ENTERPRISE AI GOVERNANCE FOR SENIOR EXECUTIVES

Whitepaper Series
April 2024

CENTRE FOR BUSINESS ANALYTICS

THE UNIVERSITY OF MELBOURNE

MELBOURNE BUSINESS SCHOOL

EXECUTIVE SUMMARY

Since late 2022, the topic of Artificial Intelligence has dominated public debate and private conversations. The rise of ChatGPT has been nothing short of astonishing, eclipsing Netflix, Uber and Facebook for the pace of adoption. Within the enterprise context, questions about how this technology will disrupt all facets of businesses have been raised. Not least of these are questions about ethical impacts, changing regulations and approaches to managing risk.

As many organisations lean into these questions, we have prepared this White Paper to assist Boards and senior executives to frame their thinking about governing this important, perhaps once-in-a-lifetime development. While the technology is new and evolving quickly, we take a historical view of these systems to help navigate the path forwards, informed by timeless principles of governance and oversight. And while there are considerable risks and known harms to be considered, we set these in the context of the enormous benefits AI presents for individuals, enterprises and our society.

The key recommendations are:

- Don't wait for new Australian AI legislation before acting; you already have legal obligations.
- Preserve your "social licence" to operate AI Systems so you can continue to innovate.
- Think about AI Governance across four dimensions: strategic alignment, value creation, operational performance and risk management.
- Undertake a systemic review to understand where your AI Systems are currently deployed.
- Collate your existing obligations – internal and external – to your customers, employees, regulators and other stakeholders.
- Ensure broad representation of people when designing, assessing or reviewing AI Systems.

The transformation of our economy to an AI-enabled future has been underway for a decade. With the growing adoption of AI systems across enterprises, we are in the early stages of a step-change in the organisation of our labour, capital and purpose of a scale not seen for a generation. New businesses will be created, new practices adopted; old ones will be supplanted overnight and over decades. No one knows how a change of this magnitude will unfold, nor who will be the winners and losers. However, open-minded, curious, empathic and informed leaders will be key to shepherding their organisations through this crucial phase.

About the Author:

DR GREGORY HILL

Qualifications

Bachelor of Engineering (Electrical) (University of Melbourne)
Bachelor of Science (Computer Science) (University of Melbourne)
Doctor of Philosophy, Analytics (University of Melbourne)

Areas of specialisation

Analytics, Data Science, Machine Learning, Statistics, Enterprise Information Systems

Experience

Greg has been leading data analytics teams since 2009 and has designed, built and led teams in Australia, NZ, UK, US and Malaysia, leveraging advanced analytics and data science to transform businesses.

Greg is passionate about connecting and developing analytics professionals with rewarding and high-impact work. His expertise is in the application of advanced analytics techniques (including machine learning, predictive modelling, mathematical optimisation, econometrics, operations research, discrete choice experiments and simulations) to solve commercial problems. These applications span forecasting, pricing, fraud, market segmentation, portfolio optimisation, customer satisfaction, propensity modelling, proposition testing and customer service across a range of industries including retailers, telecommunications, and financial services.

Greg has interacted with senior executives across organisations from a myriad of cultural backgrounds and professional disciplines, which has honed his ability to introduce and explain data analytics concepts to non-technical professionals. Greg is adept at introducing and explaining key data concepts to practitioners from a wide array of industries and corporate functions, at all levels of seniority.

He was a founding member of the Industry Advisory Board at Melbourne Business School's Centre for Business Analytics, where he helped develop the curriculum for the flagship Master of Business Analytics program. He continues to serve, advising on strategy, research priorities and industry linkages. In addition, he has designed, developed, and delivered an Executive Education three-day program targeting senior leaders new to the data and analytics space. He has also delivered several industry workshops on data, analytics, and AI to industry practitioners, ranging from consultant/analysts to team leaders and executives. He is currently an Adjunct Lecturer, teaching data leadership to MBA students at Melbourne Business School.

CENTRE FOR BUSINESS ANALYTICS

About THE CENTRE FOR BUSINESS ANALYTICS

Melbourne Business School has a proud history of advancing quality business education in Australia. The school is home to Australia's first MBA program, launched in 1963 and also the Master of Business Analytics, which is ranked 15th in the world by QS and recognised as the top program in Asia and Oceania.

The Centre for Business Analytics is at the forefront of data-informed decision making, uniting scholars, practitioners, students, and organisations driven by the challenge of leveraging data for organisational success. Since our inception in 2014, we have helped many organisations solve business challenges using data and furthered the data culture and maturity of many businesses. We have transformed the data culture and maturity of organisations through executive education, student practicums and talent acquisition, research and thought-leadership events.

Connect with the Centre for Business Analytics

cfba@mbs.edu

Acknowledgements

Thank you to Greg Cameron (Enterprise Senior Fellow, AI Platform – Faculty of Engineering and Information Technology, University of Melbourne) for his valuable contribution to this paper.


AI

CONTENTS

ENTERPRISE AI SYSTEMS

This section introduces some terminology and definitions for Artificial Intelligence that will help the reader frame current and future initiatives. Appendix 1 includes examples of how AI is presently used in the Enterprise context, while Appendix 2 provides a simple outline of the common phases of the AI system lifecycle. These can be used as checklist when planning AI Governance reviews.

The Definition of Enterprise AI

While there is no clear, widely accepted definition of Artificial Intelligence, we take the broad perspective that AI Systems are those which can perform intellectual tasks to a human standard. This can include types of decision-making and content-generation. Many organisations have been using Machine Learning – a type of AI where systems learn rules from training examples – for many years to support *decision-making*. These are types of IT systems and usually tightly integrated with other enterprise systems, like Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) systems.

More recently, Generative AI has emerged, which creates *content*, including text, images, software code and video. While deployments of Generative AI are still very new, they are more likely to be accessed through a standalone webservice, like ChatGPT. Both types of AI systems are said to be black-box systems: the specific rules that it operates are constructed by the system itself and are not explicable to humans. This contrasts with white-box systems, where clear and explicit rules have been designed by humans for the machine to execute. In the case of content, white-box means the content is retrieved and quoted from known sources, whereas black-box means it is synthesised from multiple sources, with no specific attribution possible.

To help make sense of where Enterprise AI systems fit, we introduce a simple 2x2 matrix. The columns capture the key object of the system (decisions vs content) while the rows describe how these are constructed (explicitly with white-box vs obscured with black-box).

	Decisions	**Content**
White-box (explicit)	Calculators Flowcharts Workflow tools Scoring rubrics	Archives Document repositories Databases Webpages
Black-box (implicit)	Predictors Forecasters Recommenders Optimisers	Search engines Chatbots Image generation Coding assistants

It should be noted that these categories are somewhat blurred and there are edge-cases that don't cleanly sit in one box or the other.

Types of AI Systems

While terminology can vary widely, we enumerate some of the common AI system types below. We provide enterprise use cases Appendix 1.

Classifiers and Predictors

Commonly used to group similar customers, employees, suppliers and other entities together, based on their attributes. If the commonality is an as-yet-unobserved attribute (such as, say, defaulting on a loan), we call it a prediction. The groupings (or labels) are usually defined in advance by humans to suit our pre-existing categories.

Forecasters and Estimators

These systems are used to produce a numerical score or other quantity in response to assumptions and may incorporate both extrapolation from historical examples and human expertise. Forecasting can relate to natural phenomena (such as the weather) or human behaviours (such as willingness to wait in a queue).

