Allens

In alliance with Linklaters

AI Governance Toolkit

A GUIDE FOR BOARDS AND GENERAL COUNSEL

The exponential advances in the development of generative AI, including tools such as ChatGPT, means that both the opportunities and risks of its use and deployment are increasing at scale.

AI regulation is coming. What's more, we expect that regulators, policymakers, shareholders and the public will hold companies to account for failures to address AI risks *much faster* than they have for other risks with a 'technical' component–mostly because of the nature of the risks and the groundwork laid over the past few years to address data and cyber risks.

The solution is not to ban generative AI, but rather to start applying guardrails and sharing best practice approaches, *while reliance on AI is still in its early stages*.

We understand, though, that it can be hard to know where and how to start.

That's why we've designed this AI Governance Toolkit. We hope it will help companies start managing their AI risk in a way that:

- is fit-for-purpose and proportionate to the risks, having regard to current levels of investment (which, in many cases, are still limited); and
- is an enabler to the responsible use of AI within organisations, especially as deployment becomes more widespread.

As always, if you'd like to discuss any of this in greater detail, please get in touch.

Contents

What is AI?

AI can mean different things to different people.

For the purposes of this AI Governance Toolkit...

Artificial Intelligence is any system that can perform tasks that can, for a given set of objectives, generate outputs such as predictions, recommendations or decisions influencing real or virtual environments.[1]

Generative AI is AI that can generate realistic and unique outputs (eg images, videos, software code, music or text) similar to the content used to train it.

Whereas classical computing algorithms are determinative, AI algorithms are predictive.

Classical computer algorithms		AI
Programmer sets the rules	VS	Developer provides training data
Algorithm follows rules set by programmer		Algorithm generates responses by predicting the outcome using pattern recognition

AI systems may operate with varying levels of autonomy and creativity. For a more detailed explanation of the AI landscape, see page 10.

[1] *National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework*

If you do nothing else, do this:
12 simple questions to ask your business

Highest priority actions are in bold

Chances are, there is already a lot of AI activity going on within your business–much of it unknown or uncoordinated. These 12 simple questions (and corresponding actions) will help you understand how AI is being used across the business, identify key risks and prioritise next steps.

1 **Accountability:** who is responsible for managing AI risks?

Actions:

- **Allocate responsibility for governance of AI-related decisions** (either by appointing a cross-functional AI governance committee, or nominating an existing committee) with clear allocation of roles and responsibilities for oversight, development and use of AI.

2 **Visibility:** how is AI being deployed across the business?

Actions:

- **Take an inventory of current uses and development of AI models across the business.**

  Tip! Document what should be included in the inventory and why.

- Create a process to regularly update the inventory.

3 **Regulatory compliance:** what are our legal and regulatory requirements?

Actions:

- **Obtain legal advice on current (and anticipated) regulatory requirements and broader legal considerations.**
- Ensure AI governance committee is kept up to date on these requirements and any reforms.

  Tip! Consider:

  - *anti-discrimination, privacy, intellectual property, cybersecurity, WHS, competition, consumer protection and risk management regimes;*
  - *regulatory reporting requirements in relation to serious incidents, system weaknesses or malfunctioning systems;*
  - *sector-specific laws (eg for APRA-regulated entities, operational risks under draft new CPS 230 Operational Risk Management, s912D of the Corporations Act, therapeutic goods regulation, etc);*
  - *the interplay between AI use and record retention and destruction programs;*
  - *overseas regulatory frameworks (eg the EU Artificial Intelligence Act and China's Management Measures for Generative Artificial Intelligence Services (Exposure Draft));*
  - *international human rights treaties; and*
  - *domestic AI frameworks (eg NSW Artificial Intelligence Assurance Framework).*

4 **Risk management systems:** what are our key AI risks?

- What is our risk tolerance and how does this compare to our risk profile?
- What controls have we implemented to address identified risks?

Actions:

- Update your risk management framework to document and govern AI usage and risks.
- **Create a risk scale to classify AI use cases.**

  Tip! Consider which AI Systems are classified as high risk in the EU AI Act.

