Human Technology Institute

The Rise of Agentic AI: harnessing its power and navigating the risks

Women in Tech 2026

Gaby Carney, 26 February 2026

UTS Human Technology Institute

HTI is a multidisciplinary team with lawyers, policy experts and data scientists. HTI provides independent expert advice, policy development, tools, training, and data science solutions to support human-centred technology.

The AI Corporate Governance Program works across sectors to build capability within Australian organisations in AI governance.

We publish guidance on best practice AI governance, and we advise private and public sector organisations in setting up their AI governance frameworks.

Gaby Carney, Senior Fellow, Strategic AI

Gaby advises organisations on AI governance; publishes best practice guidance to the market; and delivers workshops and seminars on responsible AI.

Before joining HTI, Gaby was a Partner at international management consulting firm, Partners in Performance; a senior executive in the NSW Government where she oversaw business intelligence, policy and legal functions; and a commercial lawyer.

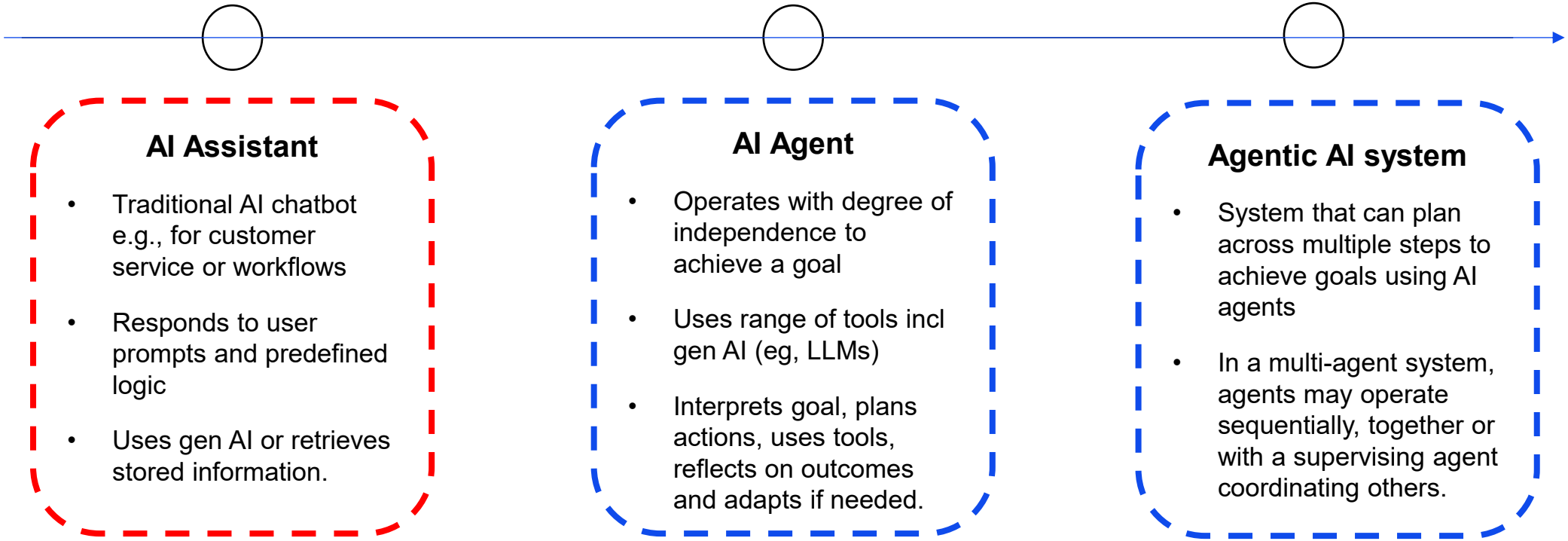
Presentation outline

1. Agentic AI: what is it and how is it being used in practice?
2. When the agent goes rogue: managing risks with agentic AI
3. How can we leverage agentic AI responsibly?
4. Q&A

01

Agentic AI: what is it and how is it being used in practice?

