Australian Chamber of Commerce and Industry

Proposals paper for introducing mandatory guardrails for AI in high-risk settings

Submission to the Department of Industry, Science and Resources, October 2024

Introduction

ACCI welcomes the opportunity to provide feedback on the proposals paper, *Introducing mandatory guardrails for AI in high-risk settings.*

ACCI is supportive of the Government's work on Safe and Responsible AI in Australia, and the initiative to establish voluntary AI safety standards / guardrails for the use of AI. Mistrust of AI is a significant barrier to widespread adoption of AI by individuals and businesses alike and it is critical that work progresses to overcome this mistrust in order to realise the productivity benefits AI offers. As part of addressing mistrust, we acknowledge that it is prudent to identify and restrict high-risk use-cases, setting appropriate conditions to protect consumers, business and society.

ACCI supports a risk-based approach, which is consistent with other international regulatory approaches to AI, including the OECD recommendations. This approach should both avoid capturing those applications which do not pose any risk to individuals or society and allow interoperability for companies providing AI products and services cross-border.

Our primary concern with the proposal as is, however, is that the mandatory guardrails in their current form appear disproportionate and require an unachievable compliance uplift for businesses of all shapes and sizes across Australia, disadvantaging all businesses across the AI supply chain seeking to develop or deploy this transformative technology. The guardrails as currently presented would impose a significant compliance burden on all businesses across Australia. This could result in a vast array of unintended consequences for industry, including requiring such a high legal and business uplift that it disincentivises businesses of all sizes to invest in AI development and/or deployment in Australia (potentially leading them to seek out other jurisdictions); and potentially creating inequities resulting in a lack of competitiveness. This is especially so for SMEs seeking to embrace the AI opportunity in Australia (currently only 23-25 per cent of SMEs are using AI in some manner) as well as those seeking to scale their businesses in line with interoperable, internationally accepted best practice risk management frameworks. In the longer-term, this could lead to significant knock-on effects for the wider

digital economy, including stifling innovation, research and development opportunities, preventing the growth of new AI-focussed businesses in Australia, missing out on wider productivity gains created by the technology, and otherwise stunting the growth of the digital economy. Consideration must be given as to how to apply a 'reasonable test' to implementation of the guardrails so that obligations for businesses, particularly SMEs, are proportionate to their relevant use cases and the level of risk such use poses.

With respect to the overarching legislative approaches considered, ACCI believes that many of the risks associated with AI, at this stage in the evolution of the technology, are adequately covered by existing regulations, including Work Health and Safety laws, Fair Work legislation, Australian consumer law, online safety, competition law and discrimination law. As such, while legislative amendments may be required to create consistent definitions and tweak relevant legislation to ensure high-risk uses associated with AI are covered, these amendments should be targeted and should not introduce unnecessary administrative burden or be duplicative.

Consultation questions

1. **Do the proposed principles adequately capture high-risk AI? Are there any principles we should add or remove? Please identify any:**
   a) **low-risk use cases that are unintentionally captured**
   b) **categories of uses that should be treated separately, such as uses for defence or national security purposes.**

The principles represent an important starting point for identifying the parameters of what could be considered high-risk AI; however, they are of little practical use for businesses seeking to classify their AI applications as high- or low-risk AI. The principles must be complemented with an easily accessible and comprehensible list of use-cases (as addressed in question 3), in order to provide legal certainty to businesses, and ease of interpretation.

Consideration should be given to the unintended capturing of private or innocuous uses of AI, which fall within the categorisation of 'high-risk' by virtue of fulfilling the principles or appearing on the list, but do not pose any real risk due to the context in which the application is deployed. Such a test is applied in the EU AI Act, as an important list of exceptions to the list of high-risk applications (cf. Article 6 and Annex III) and we would support its use here: an AI system from the list is not considered high-risk where it is intended to:

- perform a narrow procedural task;
- improve the result of a previously completed human activity;
- detect decision-making patterns or deviations and is not meant to replace or influence the previously completed human assessment, without proper human review; or
- perform a preparatory task to an assessment.

If the intent of the principles is to assist in designating high-risk then the threshold that should be used is 'significant risk' not 'risk' so as to distinguish between low, medium and high risk. This would likely then negate the need for principle f) as a standalone principle with further details of interpreting 'significant risk' provided in use cases and guidance.

