Australian Government
Productivity Commission

January 2024

Making the most of the AI opportunity

Research paper 2

The challenges of regulating AI

Contents

The Commission acknowledges and thanks the following Commissioners and staff who have worked on the Making the most of the AI opportunity research papers: Stephen King, Rosalyn Bell, Hudan Nuch, Rebecca Stoeckel, Jeremy Kamil, Rachel Burgess, and Ritaja Das.

The Productivity Commission acknowledges the Traditional Owners of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to their Cultures, Country and Elders past and present.

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

An appropriate reference for this publication is:
Productivity Commission 2024, *Making the most of the AI opportunity: The challenges of regulating AI*, Research paper, no. 2, Canberra

Publication enquiries:
Phone 03 9653 2244 | Email publications@pc.gov.au

The challenges of regulating AI

Key points

- **Approaches to regulating outcomes from uses of AI should be proportionate, effective and risk-based – enabling productivity gains from AI use while providing strong safeguards against adverse outcomes.**
  - When looking at new or existing regulation, governments should consider the nature of the potential harms and the risk of harms measured against appropriate real-world counterfactuals, and consider who has the ability and incentives to control risks.
  - As with any new technology, some consequences of AI use will only become apparent as the technology develops further and complementary technologies progress and are taken up. With general purpose technologies in particular, regulation based on 'predicted uses' or 'speculated harms' is likely to be overly broad and harm productivity.
  - Many potential harms have been encountered with past technologies and adequately dealt with by existing regulatory frameworks in areas such as consumer protection, privacy, anti-discrimination, negligence and sector-specific and profession-specific requirements. AI is no different.
- **Formal regulation is only one element of securing safe and ethical AI use. Risks of harm are also tempered by social norms, market pressure and coding architecture.**
  - Industry self-regulation (driven by a combination of industry codes, insurance and a focus on reputation) will play a role alongside formal regulation.
  - With Australia likely to import a significant amount of AI technology from overseas and domestic developers seeking sales in overseas markets, regulatory approaches in Australia's key overseas markets will by default act to regulate AI developments and outcomes in Australia.
- **A stepped approach to thinking about AI regulation is proposed, that recognises that new technology does not necessarily imply the need for new rules. In considering risks associated with new AI technologies, key steps could include:**
  - identify how the technology is already being used, or likely to be used in the immediate future (based on, for instance, stated intended uses or overseas experience)
  - determine whether this use results in heightened risks of serious harm compared to the counterfactual
  - identify which parties involved have the scope to influence risks and outcomes
  - determine whether the risk is adequately addressed by existing regulation, or whether extensions or modifications to this regulation, or improvements to its enforcement, are required. If a new regulatory instrument is needed, consider a technology-neutral approach in the first instance. If improved enforcement of existing regulations is required, it will be important to ensure regulators have the resources and skills to engage with and guide industry.

Artificial intelligence (AI) refers to a powerful set of tools with the potential to transform our economy and improve our living standards. The AI models that have emerged in recent years apply advanced machine learning to increasingly sophisticated uses, including natural language processing, image recognition, recommender systems, personalised search and social media.[1] Together, these technologies are increasingly undertaking complex tasks that were outside the scope of previous waves of automation.

But, like any tool, there are risks that without proper implementation or with little visibility, AI could be used in ways that harm individuals, businesses, the economy and/or society. Using AI can cause harm, for example from errors due to low quality technology, or from malicious or reckless use (Solomon and Davis 2023). As with any new technology, using AI has some unknown consequences, given that both the development and uptake of the technology continues to progress. This potential for harm has led AI users, experts and developers of AI to call for regulation.

So how should the Australian and State and Territory governments approach regulation in this rapidly developing field of technology?

The Australian Government's (2024) interim response to the *Safe and Responsible AI in Australia Consultation* provides a useful starting point. In this paper, the Productivity Commission presents a systematic and implementable approach to AI regulation that builds on the interim response and is designed to ensure that Australia can maximise the productivity gains from AI while providing a strong safety-net against adverse outcomes for individuals and business.

The Commission's approach recognises that regulation is only one element in securing safe, ethical AI use in Australia. Risks of harm can be tempered by social norms, market pressures and the coding architecture[2] (Lessig 1998, 2006). Industry stakeholders have consistently identified maintaining public trust as being key to AI use – and often a greater hurdle than (current) regulatory requirements.

It also recognises that new technology does not imply 'new rules'. Many of the potential harms that could be created by using AI are 'old wine in new bottles' – harms that we have previously encountered and that are adequately dealt with by existing laws and regulations. Effective implementation of AI will require our regulatory infrastructure to recognise where harms are already covered and adopt a flexible approach so that existing rules can be applied to the new context presented by AI.

Effective AI regulation needs effective regulators, who have the resources and skills to both engage with and provide guidance to industry. It also recognises that Australia will be one small part of a global AI ecosystem and regulatory landscape. This landscape is still developing and governments need to be active in making sure Australia's voice, input and interests are recognised at the international level.

[1] AI for the purpose of this paper should be distinguished from artificial *general* intelligence, which has not yet been developed and is outside the scope of this paper that considers uptake of existing AI technologies. Machine learning is a subfield of AI that describes algorithms which are capable of completing a task with minimal human instruction. Often these algorithms 'learn' to accomplish the task by being trained on example data.

[2] 'Coding architecture' refers to the coding that underlies all software and the functioning of the Internet, which inherently creates physical and technical constraints on actions that people can take.

1. What kind of regulation and accountability is needed?

To align productivity and regulatory objectives, it is vital that the regulatory approach to AI is **proportionate and effective** – creating the right incentives for those who can control risks.

