CITY OF MONASH

AI Governance: Policy & Practice

June 14, 2024

Adil Khan, Chief Information Officer

AGENDA

- AI Risks and Mitigation plan
- AI Policy
- AI Governance structure
- Key Takeaways

Key risks related to Generative AI

Data security and privacy

- Loss of data confidentiality and data integrity stemming from inputting sensitive data into the AI system or using unverified outputs from it.

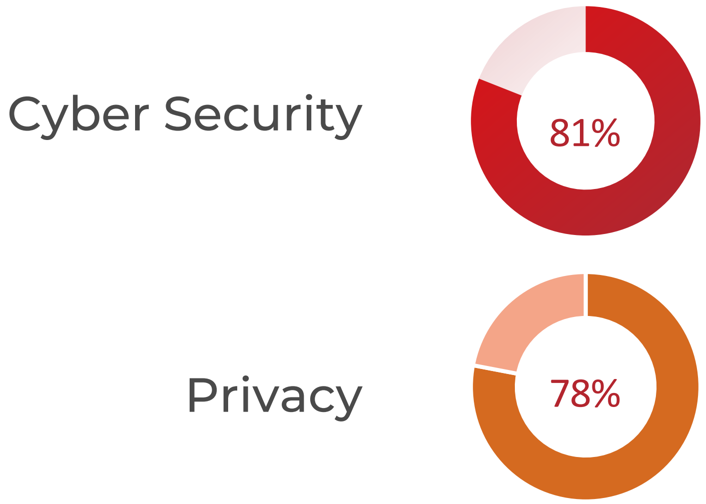
Data confidentiality

- Care must be taken when choosing whether or not to enter a given data into an AI system.
- Problems may arise if the AI model is trained using Personally Identifiable Information (PII) or Protected health information (PHI), and such information may appear in a generative AI output.

Data integrity

- The risk for data integrity comes from repeatedly using unverified generative AI outputs. A single output with faulty data may not cause much of an issue, but these low-quality outputs may end up compromising the integrity of your records over time.

