CHOICE

4 OCTOBER 2024

INTRODUCING MANDATORY GUARDRAILS FOR AI IN HIGH-RISK SETTINGS

Submission to the Department of Industry, Science and Resources

ABOUT US

Set up by consumers for consumers, CHOICE is the consumer advocate that provides Australians with information and advice, free from commercial bias. By mobilising Australia's largest and loudest consumer movement, CHOICE fights to hold industry and government accountable and achieve real change on the issues that matter most.

To find out more about CHOICE's campaign work visit www.choice.com.au/campaigns and to support our campaigns, sign up at www.choice.com.au/campaignsupporter

57 Carrington Road Marrickville NSW 2204
Phone 02 9577 3333 | Fax 02 9577 3377 | Email campaigns@choice.com.au | www.choice.com.au
The Australian Consumers' Association is a not-for-profit company limited by guarantee. ABN 72 000 281 925 ACN 000 281 925

CONTENTS

Introduction

Mandatory guardrails on the development and use of artificial intelligence (**AI**) are needed to protect consumers in Australia. While AI may present some benefits to consumers, this must be balanced against consumer harms created, or exacerbated, by AI systems.

CHOICE welcomes the Federal Government's decision to explore options for mandatory guardrails on AI in high-risk settings. CHOICE has previously advocated for robust regulations in the Department of Industry, Science and Resources' (**DISR**) Safe and Responsible AI in Australia consultation. We strongly believe that mandatory guardrails will reduce the harms associated with AI systems while increasing its value to consumers. Australia risks lagging behind the pace of technological developments and regulatory developments abroad if mandatory guardrails aren't established quickly.

Since the previous consultation, CHOICE has produced further national research demonstrating consumer expectations on AI safeguards.[1] This research shows:

- 78% of people agree businesses should have to ensure their artificial intelligence system is fair and safe before releasing it to the public;
- 77% of people agree that the government should require businesses to assess the risks of their artificial intelligence products or services before releasing them to consumers;
- 75% of people agree that the government should require businesses to prevent the risks of their artificial intelligence products or services before releasing them to consumers;
- 69% of people agree that the government should have an independent third party assess the risks of businesses' artificial intelligence products or services before releasing them to consumers; and
- 80% of people agree businesses should allow customers to speak to a person if they're unhappy with a decision made by artificial intelligence.

The Federal Government must act soon to ensure that consumer goods, services and markets are safe from harmful AI systems. This submission encourages the government to focus on three key elements in the mandatory guardrails to protect consumers: 1) a risk framework based on ethical principles including prohibitions for unacceptable risks; 2) mandatory guardrails that encompass all aspects of the AI supply chain; and 3) whole-of-economy standalone legislation

[1] CHOICE Consumer Pulse January 2024 is based on an online survey designed and analysed by CHOICE. 1,058 Australian households responded to the survey with quotas applied to ensure coverage across all age groups, genders and locations in each state and territory across metropolitan and regional areas. The data was weighted to ensure it is representative of the Australian population based on the 2021 ABS Census data. Fieldwork was conducted from the 16th of January until the 5th of February, 2024.

that establishes an AI Commissioner, complemented with reforms to consumer and privacy law, and the establishment of a digital ombudsman.

Recommendations

The Federal Government should:

1. Adopt the principles proposed by DISR for designating an AI system as high-risk;
2. Consider strengthening these principles by:
    a. Highlighting especially risky practices such as practices that may affect First Nations people;
    b. Giving regard to Australia's AI Ethics Principles;
    c. Designating all general-purpose AI (**GPAI**) systems as high-risk; and
    d. Adopting an "unacceptable risk" category for practices that should be prohibited, including certain uses of GPAI.
3. Adopt the mandatory guardrails proposed by DISR for organisations developing or deploying high-risk AI systems;
4. Consider strengthening these guardrails by:
    a. Providing details and examples that clarify the requirements under each guardrail;
    b. Clarifying and/or establishing the role of ethics and rights in the lifecycle and supply chain of an AI system;
    c. Including stronger mandates on the role of organisational governance in the AI supply chain; and
    d. Outlining explicit guidelines for developers and deployers to consider when to suspend or terminate their AI system.
5. Introduce whole-of-economy legislation to regulate the use of AI in Australia (an AI Act), including a principles-based framework to designate high-risk AI systems, a regulatory framework to assign practices as unacceptable risks, and mandatory guardrails;
6. Establish a standalone AI Commissioner under an AI Act with adequate resources and a full range of regulatory powers, including criminal and civil penalty powers and rule-making powers where appropriate;
7. Reform the *Privacy Act 1988* to better protect an individual's privacy, including the introduction of a fair and reasonable use test;
8. Introduce amendments to the Australian Consumer Law including a prohibition on unfair trade practices and a general safety provision; and
9. Establish a digital ombudsman to handle complaints, including complaints related to AI systems.

