



Submission to the Australian Government’s public consultation on introducing mandatory guardrails for artificial intelligence (AI) in High-Risk Settings

The Social Policy Group, auspicing Harmony Alliance, welcomes the opportunity to submit a response to the Department of Industry, Science and Resources’ public consultation on introducing mandatory guardrails for AI in high-risk settings. This submission draws on the insights from our consultations with migrant and refugee women, who expressed deep concerns about the rapid deployment and take up of AI in Australia and globally, and its potential to exacerbate existing inequalities.

In this submission, we address several of the consultation questions provided by the Department across:

- **Defining high-risk AI**
- **Guardrails ensuring testing, transparency and accountability of AI**
- **Regulatory options to mandate guardrails**

Introduction

Harmony Alliance: Migrant and Refugee Women for Change is one of the six National Women’s Alliances funded by the Australian Government to promote the views of all Australian women and ensure their voices are heard in decision-making processes. Harmony Alliance’s purpose is to provide a national, inclusive, and informed voice on the multiplicity of issues impacting the experiences and outcomes of migrant and refugee women, and to enable opportunities for women from migrant and refugee backgrounds to directly engage in driving positive change.

The Harmony Alliance membership comprises over 180 organisations¹ and individuals representing and/or working for the advancement and inclusion of migrant and refugee

¹ Harmony Alliance (2023) *Submission in response to International Gender Equality Strategy*



women. We acknowledge the diversity of experiences of women from migrant and refugee backgrounds and recognise the inherent value of each person, of all backgrounds, genders, ages, abilities, social standings, sexual orientations, or religions. We promote the principles of dignity, equality, autonomy, non-discrimination, and mutual respect.

This submission draws on the online consultations held with Harmony Alliance members in late September 2024. These consultations included multiple online sessions with members representing organisations and individuals.

Prior to the consultation, Harmony Alliance developed a consultation paper outlining the Department of Industry's consultation process. This paper provided a summary of what artificial intelligence is, the AI mandatory guardrails, defining high-risk settings, potential dangers of limited regulation, and case studies illustrating the issues migrant and refugee women may face with AI in real-world scenarios. The consultation paper aimed to break down this complex and highly technical subject matter to ensure representative engagement from members in AI regulation, which is likely to disproportionately impact them. Our members are more likely to be affected by the deployment of AI, which will reinforce and exacerbate existing inequalities in Australia. Additionally, they are less likely to have their voices and concerns heard and addressed due to the highly technical and complicated nature of the current approach.

Harmony Alliance members support the Government's rapid response to much needed regulation of AI in Australia but would caution that further acknowledgement and response to the trade-off's that occur is needed. Balancing speed, quality, and inclusivity often involves trade-offs, and prioritising rapid delivery can sometimes compromise the thoroughness and inclusivity of the process.



Key Themes from Consultations with Harmony Alliance members

- **Data bias:** Diverse data input is essential to prevent AI from reinforcing societal biases present in training data.
- **Underrepresentation:** There are concerns that data from culturally and linguistically diverse (CALD) backgrounds is being integrated into AI systems at a slower rate, leading to slower learning and potential bias in AI outcomes.
 - For example, in medical settings, the use of AI with unrepresentative data could result in misdiagnoses, such as missing heart attacks symptoms in women or creating disparities in treatment for women and minority groups.
- **Education and awareness of regulatory approach:** There is an opportunity to refine and simplify information about AI technology and regulation to make it more accessible, and a need to empower impacted individuals affected by AI outcomes to challenge them.
- **Inclusive AI development and policy making:** Ensuring that diverse teams develop AI and that a range of voices are represented to promote a fair and equitable consultation process.
- **Accountability in AI compliance:** Challenges arise in verifying AI adherence to guardrails. Who verifies what is high-risk and why is there no category that denies AI usage when the risks cannot be mitigated by the guardrails?
- **Human oversight:** AI should be used as a tool, with human oversight remaining central to decision-making. AI should assist, but not make the final decision, ensuring transparency and avoiding reliance on "black box" processes that lack understanding.