Top-of-mind generative AI concerns for IT leaders



– Source: KPMG, "U.S. Survey," 2023

AI Risk Mitigation Plan: Key initiatives

Phase 1
Explore the Possibilities

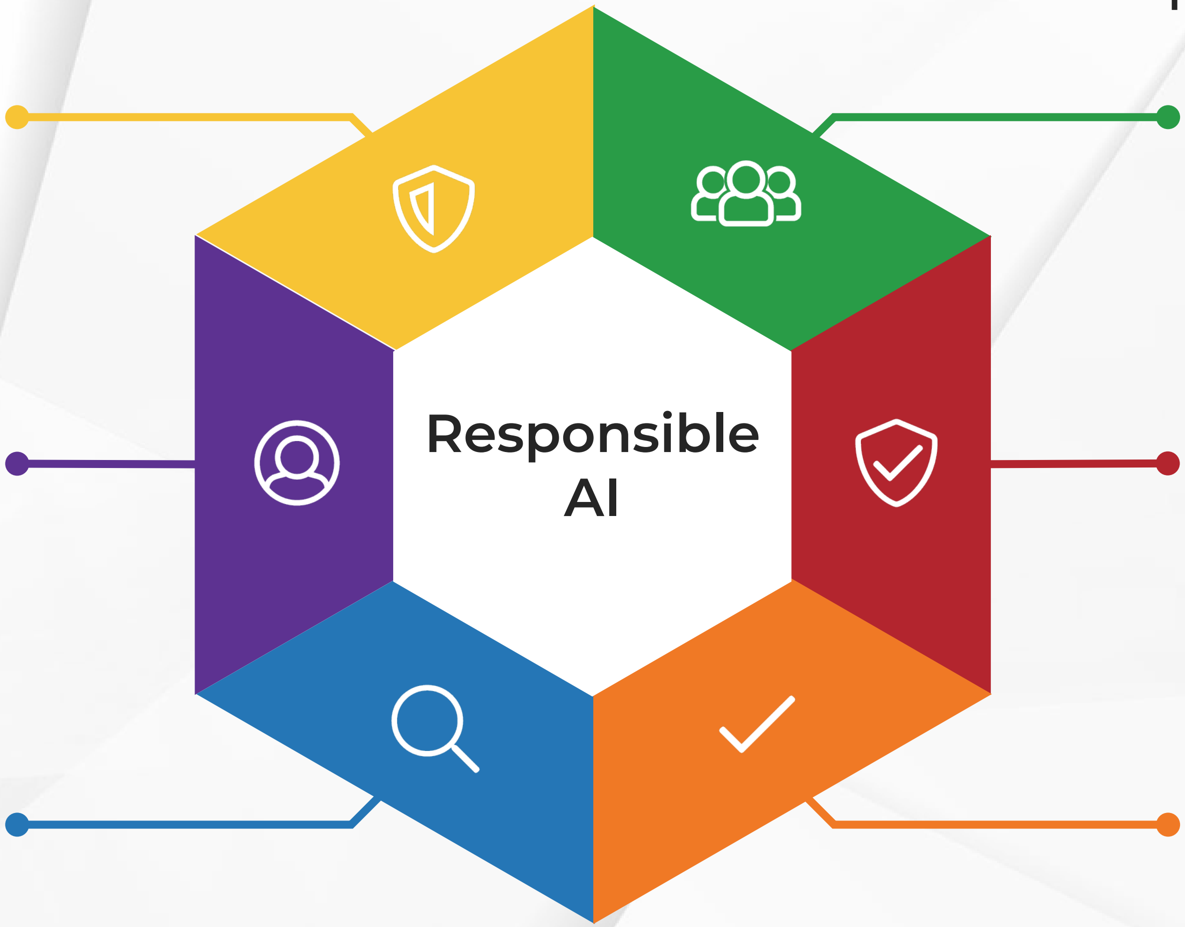
- **ChatGPT:** Restrict the use of ChatGPT.
- **AI Policy:** Develop an AI policy that aligns with the Six AI Principles.
- **MS Copilot:** Use Microsoft Azure OpenAI Service – Enterprise grade service, Microsoft Copilot.
- **Training:** Provide AI training to Employees on MS Copilot including prompt engineering.
- **MS 365 Copilot:** Expand the MS Copilot use to include MS 365 Copilot.

AI Governance frameworks globally

This list highlights global regulatory efforts to establish comprehensive policies for the ethical and responsible development and use of AI.

- **Organisation for Economic Co-operation and Development (OECD) AI Principles - May 2019**
  The OECD introduced these principles to promote innovative and trustworthy AI .
- **US Executive Order on Maintaining American Leadership in AI - February 2019**
  This order aimed to ensure the US remains a leader in AI through coordinated federal initiatives .
- **Singapore's Model AI Governance Framework - January 2020**
  Singapore published this framework to guide companies in responsible AI use.
- **UK National AI Strategy - September 2021**
  The UK outlined its approach to advancing AI innovation while ensuring safety and ethical use.
- **Japan’s AI Strategy 2022 - June 2022**
  Japan updated its AI strategy to foster innovation while addressing societal challenges.
- **NIST AI Risk Management Framework - January 2023**
  This framework provides guidelines to manage risks associated with AI, emphasising trustworthiness and accountability.
- **US Executive Order on Safe, Secure, and Trustworthy Development and Use of AI - October 2023**
  This order aims to set obligations for AI safety and trustworthiness across federal and state levels.
- **UK Pro-innovation Approach to AI Regulation - March 2023**
  The UK introduced a flexible framework to encourage AI innovation while managing risks.
- **Australia's AI Ethics Principles - May 2023**
  Australia's 8 Artificial Intelligence (AI) Ethics Principles are designed to ensure AI is safe, secure and reliable.
- **EU AI Act (final approval) - March 13, 2024**
  The EU formally approved the AI Act, with implementation from late 2024 onwards.

AI POLICY: Aligns With Six Key Principles



Data Privacy

AI applications must respect user privacy. Data must not be used outside of agreed upon terms and must be compliant with privacy norms and regulations.

Human Oversight & Accountability

AI policies must outline the human-in-the-loop or groups accountable for the planning and deployment of any AI system and must document how it will be governed.

Explainability & Transparency

AI applications will be transparent about how data is used and will provide users and key stakeholders insights into how outcomes are produced.