Achieving this requires the implementation of appropriate regulatory tools to manage the potential harms associated with the use of AI. This includes a number of steps:

- identify how the technology is already being used, or likely to be used in the immediate future (based on, for instance, stated intended uses or overseas experience)
- determine whether this use results in heightened risks of serious harm compared to the counterfactual
- identify which parties involved have the scope to influence risks and outcomes
- determine whether the risk is adequately addressed by existing regulation, or whether extensions or modifications to this regulation, or improvements to its enforcement, are required. (And if a new regulatory instrument is needed, consider a technology-neutral approach in the first instance.)

In this section we consider the first three of these steps. Issues of new regulatory instruments are discussed in section 2. Throughout the paper, 'regulatory instruments' and 'regulation' are used in their broadest sense to include any rule or directive made or enabled and enforced to achieve a social objective.

AI use and outcomes

AI can be applied in many different ways and many applications are likely to raise little if any risk for society. It is only possible to determine whether regulation is needed, and what type of regulation would be effective, by considering the outcomes of particular uses of AI, and how those outcomes are delivered.

This focus on 'outcomes' and 'use' is consistent with regulatory frameworks that apply to other technologies. However, it implies a cautionary approach to proactive regulation based on speculative future uses and harms.

For example, general purpose AI models**,** such as foundational models, by definition have a wide range of potential applications. Attempting to create ex ante regulations against potential harms is likely to be either ineffective, as many uses and harms are currently unknown; or have costs that outweigh the benefits, for example by locking in regulations that stymy innovation and investment or create vested interests who exploit the regulations and undermine competition.

Similarly, prohibitions or bans on general purpose technologies will be generally counterproductive – they may protect against harms but only by also eliminating the benefits.

Focusing on use and outcomes indicates that effective regulation of AI requires patience, to wait to see what uses actually develop and any associated potential harm.

Risk-based regulation of AI

The existence of a potential harm from the use of AI does not imply that there should be regulation to address that harm. Rather, the need for regulation depends on risk of the harm, relative to a relevant 'real world' counterfactual.

The concept of risk captures both the probability of harm and the scale of the consequences. For example, an autocomplete text function would be low risk given the consequences of error are minimal, even if the probability of error is relatively high. An algorithm governing self-driving public transport would be higher risk,

given the possible severe consequences (passenger or pedestrian injury or death) even if the probability of error is (hypothetically) modest.

Risk-based approaches to regulation are required because regulation itself is costly. It involves direct costs such as compliance and enforcement. But regulation also has indirect costs by changing behaviour. For developing technologies such as AI, regulation inevitably influences the path of development, in the extreme determining whether specific paths of research are economically viable.

A risk-based approach to AI regulation will weigh the expected harm from the use of the relevant AI with the expected cost of regulating to reduce that expected harm. Because a risk-based approach to regulation focusses on the expected net benefit of regulation, often it will not eliminate a harm. Rather, the aim is to reduce the size and likelihood of harm to acceptable levels without imposing an excessive regulatory burden on society.

Risk-based approaches to regulation have been proposed in Europe, where (it appears likely that) a tiered system of regulatory obligations will be placed on technology proportionate to its assessed risk. The risk assessment will reflect whether an AI system is likely to pose high risks to fundamental rights and safety (Meltzer and Tielemans 2022). Underlying such systems are a range of questions that governments must engage with in order to determine the need for protections – be they through formal regulation or other mechanisms.

It is misleading to measure the risk from a use of AI relative to a fictitious 'perfect world'. Rather, the appropriate benchmark for risk-based regulation is the expected harm from the use of the AI technology relative to the real-world counterfactual level of expected harm that would arise if the technology in question was not used. For example, the risk of a self-driving vehicle algorithm should be evaluated against a counterfactual of a safe, licensed, human driver, rather than a fictitious world of zero road fatalities. The risk of an AI-driven diagnostic tool in health needs to be judged against the alternative of not having such a tool to assist a general practitioner rather than a false world of perfect diagnosis.

Measuring risk relative to a real-world counterfactual avoids harmful regulation that stops technology from improving outcomes. If the counterfactual without the technology entails significant risk, then an AI application can lower risk compared to the counterfactual, even if it does not eliminate risk compared to a fictious 'perfect world'. For example, there are persistent skill gaps in parts of Australia's medical sector, particularly in rural or remote areas. The first-best option may be to fill those gaps with qualified workers over time. However, in the absence of an instant professional workforce, the best alternative could be to employ technologies that can supplement existing expertise.

The assessment of the risk of AI needs to factor in non-regulatory counters to the risk. Some harms can be reduced or eliminated at low cost by the user of the application (e.g. reading predictive text suggestions before sending an e-mail) and regulation is not needed to 'protect' the user. For business applications, competition between providers and business reputation may mitigate risk adequately. Some applications will relate to risks of harms that are reversible and compensable, in which case existing laws applying to negligence or consumer safety may be adequate.

Identifying who can influence risk and outcomes

Where there are risks from a use of AI that warrant some form of regulation, it is necessary to identify which parties (or stage of the supply chain) the regulation should focus on.

The nature of AI supply chains helps to determine the amount of control or influence different parties might have over outcomes. Brown (2023) distinguishes between three categories of AI supply chains including:

systems that are built in-house, systems relying on an application programming interface (API) and systems built or fine-tuned for a customer.

