

Singapore's Approach to Building Trustworthy AI

YEONG, Zee Kin

Assistant Chief Executive (Data Innovation & Protection)
Infocomm Media Development Authority

Deputy Commissioner
Personal Data Protection Commission Singapore

- Singapore's National AI Strategy (NAIS)
- Different responses to AI risks to build consumer trust/help industry meet int'l regs/standards
- Singapore's current approach to AI governance to support NAIS
- Latest development in AI governance testing
- Enabling use of data and PET for AI development

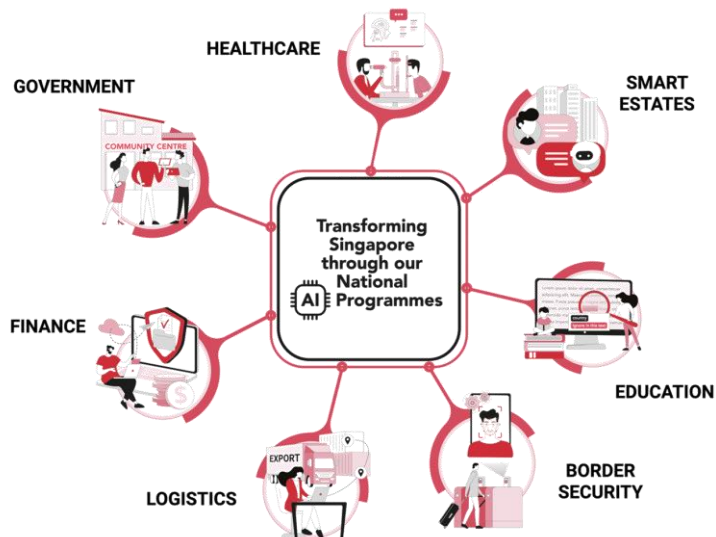
Presentation Overview

AI Governance Supports Singapore's National AI Strategy (NAIS)

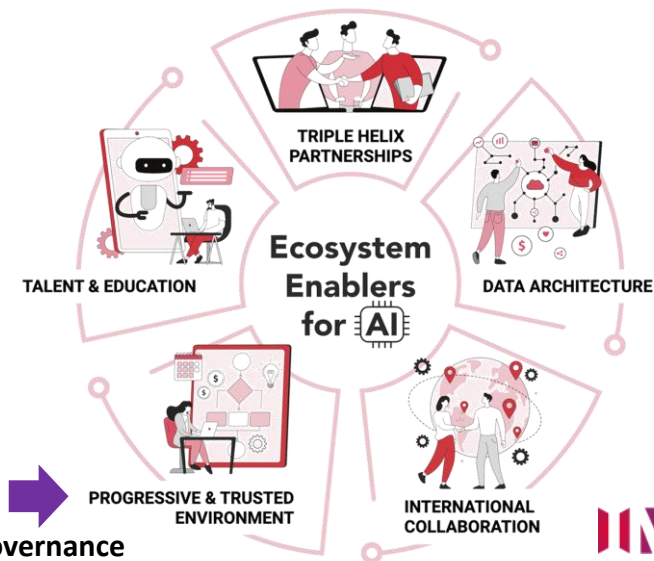
VISION

By 2030, Singapore will be a leader in developing and deploying scalable, impactful AI solutions, in key sectors of high value and relevance to our citizens and businesses

7 NATIONAL AI PROJECTS



5 ECOSYSTEM ENABLERS



AI Governance

Different Responses to AI Risks

To build consumer trust, help biz meet intl regs/standards

Consumer concerns with autonomous decision making by AI

BIAS & UNINTENDED DISCRIMINATION

Unfavorable decisions about individuals which may affect their lives / livelihoods

HARM

Harm resulting from AI misbehaving or non-proper use of AI

RECOURSE

Channels for review of decision, or ability to obtain compensation (liability)

Governments inching towards regulations (soft and hard options)

PRINCIPLES & GUIDELINES

- Ethical Principles: e.g. Australia, South Korea, China
- Guidance: e.g. Singapore, Japan, US

REGULATIONS

- EU: high-risk AI systems to undergo conformance assessment
- China: law to regulate automated decision making

INT'L PLATFORMS

- OECD Principles on AI
- UNESCO Recommendation on Ethics of AI
- GPAI responsible AI practices

Industry demonstrating responsible AI practices to regulators & stakeholders

AI GOVERNANCE TOOLS

- Big Tech: E.g. IBM Fairness 360, Google What-If, Microsoft Fairlearn
- Start-ups: E.g. Truera, Credo AI, 2021.AI, Tookitaki