Defining high-risk AI

Question 1: Do the proposed principles adequately capture high-risk AI? Are there any principles we should add or remove?

Please identify any:

- **low risk use cases that are unintentionally captured**
- **categories of uses that should be treated separately, such as uses for defence or national security purposes.**

Throughout our consultations, Harmony Alliance members consistently raised concerns regarding the complexity of AI deployment, particularly in critical areas such as healthcare, immigration, child protection, employment, and the justice system. Our members expressed with unanimous agreement that the definition of high-risk AI settings in the *Proposals Paper*² is inadequate and lacks sufficient clarity. This definition appears to have been crafted predominantly by subject matter experts, potentially neglecting practical considerations and the broader implications for affected stakeholders.

For marginalised communities, including migrant and refugee women, there is a significant fear that AI will mirror the biases that are present in our society today and perpetuate existing biases leading to unfair outcomes, particularly because AI systems often rely on historical data that reflect societal inequalities. Clearer and more comprehensive guidelines are needed to ensure a robust understanding and application of these settings. Accordingly, Harmony Alliance urges the Department of Industry, Science and Resources to define high-risk AI settings explicitly and

² Australian Government, Department of Industry, Science and Resources (2024). *Proposals paper for introducing mandatory guardrails for AI in high-risk settings*



ensure that AI applications in these environments undergo heightened scrutiny and regulation.

Question 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?

- **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?**
- **If you prefer a principles-based approach, what should we address in guidance to give the greatest clarity?**

The principles outlined in the *Proposals Paper* address key concerns such as human rights, health, legal status, and collective rights, which are critical considerations. However, our Harmony Alliance members indicated that the broadness of the proposed principles leave room for interpretation, potentially leading to inconsistent application across sectors.

A major issue identified in our consultations is the lack of understanding and engagement with AI systems, especially among marginalised groups. This educational gap increases fears that AI will be used in ways that disproportionately harm disadvantaged communities, particularly in high-risk areas like healthcare, employment, and justice, where migrant and refugee women are most vulnerable.

The principles should include clearer examples and guidance to help AI developers and users recognise high-risk systems, reducing ambiguity and ensuring consistent regulation.



Harmony Alliance members expressed a strong preference for Australia to adopt a regulatory model similar to the European Union’s AI Act, which sets clear restrictions on high-risk AI applications and mandates transparency and fairness in AI systems.

“If international developers want this market and our money, they need to adapt their product to our regulations, not the other way around.”

Given the feedback from our consultations, Harmony Alliance strongly support the European Union's AI Act as current best practice. The EU’s risk-based approach to AI regulation categorises AI systems into four levels of risk: unacceptable risk, high-risk, limited risk, and minimal or no risk. This structure is designed to protect citizens by applying different levels of regulation based on the potential harm an AI system could cause.

These systems are subject to strict obligations, such as risk assessments, high-quality datasets, traceability, human oversight, and strong security measures. We also highly value that in the EU’s regulation AI systems or products are required to undergo a third-party conformity of assessment.

Strict obligations help ensure these systems do not reinforce existing biases or create discriminatory outcomes. A similar categorisation in Australia’s regulation could benefit vulnerable groups by:

- Protecting them from discrimination in high-risk settings
- Providing accountability and oversight
- Providing transparency
- Flexibility for emerging risks



Question 4: Are there high-risk use cases that government should consider banning in its regulatory response (for example, where there is an unacceptable level of risk)? If so, how should we define these?

Yes, there are high-risk cases where the Government should consider banning in its regulatory response due to the concerning risks they pose. Specifically, the cases identified in our consultations were:

- **Employment:** AI systems used in recruitment and hiring should be banned if they show evidence of preferencing or discriminating against individuals based on names, ethnic backgrounds, or other protected characteristics. The risk of reinforcing biases in employment is too high, particularly for vulnerable groups. Migrant and refugee women may face discrimination if the AI is programmed to prefer candidates with local work experience or Western-sounding names. Similarly, consideration should be given to proximity bias, which has been shown to occur when all gender and ethnic markers are removed such as hobbies, sports, educational awards and institutions.
- **Justice system:** AI in predictive policing and decision-making, such as sentencing, should be banned due to its potential to disproportionately target and harm marginalised communities based on biased data. These systems risk reinforcing systemic inequalities, eroding trust, and undermining fairness and impartiality in the legal process. Decisions that impact people's lives and freedoms must prioritise human judgment and oversight to ensure justice is applied equitably.
- **Child protection:** AI should be banned from decisions in child protection services, where the stakes are incredibly high and biases or incorrect risk assessments could lead to devastating, life-altering consequences for families.