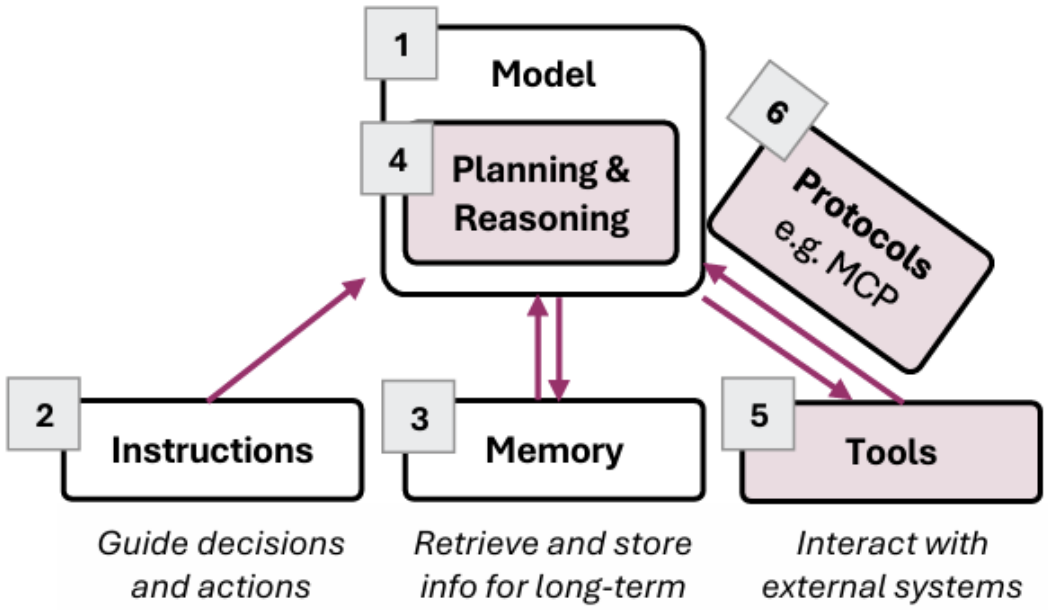
A defining feature of AI agents is their ability to operate with a degree of autonomy



Source: Singapore Government IMDA, *Model AI Governance Framework for Agentic AI* (2026); see also C Ling, L Pallaras & B Sadique, *Agentic AI: Rogue Agents, Real Liability*, KWM, 27 January 2026.

AI agents have a number of core components that enable them to plan, "reason" and take action to achieve a goal

Core components of a simple AI agent



Agentic AI systems often follow a

Plan > Act > Observe > Reflect loop

Key components of an AI agent

1. **Model:** Agent's central reasoning and planning engine (e.g, LLM, SLM etc)
2. **Instructions:** Defines the agent's role, capabilities and behavioural constraints
3. **Memory**: Information stored and accessible to the model, in short or long term storage
4. **Planning and reasoning**: The model is usually trained to reason and plan
5. **Tools**: The model uses tools to complete a task (e.g., internal data, CRMs, external tools)
6. **Protocols**: Standardised way for agents to communicate with tools and other agents.

Source: Singapore Government IMDA, *Model AI Governance Framework for Agentic AI* (2026).

Agentic AI is currently being used in a range of ways to improve efficiency and productivity within organisations

Productivity agents: Review emails and summarise content, manage diary, navigate file system and browse the web.

