ASEAN Guide on
AI Governance and Ethics

UCARE.AI

Illustration on operations management - documenting data lineage, ensuring data quality and mitigating bias

UCARE.AI is a Singapore-based deep-tech start-up, with a proprietary award-winning online ML and AI platform built on a cloud-based microservices architecture that provides real-time predictive insights, which can be applied to the healthcare sector and beyond.

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 features selected for each AI model differed for each hospital, contributing to greater accuracy in the bill estimations for patients. 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.

Modelling

Does a regulator require that the organisation can explain how the AI system arrived at its outcomes?

Does the AI system produce repeatable and reproducible results?

Is there proper documentation to track the AI model training and selection process?

During the system development process, developers should assess the approach and evaluate if AI systems are explainable, repeatable, reproducible, and robust. With the complexity of AI systems, it may not always be feasible to achieve all the above. Developers should adopt a risk-based approach to identify which model attributes are more relevant and necessary. There are two areas to consider in the risk-based approach – which features or functions of the model(s) in the AI system have the greatest impact on the users and which measures are likely to help establish more trust with users.

Explainability

Explainability is about explaining how AI systems function and how they arrive at certain decisions. To build trust in AI, it is important that humans understand how AI systems make predictions. However, in some cases such as "black box" models, it can be difficult to explain a model. Nonetheless, developers and deployers of AI systems containing such models can still work towards explainability by explaining how the predictions of AI system are used in the whole decision-making process. Explainability practices that developers and deployers can adopt when developing and deploying AI systems include:

- Auditability refers to the readiness of an AI system to undergo an assessment of its algorithms, data, and design processes. Ensuring proper documentation of the training and selection processes of the AI system, reasons for certain decisions made and measures taken to mitigate any risks found during the risk assessments will help developers keep track and remain accountable for the decisions made during the AI deployment process. Some developers may employ the use of automated machine learning which automates some or all the steps in the AI system lifecycle, such as feature engineering and hyperparameter tuning. Even for the steps that are automatically performed by automated machine learning, developers are encouraged to still consider how to incorporate transparency and explainability in such automated workflows. For example, documenting the data sources, hyperparameters, algorithms, and optimisation techniques initially selected by developers can help developers and deployers better understand how changes to these input parameters can potentially alter the machine learning outcomes.
- Where "black box" models are deployed, rendering it difficult, if not impossible to provide explanations as to the workings of the AI systems, outcome-based explanations, with a focus on explaining the impact of decision-making or results flowing from the AI systems may be relied on. Another method that can be considered is AI Model Cards, which are short documents accompanying trained machine learning models that disclose the context in which models are intended to be used, details of the performance evaluation procedures, and other relevant information[17].
- Deployers may also consider focusing on alternative aspects of the quality of the AI system or preparing information that could help demonstrate and justify the outcomes of an AI system's processing behaviour (e.g., documenting the repeatability of results produced by the AI system). If need be, they can work with developers to develop such information. Some practices to demonstrate repeatability include conducting repeatability assessments to ensure deployments in live environments are repeatable and performing counterfactual fairness testing to ensure that the AI system's decisions are the same in both the real world and in the counterfactual world.
- For AI-enabled products and devices, it is advisable for developers and deployers to include descriptions of the expected behaviour of the AI into accompanying technical specifications or product documentation. Such descriptions can include the rationale behind why certain features or models were selected during the AI system development process. Provision of such

[17] OECD.AI Policy Observatory, "Catalogue of Tools & Metrics for Trustworthy AI" (15 September 2022) <https://oecd.ai/en/catalogue/tools/model-cards>

information helps deployers remain accountable to individuals and/or regulators and helps AI systems become more explainable and transparent. As far as possible, deployers that do not develop in-house AI systems and procure them from developers should appropriately govern their relationships with these developers through contracts that allocate liability in a manner agreed between parties. Deployers can consider including an obligation for developers to help support them in meeting transparency and explainability needs in the contractual terms. This could help deployers obtain the necessary support from third-party developers on retrieving or understanding the relevant information about how the AI system functions and its expected behaviour. An example of information that deployers can obtain from third-party developers would be AI Model Cards, which provide details on how the AI model functions, model development and testing process, and limitations of the model. Some examples of AI Model Cards can be referenced from Google's Face Detection Model Card[18] and Salesforce's Einstein Optical Character Recognition (OCR) Model Card[19].