- Set your risk tolerance.
- Implement systems (including logging capabilities) to monitor and enable traceability of the performance of AI systems throughout their lifecycle.
- Conduct an AI risk assessment (on an enterprise-wide basis and/or for specific high-risk AI systems or use cases).

If you do nothing else, do this:
12 simple questions to ask your business

Highest priority actions are in bold

5 **Policies and processes:** what guidance do we give staff about their use of AI?

- Are staff aware of this guidance?

Actions:

- **Create AI Guiding Principles** to guide the ethical and responsible use and deployment of AI. Consider whether updates to the Code of Conduct are also required to reflect these principles.
- **Update your Technology Usage Policy** to address the use of AI (especially generative AI).
- **Create an AI Management Policy** to govern the use, design, development, testing, deployment and ongoing monitoring of AI, and to ensure it is appropriately documented, tested and validated.
- Consider whether to restrict access to higher risk, publicly available AI applications and/or to implement a controlled sandbox environment for use within your organisation.
- Consider which other policies may need to be updated (eg in relation to data and cyber governance).
- Internally publicise new policies and updates to existing policies.

6 **Data governance, quality and privacy:** how are we using data in connection with AI?

- What data is being used to train, validate and test AI?
- From where have we obtained that data and is this use permitted?
- How do we ensure these data sets are sufficiently relevant, representative, free of errors, complete and have appropriate statistical properties?
- Do these data sets use 'protected characteristics' or characteristics that act as proxies for protected characteristics?
- Is informed consent an adequate test for appropriate use?

Actions:

- **Consider interaction with your data governance framework.**
- **Obtain advice on the legality of data collection, use and generation.**
- Document: (i) the provenance of training data; and (ii) attribution of the decisions made by the AI system to relevant training data.
- Interrogate the quality and credibility of training, validation and testing data (and whether it includes any inherent biases or could result in unfair discrimination, infringement of third party IP rights or breach of confidentiality).
- Consider any jurisdictional issues, such as whether any data is being stored or processed overseas.

7 **Transparency, explainability and interpretability:** can (and do) we explain how the AI that we develop or use works?

- How do we ensure users know how to use AI systems and interpret outputs?
- Could we easily explain to an auditor or regulator how the AI we use or develop works?
- Could we easily explain how outcomes were derived to an affected individual or other third party?
- What avenues and recourse do affected third parties (eg staff and customers) have to make enquiries or complaints about our use of AI?

Actions:

- **Create documentation guidance and standards** to ensure AI systems are designed and developed to enable: (i) *users* to correctly interpret the output and use it appropriately; (ii) the *company, auditors and regulators* to understand the logic of the system and algorithms, the key design choices and trade-offs made by the developers/users, the provenance of data sets and how they were selected and obtained, data-cleaning methodologies, training methodologies and labelling procedures; and (iii) (where applicable) *consumers* to take meaningful action in the event of an adverse decision.

  Tip! The level of explainability required may vary depending on the impact of the AI application. Explainability may be particularly difficult to achieve for more "highly intelligent" AI systems

- Create or uplift processes to address consumer complaints and enquiries about the use of AI.

If you do nothing else, do this:
12 simple questions to ask your business

Highest priority actions are in bold

8 **Consumer engagement:** what do we say externally about our use of AI and related data?

- Is this accurate?
- Should we provide a right to opt-out of AI or choice of a non-AI product?

Actions:

- **Update consumer-facing privacy policies, collection notices and marketing collateral** to ensure they are not misleading or deceptive or unfair.
- Consider labelling rules to enable users to easily identify: (i) where content that resembles existing persons, places or events has been artificially created or manipulated; (ii) where users are interacting with an AI system; and (iii) where outputs or outcomes have been generated from a generative AI tool and how and why that may impact individuals or other third parties.

9 **Supplier risk management:** how do we identify and manage the risks involved in procuring AI tools or services from third parties?

- Is our data being used to train third party AI?