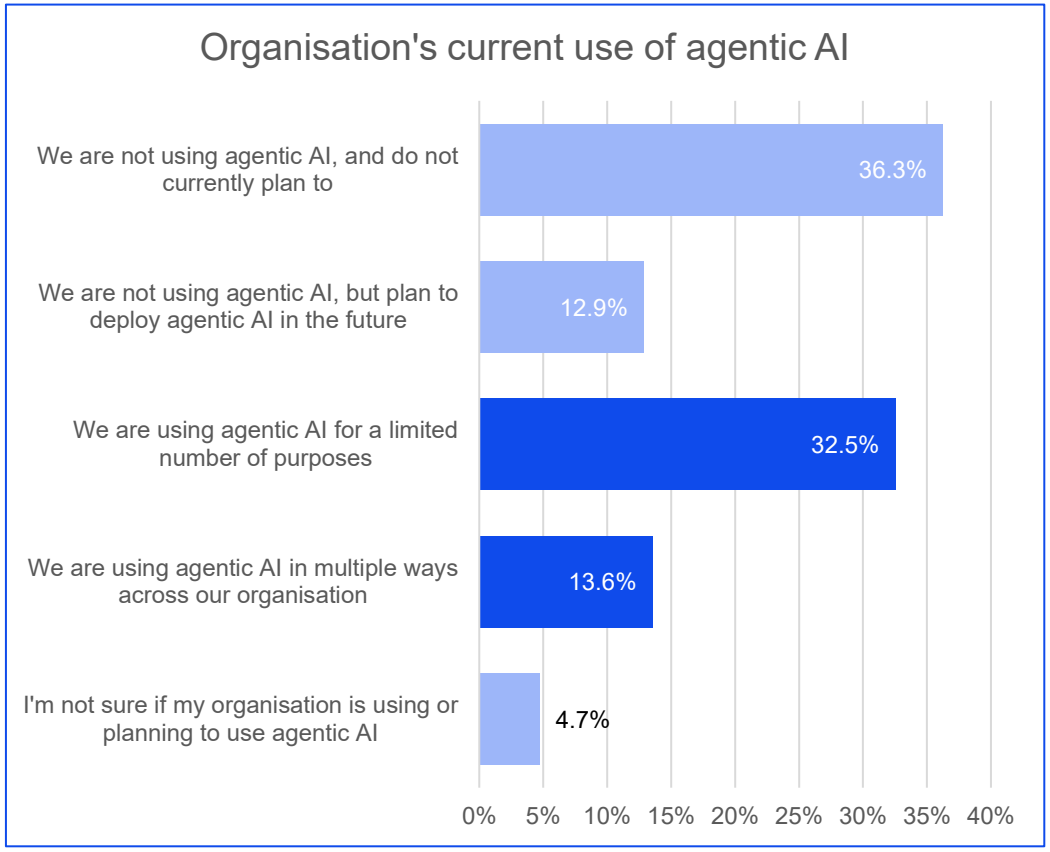
Customer service agents: Resolve complex customer issues including multi step workflows, escalation to human if necessary.

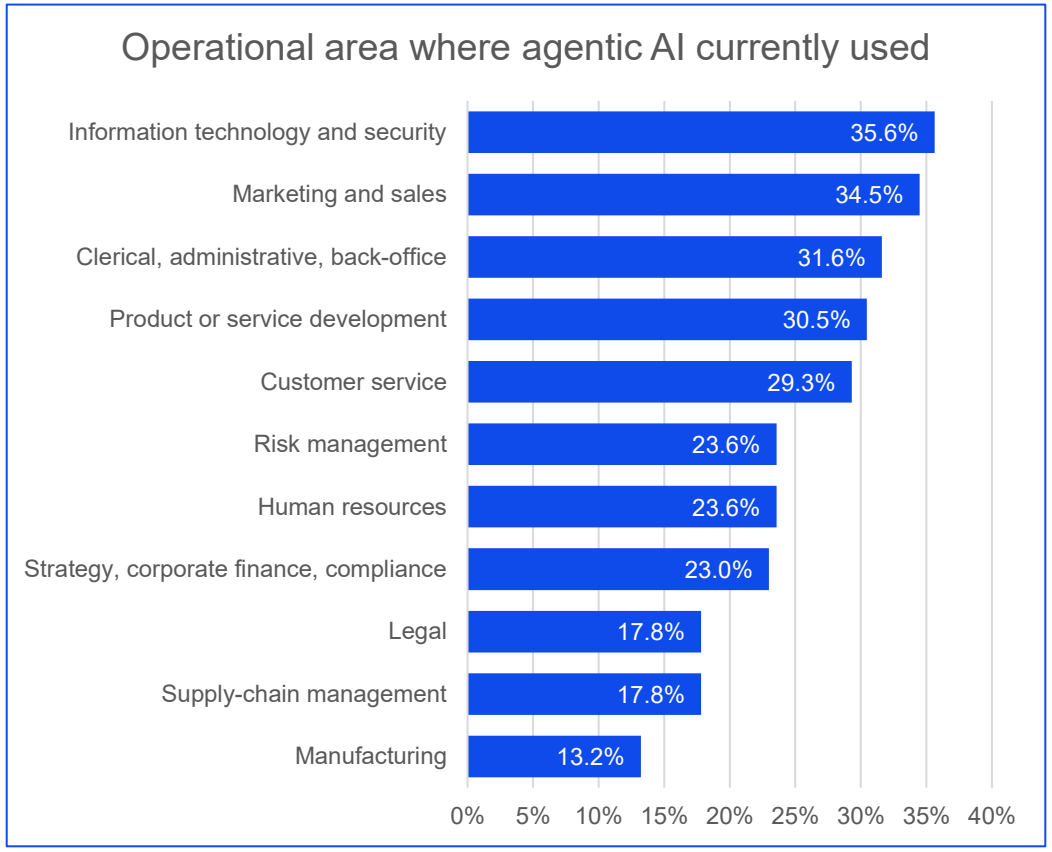
Coding agents: Generate, test and debug code, and prepare documentation.