The complexity of these cases demands human empathy, contextual understanding, and critical thinking that AI simply cannot replicate. Relying on automated systems in such critical areas risks disproportionately impacting vulnerable families, amplifying existing inequalities, and misjudging situations that require the attention and expertise of trained professionals. Decisions around the safety and well-being of children must always involve human oversight to ensure fairness, accuracy, and accountability.

- **Immigration:** AI should be banned from processing immigration applications, as it might fail to accurately recognise differences in how information, such as employment history, is recorded in other countries compared to local standards. Similar to the case raised in child protection, a wrongful rejection of an immigration application may result in life-altering consequences.
- **Health care:** AI systems used in healthcare settings should either be banned or strictly limited and must always involve human oversight. The risks associated with inherent biases in medical diagnostics, particularly regarding gender and racial differences in health conditions, could lead to misdiagnoses and subpar treatment or recommendations. Bias in AI systems could cost lives, especially in critical areas like intake and triage, where important decisions are made. One member raised concerns that AI deployment could reverse gains made by migrant and refugee communities in reproductive health and healthcare access. These communities have worked hard to secure better healthcare access, and biased AI systems risk reversing that progress, disproportionately affecting vulnerable populations who are already at a disadvantage in receiving adequate care.



These systems should be defined as high-risk due to their potential to violate fundamental rights, perpetuate discrimination, and cause significant harm.

Guardrails ensuring testing, transparency and accountability of AI

Question 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?

Not entirely. While Harmony Alliance members supported that the mandatory guardrails allow people impacted by AI to challenge its use or outcomes, they identified several gaps:

- **Expand beyond high-risk settings:**

"Why can't guardrails apply to broader contexts, not just high-risk settings?"

- **Stricter governance mechanisms:**

"Governance is not strong enough, it's just words there... How do we know organisations are actively looking for bias? What real governance is in place to ensure that?"

- **More detailed transparency:** AI systems should be required to disclose when and how it is used in decision-making, especially in employment and immigration. Additionally, our members report that women speaking out against gender-based violence or criticising foreign governments on social media are facing account bans without fair assessment, the right of reply or human intervention. AI-driven moderation often silences women's voices while



amplifying men's, even when promoting violence or coercion. This lack of transparency and accountability in AI decisions further marginalises women in the public sphere.

- **Higher penalties for non-compliance:** Penalties must be substantial enough to act as a strong deterrent, ensuring organisations take compliance seriously. Weak penalties risk encouraging negligence, while stronger consequences will drive accountability and ensure adherence to regulations.
- **Introduce an 'unacceptable' category:** High-risk uses of AI, such as in employment decisions or the justice system, should be outright prohibited due to their potential for serious harm and bias. Certain applications pose too great a risk to fairness and human rights and should be classified as unacceptable to ensure they are never implemented.

"We should push for a category of straight-up 'no' for AI in certain areas, especially in the justice system and employment... some applications of AI are just too high-risk and discriminatory."

"I personally think AI should not be used for employment decisions at all, no matter what governance or guardrails you put in place... it's too easy for biases to creep in."

Question 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?



The proposed mandatory guardrails do not distribute responsibility across the AI supply chain or throughout the AI lifecycle appropriately. A key concern raised by Harmony Alliance members is the heavy reliance on developers, many of whom may not be based in Australia, to adhere to these guardrails in high-risk situations. Concerns were raised as to whether developers can determine if AI was designed with bias. As one member pointed out,

"If we're relying on developers that largely won't be in Australia to adhere to these guardrails in high-risk situations... I've got some real queries as to whether or not we're going to end up with a more just system."