Recommenders and Rankers

Widely used in ecommerce and other online systems, recommendation engine will select the "best" product (or content) to a user based on a) attributes of the products themselves and b) choices by similar users. A related type of system seeks to rank (from best to worst) products and content and are more commonly found in social media and other services that rely on a dynamic and near-limitless algorithmic feed.

Simulators and Optimisers

These systems capture some important elements of a complex business scenario and allow for the system to explore through iteration within a sandbox environment possible responses or solutions. (This is why these are sometimes described as a "digital twin" – users can safely conduct virtual experiments.) When given some criteria to assess and compare outcomes, they can generate multiple solutions quickly and help identify the "best". Whether it's fighting a bushfire, scheduling flights at an airport or suggesting the quickest route for deliveries, these systems are closely tied to a specific business problem.

Search Engines

Modern search engines have evolved significantly from the pattern-matching document databases of the late 20th century. They are now able to use natural language processing to pre-process queries (remove spelling mistakes and ambiguous phrasing or provide additional context), assess competing content for the most relevant sources, handle a wider variety of content types (such as images and video) and incorporate user-feedback to provide further refinements.

Generative AI

As described above, Generative AI systems produce novel, original content – natural language text (such as ChatGPT), software code (like Github Copilot) or images (DALL-E) and video (Midjourney). This content is created in response to a user query, called a prompt. Increasingly, these capabilities are moving from standalone systems to becoming embedded elements within productivity software, such as document editing, coding, spreadsheet and presentation software. This sees Generative AI supplementing and supporting humans with these tasks where the content is created.

REGULATION OF AI SYSTEMS

Regulation of Artificial Intelligence systems has been a topic of niche interest since AI was first introduced to the Enterprise context. However, as the adoption and use of AI has grown and the importance of the decisions (and their impacts) has increased, its regulation has garnered more widespread interest from a diverse set of stakeholders and commentators.

Most of the risks and harms articulated in the AI regulation conversation are already extant and well understood and managed. The discussion around AI systems usually relates to how AI can accelerate or exacerbate these. Conversely, most of the prescriptions for governing AI systems responsibly are simply good practices for governing business processes generally. For example, fairness, transparency, the right to review, engaging end-users in design etc are helpful practices regardless of whether there is any AI involvement.

The Productivity Commission, in their research on AI regulations, describes this as "old wine in new bottles". While some enterprises will see AI Governance as an opportunity for a full review of all decision-making and content-generating processes across their business, others will look for a much more targeted approach.

We provide an overview of the rapidly changing regulatory environment (locally and internationally), while highlighting some key challenges for businesses. This discussion – and the preceding definitions and examples of current and future Enterprise AI use cases in Appendix 1 – can inform an early key decision for many enterprises: how to scope AI Governance in your organisation?

Sources of regulation

Like most developed nations, under Australian corporate law, all decisions and actions of corporations are the responsibility of the Board of Directors. Hence, the use of machines to automate activities – including decisions about whether and how to procure or build them, how they are configured and deployed, how they are monitored and so on – is the responsibility of the Board. This applies to Artificial Intelligence systems. (Many public sector bodies are constituted with similar roles and responsibilities; we refer throughout to corporations for simplicity.)

At present, there is no catch-all AI regulation or law operating in Australia. Instead, we have laws regulating information gathering, reporting and decision-making in specific contexts. Examples of these include consumer credit, medical devices, vehicle operations, financial planning advice, employment, food preparation and many other facets of modern life.

In some instances, the use of AI systems is explicitly considered in the relevant legislation and regulations; in others, it is inferred through case-law. We consider further Australian-specific areas of law below.

Social Licence to Operate

In addition to the law (legislation and case-law), another source of external control and oversight of new technologies is broad acceptance by the community. This is often dubbed the Social Licence to Operate (SLO). This concept is well established within the sustainability sphere, but it is a useful concept to describe the trust between enterprises and their stakeholders: customers, employees, owners, suppliers and regulators. The "licence" is of course informal, but it is nonetheless consequential should it be revoked. Enterprises that cannot leverage AI systems will find themselves at an increasing disadvantage, and playing catch-up will become more fraught.

In the context of AI systems, it is particularly important for two reasons. Firstly, there is widespread mistrust of AI systems – especially so in Australia:

> **"Australia is among the nations listed as the most fearful of AI, with many Australians believing its risks outweigh the benefits … Less than half of Australians are comfortable with the use of AI at work and only a minority of Australians believe the benefits of AI outweigh the risks." (*Trust in Artificial Intelligence: Global Insights 2023*, KPMG / University of Queensland)**

The AI Midriff Scandal

In January 2024, Channel Nine News released an image of Victorian politician, Georgie Purcell. It had been digitally manipulated by their graphics department, using Adobe Photoshop's Generative Expand AI capabilities. A cropped version of the image was expanded to include her midriff, which was portrayed as exposed (instead of covered by her dress, in the original image). She also claimed the altered image had enlarged her breasts. Channel Nine apologised – which Ms Purcell accepted – and attributed the mistake to an automation by Photoshop, while acknowledging it was a failure in their editorial processes. There was a significant national conversation about the role of AI in news production, the representation of women in politics and many other aspects of the issue.

News organisations have been manipulating images with Photoshop for decades and so have editorial processes in place. Do the new AI capabilities – Generative Expand in this case – mean the existing governance mechanisms are no longer fit-for-purpose? Or were the existing ones simply not followed? This case highlights that backlash against Generative AI can threaten the social licence to use AI.

Victorian MP Georgie Purcell criticises Nine News for 'sexist' image editing to make outfit more revealing

Adobe casts doubt on claim 'automation by Photoshop' was to blame as network apologises for image shown during bulletin

Get our morning and afternoon news emails, free app or daily news podcast

Follow our Australia news live blog for latest updates

Georgie Purcell says she had her 'body and outfit Photoshopped by the media' after the imag appeared on Nine News in Melbourne on Monday. Photograph: Nine News/Darren Howe

The Victorian upper house MP Georgie Purcell has lashed Nine News in Melbourne for using an image edited to make her breasts look bigger and expose her midriff, which the network blamed on "automation by Photoshop".

Secondly, it is during the early adoption phase of technology that mistakes are most likely and more forgiveness is asked. To create value and benefits for stakeholders, enterprises therefore do not just need a *social to licence to operate* AI, they also need a *social licence to innovate* with AI.

We propose that a key objective for Enterprise AI Governance is to secure this social licence to innovate through education, transparency, benefits-sharing, careful use of experiments and reassurance. Active participation in AI regulation development can be an important part of securing this licence.

International Perspectives

The rate of change in regulation is very high; consequently, any summary of regulatory positions and initiatives is quickly out of date. However, it is useful to present one to help understand differences around the world as a) this can inform likely developments in Australia, and b) many Australian businesses operate in foreign jurisdictions and are likely to be subject to those regulations.

The European Union

The EU has been working on their *AI Act* since it was proposed in 2021. It has been refreshed and revamped multiple times, reflecting the difficulties of regulators keeping pace with technological advancements. They have proposed a risk-based approach, with certain high-risk uses of AI systems to be banned outright. (This includes indiscriminate facial recognition, manipulation of vulnerable groups and social credit score type systems.) Lower tiers of risk would be permitted, with different levels of disclosure required of the training data, underlying models and evaluation processes.