Scenarios where the entire AI system is built in-house by a single company are likely to be relatively specialised and less common. More often, there will be multiple firms involved – in developing a foundation model, customising the model to specific uses and implementing the model in a commercial setting. It is inevitable that more than one type of supply chain will emerge in Australia, depending on the needs of particular firms and characteristics of the relevant markets.

This suggests that overall, regulation will need to fit the dynamics of different market structures. In some cases, the developer has a high degree of control over the system and an ability to monitor and mitigate risks. In other cases, that control is devolved to other firms down the supply chain that further customise the model, or to end users.

It is likely that AI regulation will focus on multiple participants in the supply chain for different risks. AI regulation may also involve a mix of ex ante and ex post regulation. We already see this approach being applied effectively for other technologies. For example, motor vehicle use has significant risks to personal and public safety. These risks are handled at both a manufacturer and a user level. Separate sets of regulations apply to the vehicle (manufacturing and import standards) and to users (licensing and road rules). Both manufacturers and users face ex ante regulations (standards for sale of a vehicle, requirement of drivers to hold a license) and ex post regulations (product recall and liability rules, criminal laws around certain acts by users). These rules create incentives for the party best placed to mitigate particular risks of harm. Efficient AI regulations will follow a similar approach.

In some situations, regulatory decisions that impact Australia will be made outside Australia. Technology supply chains are highly connected globally. As a small economy in the global AI landscape, Australia relies on global suppliers at various points in the AI supply chains, particularly for larger foundation models. Where Australia does develop technology locally, it is likely that some of these models will target international markets. Alternatively, AI systems used in Australia may be hosted and accessed outside of Australia.

Overall, the global connectedness of AI supply chains means that Australia's approach to regulating AI development and use will impact, and be impacted by, decisions and approaches in other jurisdictions.

2. Is new regulation needed?

Where regulatory intervention is justified, and there is a clear role for Australian regulation, it is important for governments to consider how existing laws apply before designing new regulations.

This section focuses on whether the use of AI technologies requires changes to Australian laws and regulations. It does not rule out the possibility that existing laws might be in need of reform regardless of AI use (i.e. due to longstanding flaws) but investigates the additional, potentially idiosyncratic, issues raised by AI.

Existing technology-neutral regulatory frameworks

In general, Australian law applies to the use of AI as it would to other technologies. Often, AI just provides a more efficient and effective way to accomplish things already being done (or things which could be done, but are already outlawed), and in these instances, introducing new regulations and laws to govern AI use would be both unnecessary and confusing.

In some instances, existing, technology-neutral regulations and laws may need to be strengthened to deal with the changing profile of risks created by AI or clarified, augmented or extended to cover new risks or harms. In these instances, existing regulations may prove insufficient to govern permissible uses of AI.

Once it has been determined that there is a risk from an outcome of the use of AI that should be regulated, the Commission suggests a three-step approach to the design of that regulation.

- Consider if existing regulatory frameworks (including regulations and regulators) adequately address the identified risks, and whether they do so without unduly constraining AI use or presenting inconsistency with equivalent international approaches. If so, there is no need for new regulation. If not:
- Consider if existing regulation can be clarified or amended to bridge any gaps (in regulation or its enforcement) associated with AI development or deployment. If so, clarify or amend existing regulations, and provide appropriate resourcing and training to regulators rather than introducing new regulations. If not:
- Consider the net benefits of new regulation using a risk-based approach. The assessment would need to take into account the relevant outcome(s) and risk(s) to be covered compared to a real-world counterfactual, any non-regulatory counters to the risk, the relevant point(s) in the supply chain where the regulation will apply, and any relevant existing international regulations that may impact the risk or limit regulatory solutions. New regulation should only be introduced if there is a net benefit from the regulation taking these factors into account.

Figure 1 provides a table of the types of issues that decision-makers need to address to apply this approach.

Do existing regulations cover the undesirable conduct or outcomes?

While the use of AI may create risks, it will often be the case these risks are covered by existing legislation and regulations (this includes both general regulations and industry-specific regulations). This is particularly the case where existing regulations have been designed to be technology-neutral.[3]

It would be a mistake for governments to presume that the risks associated with the use of AI would necessarily require new regulation, rather than the application and enforcement (potentially with amendment) of existing regulations (Solomon and Davis 2023). For instance, **existing laws** relating to privacy, consumer protection, and discrimination are likely to be among the most relevant to the use of AI. Existing laws pertaining to negligence already apply, and can establish some degree of accountability for developers, users, and other parties with regard to harms incurred as a result of technology use. As liability is determined on a case-by-case basis, it may take some time before courts can collectively clarify how specific uses of AI would be treated.

Existing sector-specific regulatory frameworks may obviate the need for technology-specific regulation. For example, the EU AI Act highlights the potential use of AI to create inaudible frequencies that can encourage truck drivers to continue driving for excessive durations (Sioli 2021). However, Australia's National Heavy Vehicle Regulator (2020) stipulates the maximum permissible driving hours for truck drivers, therefore making this particular AI use illegal.

Relying on existing regulation will provide a degree of certainty for businesses, as they can work within existing frameworks and rules rather than adapt to new rules. Importantly, it also ensures that different technologies are held to the same standard regarding their effect on users, consumers, and the public – and that regulation focuses on decisions and actions that cause harm, rather than on technology itself.

[3] Technology-neutral regulation or legislation refers to rules or laws that focus on outcomes, rather than the technology used to achieve an outcome. For example, laws in relation to price fixing are technology neutral – businesses cannot make a contract, arrangement or reach an understanding to fix prices. Of course, AI may raise new issues and there is an international debate on whether price fixing through a 'third party', such as an AI algorithm, is captured by current legal definitions.

