webb henderson
LEGAL AND REGULATORY ADVISORS

AI regulation in Australia: new road rules emerge

As organisations of all shapes and sizes continue to test and invest in the latest AI tools, the Australian Government's view on AI regulation is evolving. Previously, the government had emphasised industry- and context-specific updates to existing regulation. But this view seems to have evolved. Seemingly taking its cue from developments in the UK and Canada, the Australian Government now appears to be predisposed to a more interventionist stance. The Government's current views are outlined in its *Proposals paper for introducing mandatory guardrails for AI in high-risk settings* (the Proposals Paper).

The Australian Government states that its intent is to:

- regulate use cases that are considered 'high risk'; and
- prevent harm caused by AI through **testing**, **transparency** and **accountability** in AI development and deployment.

In parallel, the Government has followed through on its commitment to develop a voluntary AI Safety Standard that is intended to provide practical, voluntary "best-practice" guidelines. This was published in August 2024 to give organisations that are deploying AI in high-risk use cases now, a way to start adopting practices that are likely to assist in compliance with future regulations when they eventually come into effect.

The Government's desired regulatory settings

The Proposals Paper outlines a "risk-based", "preventative" approach to regulating the use of AI. A central animating concern is that AI is different to other emerging technologies in its transformative effect and rapid growth: meaning pre-emptive regulation is required to stop the genie getting out of the bottle, even if the regulation has to be written before anyone fully knows the genie's powers or tendencies to cause mischief; or how citizens, consumers and market participants might harness the promises of AI or combat its harms.

How does one regulate for such an unknown risk and preserve as-yet-unknown benefits? The Government's current answer is two-fold:

- there should be a risk filter that identifies "high-risk" AI use cases; and
- these use cases must operate within a set of AI-specific regulatory settings (the Proposals Paper describes these settings as "mandatory guardrails", which it proposes will ultimately be implemented through binding statutes and regulatory instruments).

What would the risk filter look like?

The Government proposes the filter should address two categories of high-risk AI:

Knowable

Use cases that are – by their intended or foreseeable nature – high risk (such as use in critical infrastructure, biometrics or employment)

Unknowable

General purpose AI models for which potential use cases are difficult to predict but could underpin other use cases which are high risk

To identify what falls into the first category of high-risk use cases for AI, the Proposals Paper suggests applying a set of principles — for example: does a use of AI create or increase the risk of (a) adverse impacts to an individual's rights; (b) adverse impacts to an individual's physical or mental health; or (c) adverse legal impacts. This approach deliberately resembles the EU's AI Act in setting out guiding principles for evaluating new, potentially high-risk use cases.

The Proposals Paper takes a less nuanced approach towards general purpose AI models. The Government proposes to subject these AI models to regulation regardless of whether they pose an identified risk. The key concern from the Government is around the unforeseeable risks posed by these models.

"Mandatory guardrails": ten regulatory settings for high-risk AI systems

At the core of the Proposals Paper are ten regulatory requirements that the Government proposes to apply to organisations that develop or deploy AI in ways that are considered high risk.

These requirements can be grouped into three key themes, focusing on how AI systems are developed and deployed:

Testing

Testing to ensure that AI systems perform as intended

Transparency

Transparency of AI system capabilities to end users, others in the supply chain and regulatory authorities

Accountability

Accountability in governing and managing AI system risks

At a high level, these themes reflect widespread international practice, including how high-risk AI is regulated by the EU AI Act and proposed to be regulated by Canada's *Artificial Intelligence Data Act*. They are consistent with themes in the multilateral Bletchley Declaration.

Many of the Proposals Paper's regulatory requirements – 7 out of 10 of them – are expressly linked to the EU's AI Act. This is in keeping with the Australian Government's oft-expressed desire to align any Australian obligations with comparable frameworks developed overseas and thereby promote interoperability and reduce compliance burdens. However, this has its own risks. The EU's approach itself can be criticised as onerous and overbearing. It may stifle homegrown innovation if AI users are reluctant or overly cautious in using AI, or even lead to AI developers exiting, or being slow to enter or expand into, the market in favour of more lightly regulated jurisdictions.

In light of these risks, it is encouraging that the Proposal Paper doesn't seek to import all of the EU's AI regulatory apparatus. Thus far, the Australian Government is not focused on an outright ban on 'unacceptable-risk' AI practices (though this has not been ruled out) or creating a category of 'limited-risk' AI practices that are still subject to regulation (albeit lighter touch).

On the other end of the spectrum is the New Zealand Government's approach, outlined in a Cabinet Paper earlier this year in July. The New Zealand Government is taking the view that most laws are drafted in a technology-neutral manner and so can deal with AI without the need for legislative reform, let alone a standalone statute.

Despite this difference in approach, the New Zealand Cabinet Paper says that it proposes a "risk-based" and "proportionate" regulatory response — both descriptors are used in the far more interventionist Australian proposal. While both governments are being guided by the same principle, they are headed in vastly different directions because their assessment of risk is focused on different parts of the economic value chain:

- The New Zealand government looks at the end uses of AI systems and concludes that existing laws address privacy, human rights and other important legal protections.
- The Australian Government looks further up the supply chain. The Proposals Paper:
    - looks at the development of AI models (say in a GPAI model created by a global technology company), how they are harnessed in specific AI systems (say for facial recognition by a specialised software company), and how they are ultimately deployed by specific government agencies (say by a state law enforcement agency);
    - is concerned that by the time the AI system is deployed and has an adverse impact on an individual, its core model may have been developed with biases, adapted to uses unforeseen by the model builder and poorly deployed;
    - concludes that it may be impossible to identify something has gone wrong, let alone where and how; and
    - expresses a concern that, even if all of the practical evidentiary hurdles and standards of proof can be met, it is likely that no individual actor in the supply chain will be fully responsible given the harm arises from misalignment of different actors who have no general obligation to co-ordinate.