In addition to regulating use, they are also proposing to regulate highly capable foundational models directly, outside of any specific use. The definition they use is the of amount of compute resources (measured in Floating Point Operations, or FlOps) used to create the models. These systems are subject to additional testing requirements, disclosures on their source data and performance. There are presently carve-outs for open-source systems, some aspects of security and policing use cases and existing systems. It is not expected that the laws will be in force until 2025.

The United Kingdom

The Government of the United Kingdom has sought to differentiate itself from the relatively tight and prescriptive approach of the EU with a more pro-innovation, light touch stance. They have ruled out any AI specific regulations via legislation in the short-term. Instead, they are considering reviewing individual legislation to determine if any modifications are required to account for AI capabilities.

Singapore

The Monetary Authority of Singapore – which regulates banking and financial services – has been among the earlier regulatory supporters of AI adoption through their Model AI Governance Framework from 2019. This framework provides practical steps (but not mandatory rules) for organisations to deploy AI, including testing, evaluation and controls. In 2024, it was announced that this framework has been extended to include Generative AI as well. Singapore intends to continue providing guiding principles, practical support and best-practice use cases, rather than directly regulating through legislation.

The United States

The US President announced new Federal Government regulations via an Executive Order, rather than legislation. It directs various US agencies to collaborate on guidelines, guardrails and establish the new National AI Research Resource. It builds on an early private sector voluntary code, and the National Institute of Standards and Technology AI Risk Management Framework. It is also light touch and has a focus on geopolitical competition and ensuring the US maintains its lead in AI technology. There are some requirements to mitigate specific kinds of discrimination and the Federal Government seeks to influence industry practices as a model customer for AI-based products and services it procures.

China

The Chinese Government has introduced several regulations covering all facets of AI: predictions, recommendations, content generation (Generative AI), source data, training, ethical use, privacy protection, promotion of "socialist core value" and more. There are requirements for registration, audits and complaints-handling, backed by penalties for non-compliance. It is by far the broadest, most in-depth and prescriptive regulatory regime of any major country. Much like Chinese regulation of the Internet, the focus is on stability and control of the new technology to allow people to benefit while preserving the status quo.

Global Initiatives

There four significant initiatives to help coordinate responses to AI between countries. The first (and least formal) is the World Economic Forum's AI Governance Alliance. This global thinktank prepares and disseminates content relating to AI from governments, public interest organisations, private sector consultancies and universities. While guidelines and best-practices are shared, it operates more as clearinghouse of ideas than a source of regulation.

Second, the Organisation for Economic Cooperation and Development operates a similar AI Policy Observatory with a clearer focus on providing prescriptive advice to member state regulators. This covers topics like definitions, classification and AI incidents, where the language is often used in national regulations and legislation.

Thirdly, the United Nations' UNESCO organisation has released a high-level policy document describing the ethical use of AI, grounded in four core values of human rights. This has been ratified by all 193 members of the UN in 2021.

Lastly, perhaps the most formal is the Bletchley Declaration, signed by 28 participants at the 2023 Global AI Safety Summit in the UK (including the UK, EU, US, China and Australia). It builds on earlier UNESCO agreements regarding the ethical use of AI and some OECD definitions of technologies and uses. The crux of the Declaration is an agreement from signatories to continue to work together, consult and exchange ideas as they develop their national regulatory environments.

Existing Regulations

There are many sources of AI regulation extant in legislation. These include the various anti-discrimination laws (e.g. age, race, sex and disability), Competition and Consumer Act, National Consumer Credit Protection Act, Privacy Act, Copyright Act and others. Some regulations are under development and require State cooperation, such as emerging frameworks for regulating autonomous vehicles (e.g. self-driving cars).

In Australia

Relative to international peers, Australia's response to AI regulation has been light touch and lacking in detail. There is no AI Act in Australia. Much of the practical effect of AI regulation is held in legislation and case law relating to specific uses. While the lead on policy development is the Federal Government's Department of Industry, Science and Resources, the Attorney-General's Office also has carriage of substantial policy-related questions through its work on new regulations relating to Privacy and Intellectual Property. The Communications Minister's office is also examining issues relating to the use of deepfakes, hate speech and other harmful content.

Australia's AI Ethics Principles

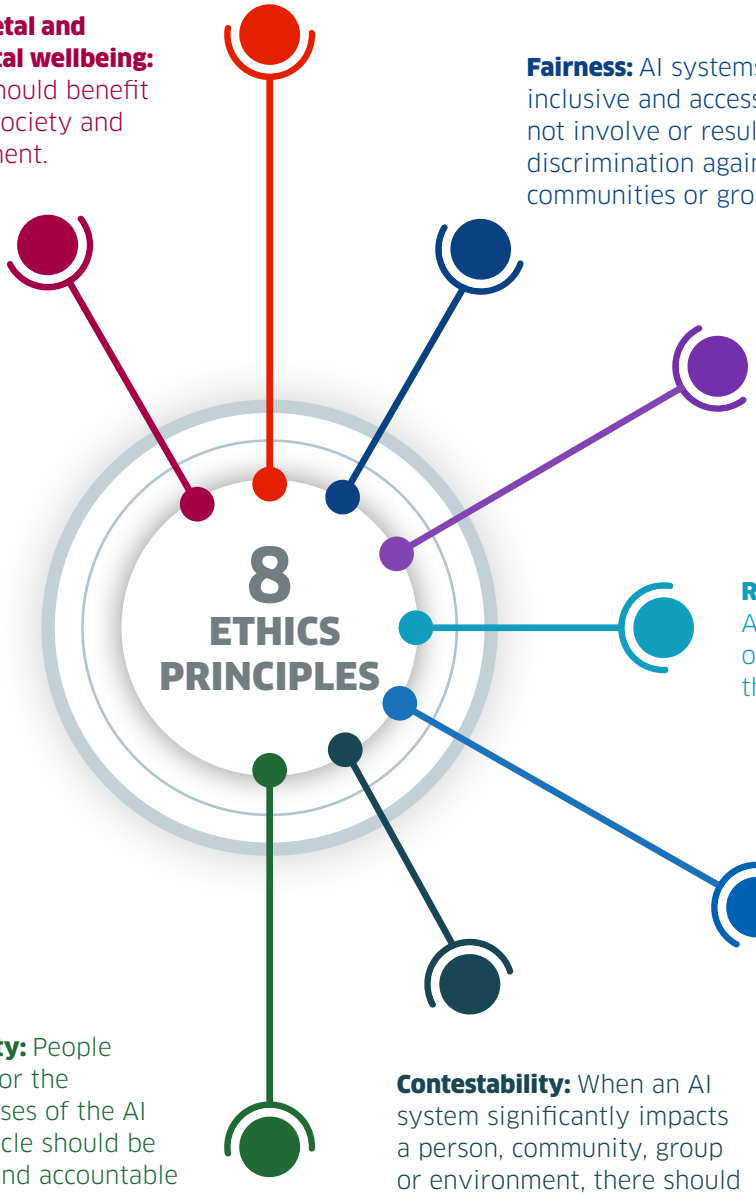
In 2019 The Department of Industry, Science and Resources (DISR) published the "*Australia's Artificial Intelligence Ethics Framework* after consultation with government agencies, industry and the general public. This document covers a set of eight voluntary principles, with guidance on how to interpret, implement and assess them, alongside case studies from leading private sector organisations.

The principles are broadly in line with language used in international forums. Thera are no reasons offered as to why only AI Systems should be designed and assessed against these principles rather than any decision-making process (automated with simple rules, or fully manual or a hybrid).

Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals.

Human, societal and environmental wellbeing: AI systems should benefit individuals, society and the environment.

Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.



Privacy protection and security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.

Reliability and safety: AI systems should reliably operate in accordance with their intended purpose.

Transparency and explainability: There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.

Accountability: People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.

AI Policy Organisations

The Australian Government coordinates AI industry development through the National AI Centre, housed within the Data61 business unit at the Commonwealth Scientific and Industrial Organisation (CSIRO). The Centre runs the Responsible AI Network, bringing together enterprises, startups, academia and government to exchange ideas and best-practices through webinars, forums and conferences. NAIC has a focus on AI at Scale, Responsible AI and Diversity and Inclusion with AI. NAIC also provides expert panels to help inform the Federal Government on a wide range of policy questions. The experts are drawn from industry, academia and public interest organisations.

Another significant actor in the regulation of AI is Standards Australia. This non-government, independent body is the local representative for the International Standards Organisation (ISO/IEC). In the context of AI, Standards Australia has produced a roadmap for the development and implantation of standards for interoperability of AI systems, while ensuring they are well-governed. Goals include harmonisation with international standards, promoting Australia's competitiveness and meeting expectations for fairness and ethics.

The Gradient Institute also plays a prominent role in Australia's AI regulatory environment. This non-profit organisation provides deep technical expertise in Machine Learning, Data Science and related fields to support policy development initiatives. They have a focus on research and practical implementation of AI systems through an ethically aware lens. They provide training and advisory services to industry, collaborate with public sector organisations on reports and receive donations from individuals and philanthropic organisations.

AI Regulation Roadmap

In mid-2022, DISR released a request for comment on a discussion paper, "*Safe and Responsible AI in Australia*". This summarised recent developments in international regulatory efforts, highlighted the potential benefits for Australia and presented some trade-offs for prescriptive and self-regulatory approaches. It outlined a risk-based approach to regulating AI, broadly following the EU and Canadian approach, with a focus on the categorisation of potential harms with a tiered response to oversight and obligations. DISR received over 500 responses from individuals and organisations.

In January 2024, DISR released an interim response. This includes discussions (but no commitments) on issues ranging from voluntary safety standards for AI, pre-deployment testing principles, watermarking and labelling of AI-generated content, and use of copyrighted materials to train models. The Federal Government has committed to spend over $40M on developing AI industry policy through NAIC, including $17M for Small and Medium Enterprises to adopt responsible AI. A temporary expert advisory group will be established to support the government's development and assessment of potential mandatory guardrails.

The Productivity Commission released three research papers relating to AI in February 2024. Broadly, the PC calls for a careful approach to regulation at this early stage, cautioning against AI-specific legislation. In January, the ASIC Chair Joe Longo outlined his thinking on AI regulation. While acknowledging that AI is "not the Wild West" and is already regulated by legislation, he queried whether this was sufficient, noting that some consumers may be harmed unknowingly by algorithms, and so struggle to seek redress.

Topics in AI Regulation

This section provides some brief definitions and discussion on terms and concepts that regularly appear in the regulation of AI and AI Governance more broadly. Like other areas of technology regulation, everyday language often lacks the precision required for practical systems implementation. This can make it challenging for policymakers and legal experts to develop regulations. Another issue is that some principles can be antagonistic in some circumstances. For example, avoiding bias might require an AI system to have more knowledge about a person than they are comfortable sharing (undermining privacy). Lastly, a broader criticism of AI regulatory principles is that they can end up as lists of overlapping near-synonyms, so anodyne as to become platitudes. This vagueness becomes a barrier to adoption, as adopters become concerned about being caught in an overly broad definition.

AI Ethics

This is an especially fraught and loaded term as it has quite different meaning to people while allowing them to assume their definitions are universal. For example, the Chinese Communist Party defines the ethical use of AI to be one that promotes socialist values and the various doctrines of the CCP. The Vatican also has a policy on AI ethics and how it can be used, consistent with Roman Catholic religious values. Their positions on topics like AI supporting IVF treatments or content describing the origins of COVID-19 are unlikely to align on ethical.

Despite these conceptual challenges, most organisations with large-scale AI research and operations capabilities – especially US-based tech giants – have invested significantly in AI Ethics. They have tended to focus on identifying harms from AI systems and supporting the development of tools to identify and combat them.

Ultimately, questions about what is considered ethical in the context of AI will sit with the Board of Directors – as they are in all contexts.

Privacy and Security

Privacy is defined formally in Australia under The Privacy Act. However, there is also a more subjective perspective, sometimes referred to as the "ick factor" or "creepy factor". This is when people feel some social norm has been transgressed by an AI system that seems to know too much. As a result, most lists of AI principles include a reference to respecting privacy. This is highly idiosyncratic and difficult to assess on a case-by-case basis. However, hiding relevant information from AI systems (or decision-making processes more broadly) will often result in sub-optimal performance for the person, organisation or society. Most organisations manage this through consents; however, consent fatigue can become an issue.

Additionally, if consent is subsequently withdrawn, this can create operational issues if users have (as in the EU) a "right to be forgotten". While "white-box" content relating to an individual can be deleted and expunged from traditional databases simply and directly, this is not possible for AI's "black-box" models. This challenge is known as unlearning and is an area of active technical research, as the only way to effect this at present is to retrain the model again with that individual's data removed – likely a substantial cost or not even technically feasible.

A further concern is that private information could be disclosed inappropriately. The risks of capturing and storing personal and sensitive information are well-known, with accounts of hackers and other cybersecurity breaches in the news. Now, with Generative AI, a whole new class of security concerns have arisen. While Generative AI systems don't typically store verbatim copies of this type of information, it may be recovered when generating output, resulting in inappropriate disclosures. (This can be either inadvertently through poor designs or deliberately through malicious or adversarial use.)

Bias, Fairness and Discrimination

Discrimination is the most useful function of any decision-making process (including AI). It is unlawful to provide credit to a child, a bankrupt or a deceased person, so a credit assessment process needs to enforce that distinction. It's also unprofitable to extend credit to someone who will not pay it back. This is desirable discrimination.

However, sometimes undesirable discrimination can be introduced unwittingly, using biased historical examples in the training process. This can result in discrimination that negatively impacts people, threatening the brand or even being considered unlawful under various anti-discrimination laws.

While there are reasonably clear tests for *unlawful* discrimination, *unfair* discrimination is harder to nail down as fairness is rarely legally defined. Within the AI technical research community, at least 14 different definitions (and metrics) have been proposed and studied. They are overlapping and not compatible with each other, making assessments highly subjective. And, again, in both law and practice, undesirable discrimination is possible in processes with no AI involvement.

AI Governance in Supply Chains

In today's complex global economy, supply chains are increasingly turning to Artificial Intelligence (AI) to streamline operations and optimise distribution networks. AI-driven predictive analytics enables precise demand forecasting, reducing overstock and stockouts, thereby significantly cutting costs and improving customer satisfaction. AI algorithms optimise routes and delivery schedules, improving fuel efficiency and reducing delivery times. AI also plays a crucial role in supplier selection and management to assess supplier performance and risk, ensuring more resilient supply chains. Additionally, AI enhances quality control processes through visual inspection systems, identifying defects more accurately and at a higher speed than human workers. Through these applications, AI not only increases efficiency and reduces costs but also supports more sustainable and resilient supply chain operations.