In regard to principle a), the language used in the principle is inconsistent with how such law is enforced. The anti-discrimination and human rights framework does not protect against risk (namely a situation where an unlawful outcome has not actually occurred) rather outcomes. Therefore, it would be impossible to use this existing framework to assess level of risk and consequently high-risk, it would only allow for the assessment of what is "lawful" or "unlawful".

In regard to principle b, this principle should not be inconsistent with existing WHS legislation. The phrasing of 'risk of adverse impacts' is inconsistent as WHS legislation does not focus on the impact of any action or inaction but rather the risk of harm. To be more consistent it should be "a *significant* risk of harm" or "a *significant* risk to the health and safety of persons". WHS legislation further defines 'health' as "physical and psychological health".

The terms 'legal effects' and adverse impacts to the 'broader Australian economy, society, environment and rule of law' are too broad-reaching and vague and are likely to result in uncertain and inconsistent application by businesses across the AI supply chain when assessing their risk. They are not workable principles and need to be further refined if they are to assist in achieving a consistent assessment by organisations.

If the intent of e) is to ensure a greater focus on AI use that has the potential to create harm at scale, then this needs to be better defined. Furthermore, it should be made clear that 'high-risk' is both the possibility of harm occurring from the use of AI but also the likelihood of it occurring being high, that is, likely or almost certain. Extensive guidance clarifying this would be needed to ensure low-risk uses or uses that have potential localised harm are not inadvertently captured.

3. Do the proposed principles, supported by examples, give enough clarity and certainty on high-risk AI settings and high-risk AI models? Is a more defined approach, with a list of illustrative uses, needed?

a) **If you prefer a list-based approach (similar to the EU and Canada), what use cases should we include? How can this list capture emerging uses of AI?**
b) **If you prefer a principles-based approach, what should we address in guidance to give the greatest clarity?**

In its present form, the list of principles to be used to classify a given AI system as high-risk is not sufficiently clear or detailed. It is not evident how such a list of principles would be applied by companies to make accurate and consistent assessments of the risk profile of AI systems, and how this principle-based approach would be enforced. If this approach were to be taken, it would need to be accompanied by extensive guidance including easy to use risk assessment templates, including those which are internationally recognised, where applicable.

While additional guidance can be helpful, the more practical approach to defining high-risk AI systems would be a list-based approach, similar to the EU and Canada. This sets out a discrete list of applications, which developers and deployers can easily consult and on this basis, classify a given AI application as high-risk or not high-risk. With respect to futureproofing a list, and guaranteeing that emerging AI applications are not de facto included or excluded without appropriate assessment, a mechanism could be introduced whereby the list could be reviewed and amended on a regular basis, with new applications added where necessary, and

applications removed, where they no longer pose a risk. Flexibility may be achieved by delegating relevant parts (i.e. a 'living list') to subordinate legislation. This may be through publishing 'Rules', a 'Code' or guidance.

Consideration should also be given to user-friendly tools to help developers and deployers use classifications (whether they be principle- or list-based), such as the OECD Framework for the Classification of AI systems.

5. Are the proposed principles flexible enough to capture new and emerging forms of high-risk AI, such as general-purpose AI (GPAI)?

6. Should mandatory guardrails apply to all GPAI models?

While GPAI represents a series of powerful AI applications, many uses of GPAI (e.g. generative AI chatbots) are deployed in low-risk scenarios, and this type of technology represents one of the most commonly used AI applications, especially among small businesses.

It would be unnecessarily heavy-handed to firstly categorise all GPAI as high-risk, and secondly, on this basis, apply all 10 guardrails in a mandatory way to use of GPAI. The EU approach deals with GPAI in a proportionate manner, applying only certain obligations (not all the obligations applicable to high-risk AI) to those GPAI applications which pose 'systemic risk', which itself has a high threshold of 'actual or reasonably foreseeable negative effects'.

The Australian approach should follow this example, to ensure that all GPAI is not unnecessarily captured in 'high-risk AI', that a risk-based approach specific to GPAI is applied, and that all guardrails are not disproportionately applied. On top of a test for systemic risk, a criterion should be added to exclude narrow procedural tasks, and applications used to improve the result of a previously completed human activity.