Actions:

- **Review and update your supplier engagement policy** to address procurement and oversight of provision of AI tools and services throughout the lifecycle of any supplier engagement.
- Consider how to risk-classify procurements of AI tools or AI-generated data.
- Update due diligence processes and supplier risk assessment templates. Ask suppliers about their own AI governance arrangements, including how their AI was trained. Test explainability (Note: this is important to ensure you can comply with your own transparency, explainability and interpretability principles).
- **Develop AI-specific clauses for inclusion in relevant contracts.**
- Identify controls to internally manage AI risks relating to supplier arrangements (eg requesting reports, auditing compliance, seeking appropriate warranties and indemnities, etc).

Tip! Consider whether your company is comfortable with its data being used to improve the overall accuracy of the supplier's AI, or whether your own quarantined instance is required.

10 **Accuracy, robustness and security:** what technical and operational controls do we have in place to address security and reliability of AI?

- How confident are we that they will help us prevent, detect, withstand, respond to and recover quickly from errors, faults, inconsistencies, unexpected situations, vulnerabilities, anomalies and malicious attacks?

Actions:

- **Consider requirements for: (i) human oversight to proactively identify errors and non-compliance; and (ii) timely human consideration, including if a system fails, produces an error or if someone wants to contest an outcome.**
- Implement measures to mitigate against attempts to manipulate training data sets (ie data poisoning), the use of inputs designed to cause the model to make a mistake (ie adversarial examples), and AI model flaws.
- **Implement tripwires and controls to override, reverse or halt the output or operation of AI systems, especially for high risk AI systems or use cases.**

11 **AI incidents and resilience:** do we have a documented AI incident response plan?

- How would we respond to, and recover from, any unintended consequences of our use of AI?
- What (if any) reporting obligations do we have in respect of AI incidents?

Actions:

- **Identify which incident response plans and playbooks need to be updated.**

Tip! Consider creating a standalone AI Incident Response Plan or uplifting your Crisis Management Plan.

- Build AI incidents into continuous disclosure assessments (where relevant) and other mandatory (regulatory and contractual) notification and reporting processes.
- Consider the impact of AI use cases on business continuity arrangements.

12 **Training:** do we provide relevant roles-based training?

Actions:

- **Create an AI training program.**
- Share experiences and learnings across the business.

Establishing your AI governance committee

Allocating responsibility for governance of AI-related decisions is a critical first step in kickstarting your AI governance program.

Who should be on your AI committee?

Representatives from key functional areas including Legal, Compliance, Risk, Product Development, Procurement, Data Science, Cyber, Marketing and Customer Service.

Tip! Consider nominating an existing committee or appointing a dedicated cross-functional AI governance committee. Either way, participants should also have diverse skillsets, backgrounds and experience. The committee may also need to leverage external resources and experience.

AI guiding principles

What is the role of the committee?

To oversee the design and rollout of your AI governance framework. This includes:

- defining **key roles and responsibilities** in relation to the oversight, design, development and use of AI across the business
- creating **AI guiding principles**
- defining and documenting the **scope of the AI governance program** (including which types of models, algorithms and systems are in and out of scope and why, and building a risk scale for in-scope use cases)
- identifying and overseeing **policies, processes and training** to enable responsible AI design, use and oversight
- identifying areas requiring **human review or oversight**, including to identify inaccuracies, identify biases and undertake other quality assurance
- developing a process to **escalate and assess high-risk** AI use cases
- **reporting** to senior management and the board
- assisting in managing **incidents** relating to AI use

When it comes to developing AI models, it matters who is 'in the room'.

UK Financial Conduct Authority

Roadmap for designing and implementing an AI governance framework

1. Appoint AI governance committee

- should be cross-functional and comprise participants from diverse skillsets, backgrounds and experience
- consider access to external experts
- define roles and responsibilities for AI governance and oversight

2. Understand current state

- create inventory of AI deployment and use of AI outputs
- identify existing arrangements with AI service providers
- obtain legal advice on AI legal and regulatory requirements
- undertake AI risk assessment

3. Create AI guiding principles and set AI risk tolerance

4. Undertake gap analysis and create roadmap for uplift

5. Create or uplift artefacts

- AI management policy for the design, development, testing, deployment, use and ongoing monitoring of AI
- technology usage policy (to address AI risks)
- supplier engagement policy (to address AI risks)
- IP and confidentiality policy (to address AI risks)
- AI incident response plan
- AI risk assessment template
- privacy policy and consumer terms
- marketing collateral
- other documentation to assist with transparency, explainability and interpretability of AI tools and outcomes

6. Create or update processes to:

- address consumer complaints and enquiries about the use of AI models
- undertake an annual AI inventory review
- interpose human oversight
- monitor compliance with AI framework
- enable early detection and escalation of issues
- brief the board

7. Implement security controls

to prevent unauthorised access to, or misuse of, AI models and training data, and to protect the value of AI-generated outputs.