Incorporating AI into supply chains offers transformative potential but also necessitates robust governance frameworks to navigate ethical, legal, and operational challenges. Effective AI governance in supply chains can ensure these technologies are used responsibly, enhancing operational efficiency without compromising ethical standards or regulatory compliance.

In supply chains, **the ethical use of AI** is crucial for ensuring that operations do not harm individuals or communities. Ethical considerations must include ensuring AI does not exacerbate existing inequalities in the supply chain, such as unfair labour practices or unsustainable sourcing that disproportionately affects vulnerable populations.

Transparency and explainability within supply chains extend to revealing the rationale behind AI-driven forecasts, procurement decisions, and inventory management. Stakeholders should have access to information about how AI systems make decisions, especially when these decisions impact human workers, suppliers, and customers.

In the context of supply chains, **data privacy and security** are especially challenging given the multiplicity of stakeholders, from suppliers and logistics providers to retailers and consumers. Governance frameworks must address the complexity of sharing sensitive information not only locally but across borders, while complying with diverse regulatory environments.

Organisations should establish clear **accountability and oversight** to assess the impact of AI-driven decisions on supply chain efficiency, worker welfare, and supplier relationships. Organisations could adopt a multi-stakeholder approach to AI governance, involving representatives from different segments of the supply chain, including minority suppliers and community representatives, to ensure diverse perspectives are considered in decision-making processes.

AI governance in supply chains with a focus on **sustainability and environmental responsibility** requires the integration of AI tools capable of analysing life cycle assessments of products, optimizing routes for lower emissions, and predicting the environmental impact of different supply chain strategies. Governance frameworks should mandate the use of AI for environmental risk assessments, ensuring that supply chain practices do not lead to ecological degradation or resource depletion. Additionally, they should encourage the adoption of AI-driven innovations that contribute to the transition towards green logistics and sustainable supply chain practices.

About the Author:

PROFESSOR YALÇIN AKÇAY

Director Centre for Business Analytics and Professor of Operations Management

Yalçın Akçay joined Melbourne Business School in 2017 and is currently the Director of the Centre for Business Analytics.

He received his dual-title PhD in Business Administration and Operations Research from the Pennsylvania State University, Smeal College of Business, in the United States.

Yalçın's research focuses on revenue management, dynamic pricing, inventory management, retail operations, stochastic modelling of service and manufacturing systems. His research has been published in leading academic journals, including Management Science, Operations Research, and Production and Operations Management. Yalçın's papers received the Wickham Skinner Best Paper Award from the Production and Operations Management Society (POMS) and the IIE Transactions Best Paper Award from the Institute of Industrial and System Engineers (IISE). His work was also selected as the Revenue Management and Pricing Section Practice Prize Finalist by the Institute for Operations Research and the Management Sciences (INFORMS).

Yalçın teaches Operations Management, Quantitative Methods for Business, Data-Driven Decision Making and Optimisation at the MBA, Executive MBA and Senior Executive MBA programs.

Yalçın's consulting work covers a broad range of applications in analytics. Some of his projects include demand forecasting with Ford (predictive analytics), assortment optimisation and rationalisation project with Unilever (prescriptive analytics), dynamic pricing with Avis (prescriptive analytics), second-hand pricing with Fleetcorp (predictive analytics), and a cross-selling with UniCredit (predictive and prescriptive analytics).

Transparency and Explainable AI

Many discussions on regulating AI suggest transparency about when it is used. While there may be some challenges in defining AI, this is reasonably straightforward. Another side of transparency of use is to disclose when humans are in fact doing the work but masquerading as AI systems. This is known as "turking" (in reference to The Mechanical Turk), and is not uncommon for validating new AI use cases and building transactional datasets for training AI. Here, a user is chatting with a bot or undertaking some other tasks, believing it is an AI system when it's a human.

The other sense of transparency relates to the introspection of the AI system's operation. This is broadly understood as the ability of an AI system to provide an explanation of why it produced a particular decision (or content, in the case of Generative AI). This is frequently mentioned because many believe it will help avoid undesirable decisions (or content) if its process can be reviewed by a human. (Further, this "right to an explanation" has been encoded in the EU's GDPR laws.) In the case of modern Machine Learning and Generative AI (black-box) systems, the underlying rules are generally not interpretable by humans.

One technical response is to use counterfactual reasoning to produce statements like "*had the inputs looked like this instead, a different result would have been produced*". This type of reasoning can be used to estimate the contribution a particular input made to a particular outcome. So, a weight score (e.g. Shapley value) can be given to each input element (in aggregate) or, in the case of an individual, the sensitivity of the outcome to each input. These "explainable AI" approaches – and related tools and techniques – have limitations and continue to be an active area of AI research.

Counter-intuitively, in many practical settings, "opening the black box" in this way is undesirable. For example, processes for assessing creditworthiness or determining identity can be gamed by adversarial users if too much detail is provided. There are also potential issues with commercial confidentiality and competitiveness. Lastly, many human-based decisions do not meet the requirements of explainability either. If an AI system is performing a task – and it can clearly perform as well as or even better than a human – it is not always obvious why this additional requirement should be imposed.

AI Safety

This is a relatively new term in the field of AI regulation. Unlike AI Ethics, which is concerned with the impacts on people today from AI-enabled decision-making, AI Safety focuses on future threats to humanity as a whole (akin to nuclear war, asteroid strikes or virus pandemics). AI Safety proponents – often drawn from the Effective Altruism movement's Longtermism wing – focus on these so-called "existential risks" arising from rogue "unaligned" AI systems, able to manipulate humans to gain access to nuclear or biological weapons and destroy humanity. It is reassuring to note that AI systems with these capabilities do not yet exist – and the AI Safety community is committed to ensuring it stays that way.

While it may seem farfetched, many leading lights at the frontier of AI research include themselves in this group. They have urged for greater regulation and, famously, in 2023 coordinated many scientists and leaders to sign an open letter urging governments to pause further AI development.

Critics claim that while some are misguided technologists who have become enamoured with their creations, others are just cynically creating marketing buzz and a regulatory moat to protect their commercial advantage. Many in the AI Ethics camp consider AI Safety concerns to be, at best, a distraction from known harms in the present day. Recently, a rival movement called e/acc has emerged seeking to counter the "doomer" narrative of human extinction, and seek instead to accelerate the adoption of AI.

Intellectual Property

Copyright law has the goal of seeing content creators sufficiently rewarded for their risk and efforts, while allowing society to access and build upon knowledge, ideas and creations. The optimal balance around how much of a monopoly to grant copyright holders has ebbed and flowed over the years and is tied to the prevailing economics of the technologies of the day.

As with the introduction of the photocopier, the VCR and the Internet, Generative AI systems are forcing a re-think on the Grand Bargain between copyright holders and society. What's clear is that Generative AI systems – whether text, images or code – require huge amounts of data for training purposes. Whether or not this constitutes fair use will play out through the various courts, commercial licensing agreements and potentially legislation. It's also likely that, in the future, we will need fewer humans in the paid content business, which further shifts the economics balance of copyright policy settings.

On the other side of the coin, what rights do AI systems have? Presently, only China's system grants copyright on content produced by Generative AI systems. In Australia, the Federal Court has ruled that AI systems cannot be an inventor on patents. The situation with AI systems defaming people is less clear, though it seems unlikely to be actionable.