This suggests the current guardrails may not hold developers, deployers, or other actors equally accountable, especially when those involved in AI development may not be subject to Australian regulations or oversight. There is a need for stronger mechanisms that distribute responsibility across the entire AI supply chain, ensuring compliance from development through to deployment, regardless of the origin of AI tool development.

Regulatory options to mandate guardrails

Question 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?

Australia is currently lagging in AI regulation, presenting significant challenges. While Harmony Alliance acknowledges the need for swift regulatory action, we emphasise that compromising comprehensiveness could have severe



consequences, particularly for minority groups including migrant and refugee women. We propose implementing a phased based approach to regulation, incorporating a second stage of consultation could ensure a more diverse and representative perspective is captured, and building in a legislative review mechanism.

A second stage of consultation would provide Government with an opportunity to refine and simplify information regarding AI technology and regulations, making it more accessible to diverse communities. Consequently, when these communities are invited to contribute ideas and insights — as they should be — they will be better prepared and informed, ensuring a fair and equitable consultation process. Additionally, a built-in legislative review should be introduced, as the rapid pace of AI development is outpacing current regulatory frameworks. This ongoing review would allow regulations to evolve and adapt, ensuring they remain effective and responsive to emerging AI risks.

Recommendations from Harmony Alliance members:

1. Call for robust governance and stringent regulation of AI technologies

Members stress the urgent need for a comprehensive governance framework that prioritises the protection of human rights and the prevention of harm. AI technologies should be subject to strict regulation that enforces transparency, accountability, and oversight. This governance must ensure that AI is not developed or deployed in ways that reinforce biases, infringe on privacy, or exacerbate inequalities, especially for marginalised communities like migrants and refugee women.

2. Ongoing dialogue and consultation to ensure AI serves the broader good

To avoid reinforcing systemic inequalities, ongoing and meaningful consultation is essential. There should be continuous dialogue with diverse communities—



particularly those most likely to be affected—before laws or regulations around AI are finalised. For example, legislation governing AI in areas like healthcare, employment, or justice should be shaped through inclusive consultations that involve civil society, advocacy groups, and individuals with lived experience to prevent AI from deepening social inequities.

3. Define clearer definitions of high-risk settings and take a cautious approach to AI deployment

Our members advocate for more precise definitions of high-risk settings where AI deployment could have life-altering consequences. Sensitive areas such as immigration, healthcare, child protection, and justice require a cautious, human-centred approach to AI. The potential for AI to misclassify individuals or make biased decisions in these areas could lead to irreversible harm, such as wrongful deportations, denied access to healthcare or misdiagnosis, or unfair sentencing. A rigorous, risk-based framework must be adopted to limit AI's role in these critical sectors.

4. Strong preference for adopting the EU AI Act as best practice

Harmony Alliance members strongly encourage the Australian Government to follow the EU AI Act as a model for best practice in regulating AI. The EU's framework classifies AI applications based on their level of risk and specifically identifies areas where AI should be banned—such as in social scoring and mass surveillance—due to their potential to harm fundamental rights. By adopting similar measures, Australia can ensure that AI is deployed ethically and responsibly, preventing the use of AI in high-risk settings that could harm vulnerable populations.

5. Invest in educating marginalised communities on AI and empowering them to challenge outcomes

Education and awareness around AI technology and regulations must be prioritised. Many marginalised communities—especially women, migrants, and refugees—are



at heightened risk of being negatively impacted by AI. It is crucial to provide accessible, culturally appropriate education on how AI affects everyday life, and how individuals can challenge unjust AI outcomes. This should include not only general AI literacy but also resources to help people understand their rights and pathways to dispute or appeal AI-based decisions, whether in employment, healthcare, or social services. Empowering these communities is vital to ensuring they are not left behind in the age of AI.

By implementing these recommendations, Australia can ensure that AI technologies are developed and deployed in a way that upholds fairness, protects vulnerable populations, and strengthens equity in society.