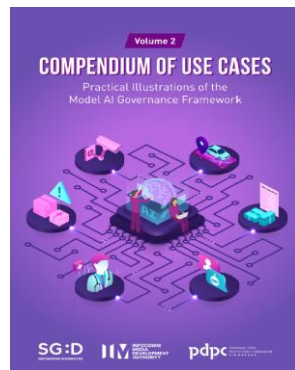
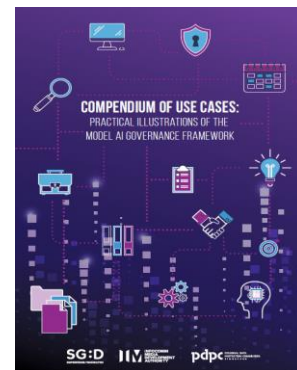
STANDARDS

- MAS Veritas
- NIST AI Risk Management F/w
- ISO SC 42, IEEE P7000 series, Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)

Singapore's Current Approach to AI Governance

Soft regs - voluntary guidance on trustworthy AI implementation

- Industry **voluntary** adoption of responsible AI with detailed **government guidance**
 - Accountability-based approach to engender trust
 - Takes AI ethics into corporate governance, risk management and operational structures
 - Practical tools for organisations
- **Government guidance**
 - Model AI Governance Framework (2nd edition)
 - Implementation & Self-Assessment Guide for Organisations
 - 2 volumes of Compendium of Use Cases
- **Multi-stakeholder approach**
 - Living documents that will evolve with tech development
 - Input and feedback from >60 companies of different sizes, from different sectors, locally and internationally
 - Worked with int'l organizations like WEF, OECD
 - Baseline for other sectors to build on



Beyond guidance - A.I. Verify MVP Helps Companies be Transparent about AI

Framework & software tool to conduct objective, verifiable tests and record process checks

INTERNATIONALLY-ALIGNED AI ETHICS PRINCIPLES CATEGORISED INTO 5 KEY AREAS CONCERNING AI SYSTEMS

TRANSPARENCY ON USE OF AI AND AI SYSTEMS

Ensuring consumer awareness on use and quality of AI systems

► **Transparency**

UNDERSTANDING HOW AI MODEL REACHES DECISION

Ensuring AI operation/ results are explainable, accurate and consistent

► **Explainability**

► **Repeatability/Reproducibility**

SAFETY & RESILIENCE OF AI SYSTEMS

Ensuring AI system is reliable and will not cause harm

► **Safety**

● **Security**

► **Robustness**

FAIRNESS/NO UNINTENDED DISCRIMINATION

Ensuring that use of AI does not unintentionally discriminate

► **Fairness**

● **Data governance**

MANAGEMENT AND OVERSIGHT OF AI

Ensuring human accountability and control

► **Accountability**

► **Human agency and oversight**

● **Inclusive growth, societal and environmental well-being**

► **8 principles in MVP**

KEY FEATURES

- **Cover key intl AI governance frameworks & guidelines**
 - EU Ethics Guidelines for Trustworthy AI, OECD Recommendation on AI, SG Model Framework
- **Validate companies' claims about AI systems' performance**
 - Does not set ethical standards
- **Single, integrated toolkit for self-test**
 - Ease of testing / recording process checks
 - Mitigates companies' concern about of commercial sensitivity
 - Working towards indep 3rd party testing svc
- **Customised testing reports to be available for diff groups of stakeholders**
 - Internal management, external partners, regulators, customers

Scope of A.I. Verify tests

Principle	Technical Test	Process Checks
Transparency		Evidence (e.g., policy, comms collaterals) of providing appropriate info to individuals who may be impacted by the AI system – intended use, limitations, risk assessment (w/o comprising IP, safety, system integrity)
Explainability	Factors contributing to AI model's output	Evidence of considerations given to choice of AI models
Repeatability/ reproducibility		Evidence of AI model provenance and data provenance
Safety		Evidence of materiality/risk assessments, identification/mitigation of known risks, evaluation of acceptable residual risks
Robustness	Model performs as expected even when encountering unexpected input	Evidence of review of factors that may affect the performance of AI model, including adversarial attack
Fairness	Protected/sensitive attributes specified by AI system owner do not contribute to algo bias of model by checking model output against ground truth	Evidence of strategy for selecting fairness metrics, definition of sensitive attributes are consistent with legislation & corporate values
Accountability		Evidence of clear internal governance mechanisms for proper management oversight of AI system's development/deployment
Human agency & oversight		Evidence that AI system is designed in a way that will not reduce human's ability to make decisions or to take control of the system (e.g.,human-in-the-loop)