Companies are responsible for their chatbots

In 2022, passenger Jake Moffatt used the Air Canada chatbot to enquire about their policies regarding bereavement fares. Based on this advice, he booked flights and then sought a refund. The airline refused, on the grounds that he did not follow their published policies. He took them to a tribunal and the Canadian court found in his favour. Air Canada argued that the chatbot was a separate legal entity "responsible for its own actions". This was rejected outright by the tribunal, which noted "It should be obvious to Air Canada that it is responsible for all the information on its website. It makes no difference whether the information comes from a static page or a chatbot."

This case makes it very clear that organisations cannot use AI systems to evade responsibility for their decisions and content – there is a clear chain of delegation from the Board to the AI system.

Air Canada ordered to pay customer who was misled by airline's chatbot

Company claimed its chatbot 'was responsible for its own actions' when giving wrong information about bereavement fare

The judge wrote that Air Canada's customers had no way of knowing which part of its website – including its chatbot – relayed the correct information. Photograph: NurPhoto/Getty Images

Canada's largest airline has been ordered to pay compensation after its chatbot gave a customer inaccurate information, misleading him into buying a full-price ticket.

Air Canada came under further criticism for later attempting to distance itself from the error by claiming that the bot was "responsible for its own actions".

Environmental, Social and Governance

Many organisations have developed or subscribed to ESG Principles. It's worth considering how the adoption and use of AI systems could impact these commitments.

Firstly, with **Environmental** impacts, it's worth noting that some AI systems – Generative AI especially – use tremendous amounts of energy and water. For most Australian organisations, this is mitigated by the fact that very few will be training their own from scratch but leveraging foundational models built by large tech firms in the US, EU and China. However, fine-tuning and inferencing will likely increase resource consumption and should be considered too.

Secondly, a number of **Social** impacts should be considered. For example, much of the training of OpenAI's hugely successful ChatGPT series of Large Language Models was undertaken by humans in low-cost countries like Kenya, which paid around $2 per hour. Enterprises whose ESG commitments extend to slavery and labour exploitation in their supply chains should consider whether these practices are acceptable. Other Social impacts include job transition and re-skilling, the dehumanising of workforce management and universal access to services.

Thirdly, with **Governance**, there are risks that some organisations could use AI – even inadvertently – to break accountability between policy-setters, decision-makers and outcomes. It should never be the case that an unlawful, harmful or even unprofitable outcome resulted for which no human is accountable. However, psychologically, there is a tendency for managers to become deferential to algorithms in some circumstances. Also, when decisions or content are produced that receive scrutiny, there will be a natural temptation to blame the algorithm to avoid criticism. Both of these need to be guarded against through clear AI Governance policies and practices, ensuring that the appropriate delegations and controls are in place.

Lastly, we should keep in mind that, with each of these factors, there are ways that AI systems can make improvements on the status quo as well. Whether it's optimisation of resources, increasing scrutiny of supply chains or automation of controls and oversight, AI systems will likely play a role in improving outcomes. Some enterprises will look to balance the positive and negative across the three categories; those with ESG policies may refresh them considering this.

TOWARDS AI GOVERNANCE

Like any question of Corporate Governance, responsibility for AI Governance starts with the Board of Directors. It is required to set in place policies and then oversee the implementation of decision-making processes that govern AI systems. We frame AI Governance more broadly than just regulation: Boards have a responsibility not just regulators and society but many other stakeholders too. In the case of publicly listed companies, they have an obligation to act in the interests of their shareholders, with a view to maximising the value while managing risks.

In this section, we outline some broad principles for AI Governance and then summarise eight key questions that we propose Boards should be able to answer on an ongoing basis. The final section outlines some considerations for implementing a plan to ensure this can be done.

Principles of Governance

There are four essential principles that allow Boards to govern AI functions.

Delegation

While Boards are ultimately accountable for the actions of the enterprise, for practicality they **delegate** decision-making authority to the senior executives and officers of the company (the C-suite), who in turn delegate to managers and other employees in a cascade. These delegations are often bound in scope in some way – for example, time or location or level of financial exposure.

In the context of AI Governance, some aspects of decision-making may be delegated to the AI system. It should be clear to all stakeholders that the AI system – as a piece of machinery – cannot be responsible for its output. That firmly remains with the humans who have built, deployed and operated the AI system in that context.

Escalation

Of course, Boards (and senior executives) are not omniscient and so cannot anticipate all scenarios – especially in a novel and fast-moving space like AI. A key mechanism for effective governance is the process of **escalation**, which sees information about a situation conveyed upwards through the enterprise to the appropriate level to provide a response. Similarly to delegation, the Board needs to put in place appropriate rules and processes to define and enforce this. As well as a hierarchical process, Boards should consider establishing information flows that bypass formal structures, like whistleblower processes.

Observability

In systems terms, **observability** refers to the ability of a system owner to understand the current state of the system. Modern enterprise AI systems produce huge amounts of observations about the system, its operations, input data, output data, usage, defects and a raft of supporting data. It is far too much information for any person to ingest, yet alone busy Board members and executives. So, thought must be given to specifying the right level of information for everyone in the enterprise to fulfil their obligations.

Controllability

The other leg of governance is **controllability**. This describes how a system owner can exert influence and direct the operation of a system. For AI systems, there are several key decisions across the AI Lifecycle (see Section 1) where control can be imposed. At the higher levels, examples include decisions about AI strategy, partnerships and major system rollouts. At the lower levels, there are system parameters and other technical measures that influence the behaviour of systems. In the case of simpler white-box AI systems, the rules themselves are available for inspection, modification and approval. In the case of black-box AI systems, it is far more challenging, requiring different approaches to evaluation and assessment.

With these concepts, we can consider a simple 2x2 of AI Governance design.

	Observability	**Controllability**
Delegation (top-down)	Specify suitable metrics, thresholds / triggers, sensitive sub-populations and testing protocols for AI systems	Specifying approval criteria for key decisions across each phase of the AI System Lifecycle and the level of the enterprise responsible
Escalation (bottom-up)	Provide formal reporting, alerting and ad hoc analysis; allow whistleblower and similar pathways for information to flow upwards in the enterprise	Design and test "failure modes" (including reversion to manual processes), human-in-the-loop reviews, spot testing, red-teaming and adversarial attacks.

Dimensions of AI Governance

When designing an AI Governance function, it's useful to think in concrete terms about the objectives and success criteria for the AI system in context. However, when dealing with many AI systems, a more abstract approach is required. We propose four dimensions for AI Governance that can help Boards and senior executives to think systematically and strategically about how well-governed these systems are. We introduce these as questions, which can be tested periodically to ensure the enterprise is making continuous improvements to its AI systems.

Strategic Alignment

Are we aligned to the enterprise's strategy?

Strategy typically informs questions like where the enterprise competes, how it's positioned in the marketplace, where resources will be deployed (or divested) and priority areas for growth and risk-taking. Pursuing AI systems in non-strategic functions of the enterprise could see unnecessary risk and wasted efforts.

Are we informing the enterprise's strategy?

Conversely, in this age of transition, the enterprise strategy should consider nascent AI capabilities in suppliers and competitors, customer expectations, changing workforce requirements and regulator demands. Sources of value will shift across the value chain and enterprises need to ensure they don't find their strategy outdated.

Value Creation

Are we clear on the business case?

In most cases, the rationale for AI system introduction is value creation. Like any initiative, good governance requires a clear business case, with an articulation of expected costs and benefits over time versus a baseline. Enterprises will have their own value metrics and evaluation programs; it's important that any AI systems are aligned to the value metrics and any assumptions have been validated by subject matter experts.