To address these supply chain concerns, the Australian Government proposes to impose obligations on actors through the AI supply chain who might otherwise unknowingly contribute to a harm that only arises when all of the pieces of supply chain are brought together by one actor at the endpoint of the supply chain. The obligation is not for all actors to foresee all potential harms, which would be impossible, but to guide the ways in which each actor behaves to minimise the likely contributions to risk, and enhance the ability to look back through the supply chain once all the pieces have been brought together, to identify latent risks or, in hindsight to identify the elements that have contributed to an actual harm.

How will the Australian approach align actors throughout the supply chain?

So how does the Australian approach align the supply chain? The Australian Government focuses its regulatory intervention on "developers" and "deployers", which is again similar to the EU approach.

According to the Australian Government:

Developers

Design, Build, Train, Adapt, or Combine AI models and applications

Deployers

Supply or use an AI application to provide a product or service

As noted above, the Government's view is that current Australian laws tend to put obligations on the deployers, with less focus on developers. The proposed regulatory requirements aim to address this potential latest source of risk. This would include rules for testing AI models and improving transparency and governance measures for how these AI models are designed and trained.

No additional regulation would apply to "end users" – that is, those that consume an AI-based product or service, interact with it or are impacted by it after deployment.

Conceptually, the model makes sense. In practice, there will be challenges. Take the example of recruitment. The Positions Paper gives the example of an abandoned Amazon AI system from 2014 that sought to filter applications for software development roles. It was trained on existing industry data. Due to the industry being overwhelmingly male dominated, the AI system "learned" to reject female candidates. The example illustrates the potential benefits of the proposed guardrails approach: whether the whole tool was developed and deployed by one organisation or many, the performers of each development and deployment function would have obligations. For example:

- the developer of the core AI model would need to describe the data being used to potential downstream users of the AI model;
- the developer of the application screening AI system that harnesses the model would need to consider how the data sources and its limitations might affect application screening in particular, having regard to protected classes of people who might face discrimination in employment scenarios and describe the risk to potential downstream users of the AI system;
- the deployer of the application would need to disclose to job applicants that an AI system is being used for application processing and how they can seek review of automated decisions made by the system; and
- all the above actors would need to seek to minimise risks and provide feedback through mandated contact points as risks are identified or occur.

Will it work?

There are three key critiques of the Australian approach that are worth considering.

- **Explanatory myth:** The examples that guide the Proposal Paper may be examples of "survivorship bias". That is — the guardrails are designed to address patterns of risk that may be well understood or identified (at least in hindsight), without necessarily being responsive to harder to identify risks. The dearth of female software developers is a known issue and an issue that is important to the internal cultural settings and external brand of large global technology companies.

  Applying the proposed AI guardrails in a different setting may prove trickier. Some types of discrimination may be more subtle or less recognised. While responsible developers will over time standardise on new industry and application-specific equivalents of Microsoft Copilot's warning that "AI-generated content may be incorrect", small to medium businesses (the majority of employers in Australia) may find it difficult to act on these warnings effectively.

- **Tunnel vision:** The Proposals Paper cites speed of AI development and deployment as compelling arguments for proactively regulating AI. However, it's striking that many of the risks raised in the Proposals Paper are not fundamentally different to risks that exist independently of AI. For example, unconscious bias in human decisions makers is known to affect employment decisions and identification of suspects in criminal proceedings. As an aside, this tunnel vision seems to affect much of the current regulatory activity affecting the digital economy. Concerns about "dark patterns" on e-commerce sites conveniently ignore how traditional subscription plans for newspapers and pay TV make it near impossible to unsubscribe.

- **Trade-off tokenism:** The Proposals Paper focuses heavily on risks but does not seek to seriously tackle the trade-offs involved in regulation. It argues that "[Regulation] will also help to attract more foreign business investment in Australia..." but it is entirely unclear how regulation will attract investment and it is bereft of examples where regulation has propelled Australia to the forefront of innovation. Given the potentially unparalleled potential of AI for consumer benefits as well as harms, a critical examination of how existing regulations have held back Australian competitiveness or consumer welfare would be useful — including in regulatory areas held up as exemplars in the Proposals Paper, like therapeutic goods and nuclear technology. In this regard, the New Zealand Cabinet Paper includes a prescient warning:

> *Regulating AI based on "predicted uses" or "speculated harms" may be overly broad in many contexts and harm productivity.*

Where to next?

The Government is considering submissions received on the Proposals Paper throughout the last month or so.

Already, in parallel with the consultation on principles we are starting to see AI-focused regulation emerge. The Privacy and Other Legislation Amendment Bill 2024 recently introduced into the

Australian Parliament proposes new transparency obligations on organisations that make "automated decisions" using individuals' personal information, where such decisions are reasonably expected to significantly affect the rights or interests of an individual. The use of automated decision making will need to be disclosed in organisations' privacy policies. This is aimed at tackling a key complaint made against AI-mediated decisions: that these decisions are opaque and difficult for individuals to engage with or challenge.

It seems we are on an inexorable journey towards broader AI regulation in Australia. The Proposals Paper is an early step, but there is no mistaking that it is a firm step towards dedicated AI regulation based on the EU model. We will provide further updates as the regulatory framework is further developed.

Angus Henderson, Partner

Ish Omar, Partner

Ara Margossian, Partner

Jordan Cox, Partner

Raymond Roca, Partner

Alan Ngo, Senior Associate

webb henderson

Level 18, 420 George Street, Sydney NSW 2000, Australia
Level 17, 188 Quay Street, Auckland 1010, New Zealand