A.I. Verify International Pilot to help enhance MVP & build industry benchmarks

- Invite
 - AI system owners/developers to pilot MVP
 - Provide feedback
 - Co-create industry benchmarks
 - Technology solution providers to build capabilities in AI governance testing
 - Framework owners/developers to explore interoperability
- 40+ companies internationally have expressed in the pilot

Find out more at
go.gov.sg/aiverify



Building AI Testing Community and sustainable Ecosystem

- **All international pilot participants are part of AI Testing Community**
 - Regulators/policymakers to share early policy thinking and seek industry feedback
 - Use case-specific workshops to build benchmarks
- **Build ecosystem to support AI governance testing**
 - 3rd-party testing service providers
 - Advisory service providers/consultancies
 - Certification body/bodies
 - Research community/tech solution providers

Enabling AI development with 'Privacy Enhancing Technologies'

- **AI systems need data but growing challenges exist in data flow**
 - Businesses concerned about losing competitive edge or risk of non-compliance if they share data
 - Public agencies and non-commercial entities holding sensitive data have stringent data governance policies which undercut their agility to adopt new AI solutions
 - Individuals can be hesitant to give consent for fear of losing privacy without fully knowing where the data is going or how it would be used
- **PET could enable flow of insights from data without disclosure of the data itself**
e.g.
 - Identifying the best features for an AI model without revealing the full dataset
 - Tuning & training model weights of an AI model without pooling data centrally
 - Testing performance metrics of an AI solution on new but confidential dataset

Definition of PET

- “PET” is an umbrella term for a group of technologies, no universal definition
- A subset of PETs are more mature

OECD/
MIT

Studying PETs from standpoint of
fulfilling its Privacy Guidelines

Canada
Privacy
Commission

Define as tech that
protects data in transit

UK CDEI

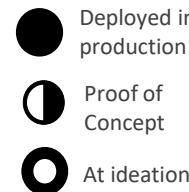
Define as tech that protects privacy or
confidentiality of sensitive
information



★ Identified as more mature
in OECD's prelim findings




PET Use Cases and State of Adoption

- 3 archetypes of commonly featured use case applications






Relevant to AI development




1 Identify Common Users

- Outcome: Count size of consumer base common to two or more coys
- Examples:
 -  Apple uses MPC to detect child abuse materials in iCloud images
 -  Temasek-led datathon to fuse public-private datasets
 -  A Singapore based mobility company and insurance firm wish to use to derive basic consumer insights
- Applicable PETs: Multiparty Computing, Anonymisation, Differential Privacy

2 Add features to enrich data

- Outcome: A data model with more features to derive new insights
- Examples:
 -  Google uses MPC to offer top ad buyers a custom Keyword Planner built on joint 1P data
 -  DataCo (ANZ Bank) building MPC-based platform to build data models w/ airlines, retailers
 -  A Singapore based Adtech firm is exploring Diff Privacy to achieve “cookie-less” tracking for brand owners & publishers
- Applicable PETs: Differential Privacy, Multiparty Computing

3 Make more data available for AI

- Outcome: Develop an AI model by either obfuscating data before ingestion or by flowing model weights instead of data
- Examples:
 -  Android device keyboards employ Fed Learning to train next word recommendation AI
 -  A Financial services MNC tried HE on image data to see performance of an open source computer vision AI
 -  A Canadian asset manager used MPC to access portfolio companies' data for its risk assessment AI
- Applicable PETs: Homomorphic Encryption, Fed Learning, Multiparty Computing

IMDA/PDPC's PET Sandbox

- Catalyze adoption by gaining hands-on experience with PETs
- When applied to real-world use cases
- To understand technical and regulatory bounds with Proofs-of-Concept (POC)

1. Use Case Owners

- Singapore registered organisations
- See delta value if the use case POC goes into production
- May or may not have an existing solution provider
- Submit essential details at go.gov.sg/petsandbox

2. Key Requirements

- Use case falls under at least 1 of 3 types
- If selected, draft detailed proposal
- Demo to IMDA the POC use case within 6 months
- Regular discussions with IMDA to extract lessons about adoption

3. Access to Funding and Tech

- Up to 50% of the cost to scope and develop POC
- On reimbursement basis *after* demo of POC
- Linkup to PET solution providers if necessary

4. Regulatory Guidance

- Regular consultations to raise questions to PDPC about compliance
- Questions specific to PET use case

Thank you