

Annex C

Fact Sheet - DISCUSSION PAPER ON GENERATIVE AI – IMPLICATIONS FOR TRUST AND GOVERNANCE

Generative AI is uncovering a myriad of use-cases and opportunities that are reshaping industries, revolutionising sectors and driving innovation. **At the same time, concerns have emerged**; from the risk of AI making gaffes to worries that it will take over the world. Amidst global discussions on Generative AI, IMDA, together with Aicadium, have co-written a discussion paper to share Singapore’s approach and ideas for building an ecosystem for trusted and responsible adoption of Generative AI, in a way that spurs innovation and taps on its opportunities.

The paper considers various methods of assessing the risks of Generative AI and approaches towards AI Governance. It serves as a starting point for policy makers who wish to ensure that Generative AI is used in a safe and responsible manner, and that the most critical outcome – trust – is sustained.

The discussion paper is available [here](#).

Overview of the Paper

The paper identifies **six key risks** that have emerged from Generative AI - 1) mistakes and hallucinations, 2) privacy and confidentiality, 3) disinformation, toxicity and cyber-threats, 4) copyright challenges, 5) embedded bias and, 6) values and alignment.

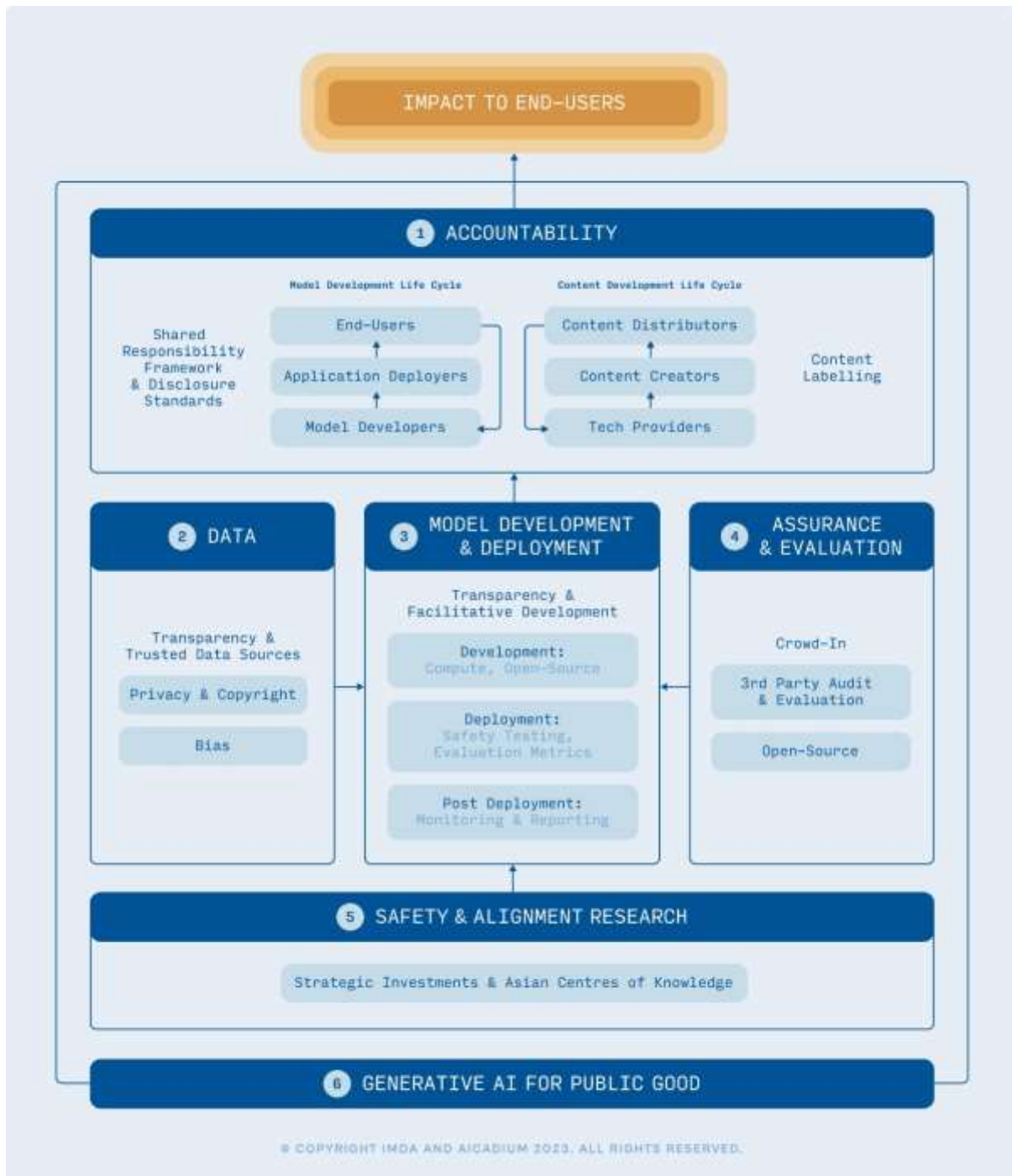
To address these challenges, Singapore adopts a **practical, risk-based and accretive** approach towards the governance of Generative AI by **building on existing AI governance principles**, such as those adopted by OECD, NIST and IMDA. Singapore’s Model AI Governance Framework, for example, is based on the key governance principles - transparency, accountability, fairness, explainability, and robustness. While these principles and practices are applicable regardless of the types of AI deployed, policy adaptations will, nevertheless, be needed to consider the unique characteristics of Generative AI:

- Generative AI will increasingly form the **foundation** upon which other models/applications are built and there are concerns that problems inherent in these models could lead to wider issues . Governance frameworks will have to provide guidance on accountability between parties and across the development lifecycle, as well as address safety concerns in model development and deployment.
- These models are **generative** it may be increasingly difficult to distinguish AI-generated content - and people may become more susceptible to misinformation and online harms. As AI potentially surpasses human capacity at some levels, there are also concerns around how to control and align AI models.

The Discussion Paper identifies six dimensions that should be looked at in totality to enable a trusted and vibrant ecosystem of Generative AI. They are (1) **Accountability** as the basis for governance; Critical components of the model lifecycle from (2) **data** to (3) **model development and deployment** to (4) 3rd party **assurance and evaluation**; Longer term (5) **safety and alignment**; and (6) **Generative AI for public good** so that no one is left behind.

Collectively, the six dimensions provides an initial framework for policy makers to (i) strengthen the foundation of AI governance provided by existing frameworks to address the unique characteristic of Generative AI, (ii) address immediate concerns and (iii) invest for longer-term governance outcomes.

The specific ideas (e.g. shared responsibility framework, disclosure standards) also seek to foster **global interoperability**, regardless of whether they are adopted as hard or soft laws.



Considerations for Regulation

There are practical considerations regarding the implementation and effectiveness of AI regulations. Technical tools, standards and technology to support regulatory implementation need to be ready before regulation can be effective. These are mostly still under development today.



Amidst the pressure to regulate, it is useful for governments to consider whether existing laws, such as sectoral legislation and data protection laws, can be tapped on and updated if necessary. Strongly interventionist regulations should also be carefully considered to tread the balance between risk mitigation and market innovation. For example, overly restrictive regulation on open-source models can stifle innovation by hindering collaboration and access. **Careful deliberation and a calibrated approach should be taken, while investing in capabilities and development of standards and tools.**

Singapore aspires to be a global leader in harnessing AI for the public good. We support the responsible development and deployment of AI so that its benefits may be enjoyed in a trusted and safe manner. If not properly governed, there is a risk that AI could be used for harms such as scams, cyberattacks, and misinformation.

We will continue to keep on top of the development of AI and will make **timely introduction and updates of targeted measures to uphold trust and safety in digital developments.** While Singapore does not currently intend to implement general AI regulation, the discussion paper itself is an example of how Singapore has taken concrete action to develop technical tools, standards and technology, which in turn lays the groundwork for clear and effective regulation in the future.

-END-

About Aicadium

Aicadium is a global technology company delivering AI-powered industrial computer vision products into the hands of enterprises. With offices in Singapore and San Diego, California, and an international team of data scientists, engineers, and business strategists, Aicadium is operationalizing AI within organizations where machine learning innovations were previously out of reach. As Temasek's AI Centre of Excellence, Aicadium identifies and develops advanced AI technologies, including areas of AI governance, regulation, and the ecosystem developments around AI assurance.