Fairness & Bias Detection

AI applications must include checks and balances to ensure results are unbiased and there is fair and equitable representation across users.

Security & Safety

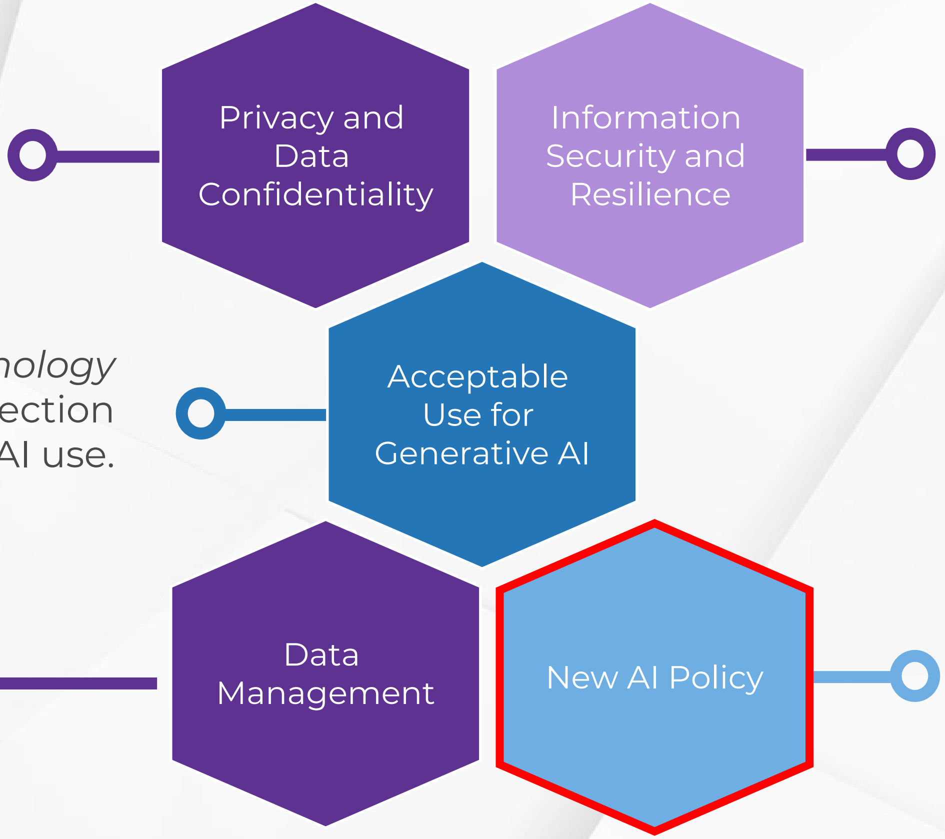
AI applications must be resilient to attacks and other risks that could provide physical or digital detriment to individuals or groups.

Validity & Reliability

AI applications must produce results that are accurate and consistent to mitigate AI risk and foster trust in the application.

AI Policy Framework

The following policy areas will be supported by a new policy and updates to existing policies.



Privacy and Data Confidentiality

Information Security and Resilience

Acceptable Use for Generative AI

Data Management

New AI Policy

Update our *Privacy Policy* to establish how PII data can be used by AI models.

Update our *Data & Information Security, Identity and Access Management*, and other policies to reflect data protection requirements for AI systems (e.g. model versions, training data).