In some cases, AI could raise the risk of harms that are already illegal – for example if the technology creates opportunities to cause harm more quickly, or at scale, in ways that are difficult to stop. Gaps could occur in **enforceable regulation**, and while this would warrant improvements in regulators' capabilities, it could also warrant amendments to, or strengthening of, existing regulations, or creation of new regulations. In other words, there may be a question as to whether existing regulations *adequately* address undesirable outcomes (as noted above, this involves judgments about appropriate thresholds and risks in the counterfactual).

Figure 1 – Regulating AI use

If not, can existing regulations be adapted to achieve the desired outcomes?

A range of tools should be tested before deciding that existing laws and regulations do not cover a specific use of AI. For instance, regulators can enhance regulatory clarity by providing **guidelines and advice**.

Some aspects of regulation may only become clear over time as laws are tested in the courts. As an example, key concepts related to employment take a common-law meaning, developed by the courts over time. Indeed, governments may face the choice of what to define in legislation, and how to balance the benefits of legislative 'certainty' with the costs of prescriptiveness. It can be argued that relying on legal precedent to extend existing regulations to new, AI-based uses is slow and uncertain. This is true. But any new legislation will equally be subject to legal interpretations and precedent over time. The uncertainty created by new regulations, regardless of how 'tightly' they are drafted, will often be greater than the uncertainty around existing rules that have already been tested in court.

To help advance understanding of the applications of existing rules to AI, there may be value in regulators undertaking **test cases**, recognising that such cases may involve significant costs to the relevant parties.

If it is found that existing regulations are inadequate for a specific use of AI, governments should consider whether existing regulations can be tightened to address gaps raised by AI (rather than drafting new rules from scratch). For example, laws defining what is considered as personal information might need to be updated as studies have found that AI is now capable of re-identifying users from data that has gone through robust de-identification process (Na et al. 2018). The laws governing the use of personal information are still sound; they just need to be updated to reflect how de-identified and re-identified data is used. This could, for example, involve redefining personal information to include some forms of re-identified data, or ensuring it is clear that re-identified data is not personal information.

What new regulation could fit the circumstances?

As a final step, if potential harms fall outside the remit of existing regulations, new regulatory safeguards may be needed.

New regulation should, where possible, be technology-neutral. **Technology-neutral regulation** allows regulators to focus on the outcomes and behaviours which are undesirable, rather than the technologies. That is, by focusing on the elements that can give rise to risks (such as use of personal data, opacity of approach or decision making, automation of high-risk decisions without human oversight) as opposed to specific models in use (e.g. large language models), technology-neutral regulation will remain applicable to technological advances as they arise, and therefore be more effective at achieving its objectives.

Technology-neutral regulation reduces scope for developers to adjust their technology in an attempt to circumvent a definition set into a law or regulation. It also avoids favour or discrimination against certain technologies, facilitating competition between incumbent and new technologies.

If it is decided that technology-neutral regulations cannot adequately address a risk from the use of AI, then technology-specific regulations *may* be required. While this should be a last resort (as they will quickly become obsolete as the technology evolves), technology-specific regulation can target the gaps in regulation that may arise through technology-neutral regulation by governing specific, permissible uses of AI.

To avoid the risk of obsolete technology-specific regulations, any regulation that is technology-specific should have a process for review and either validation or removal, at a set interval of say five years.

New regulation should also be designed to leverage **existing regulatory frameworks**, including sector-specific frameworks and existing regulators. There are risks that if technology-specific regulations were implemented economy-wide, this could complicate how those technologies are regulated in particular

sectors for specific uses. In the health sector, for example, the use of software and medical equipment are regulated by the Therapeutic Goods Administration (TGA), and there is no apparent reason why this would not continue to be the case if the latest software or medical equipment were to include advancements in AI.

3. Issues for AI regulation

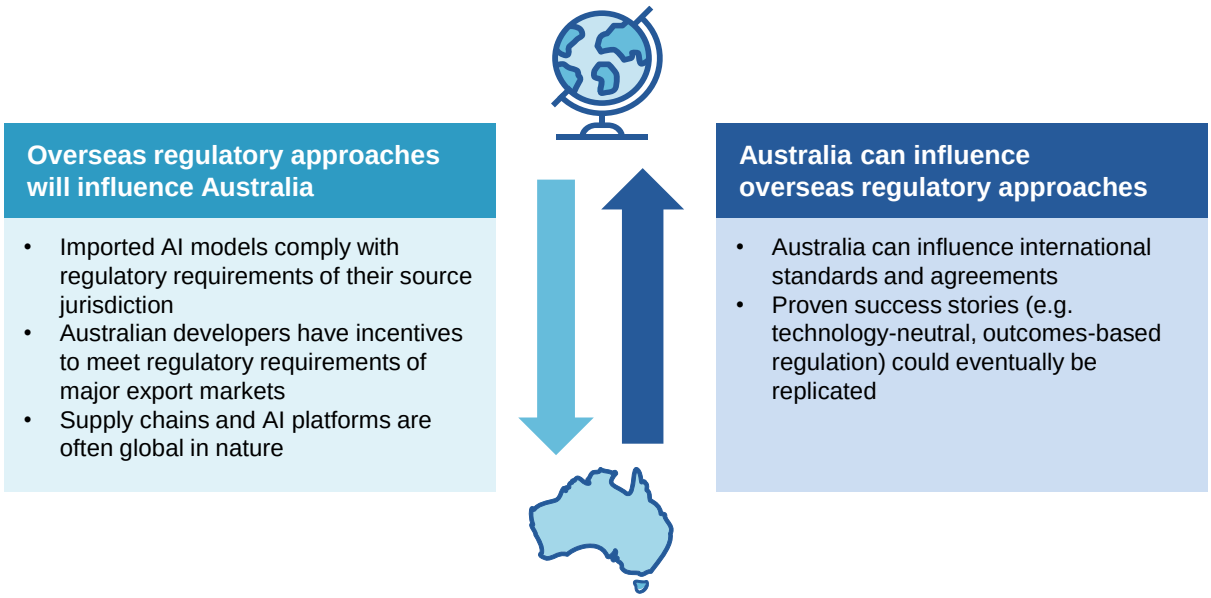
Authorities need to consider a range of specific issues when considering the effectiveness of regulation for the use of AI.

