

Volume 2

COMPENDIUM OF USE CASES

Practical Illustrations of the
Model AI Governance Framework



CONTENTS

- 3** Summary of the Model AI Governance Framework
- 5** Introduction
- 6** **AI Singapore**
 - 100 Experiments x Model AI Governance Framework
 - 10 IBM Manufacturing Solutions Pte Ltd**
Ensuring Product is Ready for Market Release with AI
 - 11 RenalTeam**
Leveraging AI to Provide Better Care for Dialysis Patients
 - 13 Sompo Holdings Asia**
Challenge Accepted — Implementing a Fraud Detection Solution Responsibly
 - 15 VersaFleet™**
Dynamic Route Solver to Optimise Business Efficiency
- 16 Darwin**
A Responsible city in CCTV Data Analytics
- 21 Google**
Celebrity Recognition with Governance in Place
- 25 Microsoft**
Ways to Implement Trustworthy Conversational AI
- 31 TAIGER**
Winning Clients with AI Governance Practices



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.

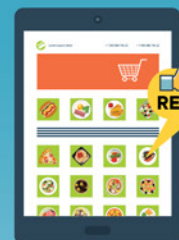
Human-in-the-loop

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

As part of Singapore's efforts to help organisations deploy AI responsibly, Singapore has released:

- Second Edition of the Model AI Governance Framework (Model Framework)
- Implementation and Self-Assessment Guide for Organisations, co-developed with the World Economic Forum Centre for the Fourth Industrial Revolution
- Volume 1: Compendium of Use Cases

Specifically, the 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 responsible use of AI. By implementing responsible AI governance practices, organisations can distinguish themselves and show that they care about building trust with their stakeholders. This will create a virtuous cycle of trust and enable organisations to continue to innovate for their customers.

In January 2020, Singapore released Volume 1: Compendium of Use Cases featuring Callsign, DBS Bank, HSBC, MSD, Ngee Ann Polytechnic, Omada Health, UCARE AI and Visa Asia Pacific.

More organisations have since come forward to share their AI governance practices with us. Volume 2: Compendium of Use Cases will feature City of Darwin (Australia), Google, Microsoft and TAIGER as well as a special section on AI Singapore's collaboration with its industry partners — IBM, RenalTeam, Sampo Holdings Asia and VersaFleet.

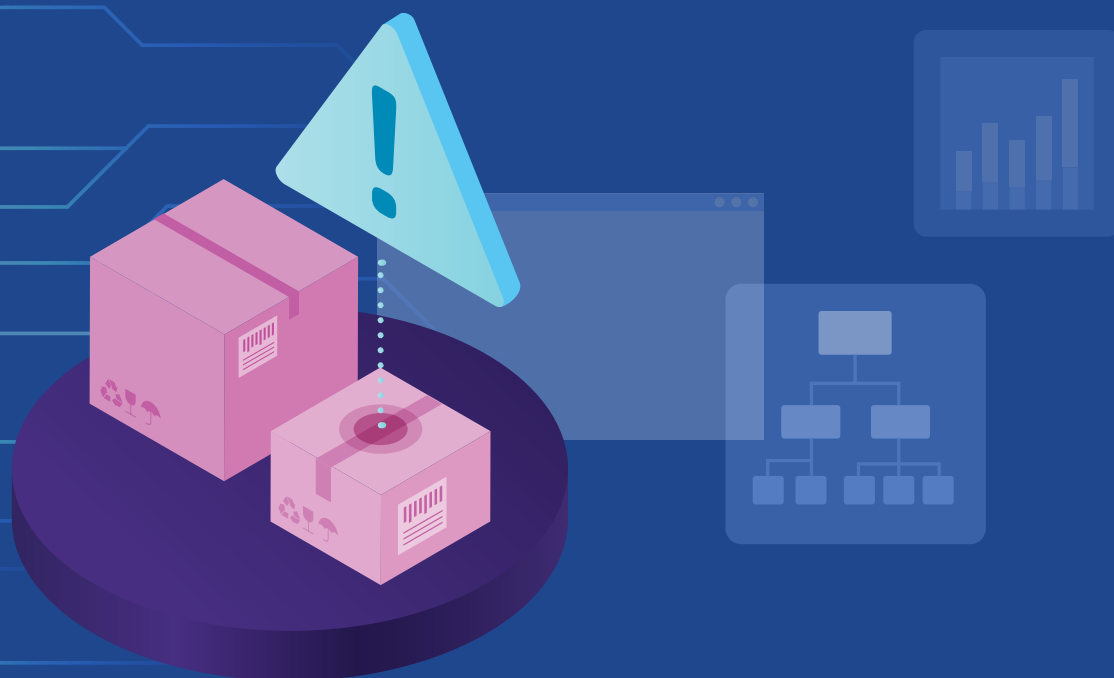
We hope that these use cases will inspire more organisations to embark on a similar journey. Here is Volume 2.

AI SINGAPORE

100 Experiments x Model AI Governance Framework

AI Singapore is a national AI programme to anchor national capabilities in AI to create social and economic impacts and build an AI ecosystem in Singapore. Launched in October 2017, “100 Experiments” (**100E**) is AI Singapore’s flagship programme to co-create AI solutions with industries. This programme allows organisations to propose real-world problems they face, where no suitable commercial off-the-shelf AI solution exists. The programme then matches the organisations with AI Singapore’s in-house researchers to build AI solutions within 9 to 18 months to address these problems. **AI Singapore also uses the Model AI Governance Framework to ensure that AI solutions are developed and used responsibly.**

To date, AI Singapore has engaged over 260 companies to understand their business needs and problems and has started over 50 projects under 100E with companies, focusing on healthcare, finance, fast-moving consumer goods (**FMCG**) and manufacturing, as well as the government sector.



Internal Governance Structures and Measures

All project proposals to develop AI solutions under 100E will undergo a multi-stage assessment process before the Director of AI Innovation tables the project for voting and approval by the management committee:

Stage 1

AI Singapore's engineering team and business development teams will jointly assess the business feasibility of the AI solution (i.e., the social or economic value of using AI over non-AI solutions such as simple data analytics or software redesign).

Stage 2

AI Singapore's engineering team will check on the data-readiness to ensure that the quality and quantity of the datasets can be meaningfully used for model development.

Stage 3

AI Singapore's engineering team and industry partners will prepare a joint technical proposal to assess whether the AI solutions are within bounds of AI Singapore's internal AI Governance and ethics protocols. These protocols are closely aligned to the Model AI Governance Framework and other academic sources. Reviewed by the engineering team every six months to ensure that they remain up to date, these protocols include data governance, integrity of AI models and risks of unlawful or unethical use of AI models. In addition, AI Singapore has developed a checklist used by 100E teams, referencing materials such as the Model AI Governance Framework and Implementation and Self-Assessment Guide for Organisations (**ISAGO**).

Stage 4

Given the increasing complexity of AI solutions, AI Singapore has established a 100E engineering approval committee with oversight of AI governance. The approval committee comprises the Director of AI Innovation and the five engineering department heads – 100E, Makerspace, AI Engineering, Data Engineering and Platform Engineering. The committee meets weekly and reviews all project applications.

A project that successfully passes through the assessment will then be voted and approved by AI Singapore's multi-stakeholder management committee, which is chaired by the Executive Chairman of AI Singapore and comprises the appointed Lead Principal Investigator (**PI**) and other members such as directors of the research and technology groups pillars.

Determining the Level of Human Involvement in AI-Augmented Decision-Making

As AI Singapore's industry partners will contribute half of the resources required for the development of the specific AI product or solution, both parties will jointly determine the extent of human involvement required in the AI-augmented decision-making process, taking into consideration the impact of the AI product or solution on the end users.

Operations Management

AI Singapore took guidance from the **Model AI Governance Framework** to establish a tactical set of protocols to define best practices on data and AI models. This step has demonstrably given AI Singapore more credibility among its industry partners.

For every AI solution developed, AI Singapore will use datasets from its industry partner to train, test and validate the AI model. In addition, AI Singapore will engage its industry partner to ensure that the datasets provided and used are pre-processed to be as accurate, complete, relevant and interpretable as possible. For example, AI Singapore will validate the datasets for accuracy and completeness of input data and labels prior to designing the AI solution. If there are limitations in the industry partner's dataset, AI Singapore will require its industry partner to state the limitation in the application of the AI model upfront.

As a consultant-vendor, AI Singapore makes efforts to ensure that personal data used in the development is protected. If data on personal attributes are required to be used by the AI model, the engineering team will perform appropriate encryption and ensure that no personally identifiable information is transmitted during any data transfer to their industry partners.

To ensure proper data lineage and accountability of the various AI solutions, AI Singapore uses Gitlab to track and manage the various stages of data transformations and modelling. In addition, AI Singapore documented the steps in the data transformation and AI modelling process as part of good governance. During modelling, AI Singapore uses stratified random sampling techniques to split the dataset into balanced and independent training and test sets, ensuring fair testing and no data leakage.

In developing the AI models, AI Singapore uses open source algorithms that have been tested in similar contexts to avoid black box solutions where they are unable to explain how the AI model functions or arrives at a certain prediction. AI Singapore will also validate the performance of AI models against a separate test and validation dataset.

Depending on the nature of the algorithm(s) used for the specific AI solutions (e.g., supervised, unsupervised or semi-supervised), AI Singapore will proactively advise its industry partner on how well the inferences made by the AI model can be explained.

Stakeholder Interaction and Communication

AI Singapore will organise regular updates and technical discussions with its industry partners at every stage of the development lifecycle. During these meetings, AI Singapore's engineering team will share how features of the AI solution are being developed and how the application can be deployed. This helps its industry partners better understand the benefits and limitations of the product, as well as how to use and deploy it in production. This also helps them to better manage product release into the markets with their customers.

Selected Use Cases

The 100E programme comprises exploratory projects between AI Singapore and its industry partners to develop the best-fit AI model. The following use cases focus on how AI Singapore developed AI solutions for IBM, RenalTeam, Sampo Holdings Asia and VersaFleet™ with measures that are aligned with the Model AI Governance Framework. These use cases also illustrate how the AI models developed exceeded the companies' expectations and were eventually adapted or deployed by the companies.

Using the Model AI Governance Framework in 100 Experiments



Internal Governance Structures & Measures

- All projects undergo a multi-stage assessment process
- Develop a checklist based on the Model AI Governance Framework and ISAGO



Human Involvement in AI-Augmented Decision-Making

- Varies for different AI applications based on impact of product or solution on end users



Operations Management

- Establish a set of protocols to define best practices on data and AI models
- Pre-process datasets to be as accurate, complete, relevant and interpretable as possible
- Ensure personal data used is protected
- Ensure proper data lineage and accountability of AI solutions
- Validate performance of AI models



Stakeholder Interaction & Communication

- Organise regular updates and technical discussions throughout the development lifecycle

Development of best-fit AI model

IBM

To assist its Quality Engineers in making more accurate, consistent and faster labelling of the risk level that every product batch possess

RenalTeam

To help its trained nurses carry out dialysis treatment for patients

Sompo Holdings Asia

To flag suspicious personal accident claims

VersaFleet™

To optimise travel routes in a timely manner and enhance business efficiency

Case 1



IBM Manufacturing Solutions Pte Ltd: Ensuring Product is Ready for Market Release with AI

IBM Manufacturing Solutions Pte Ltd (**IBM**) is a global technology company. Before releasing its manufacturing products into the market, IBM needs to ensure that they pass its quality risk assessment.