Research agents: Manage complex research including gathering information, summarising reports and preparing advice.

Legal workflow agents: Review contracts, identify high risk clauses and suggest revisions.

Australian corporate leaders report high adoption rates for agentic AI, with the highest use in IT and security functions





Source: UTS Human Technology Institute, *Survey of corporate leaders on use of AI* (2025), Graph 1, N = 295 and Graph 2, N = 174.

There is evidence that many organisations are still experimenting with agentic AI, or are using it in limited ways

A 2025 survey of ~1,900 respondents in 105 countries reflected that:

39% had begun **experimenting** with AI agents

23% said the business was **scaling** an agentic AI system

Most were scaling only in **1-2 business functions**.

Generally, it appears that many of those using agentic AI are **most likely** using:

Single AI agents with

Limited autonomy

Carrying out **pre-defined** workflows.

.

Source: McKinsey, *The State of AI in 2025: Agents, Innovation and Transformation* (2025).

Organisations are expected to expand their use of agentic AI in future as the technology develops, and trust in it is built

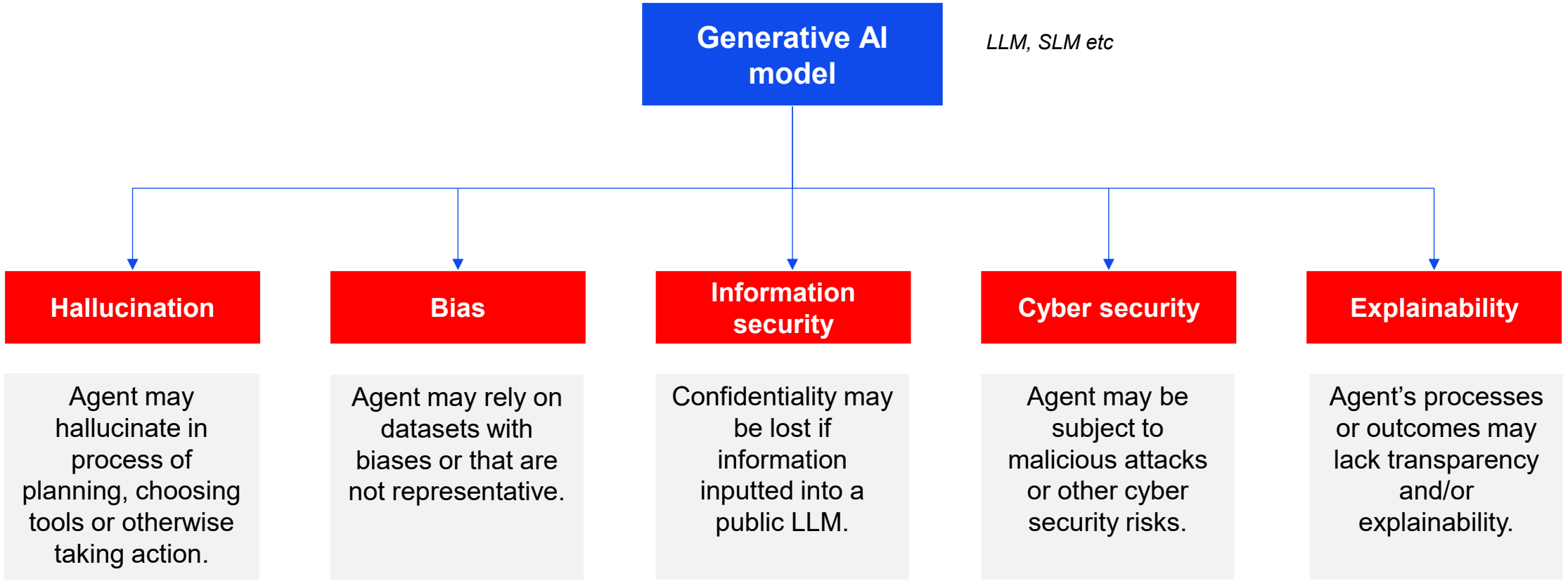
At this stage, there are key limitations to the broader rollout of agentic AI

