feasibility. With millions of cards in use worldwide, it is not operationally feasible for Visa to manually review the forecast for each of them

To address these considerations, Travel Predict’s AI method assists Visa in providing issuing banks the scoring recommendations at scale to maximise the success of their campaign.

## Ensuring Explainability

Another accountability practice that Visa applies is to follow a **Model Risk Management process** to assess materiality of AI solutions for clients. By **documenting** the technical standards, data inputs, model explanation and interpretation, methodology, fairness and quality/accuracy questions, this helps Visa to explain the development of Travel Predict to its internal stakeholders such as its internal audit teams should the need arise.

Visa also shares the top predictor scores and the AI model's key performance indicators on accuracy and precision with issuing banks to **explain how Travel Predict functions and arrives at the cards' travel propensity score**. This gives the necessary assurance to clients on the reliability of Travel Predict as a valuable tool for effective marketing.

## PROACTIVE ENGAGEMENT

As part of its efforts to develop a trusted relationship with its issuing banks, Visa also provides them ample support. In the case of Travel Predict, Visa **discloses the general AI methodology** to issuing banks. Visa also provides relevant documentation to explain Travel Predict's recommendations, so that the technical teams within the issuing banks can **explain the workings of the AI model to their respective stakeholders**.



Disclosing the general AI methodology



Providing presentations to explain Travel Predict's recommendations

## CONCLUSION

Developing a robust AI governance structure has given Visa the opportunity to demonstrate its commitment to openness and transparency and raise its standards when deploying AI solutions. The company's proactive, responsible approach to embedding principles into its AI governance processes and consistent alignment of procedures has promoted public confidence and trust in the company. This will bring closer its goal of enabling a trusted and responsible data sharing and AI ecosystem for all.



## #SGDIGITAL

Singapore Digital (SG:D) gives Singapore's digitalisation efforts a face, identifying our digital programmes and initiatives with one set of visuals, and speaking to our local and international audiences in the same language.

The SG:D logo is made up of rounded fonts that evolve from the expressive dot that is red. SG stands for Singapore and :D refers to our digital economy. The :D smiley face icon also signifies the optimism of Singaporeans moving into a digital economy. As we progress into the digital economy, it's all about the people - empathy and assurance will be at the heart of all that we do.

BROUGHT TO YOU BY



Copyright 2020 – Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission Singapore (PDPC)

This publication is intended to foster responsible development and adoption of Artificial Intelligence. The contents herein are not intended to be an authoritative statement of the law or a substitute for legal or other professional advice. The IMDA, PDPC and their members, officers and employees shall not be responsible for any inaccuracy, error or omission in this publication or liable for any damage or loss of any kind as a result of any use of or reliance on this publication.

The contents of this publication are protected by copyright, trademark or other forms of proprietary rights and may not be reproduced, republished or transmitted in any form or by any means, in whole or in part, without written permission.