The international regulatory landscape

In amending existing regulation or designing new regulation, it is appropriate to first consider the **regulations applied overseas** to the development of AI technologies.

To some extent, Australia is likely to be a 'regulation taker' in international AI markets (figure 2). With Australia likely to import a significant amount of AI technology from overseas and domestic developers seeking sales in overseas markets, regulatory approaches in Australia's key overseas markets will, by default, act to regulate AI developments and outcomes in Australia.

Figure 2 – Australia's regulatory approach will be influenced by the international landscape



Notwithstanding international agreement on high-level regulatory principles (i.e. via the Bletchley Declaration) there is currently very limited consensus on how to regulate AI among the world's major economies (box 1). Some, like the European Union, are planning to implement AI specific laws, while others, like the United States, have made guidelines for AI development and use, but have not yet introduced legislation. Approaches to regulation in major global economies will have direct effects on AI supply chains in Australia.

Box 1 – Regulatory approaches developing internationally

A range of regulatory approaches have emerged in major economies, including the United States, European Union, United Kingdom and China.

In **the European Union**, the proposed AI Act categorises AI by risks, banning systems posing an unacceptable level of risk to people's safety (such as biometric surveillance or emotional recognition systems) and subjecting other AI systems to scaled regulatory obligations, according to their assigned risk-rating. The Act also establishes complex enforcement mechanisms including infringements. Under the transparency requirements, the Act necessitates AI developers to disclose that the content was generated by AI and publish summaries of copyrighted data used for training (European Commission 2021). The proposed regulations have been criticised by some AI developers as being too burdensome, creating disproportionate liability risks and administrative responsibilities to an extent that it might stifle innovation (Mukherjee 2023).

Unlike the European Union, **the United Kingdom** has adopted a decentralised and iterative approach. The framework is intended to allow sector-specific regulators to quickly resolve sudden and unpredictable disruptions arising from AI using targeted measures (UK Department for Science, Innovation and Technology 2023a). The UK Government is considering the establishment of a central body that sector-specific regulators will be required to consult with before implementing any AI specific guidelines (UK Department for Science, Innovation and Technology 2023b).

In **the United States**, an Executive Order (EO) was passed on 30 October 2023 on the *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (U.S. President 2023). Relative to the EU approach, the EO focuses on guidelines, standards, and outlines an approach to international cooperation. And while it addresses a wide range of risks (and opportunities) the EO does not categorise applications by risk as does the EU approach. The EO builds on the previously published *Blueprint for AI Bill of Rights,* which is a non-binding list of 5 principles, intended to minimise potential harm from AI systems without stymieing innovation (The White House 2022).

The approach to AI regulation taken in **China** is characterised by state control. The recently released *Interim Measures for the Management of Generative AI Services* takes a technology based regulatory approach, requiring each individual technology to undergo security tests before release. Moreover, separate security review requirements are put in place for different recommendation algorithms (Cyberspace Administration of China 2023). These regulations specifically apply to AI technologies that can impact public opinion or have the capacity to for social mobilisation (PwC China 2023, p. 3). However, it should be noted that China's AI regulatory framework only regulates technologies whose services will be accessible to the public within China. That is, the framework excludes generative AI services developed and used by enterprises, research and academic institutions, and other public entities. Technologies targeting users outside China are also exempted.

International divergence in regulatory approaches will likely lead to varied paths in the development and uptake of AI across countries. This may create an environment where jurisdictions with less stringent regulations have competitive advantages in their ability to produce new models at faster speeds. This advantage may be reduced if demand-side pressures force companies to prioritise their AI models' reputations, so that jurisdictions with more stringent regulations provide a 'guarantee' against the risk of producing low-quality AI.

However, the international regulatory landscape is still taking shape. Even some of the most prominent examples of international regulation – such as the EU AI Act – have not yet been adopted and implemented. The European Parliament confirmed their negotiating position in June 2023 (European Parliament 2023), although it is still uncertain when the AI Act will enter into force. The regulation will only apply from 24 months following the entering into force of the regulation (European Commission 2021). It therefore remains to be seen what the EU's regulation will look like in practice.

Accounting for the international regulatory landscape

To help protect Australian interests, the Australian Government should be an active participant in international fora that consider AI regulations and standards. Australia is one of 29 signatories to the Bletchley Declaration, which encompasses high-level guidance for the international community's approach to AI regulation. Signatories to the agreement affirmed that AI should be designed, developed, deployed, and used, in a manner that is safe, in such a way as to be human-centric, trustworthy and responsible. (Agreement on more detailed standards and guidelines will likely be challenging – in part because the technology is continuing to develop and potential harms have not yet arisen at scale.)