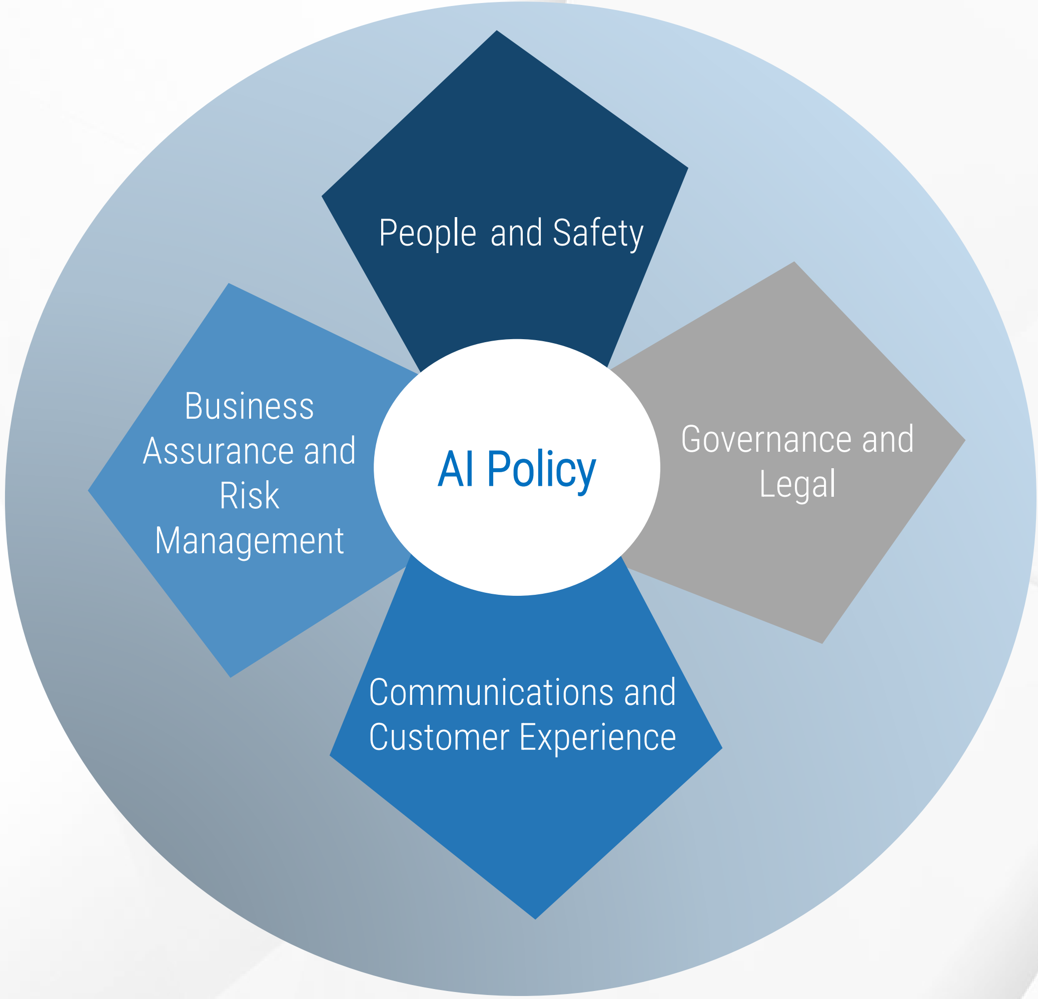
Update our *Digital & Technology Acceptable Use* policy with a section on acceptable generative AI use.

Update our existing data management policies and standards to reflect any new AI requirements.

Create a new Council AI Policy to ensure legal and ethical use of AI technology.