7. What are suitable indicators for defining GPAI models as high-risk? For example, is it enough to define GPAI as high-risk against the principles, or should it be based on technical capability such as FLOPS (e.g. 10^25 or 10^26 threshold), advice from a scientific panel, government or other indicators?

It is advisable to make a distinction between GPAI generally, and GPAI which poses a systemic risk / could be considered high-risk. While the principles are a good starting point to define high-risk AI, for the same reasons as outlined above regarding the classification of any AI system as high-risk, these should be complemented with clear parameters (e.g. a list-based model), or, in the more complex case of GPAI, a technical criterion (e.g. FLOPS) or advice from a scientific panel, as proposed in the consultation paper, following the model of Article 51 of the EU AI Act, for example, if some GPAI models are to be defined as high-risk, this could be limited to highly capable frontier models trained on more than 10^26 FLOPs.

8. Do the proposed mandatory guardrails appropriately mitigate the risks of AI used in high-risk settings? Are there any guardrails that we should add or remove?

See answer to question 12

10. Do the proposed mandatory guardrails distribute responsibility across the AI supply chain and throughout the AI lifecycle appropriately? For example, are the requirements assigned to developers and deployers appropriate?

In the proposed approach, the guardrails would apply indiscriminately across the board to both developers and deployers of AI systems and applications. Consideration should be given to different parameters, such as role (deployer v. developer), position in supply chain, and degree of customisation of AI systems – and nuances should be applied accordingly.

With respect to their role in the AI supply chain, we would encourage the Government to provide further clarity to businesses on the roles and responsibilities assigned to each actor in the complex AI supply chain. For example, developers (should be subject to different obligations by virtue of their unique role in the development of the systems themselves. For example, they are best placed to provide information about an AI system, so stakeholders can make informed choices about their use of the system. Deployers on the other hand, have a much better ability to determine whether they have implemented their AI system for high-risk use, and any appropriate risk mitigation practices that need to be implemented as a result.

Specifically, guardrails 3 (protect AI systems and implement data governance measures), and 4 (test AI models) should be mandatory only to developers of AI systems, and not to deployers. Furthermore, nuances should be added to certain guardrails (e.g. inform end-users, keep and maintain records, transparency across organisations) to more accurately reflect the different roles of developer and deployer in the AI ecosystem and the interaction of other existing domestic legislation.

Furthermore, reflecting the market reality that developers of AI are primarily established overseas (e.g. in the US), and Australian companies are principally deployers of AI technologies, it could be considered decoupling the guardrails applied to both developers and deployers, so that the rules applying to developers are consistent with international consensus (e.g. in the US, Canada, EU, UK), which may be achieved through recognition of existing international standards (e.g. ISO 42001:2023), and rules for deployers in Australia could benefit from a more rational and risk-proportionate regime.

12. Do you have suggestions for reducing the regulatory burden on small-to-medium sized businesses applying guardrails?

As a principle, the guardrails should be applied proportionately, especially with the aim of encouraging the uptake of AI applications across the Australian economy, including small-to-medium businesses. A tiered model could limit certain 'high risk' AI systems deployed by SMEs to compliance with only a reduced number of guardrails, e.g. mandating the informing of end-users about AI use (guardrail 6) and some cyber security / data protection measures (guardrail 3), but not onerous evaluations for regulatory compliance and ongoing conformity

assessments. This may be done in a way to address the scale of risk a small or medium business (with a finite pool of customers or interactions) may pose compared to that of a large company with significant market reach.