There will be benefits to explicitly aligning Australia's regulations with other major economies. Common standards could lower costs of retesting for Australian businesses that are looking to adopt AI products developed overseas or looking to export AI products developed locally. It would also create consistency for Australian entities using AI products. This does not require new regulation but involves a recognition of overseas standards for the purpose of domestic regulation. For example, an AI tool that satisfies, say, EU standards could be deemed to meet relevant Australian standards.

Broadly, regulating the design of an AI model or application either inconsistently or more harshly than in overseas markets may simply mean developers do not sell to Australia, harming the domestic market. Further, regulating less harshly may not make a difference as developers would need to meet the specifications of larger markets (such as the EU or the US). In this sense, there are advantages to Australia having regulation that is in-step with overseas regulation.

The alternative – an application of idiosyncratic local regulations on the design and development of AI technologies – could lead developers to bypass Australia. As such, attempts to unilaterally set mandatory standards for AI technologies may be useful only where they can be easily met, or where a ban would be net-beneficial (for instance, where products are temporarily recalled or banned for safety reasons). As the Australian Government (2024) considers potential 'mandatory guardrails' for AI development and deployment, any conflict with overseas regulation will need to be carefully assessed.

Market incentives and self-regulation

In considering the nature and extent of regulatory gaps, it is important to consider what role market incentives and self-regulation might play.

Risk aversion and insurance

Businesses have an incentive to self-regulate in order to maintain their reputation. Businesses consulted as part of this study commented that the value of building and maintaining trust with their customer base has, to date, ruled out use of some potentially riskier uses of AI. Businesses commented that the deterrent of losing the confidence of their customers often exceeds the deterrent of any regulation or standards. In these instances, self-regulation will potentially reduce some risky uses of AI. However, as AI becomes more widespread and emerging businesses seek to find advantages over incumbents, businesses' risk-aversion and maintaining customer trust becomes less of a deterrent.

Insurance, in combination with effective accountability mechanisms, may also reduce the use of riskier AI tools.[4] AI developers may seek to insure themselves against loss attributable to their AI models (and in some cases, they may opt to self-insure).[5] In addition, firms that use AI technologies may also insure against any additional risks posed by AI. Private insurance companies will only be willing to insure AI models that can transparently show that they have met minimum standards of safety and security. In this sense, private insurers may establish transparency and accountability frameworks for parties seeking insurance. Parties refusing to meet these standards would find their products 'uninsurable'.

In many situations, relying solely on self-regulation and private insurance markets to deter unsafe or potentially harmful uses of AI will be insufficient to protect consumers from harms or to promote confidence in AI technology among potential users. This leaves a role to play for regulators to manage harms arising from AI. However, this role, and the design and application of any regulation, needs to reflect areas of market failure, and not where businesses and customers have implicitly identified a permissible level of risk.

Industry codes

Australia is among several countries that have implemented voluntary codes of ethics and practice for AI. Such codes have been common in other areas of technology, although in Australia, the regulation of some technologies involves *enforceable* industry codes that are backed with the threat of more direct regulation.[6]

The use of industry codes recognises the importance of industry knowledge in setting appropriate standards in high-tech sectors, and the role of ongoing relationships between firms and regulators for keeping enforcement up to date. They can also provide a more flexible, light-handed approach than black letter law.

Industry codes that are voluntary in nature are a weak form of regulation. Indeed, a voluntary industry code can be used by incumbents to try and mask the need for, or delay the introduction of, more direct regulation.

This means that voluntary industry codes or other forms of industry self-regulation at best are an adjunct to rather than a replacement for formal regulation or compulsory codes.[7]

Ongoing regulatory design and review

The ability to successfully design and enforce regulation which will underpin trustworthy and effective AI depends on the quality of the regulatory environment. As the development of AI and complementary products and their uptake progress, there are several ways governments can continuously improve the regulatory environment – including collaboration between regulators and industry, upskilling regulators and testing approaches to regulation.

[4] Insurance could give rise to a moral hazard problem, where developers, once insured, take more risks with their models. However, given insurers have no incentive to insure, or continue to insure, any behaviour they deem to be a too high risk, it is more likely only safe models will be insured and unsafe models and behaviour uninsured.

[5] Note that existing cyber-risk security or product liability insurance may apply as appropriate, but these do not cover the potential legal responsibility arising from decisions made or actions taken that are informed by an algorithm.

[6] Examples include: that search engines submit an industry code to the eSafety Commissioner for registration in order to avoid having standards set by the regulator; or the News Media Bargaining Code for digital platforms, which facilitates bargaining between market participants in order to avoid intervention by the ACCC.

[7] This is recognised by industry. For example, Google (nd) notes that 'while self-regulation is vital, it is not enough'. The Royal Australian College of General Practitioners argued that reliance on a voluntary code of ethics is 'not enough where there are potential gaps in existing legislation governing high-risk AI use' (RACGP 2023, p. 1).

Industry collaboration and co-design

Given the rapidly evolving nature of AI, a relatively **close collaboration** between industry stakeholders, governments and regulators will be required. In part, this is due to the need to build technical expertise within government and regulators. An example of this is the Australian Government's (2024) intention for the National AI Centre to collaborate with industry to produce a 'best-practice and up-to-date voluntary AI risk-based safety framework for responsible adoption of AI in Australian businesses' (p. 21).

More broadly, it is likely that regulatory co-design will continue to play a role in ensuring regulations are fit for purpose over time. In part, this reflects that larger firms will play a role in setting rules and standards within their own environments. It will be vital for regulators and policymakers to understand to what extent markets are providing useful mechanisms to build trustworthiness, and what limits (or gaps) apply.