Performance and reliability	Impact of failure	Organisational readiness
Technology may not support **reliable integration** with other agents and systems.	Autonomy means **errors** may **compound** before they are noticed.	**Infrastructure** may not be ready for agentic systems, e.g., data quality.
Uses **generative AI** which can hallucinate and carries other risks.	Complex environment means **errors** may **cascade** across systems.	**Change management** required to adopt AI agents into workflows and teams.
Models can be **brittle,** experience **drift** and may not deal with **unexpected inputs**.	Scale of operation (e.g., 24/7) may **magnify harms** before detected.	**AI governance** required to manage new and amplified risk, and embed **trust**.

02

When the agent goes rogue: managing risks with agentic AI

Agentic AI uses generative AI, and so it inherits the same risks as generative AI systems



Agentic AI can also create new or enhanced risks due to its defining features—that it operates *autonomously*, in *complex environments* and *at scale*

Autonomous actions may go beyond authority

- Instructions **unclear** or agent **misinterprets** goal
- Pursues goal in **unintended ways** (e.g., modifies or deletes data, sends external communications)
- **Commits errors** when executing tasks (eg, poor coding, uses biased data)
- **Adapts away** from original goal.

Risk enhanced by complex operating environment

- **Limited common sense** and **reasoning** in ambiguous & complex contexts
- **Poor orchestration** across multiple agents or systems
- **Expanded attack surface** for malicious actors due to operation across systems
- Potential **automation bias** due to apparent capability.