Previously, IBM Quality Assurance Engineers assessed the quality of the products manually based on past return rates for similar products. This was a labour-intensive process and resulted in inconsistent labelling. To improve the quality of products and reduce the possibility of products being returned, IBM engaged AI Singapore to develop an AI solution to assist its Quality Engineers to make more accurate, consistent and faster labelling of the risk level that every product batch possesses. To do so, AI Singapore engineers built a deep learning model to learn the visual representation of the number of items that failed in a specific batch of hardware, and the types of defects that these items had.

Augmenting Engineers in Assessing Product Defects

At the outset, IBM's Quality Assurance (**QA**) engineers would review all the predictions made by the AI model regardless of their risk levels, i.e., **human-in-the-loop** approach. Subsequently, AI Singapore and IBM jointly agreed to take a **human-over-the-loop** approach, where the QA engineers would only review the product batches that were flagged out by the AI model as high-risk (i.e., likely to have defect). As the purpose of the deep learning model was to speed up the classification process of the risk level for every product batch, IBM's QA engineers were able to prioritise their inspection, focus on high-risk product batches and make the final judgement call on whether to release the batches for sale into the market.

IBM recognises that there will be cases of false positives, where non-defective batches could be labelled as defective. However, IBM prefers to review and troubleshoot these false positives than to allow defective products to be released into the market.

In this project, AI Singapore worked with IBM to **ensure that the datasets used to train the AI model are as representative as possible of the intended population in order to reduce inherent bias**. Additionally, AI Singapore conducted code walkthrough sessions with IBM engineers to ensure a common understanding of the datasets used to develop the AI solution. AI Singapore also shared with IBM a detailed and modularised code with accompanying documentation in a final repository for accountability purposes.



Two key evaluation metrics are used to determine the performance of the AI model:

- a. Consistency of prediction compared to actual defect rate. The AI model achieved 85% of prediction, higher than the specification of 80%.
- b. Time saved for IBM QA engineers — The prediction model was able to identify products that had high risks of defects and reduce the average time of 30 minutes spent by QA engineers to just few minutes.

Better detection of product defects and assurance of quality products for sale will lead to greater customer satisfaction and confidence.

IBM had signed off on the AI models developed by AISG and deployed the Minimum Viable Model into their work streams to assist their engineers. IBM also noted that the model can be adapted to analyse different products other than the one validated for the 100E programme.

Case 2



RenalTeam: Leveraging AI to Provide Better Care for Dialysis Patients

As a provider of haemodialysis services in Singapore, Malaysia and Indonesia, RenalTeam aims to provide better health outcomes for its patients with renal disease by harnessing the power of technology. With the launch of the 100E programme, RenalTeam was excited to come on board to work with AI Singapore to develop an AI solution to help its trained nurses who carry out dialysis treatment for patients.

Patients undergoing kidney dialysis have higher morbidities and high risks of hospitalisation from complications associated with kidney failure. By the time they are hospitalised, their medical conditions usually have become full-blown and their mortality risks would have increased. Even though there is research done on key predictors of hospitalisation, the current process is fuzzy and dependent on the experience of medical staff: the trained nurses decide whether to advise patients to seek medical attention based on their assessments before and after patients' dialysis sessions. The ability to predict hospitalisation risk using AI will thus allow early medical intervention.

A Win-Win Approach in Healthcare: Human-in-the-loop

RenalTeam, with the aim of improving the medical outcomes of its patients, collaborated with AI Singapore to develop an AI model that predicts the hospitalisation risk of dialysis patients. In healthcare scenarios, false results (whether negatives or positives) can have implications on patients' lives. Therefore, AI Singapore and RenalTeam jointly agreed to adopt a **human-in-the-loop** decision-making approach, where the trained nurses would make the final call on whether to proceed with the AI solution's recommendation.

The AI model can perform consistent analysis on patients' data and reduce human errors due to human fatigue. At the same time, RenalTeam's nurses can use the AI model as a support tool for a second opinion. This approach of allowing AI to augment the decision-making of nurses can help to minimise incidents where malfunctioning AI causes harm to patients as the final decision on whether a patient should be hospitalised still lies with the trained nurses.

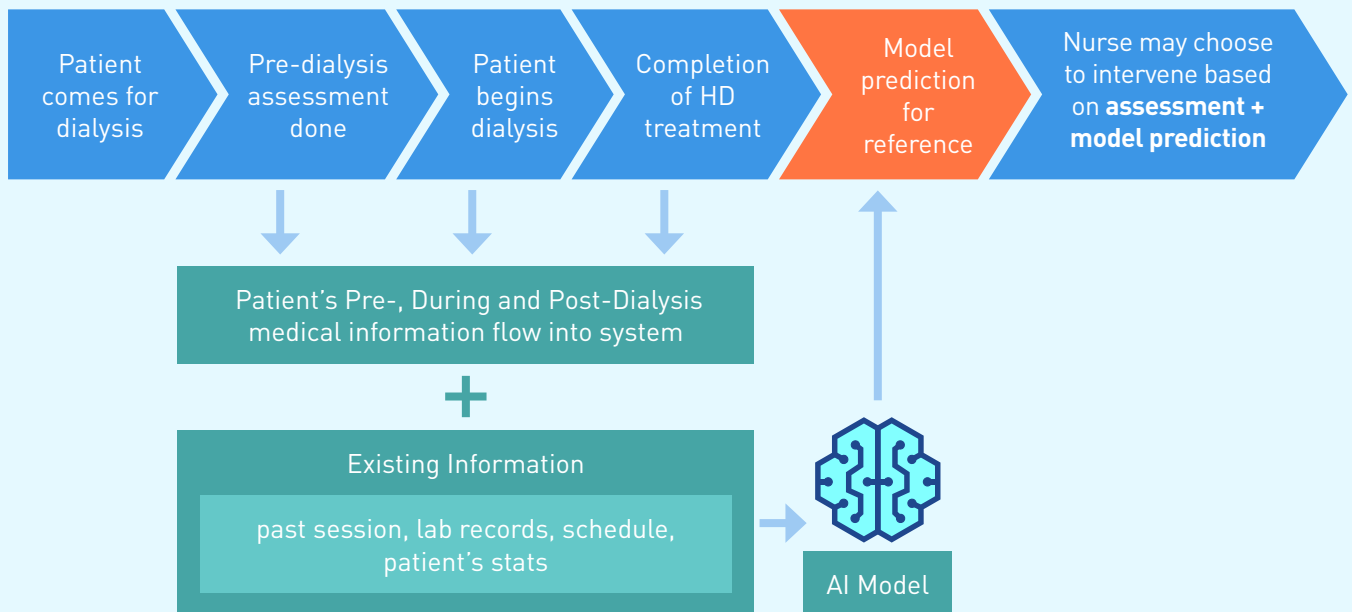
Developing an Effective AI Solution with Safeguards

AI Singapore developed and trained the AI model with patients' medical histories provided by RenalTeam. Given the sensitivity of the data, RenalTeam anonymised all personally identifiable information from the dataset before using the data for model training. AI Singapore ensured data received was kept confidential and only used for the purposes of training the AI model.

AI Singapore and RenalTeam conducted a model validation exercise to evaluate the performance of the AI model. Over the period of one month, the nurses assessed their patients as usual, made their own predictions and recorded them down.



At the end of the month, AISG used the same patients' data and ran them through the AI model. The nurses' predictions and AI predictions are then compared against whether the patients were hospitalised (which is the ground truth).



Based on the results, the AI model performed 36% better in precision (i.e., less false positive predictions). This gives RenalTeam confidence in implementing the AI solution to improve the quality of care for its patients.

RenalTeam transferred further development of the AI model to an associated company, RenalWorks Pte Ltd, in July 2019. RenalWorks, which is a medical software developer, is working to expand the capabilities of the AI model to enhance clinical care for dialysis patients.

Case 3



Sompo Holdings Asia: Challenge Accepted — Implementing a Fraud Detection Solution Responsibly

Based in Singapore, Sompo Holdings Asia (**SOMPO**) is the Asia Pacific (excluding Japan) regional headquarters of its parent company Sompo Holdings in Japan. With presence in 14 markets across the region, SOMPO provides non-life insurance solutions for corporations and individuals, such as travel, motor, personal accident insurance and more.

As part of its claims handling processes, SOMPO's claim handlers review all claims submitted manually on a daily basis, which is a laborious and time-consuming process. Keen on improving its efficiency and fraud detection, SOMPO posed a challenge to AI Singapore to develop a solution to flag suspicious personal accident claims.

AI Singapore and SOMPO jointly agreed to take a **human-over-the-loop** approach in the AI-augmented decision-making process when determining whether personal accident claims were fraudulent or eligible for payment. AI Singapore then developed an AI application that can flag potentially suspicious claims:



High-risk claims were channelled for further investigation and decision-making by SOMPO's Special Investigation Unit.



Low-risk claims that were assessed to be valid were considered as "safe-to-pay cases" and were immediately processed for payment, boosting the efficiency of the claims process.

Ensuring Proper Data Preparation

For the purposes of model development, SOMPO supplied data from its internal databases on actual claims submitted in 2018. For personally identifiable information, SOMPO either encrypted or removed them before providing the datasets to AI Singapore.

Consisting of over 2,000 claims, the training and testing data was representative of the various claims types and scenarios SOMPO encountered in the course of its daily processing of personal accident claims. Using such datasets helped to ensure that the eventual AI solution generalised well for the business.

AI Singapore made every effort in data preparation to ensure data quality. To develop the AI model, the algorithm analysed all the features extracted such as payment amount, accident location and sum insured, and modelled them against verified labels indicating the suspiciousness level of each claim. The trained AI model would then be able to identify and flag high-risk claims that require further investigation by SOMPO's Special Investigation Unit.

A large proportion of the solution development was devoted to feature engineering¹, which aimed to represent and reflect SOMPO claims experts' domain knowledge in the model. AI Singapore consulted SOMPO's business and technology leads closely to create and populate a list of possible features to test. Several of these efforts led to improvements in model accuracy.

¹ Feature engineering is the process of using domain knowledge to codify and extract features from raw data via exploratory data analysis and data mining techniques.