A risk-based approach can prevent AI harms to consumers

The development and deployment of AI systems can present significant risks to consumers. Government action on AI should centre on the prevention of harm and mitigation of risk. CHOICE welcomes DISR's interest in developing a risk-based framework, and generally supports DISR's proposed principles for designating an AI system as high-risk. The breadth of these principles will encompass the majority of high-risk applications of AI, and unlike a list-based approach, is more likely to capture future developments of high-risk AI. However, CHOICE echoes concerns from other organisations that particularly significant risks may require more specificity in the principles, such as risks to First Nations peoples.

The Federal Government and DISR should also consider explicitly referring to Australia's AI Ethics Principles in the proposed principles. These principles have been adopted by Australian government agencies as best practice[2], and are generally interoperable with AI ethics frameworks abroad. Australia's AI Ethics Principles may additionally strengthen the proposed principle regarding adverse impacts on human rights – due to Australia's lack of analogous human rights frameworks to the European Union, United States, and Canada, legislating ethics principles may help achieve intended rights-based outcomes. CHOICE has previously recommended legislating these ethics principles as part of regulations on the use of AI.

CHOICE strongly encourages the Federal Government to include an unacceptable risk category in its risk-based framework. Applications of AI that carry these risks should be legally prohibited. This would align Australia's mandatory guardrails with the European Union's approach to unacceptable risks in Article 5 of the AI Act. An unacceptable risk category could be – in contrast to the proposed principles on high-risk AI – list-based in order to more specifically target applications of AI that have been identified as posing very high risks, such as facial recognition technology. This should be a dynamic list that governments, departments, regulators and/or other key decision makers could add to, or remove from, based on consultation with stakeholders. A non-exhaustive list of practices that should be considered unacceptable risks includes:

- Facial recognition technology;
- Social scoring to deny services or discriminate against consumers; and
- Intentionally manipulative AI systems, particularly systems targeted at vulnerable people.

[2] Department of Finance (2024), "Implementing Australia's AI Ethics Principles in government", https://www.finance.gov.au/government/public-data/data-and-digital-ministers-meeting/national-framework-assurance-artificial-intelligence-government/implementing-australias-ai-ethics-principles-government.

CHOICE supports DISR's proposal to, at minimum, categorise all GPAI systems as high-risk and subject to mandatory guardrails. However, consideration should be given to whether certain GPAI platforms, designs, deployments, or uses of GPAI may present unacceptable risks.

Recommendations

The Federal Government should:

1. Adopt the principles proposed by DISR for designating an AI system as high-risk;
2. Consider strengthening these principles by:
    a. Highlighting especially risky practices such as practices that may affect First Nations people;
    b. Giving regard to Australia's AI Ethics Principles;
    c. Designating all GPAI systems as high-risk; and
    d. Adopting an "unacceptable risk" category for practices that should be prohibited, including certain uses of GPAI.

Guardrails are needed throughout the entire AI supply chain

The unique features of AI systems present new challenges for regulators and lawmakers. The highly complex, technical and often opaque nature of AI systems means that the developers and deployers of these systems hold critical responsibilities to mitigate risks and harms to end-users. However, technology firms should not be trusted to self-regulate responsibly. Therefore, *ex ante* mandatory guardrails on the entire lifecycle of an AI system should be imposed on any organisation developing or deploying high-risk AI systems.

CHOICE broadly supports the list of mandatory guardrails proposed by DISR for adoption by the Federal Government, and supports the foundational principles of testing, transparency and accountability in the AI supply chain. The final version of the mandatory guardrails should include details and examples that clarify the requirements under each guardrail to provide certainty to developers, deployers and end-users. The following recommendations for clarity will provide consumers greater protections under the mandatory guardrails.