Are we realising the value expected?

Post-implementation reviews, test and learns and retros are common tools for assessing post hoc initiatives. However, with AI systems, measuring the counterfactual – what would have happened without the AI system – is more complex and usually requires advanced statistical techniques. This often resides in the Data Science team that built the system, so thought should be given to how to avoid biases in this review.

Operational Performance

Are we meeting commitments?

As with any system, there should be clear performance objectives (Service Level Agreements, Key Performance Indicators or similar) covering things like up-time/availability, response time/throughput and accuracy-type measures. Additionally, AI system specific metrics should cover input data quality, changes in statistical properties of the inputs ("data drift") and unexpected or anomalous outputs (especially against sensitive sub-populations). Lastly, metrics relating to the business process, staff feedback and customer complaints should also be measured against targets.

Are we improving over time?

The use of AI systems in the enterprise is relatively new and most organisations are on a steep learning curve. Mistakes will be made and failures abound. However, careful measurement of operational performance should show improvements as expertise is acquired, economies of scale and scope take hold and stakeholder expectations shift. Time and resources consumed across the AI system lifecycle should fall while performance improves. If this is not occurring, it's important to understand why to rectify.

Risk Management

Do we have clarity on risk appetite?

Inserting AI systems into a business process is a risky proposition. Things will go wrong. The Board should provide clarity on the types and degrees of risk the enterprise is willing to bear in pursuing its goals. This should consider financial losses (from fines and poor system performance), broader regulatory consequences, ongoing access to talent and technologies and, perhaps most crucially, the social licence to operate (and innovate) AI systems.

Are we operating within risk tolerances?

The enterprise needs to implement mechanisms through the AI Governance matrix to ensure observability and controllability of AI systems with respect to risk appetite. In particular, Boards need to be assured that the mechanisms in place to observe and control the operation of AI systems are effective and working as intended. This requires assessments and testing, alongside scenario planning for the organisational response when a risk materialises as an issue.

Implementing AI Governance

We provide some practical (though necessarily generic) advice for Boards and executives to consider when designing, implementing and executing an AI Governance Model.

Scoping AI Governance

The first task is to form a view on what constitutes AI systems in the context of the enterprise. There are three broad approaches to consider, weighing up the time and cost to implement against the level of acceptable risk.

Narrow Scope

In this view existing AI systems – underpinned largely by Machine Learning – are governed as per existing processes and accountabilities. Only AI systems leveraging the newer Generative AI (such as OpenAI's ChatGPT or Google's Bard) fall into scope. This is a much smaller scope and ensures only greenfield initiatives are captured, significantly reducing the time and effort involved. However, some high-risk AI systems, leveraging current Machine Learning techniques, will be excluded. This could result in duplicated efforts, confusion about allocation or even governance gaps.

Broad Scope

These sees all automated decision-making systems in scope – even the white-box ones informed by direct business logic and simple calculations. For most enterprises, this would include all business processes – internal and external. While many of these would be low-risk and require cursory reviews, it would still be a formidable task. There would be efficiencies from sharing governance procedures and resources all automation systems, especially data and IT infrastructure.

Middle Scope

The definition of AI System here is framed to capture only black-box systems – that is, those whose business logic and rules were not directly created or reviewed by humans. This would include Generative AI but also Machine Learning based systems. It's likely some systems in production would be covered, increasing workloads, while the same people and tools are likely involved with both types, increasing efficiency.

AI Systems Mapping

With this scope and definition agreed, the next step is to undertake an inventory of AI systems. This should be very broad in the first instance, covering systems across the lifecycle (see Appendix 2) and business functions (see Appendix 1). Vendor products with integrated AI as well as AI-specific tools (including open-source components) and those of suppliers and contractors should be captured.

The stock-take should develop a systemic approach to recording relevant information about each AI system to allow for whole-of-enterprise analysis. As a start, consider recording:

Data: What data was used for training the model? What data is used as input when in use? Where are the outputs stored?

Models: How were the models built? By whom? With what libraries, packages and pre-trained components? What assessment, testing and reviews did they go through?

Platform: Which technical platform are the models built and executed on? Are they on-cloud or on-premises? What types of monitoring are in place? How is business continuity planning (BCP) handled? What is the security model?

Usage: What is the status of the system? How is it used and in what contexts? What change management (including training and communication) supports it?

Approvals: Who can access the source data? The models? The outputs? Who approves access? Who can make changes to the source data or models? Who approves those changes? How is this logged?

Ownership: Who is funding the AI system? Is there a business case? How is value measured? Who owns the roadmap for future development?

Lastly, the use of "shadow AI" – akin to "shadow IT" where staff bring unofficial IT tools to work – needs to be considered. It's likely ChatGPT and other free consumer-oriented tools are being used by some people. Anonymous surveys and technical firewall monitoring can help gauge the extent of use.

AI Expectations Mapping

Any enterprise must manage multiple stakeholders, often with their own expectations on a given topic. AI is no different. We suggest capturing expectations (and undertakings and obligations) from the external environment, as well as self-imposed, to ensure a comprehensive view is built and maintained.

External Sources

As outlined in Section 2.2, laws relating to AI systems are found across multiple pieces of legislation (including privacy law, Australian consumer law, anti-discrimination law and many others). Further, there may be laws and regulations specific to the industry (e.g. aviation) or use cases (e.g. credit assessments). For multi-national enterprises, there is the additional complexity of considering foreign jurisdictions.

Some enterprises are bound by codes of conduct that relate to the industry or locale they operate in. Other stakeholders include employees (via awards or enterprise bargaining agreements) and vendors (supply contracts, master service agreements and so on). Regulators may also have guidelines or best practices that, while not strictly required, should be considered.

Internal Sources

Formal contracts like Terms and Conditions or Standard Forms of Agreement should be reviewed for potential AI-related obligations. This could include undertakings about how data is stored and use, decision rights and dispute resolution. Similar commitments may be found in a customer charter or equivalent document. Many enterprises will have made commitments to shareholders through ESG and similar programs. Section 2.4 outlines potential impacts.

Lastly, and perhaps most importantly, enterprises should assess how the adoption of AI systems aligns with the brand. Understanding different stakeholders' expectations for different uses (and tolerances for potential failures) of AI in those contexts is key. As AI becomes more widely adopted and stakeholder exposure increases, those expectations (and tolerances) will shift – potentially rapidly.

AI Resourcing and Expertise

A significant challenge with governance of any technology is bringing together the expertise required to make well-informed choices. As AI is still an emerging field, there are few people who can claim mastery across the algorithms, technical platforms, regulatory and legal issues, and risk management principles. As a result, cross-functional teams are required and, depending on the organisation, external capabilities may be needed.

Technical Specialists

Expertise in Data Science in general and AI in particular will be required. As the field moves very quickly – capabilities, best practices, potential uses – these people will need to be exposed to a wide variety of project types and the ability to quickly scan and absorb and synthesis information. Many data scientists prefer to work on narrow problems and techniques, so this breadth of experience and managerial orientation will be particularly difficult to source.

Easier to find are the IT specialists with depth and breadth of expertise in building and operating complex modern data systems (including the cloud and AI-enabled enterprise applications like CRM and ERP). Many of the concepts transfer directly to the AI domain.