Guardrail	Impact on SMBs	Applicability to deployers of AI systems
1 **Establish, implement and publish an accountability process including governance, internal capability and a strategy for regulatory compliance**	Requirement to draft documents demonstrating compliance with regulations, drafting and implementing policies for data and risk management, clear roles and reporting structures for staff, details of training made available to staff. **Small businesses would be required to familiarise themselves with complex legislation (extensive guidance would be needed) and standards. Small businesses do not have the resources to undertake such preparatory measures to use AI applications, nor to alter staff structures / provide training programs.**	The threshold of 'high-risk' should be sufficiently high that most common, small business use cases of AI would not be captured and therefore they would not be subject to this guardrail. We are particularly concerned that several 'employment' use cases may be inadvertently captured that should not be deemed 'high-risk' due to their localised use and use of human oversight in decision-making. This guardrail if applied to SMEs should be scalable based on what is 'reasonably practicable' for a SME and extensive guidance provided on this.
2 **Establish and implement a risk management process to identify and mitigate risks**	Drafting of a risk management process, including processes for identifying the risks and assessing the impact of the risks, identifying and applying risk mitigation measures, and mechanisms to identify and mitigate new risks. **Businesses would need to invest in materials and guidance to support the process of developing a risk management process, which is not an intuitive process, especially when it comes to new technologies and applications. One relevant**	Clarity is needed as to whether existing WHS risk assessments would be sufficient to meet this guardrail. Government guidance would be needed on this guardrail to support industry, especially small businesses with its application. This could include checklists for drafting risk management processes, or ways to **incorporate existing risk assessments** that may have been done under a WHS or governance risk framework. We also recommend specifying which risk management

	example is the ISO standard 42001:2023, which is a costly investment for small businesses ($340), and not an intuitive document to read.	processes will apply to each member of the varied AI supply chain, for example, who should conduct an impact assessment before a high-risk AI system is deployed, and when.
3 **Protect AI systems, and implement data governance measures to manage data quality and provenance**	Requires appropriate data governance, privacy and cybersecurity measures in place – while such practices may already be in place to comply with other legislation, e.g. Privacy Act, SOCI Act, this nevertheless represents another investment requirement for businesses wishing to use AI.	It should be noted that the precise **requirements of this guardrail pertain mainly** to the **work of a developer of AI** (data provenance, training data, assessment of data quality), and that the burdensome requirements with respect to training data and quality of data should not be applied across the board to both developers and deployers. However, it has been raised by our members that deployers may further customise or modify the AI application, making them **quasi-developers** of the AI system – it is **unclear where the threshold would be** for allocating responsibilities. Clear guidance would be needed on designating responsibilities based on the proportion of risk. By the same token, Government should ensure that **businesses of all sizes are protected from disclosing proprietary information** (protected under IP law) under these obligations.
4 **Test AI models and systems to evaluate model performance and**	Requirements for post-market monitoring, and ongoing assessment of the model, would mean a deep understanding of	Testing and evaluation of **models** should not be required of deployers, who typically take pre-existing models or applications, and either use

monitor the system once deployed	the underlying AI model on which a given application is based. **A voluntary reporting mechanism could be established, whereby deployers could report issues to developers, in conjunction with guardrail 8 on supply chain.**	them directly, or build applications on top of the model. Noting comment on quasi-developers above. To accommodate small and medium businesses, this requirement should have a reasonably practicable element. It is important to avoid a one-size-fits-all model, so that requirements are scaled depending on the **specific use-case** and the **degree of customisation** undertaken by the deployer.
5 **Enable human control or intervention in an AI system to achieve meaningful human oversight**	This guardrail would require developers and deployers to understand and oversee the operation of AI systems and intervene where necessary across the AI supply chain. As above, this kind of oversight is available in a limited capacity to deployers of AI systems, as they are unable to access the 'back-room' information and systems provided by the developer. **The goal should be to foster responsible use of AI systems and healthy communication across the whole AI supply chain, consistent with guardrail 8.**	Oversight requires cooperation across the whole AI supply chain, with in-depth insights into the functioning of an AI system often unavailable to deployers of AI, especially SMBs, who may not have further developed the AI application beyond how it was marketed (particularly the case for AI embedded into existing software). It is **important to avoid a one-size-fits-all model**, so that requirements are scaled depending on the **specific use-case** and the **degree of customisation** undertaken by the deployer, or their position in the supply chain.
6 **Inform end-users regarding AI-enabled decisions, interactions with AI and AI-generated content**	Comprises 3 separate requirements: 1. Inform people when AI is used to make or inform decision relevant to them 2. Inform people when they are directly with an AI system 3. Apply best efforts to ensure AI-generated	Information to end-users about AI-based decision-making is important for consumers for building trust. This can be undertaken by SMBs, but should be supported with government guidance, e.g. model clauses and information templates.