8. Educate and increase awareness

- create an AI training program
- conduct tabletop exercises on AI incidents

9. Review, uplift and provide ongoing advice

- monitor and help shape evolving regulatory landscape
- periodically update incident response plans and playbooks, including following incidents, tabletop exercises and regulatory developments
- provide ad hoc legal and compliance advice
- conduct incident post-mortems and remediation
- advise on third-party claims and litigation

Checklist for directors

As regulators have emphasised in relation to cybersecurity, boards are ultimately responsible for overseeing AI risk.

- ☐ **Understand the company's AI strategy** and its alignment with the broader business strategy
- ☐ **Ensure AI risk owners and related roles and responsibilities are clearly defined** and that those individuals have the appropriate skill sets and resources to properly undertake those roles
- ☐ **Understand the company's AI risk profile**
- ☐ **Set or approve the tolerance for AI risks**
- ☐ **Ensure AI is a periodic board agenda item,** either at full board or risk committee meetings, **and that the board has adequate access to AI expertise**
- ☐ **Understand the legality** of the use and deployment of AI, including the collection and use of training data, across the business
- ☐ **Understand how the business ensures that ethical issues involved in AI use are identified and addressed,** especially bias and discrimination
- ☐ **Understand how AI systems and use cases are risk rated** (ie the ratings criteria and assessment process), and which have been prohibited, and why
- ☐ **Understand the critical and high risk AI systems** that are used and deployed across the business, and the nature, provenance and reliability of data used to train high-risk systems
- ☐ **Understand the trade-offs made in decisions involving AI** (eg accuracy vs fairness, interpretability vs privacy, accuracy vs privacy, accuracy vs adaptability)
- ☐ **Ensure there are processes for management to escalate and brief the board on any AI incidents**, including on the organisation's response, any impacts, the status of any investigations and learnings identified as part of the post-incident review
- ☐ **Ensure compliance with the AI risk management program is audited by the audit function** in line with its third line role
- ☐ **Ensure the AI risk owner regularly reviews the effectiveness of the AI risk management program and policies**

7 lessons learnt from managing cyber risks

You can apply many of the lessons learnt managing cyber risks, to the management of AI risks.

1 **Board oversight is critical.**
Boards will be expected to oversee AI governance—they cannot simply delegate the oversight of this risk to specific business functions or technical experts.

2 **Develop a governance framework.**
Companies should have an AI governance framework, which:

- defines roles and responsibilities;
- outlines how AI risk will be assessed and managed;
- includes processes to monitor, enforce and report on compliance with the framework; and
- includes measures to foster a culture of AI awareness.

3 **Take a cross-functional and risk-based approach** to the management of AI risks and opportunities.

4 **Know your regulatory requirements.** Even in the absence of AI-specific regulation, organisations will be expected to manage AI risks using existing risk management frameworks, in accordance with existing regulatory requirements.

AI controls to address risks should be mapped to regulatory requirements, tailored to specific businesses and assets within the group and routinely and systematically tested to ensure they are effective and remain fit for purpose.

5 **Ensure your governance framework is supported by documentation** that is routinely reviewed and updated. It is important to start creating the audit trail now to establish a track record of anticipating, assessing and addressing AI risks. This should ideally be structured so it can be easily accessed and provided to a regulator if necessary.

6 **Design effectiveness will not be sufficient.** Companies will need to ensure operational effectiveness, including by monitoring and enforcing compliance with the framework.

7 **Focus on operational resilience and war game incidents.** Companies should implement controls to ensure they can withstand and recover from disruption (whether that is a faulty AI system or the activation of a 'kill switch'), with as little impact as possible on the business and customers.