Being Transparent in the Development of an Explainable AI Model

The development and evaluation of the AI model was done in full consultation with SOMPO's key business and technical stakeholders. Several workshop sessions were organised for AI Singapore's engineers to explain the intuition behind the chosen algorithm, analysis of the importance of selected features and evaluation of the model's accuracy. In particular, an evaluation included an in-depth error analysis of false positives and false negatives. The discussions resulted in a few key outcomes:

- a. A proportion of the original labels were identified as wrongly labelled and sent back for relabelling.
- b. Reasons for the misclassifications were hypothesised and additional features were engineered.

Following the in-depth evaluation and revisions, the AI model saw a 12% increase in accuracy in detecting fraudulent personal accident claims.

AI Singapore found that providing necessary documentation worked well with its industry partners as it continues to reflect the trustworthiness of AI Singapore. AI Singapore documented its entire model development process and shared it with SOMPO. To enhance transparency, AI Singapore chose an open source machine learning model that came with an explainability module, and released the full codebase with clear user guidance to SOMPO's digital team.

Achieving Success in Fraud Detection

As the final step of the 100E programme, AI Singapore helped SOMPO deploy the AI solution into its daily claims processes by packaging it in a Docker container hosted on SOMPO's cloud environment. AI Singapore also partnered with SOMPO's IT vendor, Hashcom, to iron out changes required for its production processes. For instance, initial testing revealed that the pipeline of claims submitted to the AI solution for processing tended to break over the weekend due to gaps in the scheduling of data transfer. Hence, AI Singapore and Hashcom instituted a further update to ensure fresh claims were being populated in the system. The final production workflow involved the prediction engine running twice daily, processing the claims in two scheduled batches.

After deploying the AI solution, SOMPO achieved the following outcomes:



100% fraud detection coverage whereby all personal accident claims are processed, thus reducing potential undetected fraudulent cases.



SOMPO's Special Investigation Unit can prioritise and focus on high-risk claim cases.



Significant reduction in time taken to handle low-risk claim cases.



10 to 20% of SOMPO's customers received their payments within minutes, via the identification of safe-to-pay cases.

Case 4



VersaFleet™ – Dynamic Route Solver to Optimise Business Efficiency

VersaFleet™ is a transport management Software-as-a-Service (**TMS SaaS**) that automates logistics operations with route optimisation, electronic Proof-Of-Delivery, instant notifications and real-time job status tracking. Designed for the everyman, VersaFleet powers thousands of users worldwide, from Fortune 100 brands of consumer goods and multi-national logistics conglomerates to SME transporters and even ambulances, limousine drivers, passenger buses and minivans.

VersaFleet's clients include principal 'brand owners' like Watsons Malaysia, Kara, King Koil and Johnson & Johnson, large logistics players like XPO Logistics and Agility Logistics, as well as SME transporters like LLMS Logistics, UDL and S&P Logistics.

VersaFleet has been employing a heuristics-based model to provide route optimisation capabilities to its clients. As VersaFleet's clientele are regional in nature, they need to comply with various local laws and labour guidelines when dispatching to a transporter or driver. For example, some cities require drivers to take a 30-minute break for every two hours on the road. Furthermore, accommodating changes in route-plans due to sudden driver or vehicle unavailability requires re-computation of optimal routes. Towards this, VersaFleet worked with AI Singapore to develop an improved heuristics model (i.e., the dynamic route solver) to dynamically adapt routes in a timely manner.

In the project with VersaFleet, all datasets used were anonymised and redacted. Additionally, only VersaFleet maintained and hosted the master data, with limited access to scrubbed subsets of data only provided to the assigned AI Singapore engineers. The data was used to test the algorithms in the dynamic route solver. Using the supplied datasets, AI Singapore built a Minimum Viable Model (**MVM**).

Engaging Stakeholders Throughout the Model Development Process

Clients of VersaFleet typically utilise route optimisation before dispatch. Every user can override recommendations from the solver at any time before the actual dispatch.

To assist VersaFleet's stakeholders with quick understanding and internally managing development progress, AI Singapore organised regular updates and technical discussions with VersaFleet's stakeholders at every stage of the development lifecycle.

During these meetings, AI Singapore's engineering team would demonstrate how features of the AI solution were being developed and how the MVM might be deployed. This helped VersaFleet's product development team better understand the benefits and limitations of the MVM, as well as how to potentially use and deploy viable improvements into production.

Improvement to Routing Service

Comparing the results of the existing model and the MVM, the enhanced dynamic route solver was designed to allow VersaFleet to help its clients:

- a. Achieve greater operational flexibility in optimising cost savings per route
- b. Minimise vehicle utilisation in a wider variety of test scenarios

VersaFleet might potentially incorporate aspects of the MVM developed into the VersaFleet TMS product suite. As part of the 100E programme, AI Singapore delivered the prototype source code to VersaFleet. VersaFleet aims to further refine the prototype source code and re-engineer it for potential deployment to production.

DARWIN

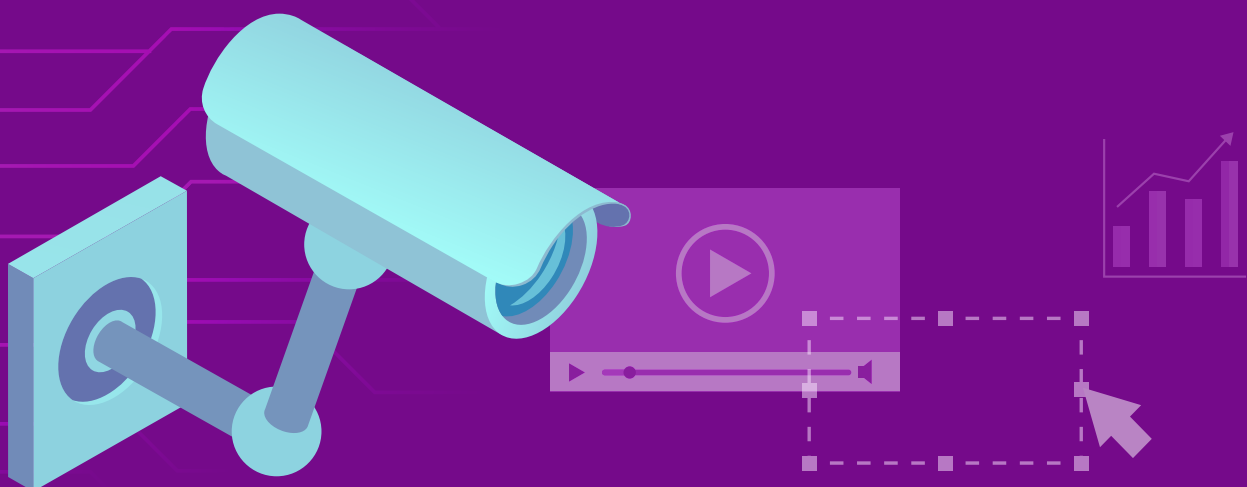
A Responsible city in CCTV Data Analytics

Founded in 1869, Darwin is the capital of Northern Australia, a thriving modern capital city. Seen as Australia's gateway to South East Asia, Darwin aims to be a smart, liveable, productive and sustainable global city.

Through the Smart Cities and Suburbs Program, City of Darwin has led the delivery of Smart City technology through the \$10M "Switching on Darwin" project; funded with \$5M from the Federal Government and \$2.5M from both the Northern Territory Government and City of Darwin. This project entailed the roll out of new technologies, which provides significant data to assist City of Darwin in improving service delivery and future planning for the city.

The 'Switching on Darwin' project delivered:

- a. Smart LED street lighting
- b. Free public Wi-Fi
- c. Weather and particulate sensors
- d. Parking bay sensors
- e. CCTV with smart analytics
- f. Community audio in the Mall
- g. City wayfinding kiosks



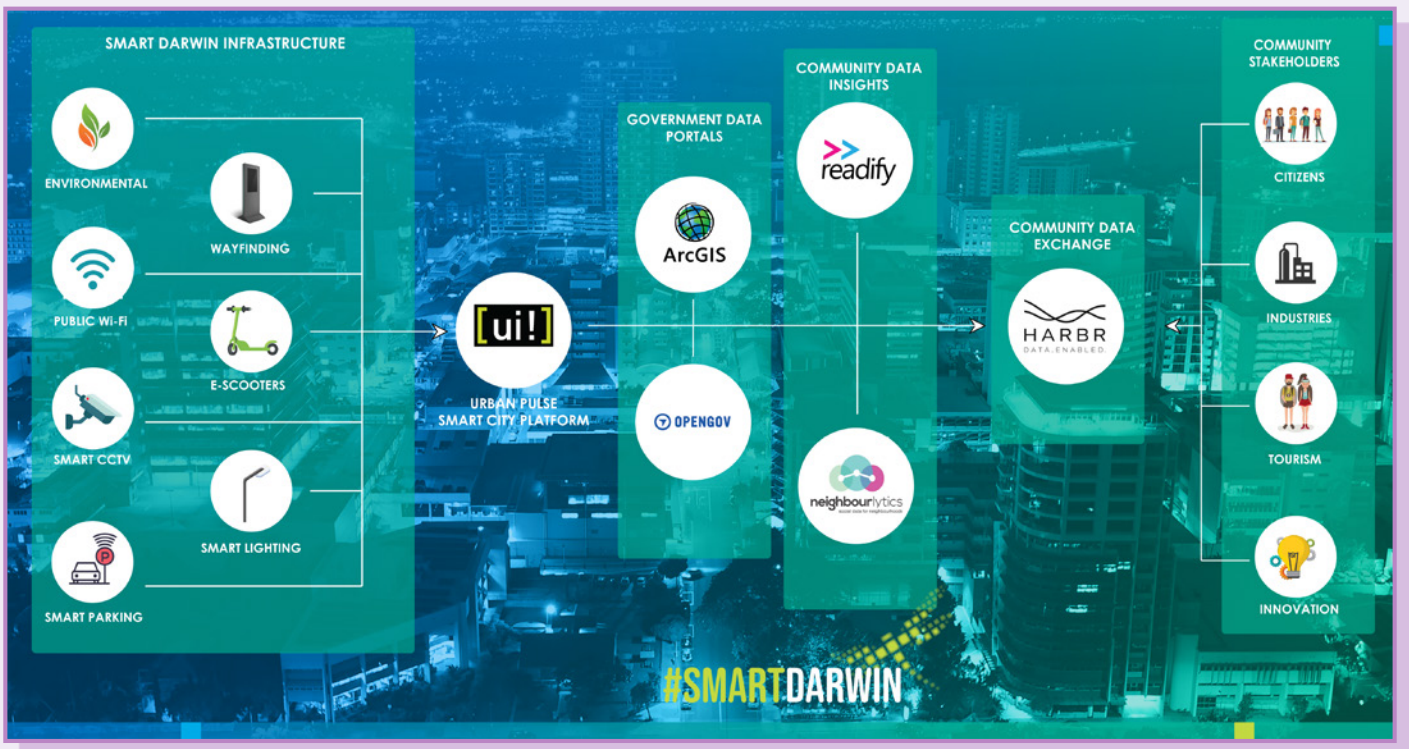


Figure 1 – The “Switching on Darwin” Information Architecture

City of Darwin led the project and rolled out innovative solutions and digital infrastructure to enhance community safety through the use of data-driven technologies in various areas such as street lighting, CCTV, environmental IoT and public Wi-Fi. This is in line with City of Darwin’s core values and vision: to create a vibrant, creative, innovative, connected, healthy and environmentally responsible city by 2030. In particular, City of Darwin has identified digital and data technology as key to support its smart city journey, and has started to leverage AI from video analytics within the city centre. The AI used in this instance is derived from anonymised data collected from people and vehicle movement. Facial recognition technology is not available through this analytic tool.