	outputs can be detected as artificially generated The informational requirements (parts 1 and 2) would be easily implementable by deployers, including SMBs – these should be provided by government guidance on model information clauses. The third aspect is a very specific requirement, which should not be applied across the board, but should be applicable only to very specific types of business and AI system (which rely on AI-generated outputs).	
7 **Establish processes for people impacted by AI systems to challenge use or outcomes**	Process will include establishing internal complaint handling functions, assignment of responsibility for dealing with complaints, and the provision of redress mechanisms. **Such procedures are already covered by sector-specific regulations, e.g. consumer law, privacy law, employment law.** **Any additional nuances on complaints should be embedded in existing complaint-handling procedures, and care should be taken to clearly delineate issues to be resolved by the deployer versus the developer.**	**Further clarifications are needed to delineate precise types of issue / complaint which should be addressed by the deployer (in the first instance), or when the issue is specifically related to the system itself, the developer.**
8 **Be transparent with other organisations across the AI supply chain about data, models and systems to help them effectively address risks**	The transparency of information across the supply chain (i.e. the provision of information on the operation of the system, and interpretation of outputs, from developer to deployer) is an important topic, and is essential in order for deployers to be fully aware of how to resolve issues with the system, and to comply	This is an important aspect to the AI ecosystem, however further thought should be given to the specific responsibilities along the AI supply chain, including types of information to be communicated, timelines, accessibility / machine-readable and interoperability of information. Clarity is also

	with the other guardrails, where oversight is required. **Instead of making this a strict guardrail, a transparent environment should be encouraged, with issue reporting a commonplace procedure in the deployment of AI systems. Furthermore, providing key transparency information about how the LLMs have been evaluated for veracity (i.e. the likelihood of hallucination), safety (including efforts to red team the model), and controllability and similar would be more useful than technical information.**	needed on the implications of information provisions for liability.
9 **Keep and maintain records to allow third parties to assess compliance with guardrails**	This requires documentation consisting of a general description of the AI system, design specification from the development phase (including testing methodology and results), a description of the datasets and their provenance, an assessment of human oversight measures, a detailed description of the capabilities and limitations of the system, and the risk management processes and mitigation measures implemented. **This requirement should be fulfilled first and foremost by the <u>developer</u>, as these technical details can only be supplied by the party with technical oversight of the development and testing phase, consistent with guardrails 1 and 2.**	**This guardrail should be decoupled into:** • **A requirement for <u>developers</u> to provide all relevant information about the development and testing (including risk management) to deployers** • **A requirement for <u>deployers</u> to keep and maintain records only about their own internal risk management procedures**
10 **Undertake conformity**	If conformity assessments are conducted on deployers and developers by third parties, this	

assessments to demonstrate and certify compliance with the guardrails	will represent a major cost in terms of time (preparation), cost (to pay the third-party conformity assessor), and implementation of recommendations. Furthermore, we are unaware of how many service providers are currently able to provide this service, and are concerned that there would be a shortage of providers, should this guardrail be made mandatory, which would further stifle adoption.	

13. Which legislative option do you feel will best address the use of AI in high-risk settings? What opportunities should the government take into account in considering each approach?

The third option – a whole-of-economy approach, introducing a new AI Act – would be the most efficacious approach to applying the guardrails across all sectors of the economy, and ensuring consistency of application, without the risk of diverging approaches in sector-specific regulations.

Nevertheless, taking into account the interplay between Commonwealth and State/Territory regulations on certain issues, e.g. Work Health & Safety (WHS), Health, and Employment law, which are rigorously regulated and have their own regulatory oversight bodies and mechanisms (e.g. TGA for health), these should be exempted from the horizontal AI Act. Definitions may be adopted where practicable for greater consistency across legislation. Please see question 16 for a more detailed response to concerns around intersection with employment law.

14. Are there any additional limitations of options outlined in this section which the Australian Government should consider?

Given the constant evolution of the state of the technology, and the AI market both globally in Australia, and the need for Australian businesses to develop and compete, regulatory sandboxing mechanisms should be made available in the legislation to encourage innovative development and deployment of new AI solutions.