CHOICE encourages DISR to clarify or establish the role of ethics and rights in the AI lifecycle and supply chain. While ethics may be implicit in the guardrails required to consider risk and accountability, mandating the role of ethics as a key measure for the AI supply chain will require developers and deployers to consider the safety, fair treatment and rights of their end-users and the public throughout the lifecycle of the system. CHOICE notes that Article 10 of the EU's AI Act makes explicit mention of risks to fundamental rights when articulating appropriate data governance and management practices, and the details of Australia's mandatory guardrails should make similar directives.

CHOICE also encourages DISR to include stronger mandates on the role of organisational governance in the AI supply chain. Organisations operating AI systems should have clear roles and responsibilities at a governance level to ensure accountability to all stakeholders, such as shareholders, end-users and the broader public. The autonomous nature of AI systems should not act as a defence if harms occur, and reinforcing governance responsibilities in the mandatory guardrails will create strong incentives for careful supervision of AI systems across all levels of an organisation. The increasing role of governance in preventing harms to consumers has been raised in recent scams prevention frameworks and the Hayne Royal Commission[3], and ongoing issues with data collection and privacy in the partnership between I-MED and harrison.ai highlight how unclear governance processes may emerge in AI supply chains.[4]

Finally, DISR should outline explicit guidelines for developers and deployers to consider when to suspend or terminate their AI system. While this may be implicit in guardrails around risk mitigation, developers and deployers of AI systems should be explicitly required to consider whether the risks of a system are too great to continue operations, rather than simply find ways to mitigate these risks. As these guardrails are designed for systems already identified as high-risk, greater onus should be placed on organisations to justify the continuity of their systems, and should be wholly accountable to regulators and the public as to why they continued operations. Developers and deployers of high-risk AI systems should not assume that once they have satisfied minimum requirements that their systems will always be safe; this is particularly true for AI systems due to its autonomous nature and unforeseeable use cases and adaptations from deployers and end-users. This guardrail should also refer to unacceptable risks recommended in the previous section.

Recommendations

The Federal Government should:

3. Adopt the mandatory guardrails proposed by DISR for organisations developing or deploying high-risk AI systems;
4. Consider strengthening these guardrails by:
    a. Providing details and examples that clarify the requirements under each guardrail;

[3] *Treasury Laws Amendment Bill 2024: Scams Prevention Framework* (Cth), sub-div B; Financial Services Royal Commission, *Final Report of the Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry,* pp. 35-37.
[4] Wilson (2024), "Leaked harrison.ai email shifts blame to IMED over patient consent", Crikey, https://www.crikey.com.au/2024/09/20/leaked-harrison-ai-email-i-med-patient-scan-training-ai.

b. Clarifying and/or establishing the role of ethics and rights in the lifecycle and supply chain of an AI system;
c. Including stronger mandates on the role of organisational governance in the AI supply chain; and
d. Outlining explicit guidelines for developers and deployers to consider when to suspend or terminate their AI system.

Whole-of-economy laws can keep our community safe from harmful AI

Consumers are best protected when laws are applied consistently across markets. Relying on sector-specific legislation, voluntary industry codes and standards alone will likely leave consumers vulnerable to uneven protections and rights. Consumers may also be left unaware of, or confused about, how to seek redress for harms.

CHOICE supports a whole-of-economy approach including a new cross-economy AI-specific Act (option 3) as the preferred regulatory option to mandate guardrails. CHOICE has previously supported an Australian *Artificial Intelligence Act* in the Safe and Responsible AI in Australia consultation. With appropriate legislative design to avoid duplication with other laws, an *Artificial Intelligence Act* would centralise and simplify the regulations that businesses are required to follow and the rights that consumers can assert. A centralised Act would also help successive governments streamline changes to AI legislation as laws require adaptation to future developments in technology. Whole-of-economy legislation is also closely aligned with the approaches in Canada and its proposed *Artificial Intelligence and Data Act* and the European Union and its *Artificial Intelligence Act*.

CHOICE believes whole-of-economy legislation can deliver the same benefits of framework legislation while avoiding the potential limitations raised in the paper, such as gaps across regimes and regulatory enforcement. Whole-of-economy legislation is also more analogous to other approaches in industry regulation and consumer safety in Australia (e.g. the *Therapeutic Goods Act 1989*, *Civil Aviation Act 1988*, *Insurance Act 1973*, and *Gene Technology Act 2000*), as well as other proposed approaches to AI regulation abroad (e.g. Canada, the European Union). Unlike the *Regulatory Powers (Standard Provisions) Act 2014* that brought clarity and consistency to existing yet divergent laws, the new technologies, markets, and challenges of AI require new laws and enforcement strategies. CHOICE believes that whole-of-economy legislation is more likely to cohere all elements of AI regulations – such as creating consistency between AI guardrails and principles with enforcement and compliance – than framework legislation which may create unnecessary complexity or gaps.