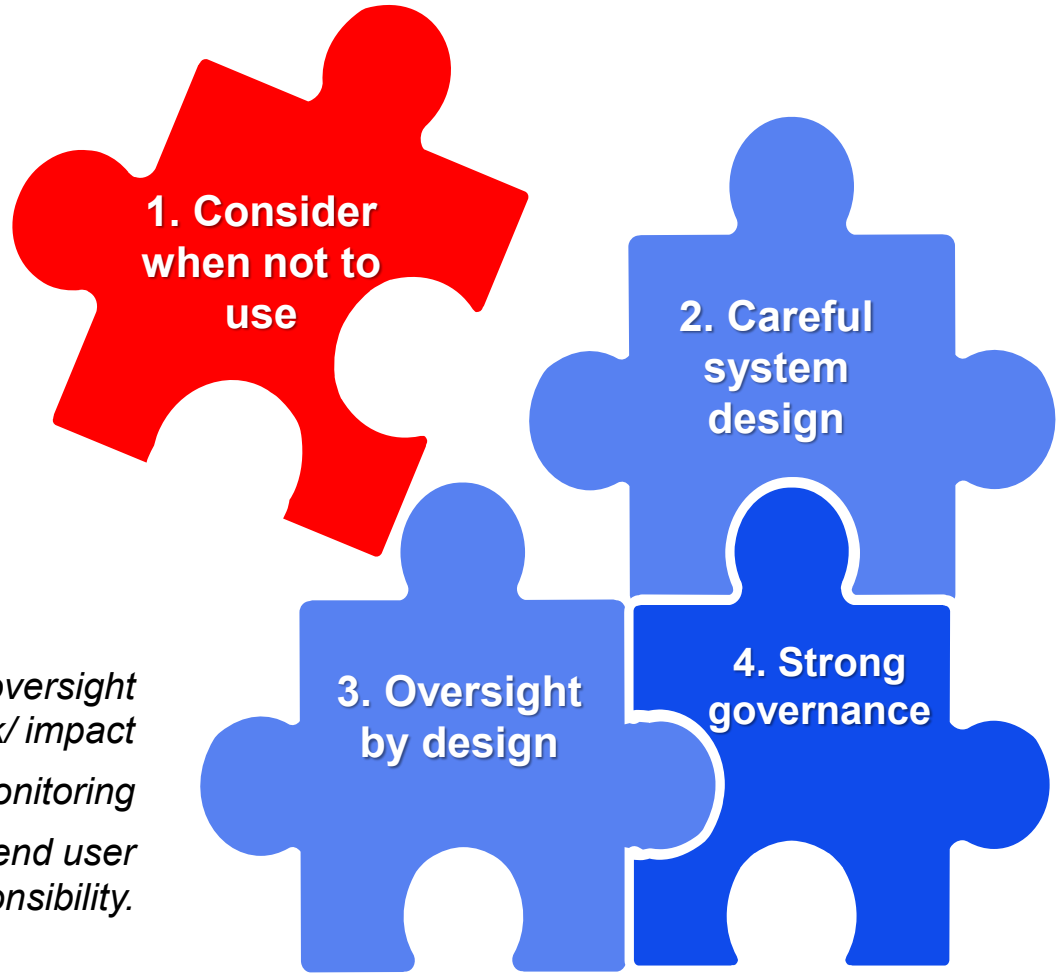
Impact compounded by operating at scale

- **Autonomous operation** means errors may not be detected, and may compound
- **Errors compound** and **expand** as agentic AI:
  - Engages with other agents or systems
  - Operates at scale (e.g., 24/7 operation).

Agentic AI risks should be mitigated through careful design, continuous oversight and ongoing governance

Organisations should also improve staff AI literacy to help manage risk of "shadow" use of AI agents.

Don't use agentic AI if it is not necessary or feasible, or it is too costly



Clear instructions re goal, autonomy, (un)approved actions

Identify risks early and address in design

Optimise tools, data quality, systems

Test rigorously before deploying.

Incorporate human oversight according to risk/ impact

Implement ongoing monitoring

Train staff in end user responsibility.