To ensure an accountable deployment of AI, City of Darwin has adopted the Model AI Governance Framework and piloted the Implementation and Self-Assessment Guide for Organisations (**ISAGO**) to assess the alignment of their AI governance practices with the Model AI Governance Framework.

Deploying CCTVs

With public safety a high priority for the community, the installation of 138 new CCTV cameras across the Central Business District (**CBD**) as part of the ‘Switching on Darwin’ project supports this objective and provides law enforcement with additional tools to investigate and prevent crime. According to the Northern Territory Police, Fire and Emergency Services website, commercial break-ins in Darwin between August 2019 and July 2020 fell by 38% from the previous year. Additionally, anonymised CCTV data provides key insights into vehicle and pedestrian movement, which underpins city planning and drives service delivery improvements.

The Responsible Implementation of AI

The Federal and Northern Territory Governments and City of Darwin funded the “Switching on Darwin” project.

As a local government, City of Darwin:

- Sets the strategic direction of the organisation’s strategic goals as per its “*Darwin 2030 – City for People. City of Colour.*” plan.
- Planned and operationalised the “Switching on Darwin” project.
- Engaged a Smart Video Analytics vendor to deploy the CCTV AI Analytics solution.

Ensuring Human Oversight

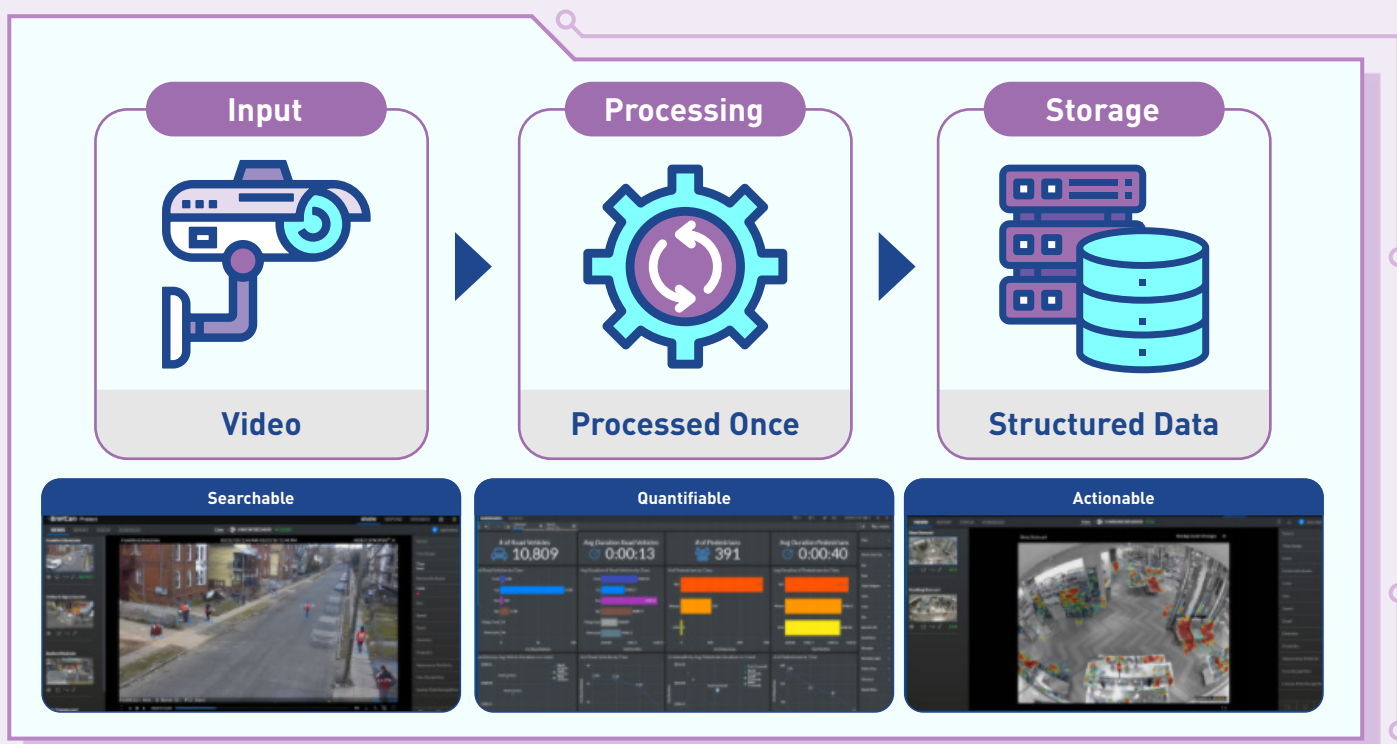
The vendor’s CCTV AI Analytics solution provides reporting using anonymised data based on people and vehicle movements across the CBD, captured through the ‘Switching on Darwin’ CCTV network. In particular, the solution leverages AI to detect, extract and classify video data to capture pedestrian numbers, pace of movement and vehicle details including type, colour and speed using its Deep Learning technology. Video data captured through the network is stored on secured on-site servers.

City of Darwin adopted a **human-in-the-loop approach** to ensure the system does not make any autonomous decisions, with all analytic reports scheduled and reviewed.

City of Darwin will conduct quarterly reviews of the data to identify any anomalies or inconsistencies and report these to the vendor.

Understanding the Use of AI

City of Darwin is utilising CCTV AI Analytics in order to improve public safety and underpin city planning.



The system's AI processes, analyses, reports and stores video footage obtained from CCTV cameras across a number of phases:

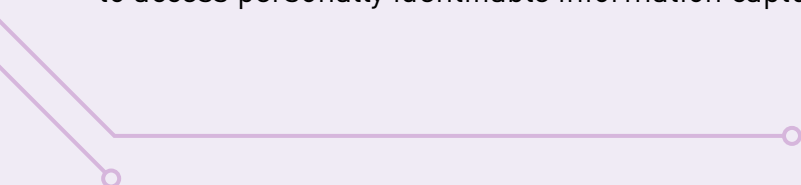
- a. Input phase: As the first step of its video-to-insight pipeline, a user will assign relevant CCTVs in order to build a custom report.
- b. Processing phase: In order to produce custom reports, the application schedules CCTV cameras to collect data using supervised learning and deep learning techniques, detecting objects and determining what they are (e.g., vehicle, animal and person). This information is processed into anonymised data.
- c. Object extraction: The application then detects and tracks objects through the CCTV data collection, which identifies, classifies and analyses each object's behaviour and attributes to determine what the object is (e.g., vehicle type, colour and speed).
- d. Background subtraction: The application applies background subtraction to the image to separate the foreground from the background on a per pixel level.
- e. Detection: Once the background subtraction is complete, it assigns colour labelling to objects in the foreground to differentiate them. The application also filters items that are not detected as objects, such as shadows and lighting.
- f. Tracking: The application tracks objects once detected, using tracking algorithms to assign unique object IDs to those objects in order to track them as they move (e.g., a Sports Utility Vehicle).
- g. Display Masks and Metadata Extraction: The application will conduct deep learning classification to extract relevant metadata in order to create accurate display masks for each object.
- h. Storage: The CCTV network collects and stores structured metadata in City of Darwin's secured on-site storage database.

City of Darwin performs periodical reviews of reports provided through the AI solution, informing the vendor of any issues or errors. The vendor continues to invest in the research and development of its AI model including supervised and deep learning techniques with a view to providing even greater accuracy in reporting.

Protecting Personal Data Whilst Ensuring Innovation

The video analytics application is hosted on secure on-site servers. Only specifically trained City of Darwin staff have access to the Video Analytics system. Internal processes ensure strict security standards are adhered to.

City of Darwin conducted a privacy impact assessment through an independent privacy consultant in December 2019. The consultant confirmed City of Darwin does not use facial recognition as part of its smart CCTV data collection and acknowledged the stringent physical, technical and administrative measures taken to ensure individual privacy is protected. Only law enforcement agencies are able to access personally identifiable information captured through the CCTV network.



Addressing Community Concerns

As part of its stakeholder engagement undertaken during the ‘Switching on Darwin’ rollout, key stakeholders including community members, industry partners, government agencies and academia were identified. Recognising the importance of being transparent in the use of AI capabilities in video analytics, City of Darwin shared information with these stakeholders on the use of CCTV technology, data collection and analysis through a wide range of promotional activities, including roadshows, demonstrations, media events and website updates. City of Darwin also held public information sessions to address community concerns about CCTV cameras installed in the city centre.

During the stakeholder engagement process, City of Darwin:



Used plain, non-technical language to ensure the community understood the use of CCTV, data collection and video analytics.



Provided a feedback channel through the City of Darwin’s privacy statement.



Made available on its website its privacy policy and provided details of the “Switching on Darwin” project as well as key FAQs relating to the use of CCTV.

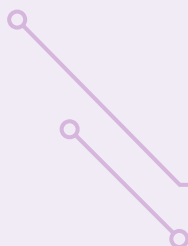


The Northern Territory Police placed signage around light poles where the CCTV cameras are mounted on, to ensure the community is aware of their location and when they enter active CCTV camera locations, in line with Australian legislative requirements.

Conclusion

The Model AI Governance Framework has been used as a tool for City of Darwin to understand its AI readiness. The guidance in the Model AI Governance Framework has provided City of Darwin with a detailed understanding of the potential risks involved. Using ISAGO has ensured City of Darwin adhered to stringent best practice standards, protected individual privacy and maintained public trust.

The Model AI Governance Framework and ISAGO have paved the way for City of Darwin’s “AI Quick start Guide” project. This website, which is currently being developed, contains a short questionnaire to help provide local government agencies and organisations with an insight into their AI readiness.