Domain-specific reforms (option 1) are also likely to be insufficient and inefficient in regulating AI. As outlined in the proposal paper, domain-specific reforms are likely to create inconsistencies between markets, and regulators will likely have varying priorities. For consumers facing new technologies that include unforeseeable harms, baseline *ex ante* guardrails, irrespective of economic domains, are more likely to prevent harms occurring. It will also ensure that harmful uses of AI do not occur in the gaps between domains, particularly as new enterprises form that may have ambiguous domains. CHOICE also agrees with concerns in the proposals paper that domain-specific reforms may be time-consuming – delays in legislation may create inconsistent implementations of the mandatory guardrails, and leave consumers protected in some markets but not others as the legislative process unfolds.

CHOICE also recommends that the Federal Government establish a standalone AI Commissioner to enforce compliance with the mandatory guardrails and the AI Act more generally. Given the complexity of AI and the complexity of harm experienced by consumers, a specialist and well-resourced regulator is necessary. This will future-proof Australia from emerging AI harms as the technology evolves and its impact on our community increases. The AI Commissioner could also be empowered with appropriate rule-making or advisory powers to ensure the risk-based framework and mandatory guardrails are kept fit for purpose. The proposed mandatory guardrails and any explanatory guidelines should be updated with reference to the regulator to ensure organisations are aware of their accountability and transparency obligations to the AI Commissioner.

CHOICE also recommends that the Federal Government modernise the *Privacy Act 1988* and the Australian Consumer Law (**ACL**) to recognise the growing risks from AI and other emerging technology. Reforming the Privacy Act is critical to any regulatory regime on AI – guardrails on data provenance, for instance, may prove inadequate without substantial reforms such as a fair and reasonable use test that will create baseline obligations on businesses to collect data responsibly. Businesses that train their AI models on user data without express consent pose unacceptable risks to these users; this in turn ethically undermines the entire AI supply chain.

CHOICE strongly disapproves of business models that rely on scraping consumer data without consent, care and compensation for their AI systems[5], and overhauling the *Privacy Act 1988* is the only sufficient way to prevent this breach of consumer privacy. Additionally, prohibiting unfair trading practices and establishing a general safety provision in the ACL will further strengthen protections for consumers from unfair and unsafe products and services using AI beyond whole-of-economy or framework legislation. CHOICE looks forward to the Treasury's review of the ACL in light of developments in AI, as indicated in the Budget and proposals paper.

[5] Taylor (2024), "Meta's AI is scraping users' photos and posts. Europeans can opt out, but Australians cannot", *The Guardian*, https://www.theguardian.com/technology/article/2024/sep/11/meta-ai-post-scraping-security-opt-out-privacy-laws.

Consumers who have been harmed by AI need access to redress. The Federal Government should also establish a digital ombudsman to handle complaints, including complaints related to AI systems. Digital harms often overlap multiple areas of harm, or may be ambiguous. If a consumer has a complaint that has arisen in a digital context, it may be easier for them to identify the digital ombudsman as a place to seek redress than to identify whether it involves a breach of privacy, human rights or consumer protection law. The digital ombudsman should take responsibility for referring complaints to other regulators or complaints handling bodies where appropriate.

Recommendations

The Federal Government should:

5. Introduce whole-of-economy legislation to regulate the use of AI in Australia (an AI Act), including a principles-based framework to designate high-risk AI systems, a regulatory framework to assign practices as unacceptable risks, and mandatory guardrails;
6. Establish a standalone AI Commissioner under an AI Act with adequate resources and a full range of regulatory powers, including criminal and civil penalty powers and rule-making powers where appropriate;
7. Reform the *Privacy Act 1988* to better protect an individual's privacy, including the introduction of a fair and reasonable use test;
8. Introduce amendments to the Australian Consumer Law including a prohibition on unfair trade practices and a general safety provision; and
9. Establish a digital ombudsman to handle complaints, including complaints related to AI systems.