Make humans accountable

Implement dynamic controls

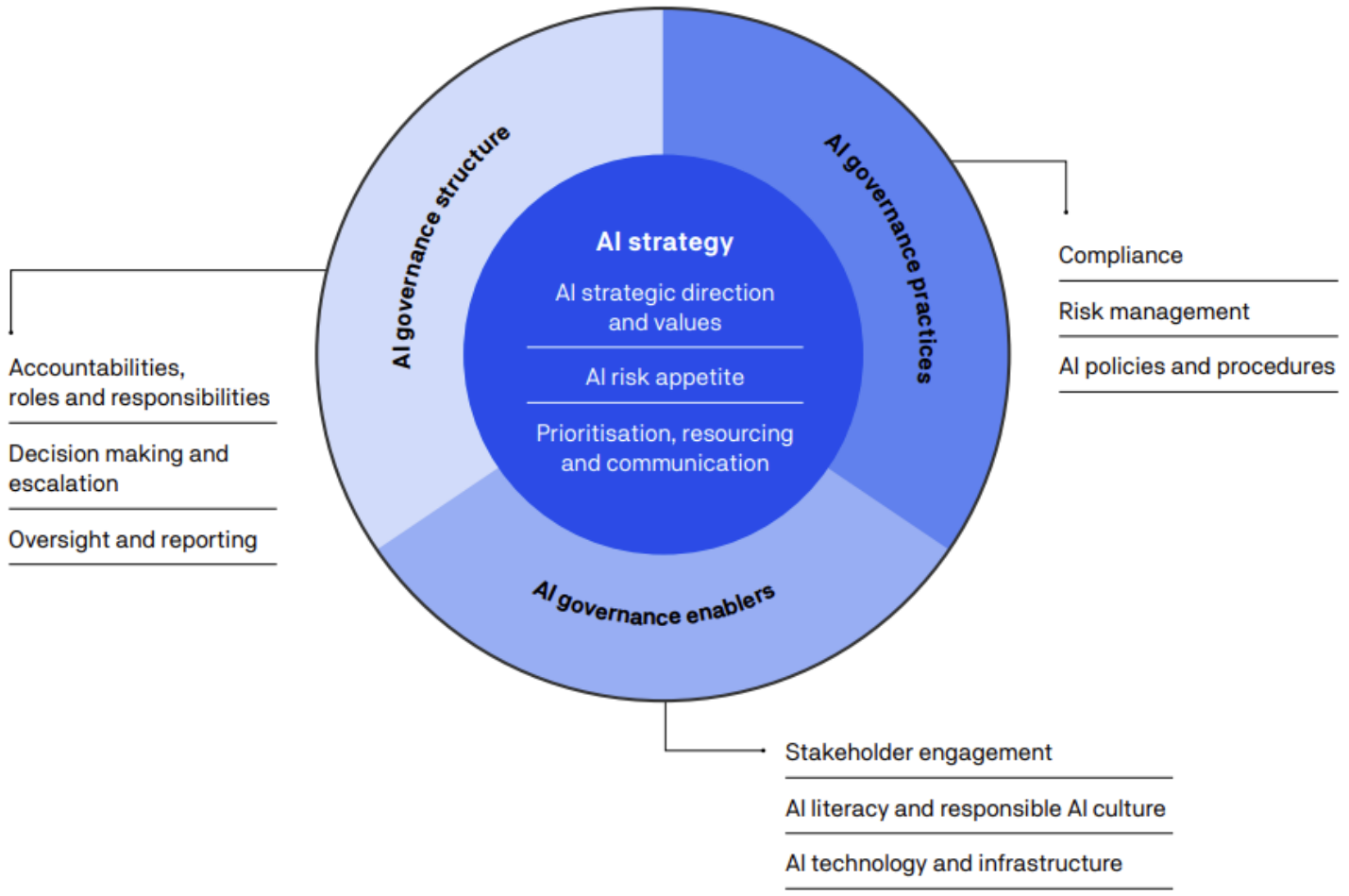
Design processes to override agent behaviour, where necessary.

03

How can we leverage agentic AI responsibly?

Implement strong AI governance for agentic AI, including system level governance initiatives

Human Technology Institute has developed an AI governance operating model for best practice AI corporate governance



AI governance practices include system level policies and controls for agentic AI systems

Engage with workers in the adoption process to identify the best opportunities for agentic AI systems

There is a range of ways agentic AI could be implemented, and each can impact differently for workers

Redesign processes to use AI agents

- Agentic AI used for simple tasks or end-to-end processes
- May free up staff for work that requires human skills (e.g., creativity, judgement etc)

Redesign teams to incorporate AI agents

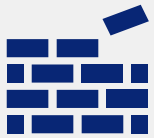
- Agentic AI systems used to augment human teams
- Workers may supervise and/or coordinate AI agents

Redesign operating models to leverage agentic/ human contributions

- Organisation redesigns functions and roles more broadly
- Develops new operating model that leverages AI agents and incorporates new roles for workers.

Planning should consider the sustainability risks of agentic AI given its autonomous nature, and its ability to operate 24/7

Generative AI has substantial energy and resource needs across its lifecycle—including electricity, water and minerals



AI infrastructure

- Energy for data centre construction and operation
- Minerals, water and energy to develop chips for AI models.