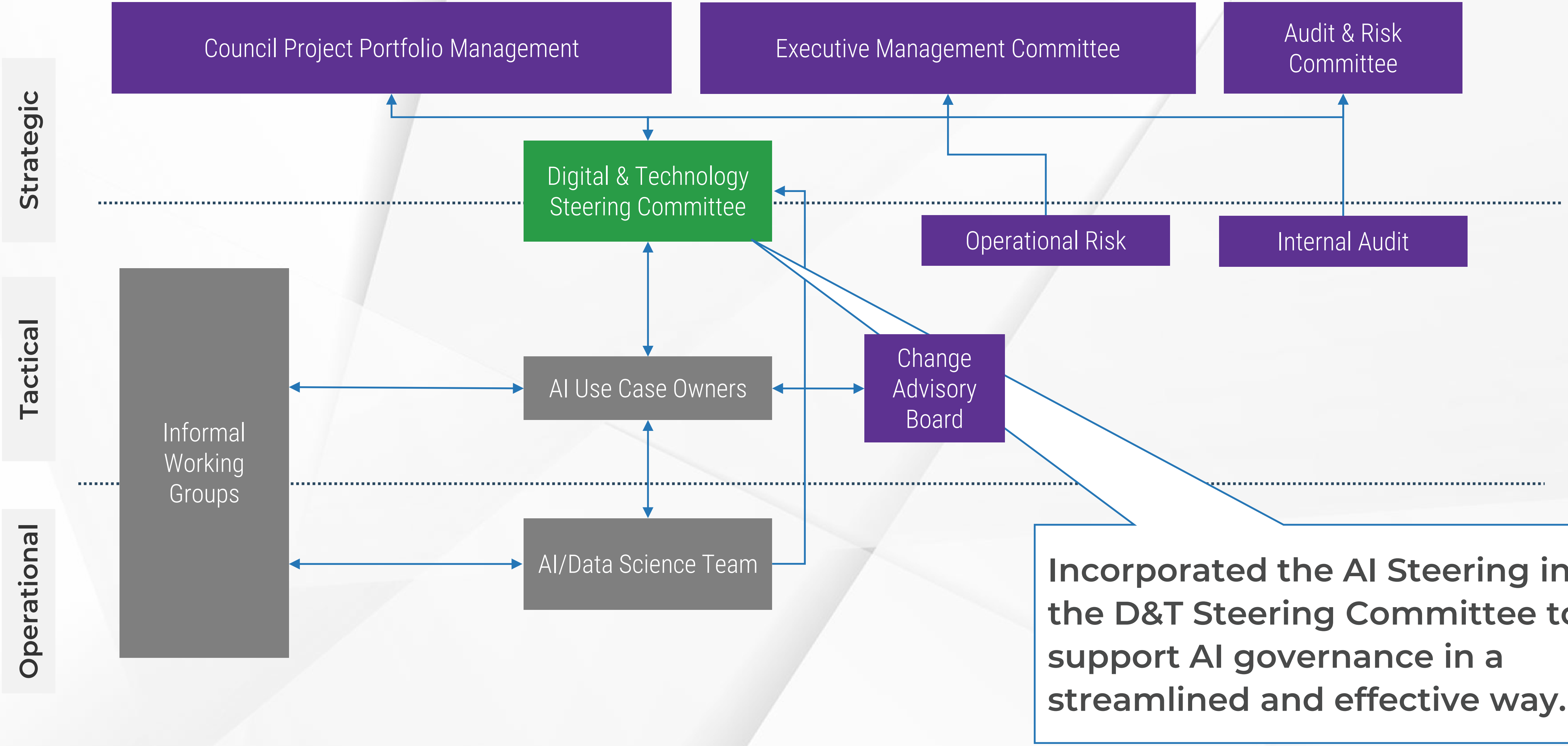
AI POLICY: Working group

We collaborated with four business areas to develop the new AI policy.



People and Safety
Business Assurance and Risk Management
AI Policy
Governance and Legal
Communications and Customer Experience

AI Governance structure



Incorporated the AI Steering into the D&T Steering Committee to support AI governance in a streamlined and effective way.

AI Governance Roadmap: Next Steps

Phase 1
Explore the Possibilities

ChatGPT: Restrict the use of ChatGPT.

AI Policy: Develop an AI policy that aligns with the Six AI Principles.

MS Copilot: Use Microsoft Azure OpenAI Service – Enterprise grade service, Microsoft Copilot.

Training: Provide AI training to Employees on MS Copilot including prompt engineering.

MS 365 Copilot: Expand the MS Copilot use to include MS 365 Copilot.

Phase 2
First AI PoC (Private AI System)

Data Loss Prevention: Implement a DLP (Data Loss Prevention) Solution This will provide complete control over sharing sensitive and personally identifiable information (PII) on MS Copilot or any other Gen-AI tool

Risk Management: Integrate AI risk management with the Council risk management framework.

Private AI Pilot: Document dependencies and determine feasibility.
Select first PoC.

Phase 3
Data Governance

Strategy: Develop a new Data & AI Strategy.

AI Policy: Incorporate the learning from the two phases and update the AI policy.

Key Takeaways

- Use AI Policy and Governance best practices for better AI outcomes.
- Think of your AI governance policy as an insurance policy for the modern enterprise.
- Align your AI governance with AI principles to best prepare for new regulations as they develop.
- AI policies should not be static documents; they should dynamically change as tooling, user expectations, industry trends and regulations, and other factors shift over time.
- AI Impact Assessments – your new best friend!
    - Just like Privacy Impact Assessments (PIAs), an AI Impact Assessment (AIA) is all about systematically identifying potential risks and then working out appropriate ways to mitigate them.

Questions ?