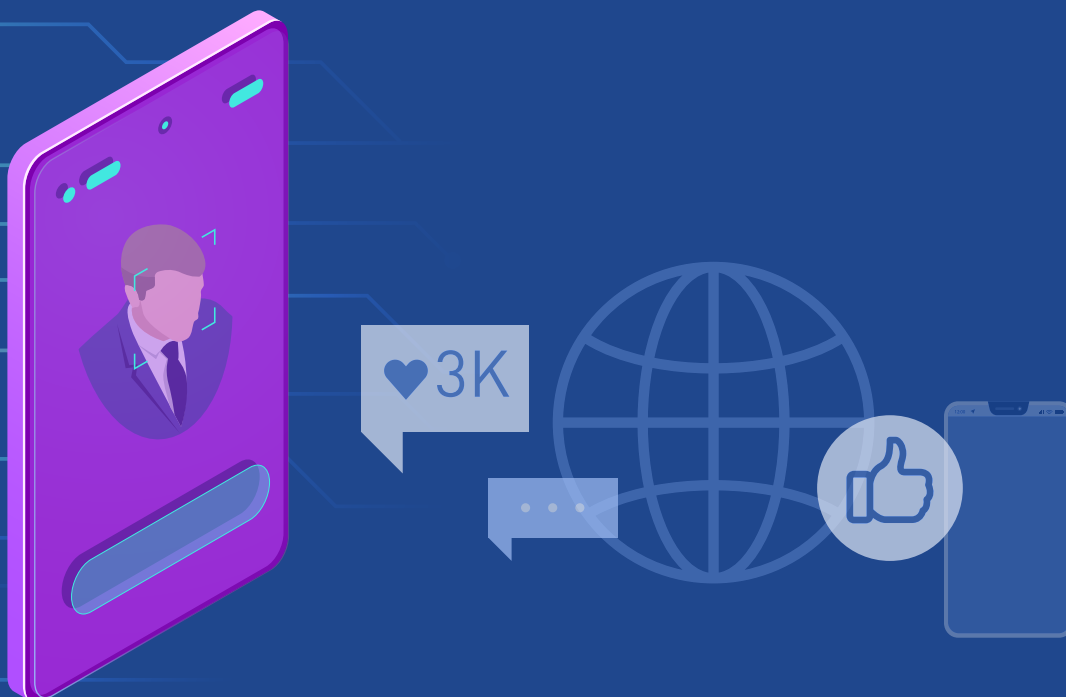
GOOGLE

Celebrity Recognition with Governance in Place

Google's mission is to organise the world's information and make it universally accessible and useful. As part of this mission, Google offers products designed to be helpful to people such as Google Search, Translate and Gmail. It also offers business solutions such as the Celebrity Recognition Application Programming Interface (**API**) through Google Cloud.

As part of Google's Video Intelligence products, the Celebrity Recognition API is a limited-availability tool that helps its clients detect and track an international roster of widely-known celebrities. In addition to searching photos and video footage of basic visual concepts like "city street" or "railroad crossing", Google's approved customers can use this tool to identify and search for professionally-produced content of celebrities.

In today's digital age, new movies, documentaries and series are being created at an unprecedented rate, joining decades of existing libraries, sports broadcasts and a vibrant influx of international works. Without an expensive and labour-intensive tagging process, much of this video content is unsearchable. This makes it difficult for media and entertainment companies to organise, search and fully understand the contents of their media catalogues. This Celebrity Recognition API thus helps companies address this challenge of managing their huge video databases, and in turn companies can better cater to the increasing demand for personalised experiences.



Adhering to AI Principles and a Risk-Based Governance Process

The Celebrity Recognition API adheres to Google's AI principles. Released in 2018, these principles set out Google's commitment to developing advanced technology responsibly:

- | | |
|--|---|
| a. Be socially beneficial | e. Incorporate privacy design principles |
| b. Avoid creating or reinforcing unfair bias | f. Uphold high standards of scientific excellence |
| c. Be built and tested for safety | g. Be made available for uses that accord with these principles |
| d. Be accountable to people | |

Google's Principles also detail applications it will not pursue:

- Technologies that cause or are likely to cause overall harm
- Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people
- Technologies that gather or use information for surveillance, violating internationally accepted norms
- Technologies whose purpose contravenes widely accepted principles of international law and human rights

Google believes that these principles are the right foundation for its company and the future development of AI. To ensure alignment with these principles and address questions that may arise in third party enterprise AI deals, Google Cloud established a governance process. This process comprises two key governance structures:

A A decision-making body with a clear quorum structure that reviews custom deals for AI Principles risks.

B A committee that meets regularly to review AI-related Cloud projects in development, and works with Cloud engineering and product teams to undertake rigorous ethical risk analysis. This review body identifies potential risks for products in development, recommends mitigation strategies to address the risks and ensures alignment with its AI principles. Reviewers from Google's central AI Principles review team also participate to ensure consistency.

The Celebrity Recognition API went through Cloud's AI Principles review process to ensure alignment with its principles. Additionally, the project was reviewed for conformity to Google's approach to face-related technologies and consistency with its due diligence practices.

The Celebrity Recognition API works by identifying faces and comparing them against an indexed gallery of thousands of celebrities collated by Google. If a match is found, the Vision API provides the Knowledge Graph Machine ID (**MID**) of the celebrity, their name and a bounding box indicating where the face appears in the image. This capability can be integrated into a range of workflows, and the extent to which human decision makers rely on the tool depends on the developers and end users and their risk assessment.

Incorporating Safeguards into Development to Mitigate Bias and Protect Privacy

During the development of the Celebrity Recognition API tool, Google obtained expert guidance from human rights non-profit Business for Social Responsibility (**BSR**) to identify and mitigate potential human rights impacts. This included accepting the following recommendations:

Investigating and correcting the AI model for unfair bias through intersectional fairness testing

As opposed to a general-purpose facial recognition API, the tool focused on a specific business use case — celebrity recognition is a pre-trained AI model that is able to spot thousands of popular actors and athletes from around the world, and it is based on licensed images so media and entertainment customers can now search their professionally produced content for celebrities. This narrow scope meant that it was possible to review all the images in the entire dataset individually to determine the possible cause of any skewed results.



This intersectional fairness testing played a crucial role in investigating and correcting the model for unfair bias. For example, Google found a discrepancy in its training data sets falling on skin tone lines. Because Google had done the work to understand the societal context, Google knew that it needed to investigate more deeply to understand what was driving this. Google determined that the error rates across skin tone lines were partly attributed to inaccurate skin tone labels in its benchmark datasets. To correct for this, Google changed how it categorised skin tone, using the dermatological Fitzpatrick skin type scale, which improved the model's performance. Google also used manual labelling to further close the performance gap.



Defining “celebrity” and restricting to a predefined list

Among other criteria, “celebrity” refers to professional actors or athletes who make their primary living by voluntarily appearing on TV or in movies.



Using this definition, Google pre-loaded the model to recognise a limited and curated list of thousands of celebrity figures from across the world, based on licensed images.



Ensuring that this is not a generally available tool and implementing a whitelisting process

An interested customer must pass a manual review process to ensure they are an established media or entertainment company or partner with an approved use case applying only to professionally-produced video content like movies, TV shows and sporting events.



Designing the tool such that its customers do not have the ability to add individuals to the list—even for private use

This ensures that this tool will not be able to expand to behave more generally. While such constraints limit the tool's flexibility, Google assessed that these safeguards were needed to reduce the risk of misuse for surveillance.

Interacting and Communicating with Stakeholders

Google consulted a wide range of internal and external stakeholders including experts to seek feedback. As a result of its consultation process, Google implemented various measures to allow for feedback and communication:



Created an expanded terms of service to address unique concerns raised by this capability, and to ensure that their customers follow the same principles that guide its approach to AI deployment.



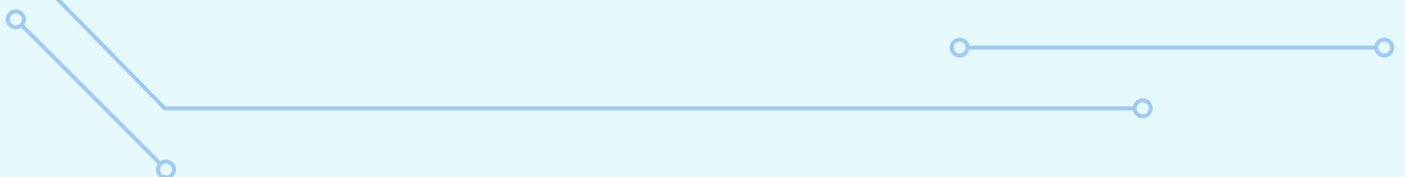
Establishing an opt-out process for celebrities who would like assurances that they will not be recognised by the API. Google made available an opt-out request as part of the API documentation.



Included a function to report misuse of the tool.

Conclusion

Google supports Singapore's Model AI Governance Framework. As a leader in AI, Google prioritises the importance of understanding its societal implications and developing its solutions in a way that gets it right for everyone. This is why Google released its AI principles and have since worked to build the processes, teams, tools and training necessary to operationalise the principles. In shepherding the Celebrity Recognition API through its internal governance processes, Google was able to leverage the expertise of product experts, social scientists, human rights specialists, legal experts and privacy advisors to put together a comprehensive review for the Celebrity Recognition API.

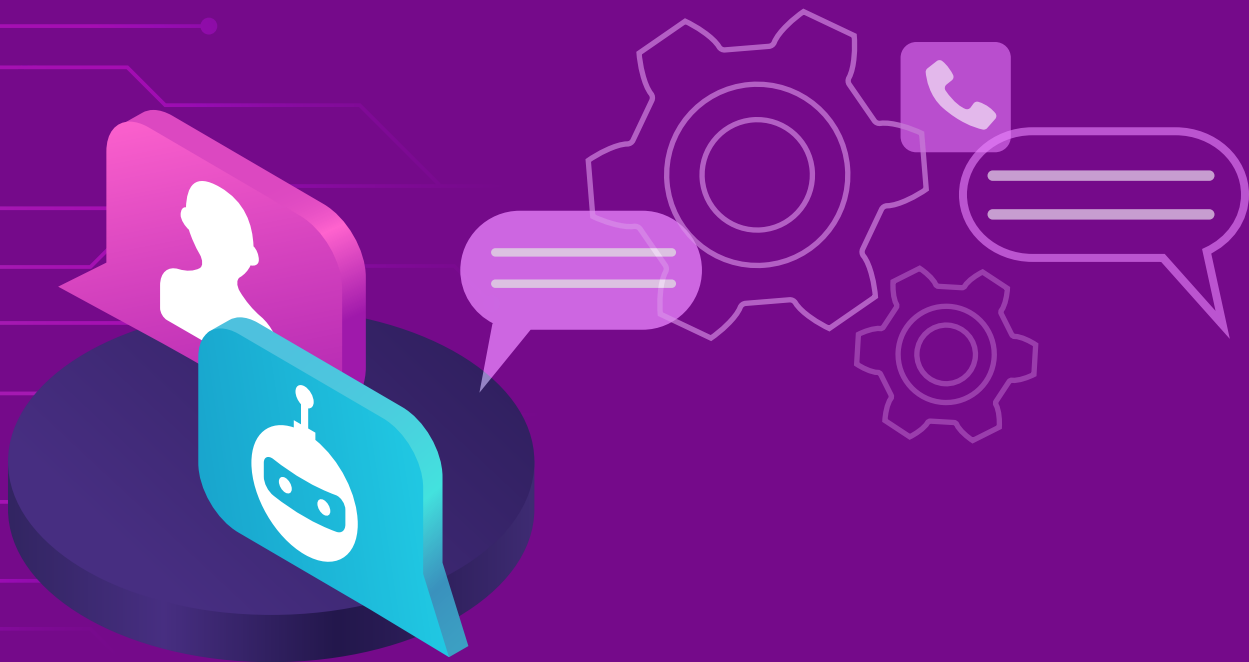


MICROSOFT

Ways to Implement Trustworthy Conversational AI

Headquartered in Redmond, USA, Microsoft is a platform provider and developer of computer software, hardware and related products. One of the most widely-used AI technologies that Microsoft sees being deployed is conversational AI, often used in FAQ chatbots and virtual assistants embedded in consumer devices. Conversational AI is able to bring about opportunities for improved efficiency, scalable and filtered customer service and the ability for all humans to communicate naturally with computers – making conversational AI the new user interface.