- Deployers can also collaborate with developers to conduct joint audits and assessments for transparency and explainability.
- Deployers can also use explainability tools to evaluate the quality of their explanations. Some examples of such tools include AI Verify's Toolkit[20], OmniXAI[21], AIX360, Shapley Additive Explanations (SHAP)[22], and Local Interpretable Model-agnostic Explanations (LIME)[23]. It should be noted that most of the tools mentioned provide technical (explicit) explanation and are meant to be used by technical users (i.e., data scientists).

Even with the use of tools to explain AI systems, it may be difficult for the layman, and even experts, to understand exactly how AI systems work. In these cases, deployers can consider using implicit explanations of AI instead. For example, deployers could use comparisons such as "users with similar profiles as you were recommended these similar products" to help users understand to some degree how the AI system uses other users' data and their purchase history to recommend products to them. Counterfactuals may also be a useful means to explain how changes in variables or data affect outcomes of AI systems. For example, informing users that "you would have been approved if your average debt was 15% lower".

Although explainability is generally encouraged to promote transparency and build trust in AI systems, there may be situations where it does not make sense to disclose important information about how AI systems function. For example, the workings of fraud detection AI systems need to remain confidential to the organisation so that bad actors are not able to circumvent them. There also needs to be an appropriate balance between transparency and ensuring that companies' Intellectual Property (IP) is protected. Certain algorithms may be essential to business operations and bottom line, such as trading algorithms for robo-advisors, and disclosing such information will put confidential or proprietary business information at stake.

[18] Google, "Face Detection Model Card v0" < https://modelcards.withgoogle.com/face-detection >

[19] Salesforce, "Salesforce Einstein Model Cards" < https://resources.docs.salesforce.com/latest/latest/en-us/sfdc/pdf/salesforce_ai_model_cards.pdf>

[20] Infocomm Media Development Authority, "Fact Sheet – Open-sourcing of AI Verify and set up of AI Verify Foundation" (2023) < https://www.imda.gov.sg/-/media/imda/files/news-and-events/media-room/media-releases/2023/06/7-jun---ai-annoucements---annex-a.pdf >

[21] Salesforce, "Welcome to OmniXAI's documentation!" (2022) < https://opensource.salesforce.com/OmniXAI/latest/index.html >

[22] SHAP, "Welcome to the SHAP documentation" (2018) < https://shap.readthedocs.io/en/latest/ >

[23] C3.ai, "What is Local Interpretable Model-Agnostic Explanations (LIME)" < https://c3.ai/glossary/data-science/lime-local-interpretable-model-agnostic-explanations/#:~:text=What%20is%20Local%20Interpretable%20Model,to%20explain%20each%20individual%20prediction. >

UCARE.AI

Illustration on explainability in the model lifecycle

UCARE.AI deployed its AI-Powered Cost Predictor (AlgoExpect^{TM}) in Parkway's four Singapore hospitals to provide dynamic real-time predictions of bill size at pre-admission at 82% accuracy.

UCARE.AI has incorporated explainability directly into the AI Cost Predictor model. Along with providing prediction, it is also able to tell on demand what are the important features that contribute to each prediction result. Client applications consuming the model's prediction service are also able to ask the model to "explain" each prediction result without going through UCARE.AI's support team. This helps users to understand the model predictions and provides greater transparency for the model's performance and instils greater trust.

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.

Robustness

Robustness helps establish trust in the performance of AI systems, especially in unpredictable circumstances. Robustness refers to the ability of the AI system to still function as intended or in a safe manner in the face of errors during execution or erroneous input. It can be measured by the extent to which the AI system can function correctly with invalid input or environmental stress. Robustness is an especially important attribute for building trust with users as it shows that the AI system can withstand a range of input environments.

AI systems are only as good as the data used to train and test them. It is difficult for the AI system to be trained on every possible precondition and scenario that it will face, especially when it is deployed in the real-world with dynamic human interactions. When faced with an unfamiliar input, the system might produce insensible or unintended outputs.

Developers and/or deployers can assess robustness by testing the AI system on various foreseeable erroneous input and scenarios. This can be done via adversarial testing, which is a series of tests to expose the system to a broad range of unexpected inputs and mitigate any unintended behaviour before the deployment of the AI system in live environments.

Even with the use of adversarial testing, AI systems are not immune to changes in inputs and operation environments that occur over time. To address this, some deployers or developers may choose to adopt continuous learning practices, where the learned parameters of the AI system are not fixed and can continue to change as the AI system is deployed in the live environment and learns from data it receives. However, it is still important for deployers to closely monitor the AI system and ensure that it does not learn unintended behaviour in the process.

Outcome analysis

How does the organisation determine if the AI system developed is fit-for-purpose?

How does the performance of the AI system compare to industry standards?

