INFOCOMM MEDIA DEVELOPMENT AUTHORITY

Fact Sheet

1 June 2022

IMDA Invites Companies to Pilot and Contribute to International Standards Development as Part of Next Step in Singapore's AI Governance Work

Singapore has launched A.I. Verify – the world's first AI Governance Testing Framework and Toolkit Minimum Viable Product (MVP) for companies who want to demonstrate responsible AI in an objective and verifiable manner. A.I. Verify aims to promote transparency between companies and their stakeholders through a combination of technical tests and process checks. The MVP was launched by Singapore's Minister for Communications and Information Mrs Josephine Teo at the World Economic Forum Annual Meeting in Davos on 25 May 2022.

Background

Artificial Intelligence ("AI") has been identified as a key step of Singapore's Smart Nation journey. While AI bring about benefits, its risks have been an on-going discussion in international fora and among governments, international organisations, industry, academia, and civil society. As more companies use AI in their products and services, fostering public's trust in AI technologies remains key in unlocking the transformative opportunities of AI. There are public concerns with issues relating to AI systems' transparency, explainability, safety, robustness, unintended bias, and accountability. Testing for the trustworthiness for AI systems remains an emergent space globally.

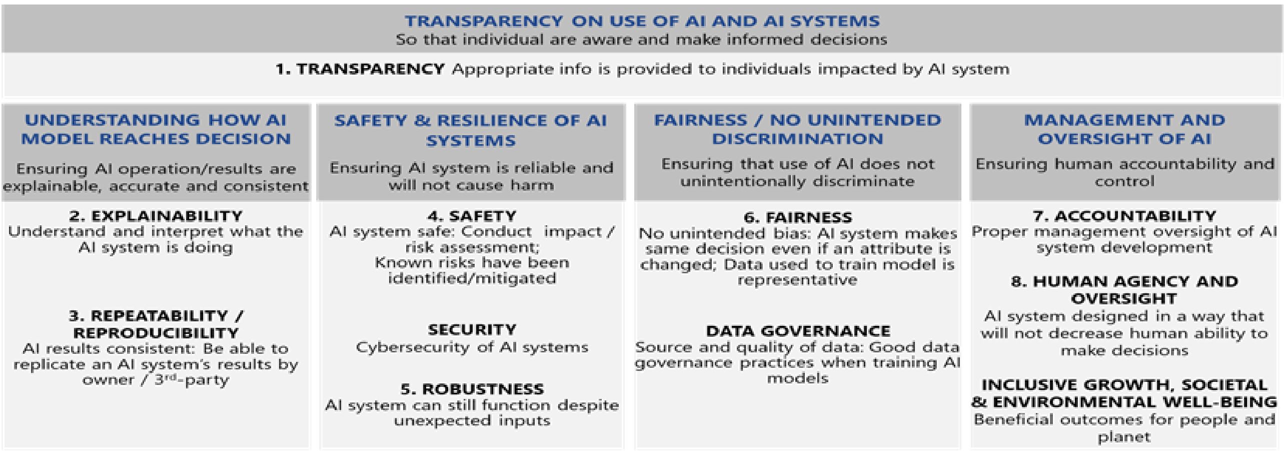
The launch of A.I. Verify follows Singapore's launch of the Model AI Governance Framework (second edition) in Davos in 2020, and the National AI Strategy in November 2019. Having provided practical guidance to industry on implementing responsible AI, A.I. Verify is Singapore's next step in helping companies be more transparent about their AI products and services, to build trust with their stakeholders.

Development of an AI Governance Testing Framework and Toolkit

The aim of the Testing Framework and Toolkit (A.I. Verify) is to help AI system developers and owners be more transparent about their systems by verifying the performance of their AI systems against a set of AI ethics principles through a combination of technical tests and process checks. Globally, countries are coalescing around 11 key AI ethics principles grouped into 5 pillars (See Figure 1 below)

The approach is to allow transparency, of what the AI model claims to do vis-à-vis the test results, and covers areas such as:

a) Transparency:
   i. On the use of AI to achieve what stated outcome
   ii. Understanding how the AI model reaches a decision
   iii. Whether the decisions predicted by the AI show unintended bias
b) Safety and resilience of AI system
c) Accountability and oversight of AI systems