As a developer and platform provider of such bots, Microsoft is well-positioned to learn from its experiences and interactions with its customers on responsible conversational AI development and deployment.



Overseeing the Responsible Use of AI

Microsoft's approach is centred around building an ecosystem and company-wide culture in operationalising responsible AI, rather than having a single team leading the way. As a first step, Microsoft developed six core principles to guide its approach to responsible AI, which are aligned to Singapore's Model AI Governance Framework:



As the second step in putting its principles into practices, Microsoft established an AI, Ethics, and Effects in Engineering and Research (**Aether**) Committee and an Office of Responsible AI. The Aether Committee includes members across Microsoft's engineering, field, consulting, legal and research teams, determines key responsible AI challenges and provides advice and recommendations to Microsoft leadership. The Office of Responsible AI executes cross-company governance and public policy work based on the Committee's guidance.

The Aether Committee and Office work closely with teams across the company to develop guidelines and implement responsible AI best practices in day-to-day work. This is supported by Responsible AI Champions within each team to help communicate and align responsible AI practices with its developers and users. As part of its organisational-wide engagement, relevant teams undergo training on how to leverage these resources, which are reviewed periodically.

Varying Level of Human Involvement in AI-Augmented Decision-Making

For any deployment of AI, Microsoft assesses the sensitivity of the use case to determine the level of human involvement required. This involves reviewing potential risks and impacts such as the denial of consequential services for individuals or significant risk of harm to society. While conversational AI is generally unlikely to have such an impact, the extent of human involvement in the AI-augmented decision-making process will ultimately depend on the tasks and role of the AI bot.

In the context of an insurance claims process, for example, a **human-over-the-loop** mechanism may be appropriate for a bot that acts as a filter to answer simple questions and refers any complex requests to humans. On the other hand, if a bot is deployed to ultimately process the insurance claim, a **human-in-the-loop** mechanism may be needed as the claim could be complex or cause potential harm to the person submitting the insurance claim should it be rejected. In this case, while the AI bot can provide recommendations, a human should make any decision.

Microsoft also makes efforts to inform and educate customers about responsible use by providing technical documentation and non-technical documentation such as guidelines, practical demos and training materials. Developers using Microsoft's bot technology are contractually required to comply with specific requirements that prevent potential harmful and unlawful uses, as well as high-risk use of its technology without reasonable safeguards. For example, Microsoft's terms of service require that its Healthcare Bot service is not used as a medical device or substitute for professional medical advice. Its customers are also required to comply with all relevant laws and regulations.

Reflecting On Key Practices To Build A Safe and Accountable AI Bot

The implementation of good data, algorithm and model accountability practices such as those identified in the Model AI Governance Framework are critical to achieving responsible conversational AI development and deployment. For a bot to be useful and trusted, it must be sufficiently reliable and robust as well as explainable and transparent.

One of the challenges identified is that bias can often occur unintentionally. AI systems may perpetuate existing and new biases found in the data used to train the AI model as well as bias occurring in the algorithm design itself. For example, it was seen in Microsoft's Tay chatbot how malicious or ignorant users can quickly train an AI-powered chatbot to exhibit negative behaviours. This had resulted in Microsoft taking Tay offline soon after its launch. As AI may feed off positive and negative interactions with people, the challenge of reflecting human-centred values in the AI design can be both technical and social.

Based on its experience, **Microsoft has found the following practices useful for identifying and managing bias:**

- a. Employing a diverse development team focused on the design, development and testing of bot technology as well as upper management involved in AI adoption and deployment decisions will help combat bias and ensure different perspectives and backgrounds are accounted for.
- b. Ensuring clear understanding of data lineage and relevant attributes in training data is especially important for identifying and managing bias. This requires systematically assessing data used for training for appropriate representativeness and quality.
- c. Applying machine learning techniques and keyword filtering mechanisms to enable bots to detect and respond appropriately to sensitive or offensive input from users.
- d. Using technical tools to assess model fairness and mitigate the negative impacts of bias under protected attributes such as race, gender, age or disability status.
- e. Using its AI Fairness Checklist to assess bias.
- f. Referencing international and commonly used accessibility standards, such as WCAG 2.1 to ensure people with disabilities are able to use conversational AI solutions. This is because bias may occur through non-inclusivity or inaccessibility.

Based on Microsoft's experience, establishing reliability metrics and reviewing them periodically will help improve AI system robustness and performance over time. These metrics can include determining the acceptable error rate for the bot, the desirable ratio of positive to negative interactions and the reasoning behind these. While the relevant error and reliability threshold will vary according to use case (e.g., lower margin of error in a bot used for army recruitment than one used to purchase socks), the ideal error rate for any use case should be one closest to zero.

Microsoft also notes the importance of building traceability capabilities into the AI bot for monitoring and auditing purposes and to pinpoint any issues that need to be addressed (e.g., detecting performance anomalies). To track the performance of the AI model and detect errors over time, Microsoft has developed a tool for Machine Learning DevOps, which collects performance statistics and feeds them back into the operation of the model to improve reliability.

Ensuring Transparency And Building Trust With Stakeholders

In developing and helping customers deploy conversational AI bots, Microsoft has learnt that being transparent means providing different, meaningful information, like the capabilities and limitations of the technology, to different stakeholders. These may include:



Business customers

As the actions of the AI bot will directly impact the entity's reputation, it is a good practice to provide information on accuracy and reliability metric performance and error rates to business customers.



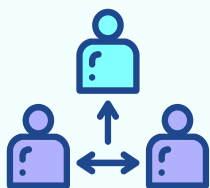
Auditors/Regulators

These stakeholders may want evidence of accountability through understanding the AI model, what data was used to train the conversational AI, decision-making processes etc.



Direct user (i.e., individuals interacting with the bot)

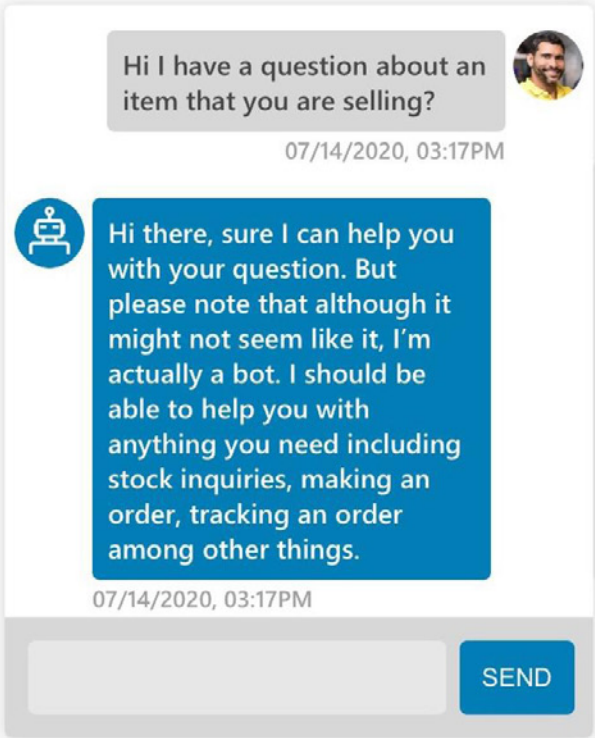
Users will often be concerned with knowing when they are interacting with a bot, how the bot functionally makes decisions and its capabilities and limitations.



Third parties

A bot deployment may involve others by design e.g., outputs from social media may be accessed by multiple third parties if made public, and therefore if bots are deployed on platforms where it's reasonably foreseeable that third parties may be involved (such as social media) then information provided to direct users should be made easily available to third parties as well.

As a matter of responsible practice, it is important to **clearly communicate to users when they are interacting with a bot**. Today, machines that use conversational AI are capable of passing the Turing test. Developers may endow their bots with “personality” and natural language capabilities and users may be easily unaware that they are interacting with an AI bot and believe they are communicating with another human being instead. When a user thinks that they were interacting with a human, and was instead communicating with a bot it can undermine that consumer’s trust.



Hi I have a question about an item that you are selling?

07/14/2020, 03:17PM

Hi there, sure I can help you with your question. But please note that although it might not seem like it, I'm actually a bot. I should be able to help you with anything you need including stock inquiries, making an order, tracking an order among other things.

07/14/2020, 03:17PM

SEND

Make it clear they are talking to a bot

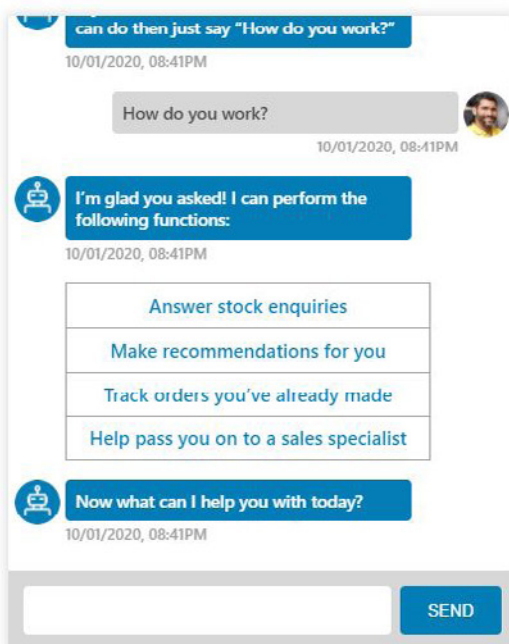
Although bot technology is getting better and better everyday it is good practice to make sure your users know they are working with a bot. This is not only about being transparent, but some users simply won't feel comfortable talking to a bot, and you should make sure they are given an option to speak to a real user, or use another user interface like a regular website.

Users are more likely to find a bot trustworthy if they **understand the purpose of the bot and have reasonable expectations of what it can and cannot do well**. Users should be able to easily find information about the limitations of the bot, including the possibility of errors and the consequences that can flow from such errors. Microsoft has found that making available detailed explanations of the purpose and operation of the bot and establishing metrics to assess user satisfaction can help improve the bot experience and build trust with users.