AI model training

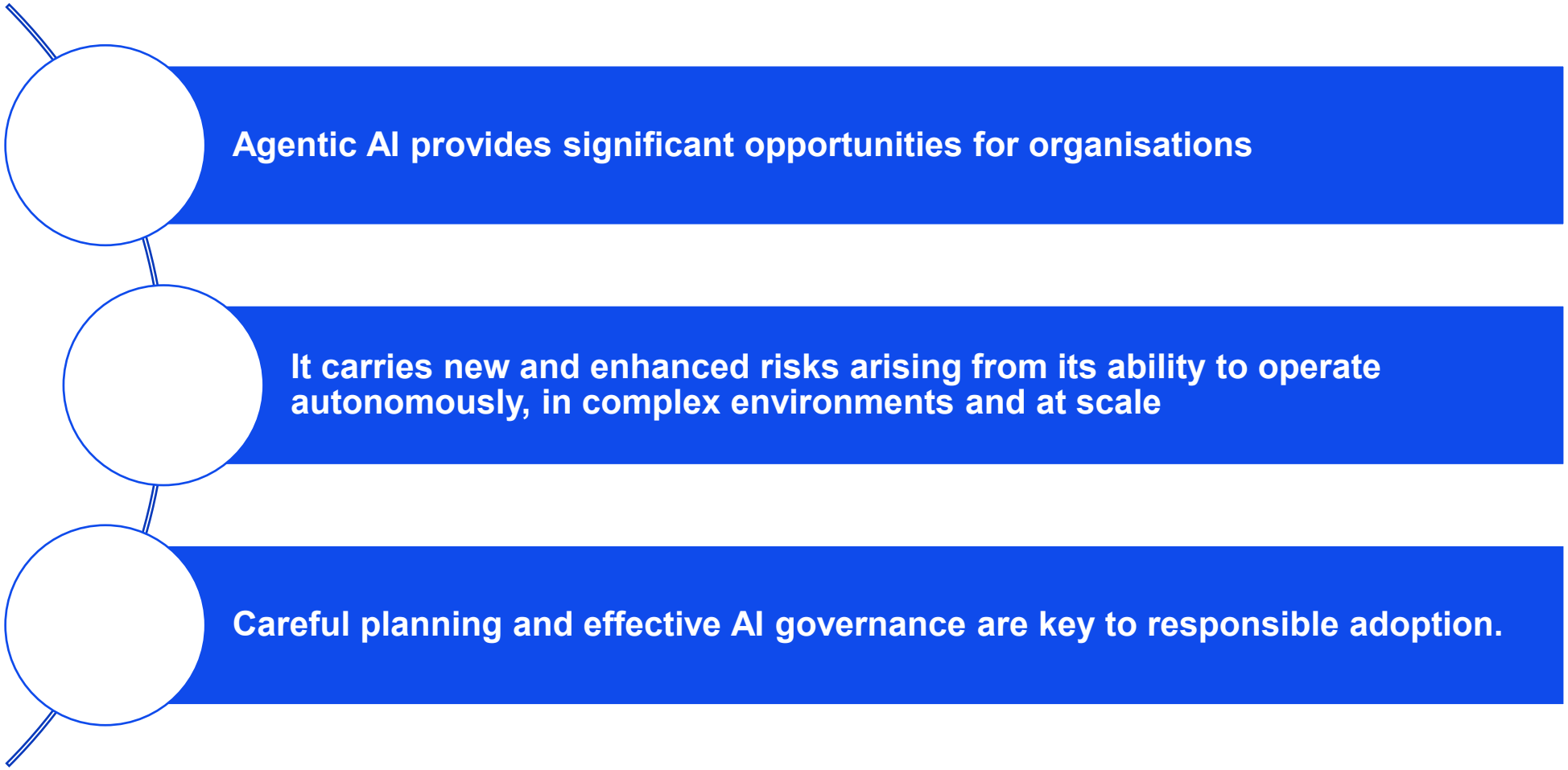
- Energy and water used in training an AI model
- Larger the model, the more energy required to train it.

AI inference

- Energy and water are used for computation to generate responses (i.e., ‘inference’)
- Can be more energy intensive than training.

Energy and resource needs of generative AI models

Agentic AI provides unique opportunities to improve efficiencies and ways of working provided it is used responsibly



05

Questions and answers

Human Technology Institute

For more information about HTI, please contact:

Gaby Carney
Senior Fellow, HTI
Gaby.Carney@uts.edu.au
Phone: 0401 355 784