15. Which regulatory option/s will best ensure that guardrails for high-risk AI can adapt and respond to step-changes in technology?

As mentioned above, in order for the regulations to keep up with the state of play of the technology, the list of high-risk AI applications needs to be equipped with a mechanism whereby the regulator or the legislator can amend that list, taking into account technological

evolutions and newly-identified risks, or risks which are no longer to be classified as such, and therefore can be removed from the high-risk list of applications to which the guardrails apply.

16. Where do you see the greatest risks of gaps or inconsistencies with Australia's existing laws for the development and deployment of AI? Which regulatory option best addresses this, and why?

As mentioned above, the biggest regulatory issues lie with consistency of regulations at the Commonwealth and State/Territory levels: this would be most keenly felt in areas such as Work Health & Safety (WHS), and in Employment law.

The model WHS laws are risk-based and flexible enough to address emerging AI risks. The risk management process set up under this legislation is well -known and methodical. The proposals in this paper would create a parallel assessment process that is duplicative in parts and may result in two risk assessments with two different risk ratings and outcomes. This would ultimately undermine the objectives of both pieces of legislation and result in only further confusion from SMEs. Existing WHS risk identification and assessment processes as it relates to AI use and WHS should prevail with any bespoke AI legislation and guidance mirroring this. Attempting to implement these guardrails in employment settings, with particular reference to obligations and entitlements arising under the Fair Work Act or discrimination frameworks, would be immensely difficult due to the high level of duplication that would ensue. ACCI strongly recommends against imposing the guardrails on employment through designating employment uses as automatically high-risk. The textbox on page 27 makes particular reference to several instances where AI might be used in workplace settings:

CV-scanning services / AI applications used in hiring

The paper suggests that CV-scanning services may create the **risk of discrimination** in assessing individuals' suitability for employment opportunities. The legislative framework already protects workers from discrimination in recruitment, across multiple acts and multiple jurisdictions.[1]

Under the Australian anti-discrimination framework, both direct and indirect discrimination are unlawful. Direct discrimination occurs when a person, or a group of people, is treated less favourably than another person or group because of their background or certain personal characteristics. Indirect discrimination occurs when there is an unreasonable rule or policy that is the same for everyone but has an unfair effect on people who share a particular attribute. In an Australian Human Rights Commission (AHRC) Technical Paper from 2020, it was established that the use of AI would already lead to "unlawful discrimination and other forms of human rights violation". Meaning the existing legal infrastructure is already effective.

[1] Australia's anti-discrimination framework is broadly governed by the following regimes: the Age Discrimination Act 2004 (Cth), the Disability Discrimination Act 1992 (Cth), the Racial Discrimination Act 1975 (Cth), the Sex Discrimination Act 1984 (Cth), the Anti-Discrimination Act 1977 (NSW), the Equal Opportunity Act 2010 (VIC), the Anti-Discrimination Act 1991 (QLD), the Equal Opportunity Act 1984 (WA), the Equal Opportunity Act 1984 (SA), the Ant-Discrimination Act 1991 (TAS), and the Anti-Discrimination Act 1992 (NT).

The Fair Work Act 2009 (FW Act) specifically deals with discrimination in the hiring process. Under the general protections in section 351, an employer must not take adverse action against an employee or a prospective employee based on any of the following characteristics: race or colour, sex, gender identity, intersex status, or sexual orientation, marital status, pregnancy, breastfeeding, or family or caring responsibilities, age, physical or mental disability, religion, political opinion, national extraction or social origin, and because they are experiencing family and domestic violence. Section 342 of the FW Act clearly states that adverse action against a prospective employee includes refusing to employ the prospective employee. This means that it is already unlawful for a business to not hire someone because of the above characteristics. Importantly, the FW Act covers any action taken by an employer, including the implementation of AI technologies. The FW Act mandates that the person alleged to have engaged in the action is presumed to have engaged in the alleged action with the alleged intent unless they can prove otherwise. If AI technologies have caused an incident of discrimination to occur during recruitment, then they are captured by the FW Act and may be dealt with accordingly. The requirement of a person to prove they have not engaged in adverse action is an additional element of accountability. Employers are not absolved if a contravention occurs through the use of AI. Hence, where risks to employers' obligations arise from using AI, then there must be oversight and review systems in place to ensure that there is not a failure to comply with legislation, such as the discrimination framework (because it is already effective). The approach outlined in relation to this possibility at page 27 is hence unwarranted – the current framework is rigorous.