Microsoft also emphasises putting in place **various communication policies** surrounding particular issues such as privacy, seeking human review and providing feedback:

- a. To protect users' privacy, Microsoft encourages **informing users upfront about the data that is collected and how it is used, and obtain their consent beforehand**. Easy access should be provided to any valid privacy statements, applicable service agreements, and including a “profile page” for users to manage privacy settings and other relevant legal information. For bots that store users' personal information, privacy-protecting user controls can also be implemented, such as including an easy-to-find “Show me all you know about me” button or “Forget my last interaction” or “Delete all you know about me” options.

- b. Users will feel more comfortable with bots if they can provide feedback on their operation or report/challenge incidences of bias, misuse or abuse. Microsoft finds building in **feedback mechanisms** can provide critical and timely information on bot performance and satisfaction, which helps in managing customer expectations. User interfaces should also be designed in a manner that **allows users to seek redress or help** in the event of inaccurate or unexpected outcomes.
- c. Microsoft also notes that it is best to **communicate information concerning the reliability of the bot**, such as summaries of general statistical performance as well as performance under particular circumstances. This ensures accountability and builds user trust.



Make explicit what your bot can do

Just like building any application, you should make the functionality of your bot discoverable.

Conclusion

Since Microsoft announced the set of six principles to guide the development and deployment of trustworthy and ethical AI in 2016, it has been on a journey to further refine those principles and operationalise them in the development and deployment of responsible AI because principles have value only if they are lived by. Microsoft's governance mechanisms include the Aether Committee to advise its leadership on questions in the development and deployment of AI innovations, as well as an Office of Responsible AI to implement AI governance and enablement. Together, the Aether Committee and the Office of Responsible AI help its engineering and sales teams uphold Microsoft's AI principles in its day-to-day work.

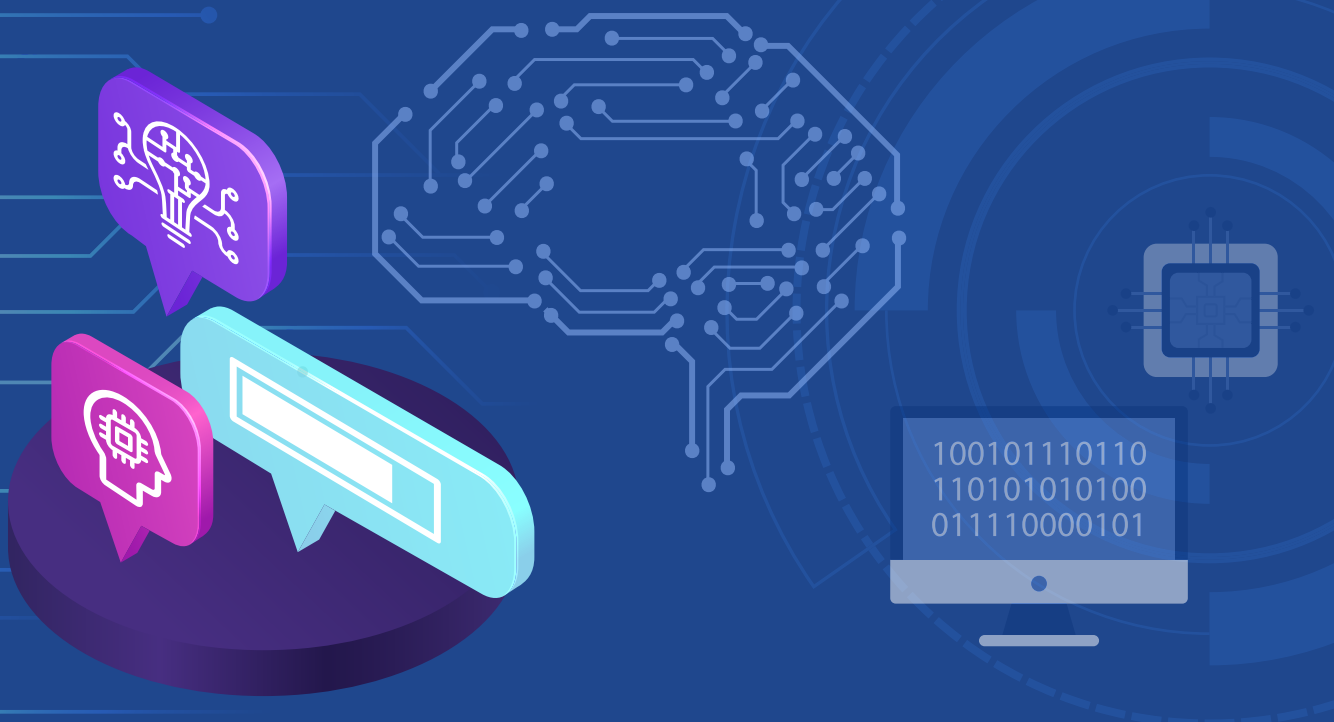
The Model AI Governance Framework provides many useful practical tips and guidance that could be considered by organisations in establishing AI governance frameworks, which from Microsoft's experience is a critical first step in ensuring responsible AI development and deployment. Given the enormous benefits of AI on people and society, but also the risks that AI can create if not carefully designed and deployed, AI governance is vital in building AI's trustworthiness and societal acceptance.

TAIGER

Winning Clients with AI Governance Practices

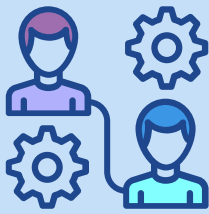
TAIGER is a Singapore-based company that develops AI solutions to automate tasks and simplify processes. Its solutions leverage symbolic and non-symbolic AI (e.g., Natural Language Processing (**NLP**), Machine Learning, Knowledge Representation, Automated Reasoning and Computer Vision) to understand large amounts of unstructured information. It results in increased efficiency and decreased costs for its clients. TAIGER's clients include financial institutions and public sector agencies from various countries such as Singapore, Spain, Mexico and Russia.

TAIGER has put in place AI governance practices that are aligned to the Model AI Governance Framework in its AI development and deployment.



Clear Responsibilities and Processes for AI Solutions Deployed In-House and Externally

In 2019, TAIGER embarked on an organisational restructuring effort to define clear roles and responsibilities for every stakeholder within the company. With these clearly defined Internal Governance Structures, TAIGER was better able to implement measures such as Objectives and Key Results, Key Performance Indicators and Statements of Purpose to establish clear goals and tactics for each team. This clarity has helped to:



Identify knowledge gaps and train employees



Measure productivity and profit-margins at team, project and product levels



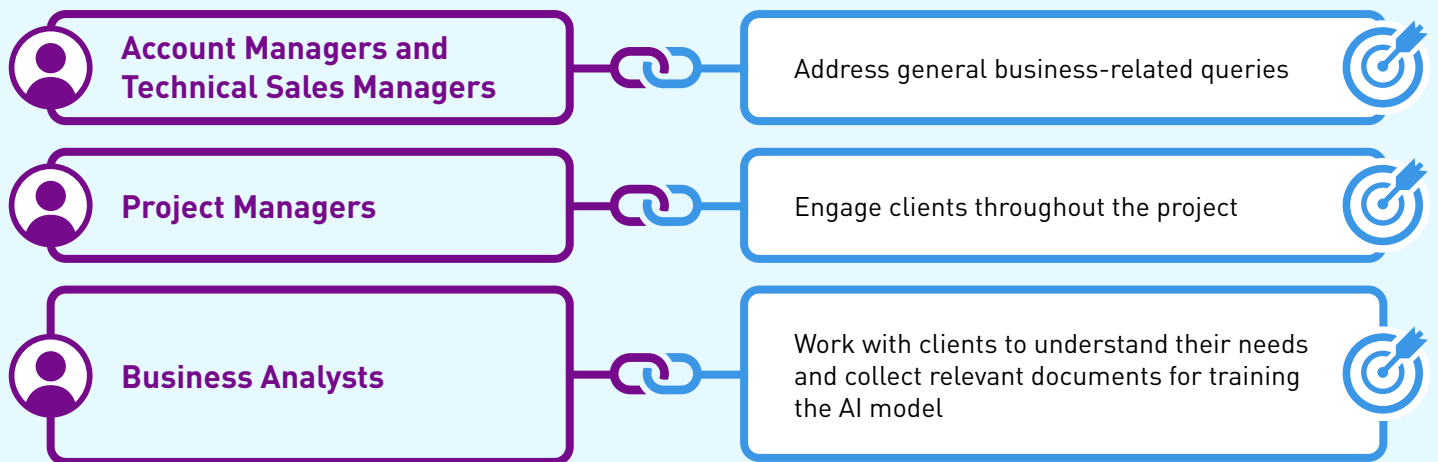
Anticipate risks such as lack of resources and timeline delays, and resolve them in a timely manner

As part of its restructuring efforts, TAIGER has since established dedicated Business and Engineering Operations teams to clearly define AI methodologies, assess the degree of human involvement required for its AI solutions and put in place user feedback loops. These measures have a direct positive impact on its clients' experience when working with TAIGER.