Regulatory guidelines will help in this regard. Preparing guidelines (including public drafts) facilitates conversations between industry and regulators. Guidelines allow policy makers to effectively communicate not only the policy decisions, but also the underlying rationale to stakeholders, along with clear and unambiguous compliance guidance. And guidelines can be regularly updated, aiding ongoing regulator-industry dialogue and alerting all parties to emerging risks.

Guidelines help governments and regulators provide greater **regulatory clarity** for firms. Regulatory uncertainty has been raised in discussions with the Commission as a possible barrier to adoption, so regulators can help businesses by clarifying the practical application of a regulatory regime. Any guidelines should aim to provide AI developers, deployers and users with clear requirements and obligations regarding the development and use of AI.

Coordination and consistency between Australian governments

There is a strong case for **consistent AI regulations across Australia**.

Inconsistent or overlapping regulation surrounding the use of AI will increase compliance costs for companies that operate across multiple jurisdictions, as firms are required to adapt to different regulatory requirements in each jurisdiction. Similarly, inconsistent data-sharing approaches (or regulation) across state borders may create an impediment to sharing data.

Inconsistencies in how AI is used by different levels of government could create confusion for individuals who have a reasonable expectation that common principles and objectives would apply in their interactions with government administration and services, regardless of whether they are from Australian, state or territory government agencies.

A clear difficulty in achieving consistency is that regulation is still evolving. State and territory governments may differ in opinion, or simply vary in terms of progress in their reform agendas. However, history in Australia, for example in labour markets and transport, show that having different regulations in different jurisdictions can significantly impact competition and harm workers, business and consumers.

Given the costs of inconsistency, governments should consider forms of coordinated regulatory experimentation to innovate, weed out undesirable features and identify effective approaches to regulation.

Coordinated experimentation through regulatory sandboxes

Regulatory sandboxes are used (most commonly in the financial sector) to give a temporary and conditional reprieve from existing regulations that allows business to develop innovative technologies and practices by live testing cutting edge technologies in the market with real consumers. The sandboxes allow regulators to understand and assess unforeseen risks and harms associated with the new technology, and design regulations which address them.

Regulatory sandboxes could allow regulators to understand more about how AI learns and makes decisions. By observing the practical application of the technology, regulators could identify the underlying causes of the generated outputs. They could also be beneficial in providing an environment for innovators to explore and experiment, and enhance incentives for AI uptake and development of complementary innovations.

It should be noted that there would exist significant challenges to the effective implementation of AI regulatory sandboxes. One challenge would be to ensure sufficient technical expertise among the regulators responsible. Another would be to design sandboxes that allow sufficiently broad participation so as not to distort competition.[8]

Many jurisdictions overseas are creating regulatory sandboxes related to AI, including in Spain and France as part of EU AI Act[9] (CNIL 2023; European Commission 2022) and in the United Kingdom[10] (UK Department for Science, Innovation and Technology 2023a). While these are at early stages, they may prove useful examples of sandbox design.

Improving regulators' capabilities

Highly skilled regulators will improve the standards of regulation. Globally, concerns have been raised that regulators generally lack the technical skills to effectively regulate AI (UNESCO 2022). A lack of technical expertise can mean that regulators may be unable to assess risk regarding novel AI models and applications, or to identify appropriate courses of action. It can also lead to either higher regulatory burden or regulatory capture as regulators rely on firms to provide more information. Strategies to attract, develop and retain AI specialists may enable AI policy makers to expand their knowledge and understanding of AI and assist with structuring effective AI regulations and taking a risk-based (rather than one-size-fits-all) approach to regulatory implementation.

Monitoring and stress-testing

In sectors where AI systems play a role in large, networked systems, there may be value in regulators stress-testing systems on an ongoing basis.

Stress-testing of systems to key risks is an approach taken by regulators in the financial sector. The Australian Securities and Investments Commission (ASIC), for example, mandates that fund operators perform stress

[8] The Australian Government has, in the past, created sandboxes for multiple businesses to participate (ASIC 2020a).
[9] In the EU, small and medium sized enterprises will be given priority access to the sandbox to support the reduction of barriers that these companies might face when launching their AI systems under the new regulation. Regulators will document participants' obligations (such as the liability structure or testing standards) and outline clear methods for monitoring the technology and how they will follow up with developers. However, those that participate in the sandboxes are still liable for any harm inflicted on third parties that results from their activities – a feature which has been criticised as it will reduce the incentive to participate.
[10] In the United Kingdom, the sandbox pilot is sector specific, primarily focusing on sectors where there is a high degree of AI investment and demand for sandboxes. The UK sandbox will also prioritise small-to-medium enterprises and allow innovators to trial new products under relaxed regulatory environment for a limited period of time.

tests or scenario analysis[11] to assess the liquidity strength to withstand 'potential disruptions'.[12] Stress testing of the banking system occurs through the Australian Prudential Regulation Authority (APRA) led industry stress tests, banks own stress tests and through APRA's own internal testing models.

Stress testing involves subjecting a system to hypothetical yet well targeted, plausible and sufficiently adverse scenarios – such as adverse macroeconomic and financial conditions to test their resilience (hypothetical scenarios are based on experience and historical events). APRA notes that the primary objective of its testing is to 'provide assurance of the banking system to a severe shock' (APRA 2020, p. 4). ASIC uses test results to inform policy decisions – such as setting capital requirements on fund operators.