Not only is any risk already dealt with appropriately under anti-discrimination laws, but the proposals paper's desire to designate the use of AI in recruiting as high-risk might make it less attractive to employers thereby dismantling the potential benefits that might be obtained from its use. The use of AI in the hiring process may have material impacts on driving down inequalities in the employment system. Studies which examined the biases of recruiters and managers with respect to ethnic groups have shown that those with non-English names are less likely to progress through a hiring process than those with an English name. AI presents an opportunity to overcome the challenges associated with the unconscious bias in human decision making, which research from the UK suggests is present in hiring processes. ACCI has also previously heard feedback to that effect from its membership network. One member observed earlier this year that "the use of this technology in hiring which blindly-screens candidates has secured results of more diverse staff being hired within my organisation than ever-before." Naturally, where blind screening tools can be used through AI, the unconscious bias that would be present in a human's decision making will not be present. Research has shown that a recruiter may only scan a CV for seconds, meaning that their preconceptions may play a strong role in who they invite for an interview. AI can help remove these preconceptions by only dealing with specific criteria and inputted information.

Automated rostering systems

The paper suggests the case of automated rostering systems not adequately factor in caring responsibilities of employees. Implementing such systems does not abrogate employers of their obligations to consult with employees about changes which impact their rosters under their award or EA as is required by sections 145A and 205 of the Fair Work Act 2009 (FW Act).

Nor could it possibly overcome any flexible working arrangement made under the National Employment Standards. As the FWO outlines, where an employer wants to change an employee's regular roster or ordinary hours of work, they have to discuss it with them first, provide information about the change, invite their views, and consider those views. Employers are still required to meet these consultation obligations regardless of whether they used an AI application to create the roster. Naturally, where risks do arise with respect to employers' obligations, then there must be oversight and review systems in place to ensure that there is not a failure to comply. Importantly, automated rostering can be programmed to consider compliance rules, reducing the risk of human error and ensuring that staff are scheduled accordingly.

AI is already being used lawfully and carefully by employers to implement more efficient processes with respect to rostering arrangements. Instituting these mechanisms increases efficiency and reduces administrative burdens on employers and managers at an economy-wide scale. Furthermore, these systems may help to decrease unconscious bias. Some employers or managers may unconsciously roster others more frequently, or at times more often, due to personal relationships, for example, and AI makes decisions based on pre-defined rules and data, mitigating the potential for bias in scheduling. One business from ACCI's membership network made the following observation: "Managers within my organisation will roster shifts for 200-300 employees, this was not at all possible before the use of this rostering technology. This technology allows managers to not be bogged down in these administrative duties, and instead to free up time for other employees within the team. Employees really benefit from this technology as they can simply input availability and preferences, and the technology can effectively accommodate staff needs".

Monitoring and surveillance

This example is twofold – one that it might impose unlawful monitoring, and two it might lead to an unfair determination about the employee's employment. These are problems with several existing solutions that do not need to be further duplicated by way of the guardrails. Firstly, in relation to monitoring and surveillance as a practice. Businesses invest significant amounts of money into employees and their equipment, they have a right to assure themselves that they are being used in the appropriate way. The Fair Work Ombudsman, on its website, elucidates on the use of monitoring technology, advising that businesses are increasingly using technology to lawfully monitor such things as work output, how business property is being used and employee attendance. The mechanism through which monitoring (i.e., AI applications) occurs is not necessarily the key factor: as the Australian Privacy Principles are technology-neutral, the principal factor is whether these actions are occurring in a lawful manner and that the employer is conforming to their privacy obligations. Privacy law applies, regardless of how monitoring occurs (as such employers should always have review and oversight mechanisms in place to ensure compliance).

Secondly, in relation to decisions taken about an employee's employment due to monitoring and evaluation by an AI, ACCI would draw attention to the rigorous protections and unfair dismissal laws present in the Fair Work Act. Simply because an AI application was used in a decision, employers cannot obviate their obligations and entitlements under the Fair Work Act.