Legal expertise – checking contracts and other agreements and interpreting legislation, case law and other sources – is in short supply too when it comes to AI. However, properly informed by AI experts, they can apply their general analytical skills to this new domain.

Lastly, Risk Management specialists are adept at taking generic risk frameworks and tools, like the Three Lines of Defence model, and applying them to a specific enterprise context. It may be worth considering actuaries for this as some have both the algorithmic and risk expertise and are able to perform algorithm audits on critical processes.

Managerial Generalists

Expertise and resourcing are also required for setting up and running an enterprise-wide review, development and implementation. Capabilities relating to business case development and assessment, change management (including communication), organisational design, process design and leadership of cross-functional teams are all needed.

Where possible, these should be sourced internally to ensure fluency with the enterprise's key people, processes and culture.

Training and Upskilling

Many organisations have data literacy training programs in place to upskill their workforce on topics in data. These should be expanded to include AI Governance concepts, ensuring that workers have a role-specific understanding of AI concepts, policies and your organisation's governance model. This will help realise value while reducing risk, without becoming dependent on external providers.

Membership and Engagement

When establishing a working group, ensure there is clarity on its mandate (including the executive sponsor, decision rights and reporting lines), access to adequate expertise and resources (technical and managerial, ad hoc and dedicated, internal and external) and aligned representation across divisions, functions and geographies. Striking the right balance between too broad a membership and too narrow is critical.

Lastly, it is important to ensure that teams, panels, review boards and other bodies are representative of the community, with a variety of backgrounds, perspectives and experiences. It is also a good practice to involve end-users and other stakeholders during design, assessment and review. This will increase the chances of detecting issues early.

APPENDICES

APPENDIX 1: COMMON USES OF AI SYSTEMS

Many enterprises have been leveraging AI (as Machine Learning) for over a decade. Often, these use cases will be embedded in other products (like CRM or ERP systems). We provide a brief survey of some common uses for Machine Learning, alongside some putative uses for emerging use cases for Generative AI. Broadly, Machine Learning (and mathematical optimisation) can provide custom, narrow information to support analysts and decision-makers; Generative AI provides open-ended, interactive and highly flexible knowledge to both employees and customers.

Human Resources

Now: Resume screening, employee churn prediction, scheduling/rostering

Next: Position description authoring, application assessment, performance evaluation, training plan development, training material development (text, imagery, video)

Customer Operations

Now: Call triaging and routing, task/job allocation, performance measurement, online search/FAQs, simple chatbots, online advertising, complaints categorisation and sentiment analysis

Next: Sophisticated interactive agents (able to perform many customer service representative tasks)

Finance and Risk

Now: Budget forecasting, scenario planning, expense categorisation, cost optimisation, fraud detection, credit risk assessments

Next: Sophisticated fraud detection, auditing and compliance (incorporating use of natural language to assess and query transactions)

Retail and Supply Chain

Now: Demand planning, ranging and assortments, inventory optimisation, store location planning, discounting and price optimisation

Next: Sophisticated product recommendations (including bundling, cross-selling, discounting), latent demand assessment

Marketing

Now: Customer segmentation, Lifetime Value forecasting, direct marketing campaigns, marketing mix and attribution, discrete choice experiments

Next: Personalised product descriptions and imagery, personalised offers, customer experience feedback elicitation, competitor intelligence activities (new products, pricing, promotions)

Technology and Operations

Now: Resource utilisation, cybersecurity monitoring, software code generation (emerging) and testing, predictive maintenance, production planning and scheduling, warehousing and logistics optimisation

Next: Software code generation and testing (embedded), cybersecurity penetration testing, autonomous vehicles, robotics

General Administration

Now: Spelling and grammar, translation, document layout suggestion, related document matching, document classification

Next: Procurement documents (e.g. RFI/RFP/RFT and their responses), legal contracts (reviews, summarisation, customisation, generation), internal communications (newsletters, memos, blog posts), meeting minutes and summarisation, technical writing tasks, regulatory reports and filings, B2B sales outreach and proposals (and their evaluation), online research and monitoring of industry news and events

APPENDIX 2: AI SYSTEMS LIFECYCLE

Like many types of technologies, there is convergence on a set of common phases organisations typically go through in the end-to-end lifecycle. This is useful for governance purposes, as it allows for oversight, gating, benchmarking and comparison across initiatives. Leading technology vendors offer templated patterns for aspects of AI system project management.

Design

This phase sees the objectives and scope of the project articulated and key metrics defined; business-case developed; input data for training purposes; algorithms and models used; target technology architecture; change management planning; project oversight and governance arrangements. From a governance perspective, this is a critical phase to ensure success and avoid harms.

Development

The construction of the core models takes place during this phase, along with related user-interfaces and data pipelines. Considerable amounts of compute resources are required for model training and evaluation, along with the deepest technical and mathematical expertise. Testing also takes place in this phase.

Deployment

This is where the critical go/no-go decision takes place, along with appropriate change management activities (including training, communication and contingency planning). In most enterprises, AI projects are deployed into an existing business and technology context, so impacts on other production systems need to be considered.

Operations

This phase is usually the longest-running phase and where the value is delivered through usage. In contrast to training models from many examples, here we use inferencing (or scoring) to produce output for a particular input. This phase has been a key focus within the AI community these past five years, as organisations have historically struggled to move AI projects "from the lab to the factory". New roles (like ML Engineers) and functions (like MLOps) have emerged to manage the complexity of running AI systems in production. This includes monitoring performance (and harms), ongoing costs (especially cloud-related), security, versioning, release management, test and learn, oversight and audit. These activities increasingly resemble the "DevOps" frameworks used to govern other large, mission-critical IT production systems.

Retirement

Often overlooked, the lifecycle of an AI system is not complete unless the end-of-life and decommissioning are considered. Similar to the Deployment phase, the Retirement phase considers impacts on other systems and change management activities but also includes retention of critical information (such as models, data and documentation to support future audits), retrospectives and reflections for ongoing learning and continuous improvement.

APPENDIX 3: FURTHER READING

- "Supporting responsible AI: discussion paper", Department of Industry, Science and Resources
- "Making the most of the AI opportunity: The challenges of regulating AI", Productivity Commission
- "AI Policy Briefs", Massachusetts Institute of Technology
- "The economic potential of generative AI: The next productivity frontier", McKinsey Global Institute
- "AI Governance for Directors Webinar Series", Australian Institute of Company Directors
- "Empowering AI Leadership: An Oversight Toolkit for Boards of Directors", World Economic Forum

APPENDIX 4: GLOSSARY

AGI	Artificial General Intelligence
AI	Artificial Intelligence
ASIC	Australian Securities and Investments Commission
BCP	Business Continuity Planning
CRM	Customer Relationship Management
DISR	Department of Industry, Science and Resources
ERP	Enterprise Resource Planning
ESG	Environmental, Social and Governance
GDPR	General Data Protection Regulation
GenAI	Generative Artificial Intelligence
ISO/IEC	International Standards Organisation / International Electrotechnical Commission
MLOps	Machine Learning Operations
NAIC	National AI Centre
PC	Productivity Commission
SLO	Social Licence to Operate

Connect with the Centre for Business Analytics

cfba@mbs.edu

CONNECT WITH US

@MelbBSchool

/MelbourneBusinessSchool

melbourne-business-school

/MelbBSchool

@melbbschool

CENTRE FOR BUSINESS ANALYTICS

THE UNIVERSITY OF MELBOURNE

MELBOURNE BUSINESS SCHOOL