TRANSPARENCY ON USE OF AI AND AI SYSTEMS
So that individual are aware and make informed decisions

1. TRANSPARENCY Appropriate info is provided to individuals impacted by AI system

UNDERSTANDING HOW AI MODEL REACHES DECISION Ensuring AI operation/results are explainable, accurate and consistent	**SAFETY & RESILIENCE OF AI SYSTEMS** Ensuring AI system is reliable and will not cause harm	**FAIRNESS / NO UNINTENDED DISCRIMINATION** Ensuring that use of AI does not unintentionally discriminate	**MANAGEMENT AND OVERSIGHT OF AI** Ensuring human accountability and control
2. EXPLAINABILITY Understand and interpret what the AI system is doing **3. REPEATABILITY / REPRODUCIBILITY** AI results consistent: Be able to replicate an AI system's results by owner / 3rd-party	**4. SAFETY** AI system safe: Conduct impact / risk assessment; Known risks have been identified/mitigated **SECURITY** Cybersecurity of AI systems **5. ROBUSTNESS** AI system can still function despite unexpected inputs	**6. FAIRNESS** No unintended bias: AI system makes same decision even if an attribute is changed; Data used to train model is representative **DATA GOVERNANCE** Source and quality of data: Good data governance practices when training AI models	**7. ACCOUNTABILITY** Proper management oversight of AI system development **8. HUMAN AGENCY AND OVERSIGHT** AI system designed in a way that will not decrease human ability to make decisions **INCLUSIVE GROWTH, SOCIETAL & ENVIRONMENTAL WELL-BEING** Beneficial outcomes for people and planet

Figure 1

The five pillars describe how system owners and developers can build trust with customers and consumers by demonstrating the following:

a) Transparency on use of AI & AI system: By disclosing to individuals that AI is used in the system, individuals will become aware and can make an informed choice of whether to use the AI-enabled system.

b) Understanding how an AI model reaches a decision: This allows individuals to know the factors contributing to the AI model's output, which can be a decision or a recommendation. Individuals will also know that the AI model's output will be consistent and performs at the level of claimed accuracy given similar conditions.

c) Ensuring safety and resilience of AI system: Individuals know that the AI system will not cause harm, is reliable and will perform according to intended purpose even when encountering unexpected inputs.

d) Ensuring fairness i.e., no unintended discrimination: Individuals know that the data used to train the AI model is sufficiently representative, and that the AI system does not unintentionally discriminate.

e) Ensuring proper management and oversight of AI system: Individuals know that there is human accountability and control in the development and/or deployment of AI system and the AI system is for the good of humans and society.

Scope and Limitations of Testing Framework and Toolkit

As we are in the early stages of development and iteration, the Toolkit currently has the following features and limitations:

a) Works with commonly used AI frameworks and software library such as, TensorFlow, XGBoost, and Scikit-Learn;

b) Is able to test most supervised AI, i.e., common AI models such as classification and regression, but not deep learning models; and

c) Can handle tabular and selected image datasets.

IMDA/PDPC will continue to work with industry, including technology solutions providers, to enhance the applicability of the Testing Framework and Toolkit to a broader range of AI models and systems. We are inviting companies to join us and participate in the MVP. Beyond the pilot stage of the MVP, Singapore aims to work with AI system owners/developers globally to collate and build industry benchmarks. This enables Singapore to continue to contribute to the development of international standards on AI governance.

About Infocomm Media Development Authority

The Infocomm Media Development Authority (IMDA) leads Singapore's digital transformation by developing a vibrant digital economy and an inclusive digital society. As Architects of Singapore's Digital Future, we foster growth in Infocomm Technology and Media sectors in concert with progressive regulations, harnessing frontier technologies, and developing local talent and digital infrastructure ecosystems to establish Singapore as a digital metropolis.

For more news and information, visit www.imda.gov.sg or follow IMDA on Facebook (IMDAsg) and Twitter (@IMDAsg).

For media clarifications, please contact:

Choo Hong Xian (Mr)
Manager, Communications and Marketing, IMDA
DID: (65) 6955 0221
Email: choo_hong_xian@imda.gov.sg