Section 387(e) of the Fair Work Act is clear: a consideration that must be taken into account as to whether a dismissal was unfair (harsh, unjust or unreasonable) when it occurred because of unsatisfactory performance, is whether the person had been previously warned about their performance. This consideration does not evaporate because AI was used to assess that performance. Hence, where an AI application was used to monitor performance and take a decision about an employee's employment, it still needs to be fair in that context (i.e., the employees should be given a chance to ameliorate their performance issues). If the proposed guardrails also seek to deal with these obligations, they will introduce duplication. For example, as recently as a few months ago, the FWC ruled a dismissal unfair after an employer used generative AI to draft a text terminating an employment relationship. The employer was not able to avoid their obligations by using AI in the workplace. The guardrails hence would have very little work to do in employment settings beyond imposing unnecessary new burdens on businesses. This would serve to make AI less attractive to employers (thereby reducing their proliferation) when they might actually be of some benefit to employees. Some employers or managers may unconsciously promote employees due to personal relationships, for example, and AI could make decisions based on pre-defined rules and data, stemming from such monitoring. This would actually mitigate the potential for bias.

Other employment-specific situations

In addition to the three examples listed above, the paper makes further reference to specific employment settings that might give rise to a high-risk designation for the purposes of the guardrails and/or potential legislation. Those settings are "recruitment and hiring, promotions, transfers, pay and termination." Recruitment, promotions, and termination have already been addressed in this submission and it has been evidenced that there is no new risk or challenge that AI might create, which is not already addressed by anti-discrimination laws and the Fair Work Act. With respect to pay (wage setting), ACCI submits that this is not a high risk setting as the proposals paper argues.

Wage setting

With respect to the use of wage setting in employment contexts, the risks of AI are minimal. The guardrails will only add to the complexity of the industrial relations system if it seeks to impose on how AI might interact with pay. Australia has a substantial minimum safety net of terms and conditions through the National Minimum Wage (NMW) and modern award minimum wages, which are reviewed annually to assess their adequacy. An employee cannot be paid less than their applicable modern award minimum wage, even if they agree to it, and if no modern award applies, they cannot be paid less than the NMW. Furthermore, in contexts where an identifiable pay point is available such as in modern award-reliant businesses and those with an enterprise agreement, the risks of AI are minimal and may simply present an opportunity to increase the efficiency of clerical tasks. Most independent contracting arrangements also bear minimal risk as the contractor can set or negotiate their own pay.

Transfers

While the reference to transfers is vague, it could be taken to have its meaning in section 311 of the Fair Work Act, a transfer to a new position, or it could be taken to mean physical location (geographical transfer). If it is the latter, although this is an operational decision that is firmly within managerial prerogative, certain relocations would be taken to be a repudiation of the contract and would allow the employee to receive redundancy pay. For example, as recently as September, the Fair Work Commission (FWC) has rendered workers entitled to genuine redundancy because an employer sought to transfer their location of work. If it is taken to mean a transfer to a new position, then protections already exist in that instance as well. The FWC has previously held that an employer does not have the right to transfer an employee from one position to another without his or her consent and forcing them to do so would thereby enliven section 386(1)(b) of the Fair Work Act (namely that a dismissal has occurred). Finally, if it is within the meaning of section 311 then subclause (1)(a) clarifies that a termination must have occurred, which at the outset means the unfair dismissal regime is in effect. Hence, it is clear that the guardrails do not have any work to do with transfers – it does not matter if AI is used in a decision about a transfer, the relevant protections apply.

In conclusion, ACCI submits that the guardrails do not offer any additional protections not already provided to employees under employment law in Australia. The application of the guardrails to the employment context, which is already highly regulated, would cause duplication, leading to new complexities and unnecessary administrative burdens.

We thank you for your consideration of our feedback. Should you require any additional information or clarification of any points contained within, please contact Jennifer Low, Director Health, Safety, Resilience and Digital Policy at jennifer.low@acci.com.au.

Telephone 02 6270 8000 | Email info@acci.com.au | Website www.acci.com.au
Media enquiries: Telephone 02 6270 8020 | Email media@acci.com.au