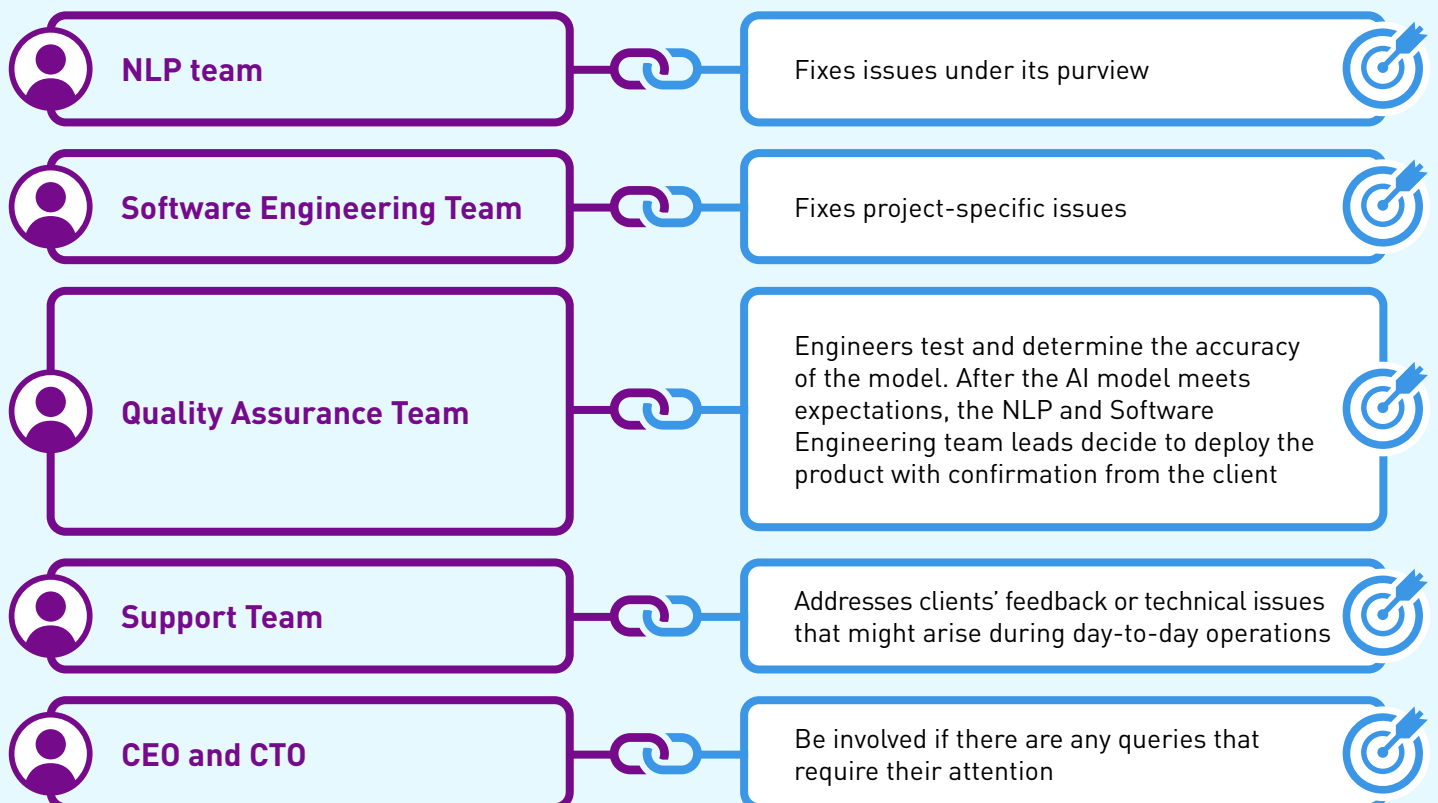
TAIGER has clear processes for the use of AI within its company and for its clients. Before deploying an AI solution for use within its company, TAIGER adopts a four-step governance process internally, with responsibilities of its staff clearly defined:

1. The relevant team to submit a proposal to justify the need for the AI solution.
2. Subject matter experts to research and conduct functional analysis of the solution, and engineering team to conduct a technical analysis of the implementation plan. This joint assessment by the various teams would cover expected impact and value-add, fit to the business needs, potential risk to customers, cost, amount of training required for the staff, effort required for usage (e.g., training of the AI models), resources to maintain the solution and the analysis of vendor options.
3. The Director of IT to sign off on the implementation plan and the Director of Security to sign off on the security compliance of the proposed AI solution.
4. The Chief Executive Officer (**CEO**) and Chief Technology Officer to jointly approve the deployment of the proposed AI solution and the vendor.

Similarly, TAIGER has put in place a clear governance structure when deploying AI solutions for its clients:



Various technical teams collaborate to develop the customised AI solution for clients, such as:



Besides its recent restructuring efforts, TAIGER intends to review its processes every quarter year. Should an issue be identified, a proposal will be presented for evaluation and subsequently be approved or rejected by the CEO. If approved, the new policy is documented and circulated with immediate effect.

Determining the Level of Human Involvement in AI-Augmented Decision-Making

TAIGER typically designs a **human-in-the-loop decision-making process** for its AI models at the initial phases of the production or proof-of-concept stage. For example, one of TAIGER's solutions is to identify, extract, validate and store key pieces of information from documents. When a client user uploads documents, the AI solution will extract the unstructured information and convert them into structured data fields. During the initial phases, TAIGER will require the client to verify that the information has been converted into accurate data fields (e.g., is "apple" correctly extracted as the name of a fruit). If there is an error, the client makes the correction and indicates the correct value through an in-built reporting tab by TAIGER. TAIGER will then check the reporting tab and fix the relevant issues. Subsequently, a **human-over-the-loop approach** will be implemented where the clients check the results extracted by the AI model on a scheduled basis to ensure that the extraction works well.

Using Data Securely and Removing Bias

TAIGER takes a serious view of the data it receives from its clients. Hence, TAIGER only shares the data with staff who have signed a Non-Disclosure Agreement (**NDA**) with the client. In addition, TAIGER removes all personally identifiable information from the client's data before it is being processed.

For model development, TAIGER **uses different datasets for training and testing of the AI models** to ensure that the eventual **AI model can generalise on new data**.

TAIGER uses multiple algorithms to understand the bias in data. Using n-gram analysis can help to find words that are biased and understand how these words appear in the text. After determining the words that are biased, TAIGER analyses the statistics of such words (e.g., how many times a word appears in the text and the kind of context it fits with). After multiple steps of analysis, TAIGER would attempt to remove such biases without reducing the data size.

Ensuring Explainability and Repeatability of the AI models

TAIGER will discuss with its clients which algorithms to use for its AI model. This is because different clients have different requirements and not all problems can be solved by a single algorithm. Hence, TAIGER would work with its clients to evaluate each algorithm based on accuracy, the infrastructure required and various project-specific parameters before finalising the algorithm to be used.

As it could be difficult to explain how some of the algorithms in the AI model function and make their predictions, TAIGER uses multiple methods to explain this to its senior management and clients. Specifically, the various technical teams would provide information on their methods and explain how they:



Analyse step-by-step predictions of an end-to-end system. As a system consists of multiple subsystems, the engineers would analyse the predictions of each sub-system so as to understand which sub-system made the first error and subsequently led to the final error.



Analyse statistics of words used in the sentence (e.g., how often the two words "thank you" appear together in the training data).



Visualise word embedding to shed light on whether two words that have similar meanings (e.g., sorry and apology) would appear together.



Check the confidence score of each prediction. If the model is confused with two predictions, both of them will have similar confidence scores. Once it identifies which two predictions are conflicting, TAIGER will conduct further analysis to identify the reasons for the conflicts (e.g., the AI model thinks two words or predictions have the same meaning).

TAIGER's AI solutions typically use a common base of algorithms that work very well for various problem statements. To tailor the AI solutions to meet the needs of its clients, TAIGER's engineers would fine-tune its AI models. The engineers would also check the end-to-end system to ensure the results or predictions by the AI models are correct and repeatable.

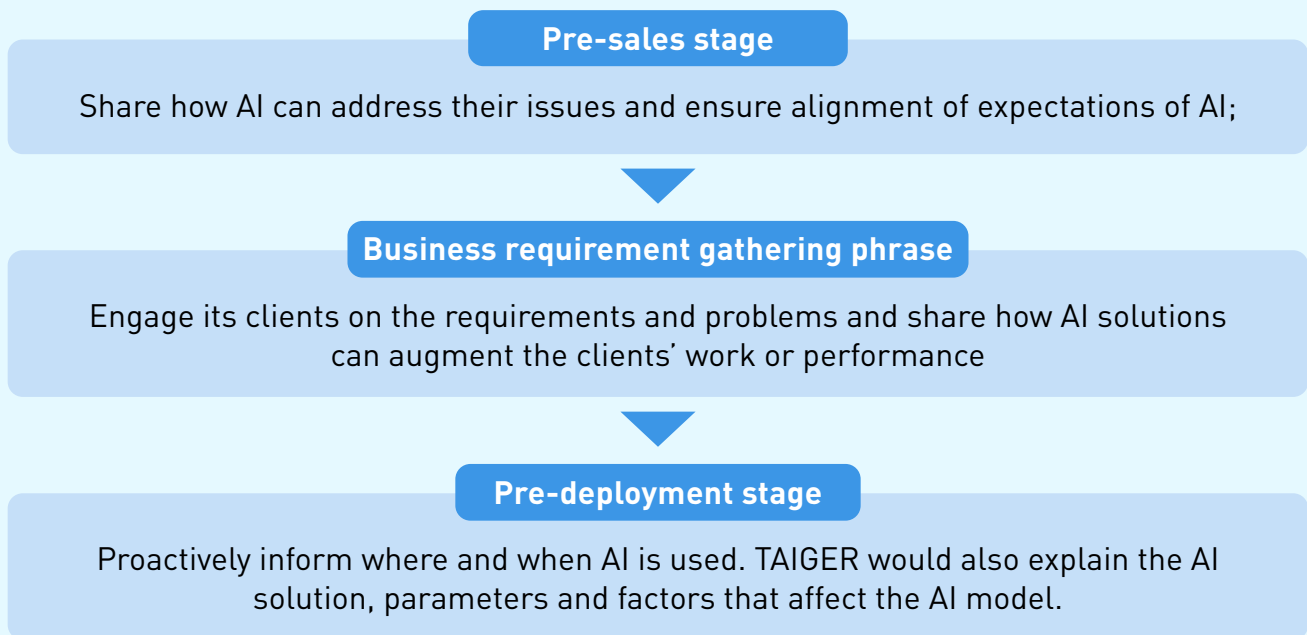
To validate the accuracy and enhance the transparency of the AI models, TAIGER would **conduct repeatability assessments of its AI model and document the results**. This provides better assurance of consistency in the performance of the AI model. To further improve the accuracy and allow the model to learn on new data, TAIGER puts in place a feedback mechanism. TAIGER uses multiple algorithms to test this feedback-learning methodology as well as the confidence level of the prediction made by the AI model. This helps TAIGER better understand the types of data that the AI model is not able to process well. Based on TAIGER's analysis, insufficient data or contradicting data would make it hard for an AI model to predict with a high confidence level.

Tailoring Communication for its Audience

TAIGER puts in place **appropriate communication for its stakeholders** so as to build trust. As different stakeholders have different information needs, TAIGER uses a framework to first identify its audience. TAIGER formulated and aligned this framework to the Model AI Governance Framework.



TAIGER generally keeps any proprietary information internal. For its clients, TAIGER engages them through three stages:



TAIGER recognises that managing clients' expectations at the outset is key to building trust with them. Hence, TAIGER puts in place feedback channels and ensures regular communication with its prospective and current clients. For example, TAIGER has:

- a. Account managers to address general business-related queries;
- b. Project managers and business analysts to engage clients during the project;
- c. A support team to address clients' feedback that might arise from using the AI solution during day-to-day operations.

For the general public and investors, TAIGER discloses to them general information such as the purpose of its AI solutions and how they are built. The company only provides more in-depth information with interested investors after NDAs are signed.

Conclusion

Putting in place practices that are aligned to the Model AI Governance Framework helps TAIGER assure customers that the AI products they purchased are produced by a company that understands its own technology and takes measures to ensure that its AI models are explainable, predictable and transparent. With AI still evolving, TAIGER believes that it is important to continuously strengthen its governance structure to enhance the trustworthiness of its AI models.

Additionally, TAIGER's restructuring efforts to put in place a proper internal governance structure that helps improve the explainability of AI models as well as communicate the implications and potential risks of its AI solution have benefitted them in many ways. In certain cases, adopting responsible AI governance practices has helped TAIGER win client projects from competitors, as customers appreciate transparent and structured processes, both implementation-wise and management-wise, despite working with relatively new AI solutions.

#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.