Stress testing is most beneficial where 'tail' outcomes create concern even when 'average' outcomes are adequate. This may apply, for example, to some general AI technologies where there are small probabilities of producing socially-adverse outcomes. A stress test framework could subject relevant AI models to 'extreme' scenarios that reveal these adverse outcomes. Research has demonstrated that the process of stress testing AI models can be an useful tool to identify hidden limitations – such as reduction in accuracy when deployed on independent datasets that differ from the data used for training (Young et al. 2021). Identifying and remedying these gaps can be vital for preventing the potential for adverse outcomes.

References

APRA 2020, Stress testing banks during COVID-19: Stress testing banks during COVID-19, Information Paper.

ASIC (Australian Security & Investments Commission) 2020a, Enhanced regulatory sandbox, https://asic.gov.au/for-business/innovation-hub/enhanced-regulatory-sandbox/ (accessed 20 October 2023).

—— 2020b, Operational resilience of market intermediaries during the COVID-19 pandemic, https://asic.gov.au/regulatory-resources/markets/market-supervision/operational-resilience-of-market-intermediaries-during-the-covid-19-pandemic/ (accessed 4 December 2023).

Australian Government 2024, Safe and responsible AI in Australia consultation: Australian Government's interim response, Canberra.

Brown, I. 2023, Expert explainer: Allocating accountability in AI supply chains, Ada Lovelace Institute.

CNIL 2023, "Sandbox": CNIL launches call for projects on artificial intelligence in public services, https://www.cnil.fr/en/sandbox-cnil-launches-call-projects-artificial-intelligence-public-services (accessed 12 October 2023).

Cyberspace Administration of China 2023, Interim Measures for the Management of Generative Artificial Intelligence, http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (accessed 19 September 2023).

European Commission 2021, The AI Act, 10 February, https://artificialintelligenceact.eu/the-act/ (accessed 19 September 2023).

—— 2022, First regulatory sandbox on Artificial Intelligence presented, https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented (accessed 9 November 2023).

European Parliament 2023, Parliament's negotiating position on the artificial intelligence act, June.

Google nd, Recommendations for Regulating AI, p. 1, https://ai.google/static/documents/recommendations-for-regulating-ai.pdf (accessed 18 December 2023).

Lessig, L. 1998, 'The New Chicago School', Journal of Legal Studies, vol. 27, no. S2, pp. 661–691.

—— 2006, Code 2.0, Basic Books.

[11] ASIC mandates entities to perform stress testing of liquidity risks as often as possible depending on the nature, scale and complexity of the business and update arrangements as necessary in response to stress test results. Operators not performing stress tests must document their reasons and review this decision regularly. The structure and process of conducting stress tests is also expected to be reviewed at regular intervals (at a minimum, annually) to ensure the nature, currency and severity of the tested scenarios are relevant and appropriate. There is no common methodology for stress testing and it is tailored on a case-by-case basis according to the specific needs of each operator. Stress testing material risks other than liquidity is not mandatory for operators but is proposed as a good practice.

[12] As an example, through the process of stress testing financial institutions during the COVID-19 pandemic, regulatory bodies were able to troubleshoot the areas where institutions exhibited a lower risk tolerance. The key risk areas identified included off-premises trading, fraud detection, structured financial products, credit risk management, and operational resilience (ASIC 2020b).

Meltzer, J. and Tielemans, A. 2022, The European Union AI Act: Next steps and issues for building international cooperation, May, The Brookings Institute.

Mukherjee, S. 2023, 'Draft EU artificial Intelligence rules could hurt Europe, executives say', Reuters, July.

Na, L., Yang, C., Lo, C.-C., Zhao, F., Fukuoka, Y. and Aswani, A. 2018, 'Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning', JAMA Network Open, vol. 1, no. 8, p. e186040.

National Heavy Vehicle Regulator 2020, Standard hours | NHVR, https://www.nhvr.gov.au/safety-accreditation-compliance/fatigue-management/work-and-rest-requirements/standard-hours (accessed 10 November 2023).

PwC China (Pricewaterhouse Coopers China) 2023, Regulatory and legislation: China's Intermi Measures for the Management of Generative Artificial Intelligence Services officially implemented, August.

RACGP (Royal Australian College of General Practitioners) 2023, Submission to Supporting Responsible AI consultation, Submission to the Department of Industry, Science and Resources.

Sioli, L. 2021, 'A European Strategy for Artificial Intelligence'.

Solomon, L. and Davis, P.N. 2023, The State of AI Governance in Australia, May, Human Technology Institute, The University of Technology Sydney, https://www.uts.edu.au/human-technology-institute/news/report-launch-state-ai-governance-australia (accessed 24 July 2023).

The White House 2022, Blueprint for an AI Bill of Rights | OSTP, The White House, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed 19 September 2023).

UK Department for Science, Innovation and Technology 2023a, A pro-innovation approach to AI regulation.

—— 2023b, UK Artificial Intelligence Regulation Impact Assessment.

UNESCO 2022, 'Artificial intelligence and digital transformation: competencies for civil servants - UNESCO Digital Library', https://unesdoc.unesco.org/ark:/48223/pf0000383325 (accessed 20 October 2023).

U.S. President 2023, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Federal Register, vol. 88.

Young, A.T., Fernandez, K., Pfau, J., Reddy, R., Cao, N.A., von Franque, M.Y., Johal, A., Wu, B.V., Wu, R.R., Chen, J.Y., Fadadu, R.P., Vasquez, J.A., Tam, A., Keiser, M.J. and Wei, M.L. 2021, 'Stress testing reveals gaps in clinic readiness of image-based diagnostic artificial intelligence models', Nature Publishing Group, npj Digital Medicine, vol. 4, no. 1, pp. 1–8.