After the design and development process, deployers need to confirm that the AI system's outcomes are fit for purpose, achieve the desired level of precision and consistency, and are aligned with ethical, lawful, and fair design criteria.

During the project governance and problem statement definition phase, deployers should have clearly documented the intended purpose of the AI system and assessed that the risks associated with the AI system will be mitigated with appropriate measures. Risk identification and analysis are also necessary to address the root cause of the risks encountered.

In the outcome analysis phase, business stakeholders should be involved to observe the performance of the AI system and validate that it fulfils the purposes laid out at the start. Organisations can also consider conducting acceptance tests, which may cover functional and non-functional aspects, including security and performance evaluations. In some cases, developers and deployers may also choose to compare the actual Return on Investment (ROI) of the AI system against the planned ROI that was estimated during the planning phase.

To facilitate the outcome analysis process, it is important to establish a clear communication channel between the technical team and other stakeholders to ensure mutual understanding of AI system performance and potential improvements.

Fairness testing should also be conducted at this stage to ensure that the AI system does not make decisions that can result in unintended discrimination of certain demographics of users.

UCARE.AI

Illustration on operations management – robust model testing before deployment

UCARE.AI deployed its AI-Powered Cost Predictor (AlgoExpect™) in Parkway's four Singapore hospitals to provide dynamic real-time predictions of bill size at pre-admission at 82% accuracy.

UCARE.AI 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.

Annex B: Use Cases

UCARE.AI
unlock the impossible

[Use case 1] Data Security, Privacy & Transparency

UCARE.AI (https://www.ucare.ai/home/news/) is a Singapore-based deep-tech start-up, with a proprietary award-winning online ML and AI platform built on a cloud-based microservices architecture that provides real-time predictive insights, which can be applied to the healthcare sector and beyond. UCARE.AI's solutions have been used by customers such as Parkway Pantai, Singapore's Ministry of Health, Grab and Great Eastern Life Assurance to manage risk, contain cost and maximise efficiency.

Among their various solutions, UCARE.AI's AI-powered Cost Predictor (AlgoExpect™) works with hospitals to deliver accurate estimations of hospital bills to patients.

Illustration on operations management - good data accountability and adhering to privacy and data protection policies

UCARE.AI invested its 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.

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 PDPA (2012) is 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 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.

The company also actively reinforced its commitment to data protection, cataloguing and evaluating every use of data that could be accessed by clients.

Illustration on operations management - documenting data lineage, ensuring data quality and mitigating bias

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 features selected for each AI model differed for each hospital, contributing to greater accuracy in the bill estimations for patients.

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.

Illustration on Transparency 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. 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.

Illustration on stakeholder interaction and communication

Clients with concerns about bill predictions were also encouraged to highlight them through UCARE.AI's communication channels. This gave stakeholders the necessary assurance on UCARE.AI's policies and processes for responsible AI use.

[Use case 2] Oversight, Validation & Monitoring

UCARE.AI, a deep-tech start-up with a proprietary award-winning online ML and AI platform built on a cloud-based microservices architecture that provides real-time predictive insights to help insurers, hospitals, pharmacies and governments manage risk, reduce medical cost, and maximise efficiency, with the end goal of making healthcare affordable to all.

UCARE.AI deployed its AI-Powered Cost Predictor (AlgoExpect™) in Parkway's four Singapore hospitals to provide dynamic real-time predictions of bill size at pre-admission at 82% accuracy. Based on this success, Parkway launched the Price Guarantee Programme for six hospital procedures. In its commitment to help patients 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.

Illustration on internal governance structures and measures - defining clear roles for internal oversight of AI

As a critical part of AI governance is oversight, UCARE.AI put in place certain internal governance measures, which includes **assigning clear roles and responsibilities for the ethical development and deployment of AI**. UCARE.AI's projects all include a primary and secondary data science lead to concurrently develop AI models for the same problem statement. Once completed, the data science leads would present their results to UCARE.AI's internal team, which consists of the CEO, CTO, CSO, project managers and the client services team for validation. During 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 QA team reviewed and ensured that the model was production-ready before deployment.

Illustration on operations management - robust model testing and continuous monitoring of deployed AI models

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.

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 AlgoExpect™'s predictive insights.

Illustration on Explainability in the model lifecycle

UCARE.AI has incorporated explainability directly into the AI Cost Predictor model. Along with providing prediction, it is also able to tell on demand what are the important features that contribute to each prediction result. Client applications consuming the model's prediction service are also able to ask the model to "explain" each prediction result without going through UCARE.AI's support team. This helps users to understand the model predictions and provides greater transparency for the model's performance and instils greater trust.