Conducting tabletop or war gaming exercises is an excellent way to identify decisions that can be made in advance of an incident, identify gaps in response processes and prioritise uplifts.

AI Landscape

No intelligence → **High intelligence**

Determinative | **Predictive**

	Classical algorithm	Narrow or Weak AI	General-purpose AI Generative AI *Note: some Generative AI systems may be more narrowly focused than General-purpose AI*
Examples	▪ Simple rule-based chatbots that answer basic customer service queries ▪ Spam and fraud detection systems (eg Gmail or Outlook's scanning and auto filtering of emails into spam folders)	▪ Speech recognition (eg Siri) ▪ Image classification (eg Google Cloud Vision API) ▪ Facial recognition (eg Apple Face ID and Clearview AI) ▪ Recommendation systems (eg Amazon, Spotify and Netflix personalised content)	For generation of: ▪ text – ChatGPT and Bard ▪ images – Dall-E 2 and Midjourney ▪ software code – Co-Pilot ▪ music and sounds – Jukebox, Music LM, Aimi and VALL-E
Functionality	▪ **Rule-based or predefined** algorithm designed to perform **specific tasks or solve specific problems.** ▪ Requires explicit programming and modifications to accommodate different tasks or domains.	▪ Can replicate or outperform human intelligence* but only in **specific narrowly defined and structured tasks.**	▪ *General-purpose AI* – may in some cases replicate or outperform human intelligence* in relation to a **range of tasks within specific domains.** ▪ *Generative AI* – can **generate realistic and unique outputs** (eg images, videos, software code, music or text) similar to the content used to train it. *Note: many (but not all) applications of Generative AI are powered by large language models (LLMs), which can generate human-like, coherent and contextually appropriate responses across a wide range of topics in response to prompts (eg ChatGPT is a chatbot that allows users to query the GPT-3.5 LLM).*
Learning approach	▪ Follows the rules set by the programmer. ▪ Does not involve automated learning from data.	Can employ various learning techniques (which may include machine learning, natural language processing, computer vision, deep learning, and probabilistic modelling) which predict the outcome using pattern recognition.	
		▪ Typically learns from **labelled data relevant to its specialised task** using **supervised learning.** *Note: labelled data is data that has been annotated or tagged with predefined labels or categories, indicating the correct answer or desired output for a given input.*	Can learn from **diverse labelled or unlabelled data sets** using **supervised or unsupervised** deep learning techniques. *Note: in* **supervised** *learning, the model learns to map inputs to corresponding outputs by minimising the discrepancy between its predictions and human-provided labels. In* **unsupervised** *learning, the model autonomously identifies patterns, relationships and structures in the data without human-provided labels.*

General AI (also known as **artificial general intelligence**) represents a more advanced, autonomous and complex level of artificial intelligence that is still largely hypothetical and does not yet exist in practice. General AI would possess the ability to understand, learn, adapt to and apply its intelligence across **multiple tasks and domains** that require a range of skills and knowledge, similar to the broad cognitive abilities of a human being.

*With the advantage of computational power and access to vast amounts of information.

Key contacts

HEAD OF CYBER

Valeska Bloch
Partner
T +61 2 9230 4030
Valeska.Bloch@allens.com.au

HEAD OF TMT

Gavin Smith
Partner
T +61 2 9230 4891
Gavin.Smith@allens.com.au

Lisa Kozaris
Chief Innovation & Legal Solutions Officer
T +61 3 9613 8944
Lisa.Kozaris@allens.com.au

Phil O'Sullivan
Partner
T +61 2 9230 4393
Phil.O'Sullivan@allens.com.au

David Rountree
Partner
T +61 7 3334 3368
David.Rountree@allens.com.au

Jessica Mottau
Partner
T +61 2 9230 4587
Jessica.Mottau@allens.com.au

Elyse Adams
Partner
T +61 3 9613 8534
Elyse.Adams@allens.com.au

Miriam Stiel
Partner
T +61 2 9230 4614
Miriam.Stiel@allens.com.au

Dominic Anderson
Partner
T +61 2 9230 4099
Dominic.Anderson@allens.com.au