MEAA Submission

Mandatory Guardrails for AI in High-Risk Settings

The Media, Entertainment and Arts Alliance (MEAA) welcomes the opportunity to make a submission to the Department of Industry, Science, and Resources' consultation process for its proposal paper on the introduction of mandatory guardrails for AI in high-risk settings.

This submission engages with questions raised in the Department's proposal paper. We have opted to respond to questions 1-4 and 6 in section 1 of our submission, 8-9 in section 2, and 13-16 in section 3.

MEAA is the largest and most established union and industry advocate for workers in the media, film, music, and arts industries, with a history going back more than 110 years. We endorse the submission of the Australian Council of Trade Unions (ACTU), of which we are an affiliate, and seek to provide additional feedback specific to the creative and media industries.

Introduction

The emergence of generative Artificial Intelligence comes with significant potential risks for Australia's creative industries. MEAA's recent member survey showed that a majority of members (56%) are extremely concerned about the rise of AI, whereas only one in fifty members (2%) are not at all concerned about the rise of AI.

If left unchecked, it is conceivable that the increased use of AI tools could lead to a significant loss of jobs and the degradation of conditions for Australia's media and creative sector workers. AI also poses unique threats in its potential to digitally replicate a performers' voice or likeness, as well as in its ability to replicate a creators' technical and/or creative style. This has the capacity to alienate that persons' unique style, likeness, or voice from them without compensation, credit, or consent.

The basis for the development of these tools is the mass ingestion of creative and intellectual works by AI machines. Creatives, writers, and journalists from Australia and around the globe have had their work scraped, largely without any compensation or consent, and used to train AI. In this sense, the theft of creatives' and journalists' work is directly related to the development of the very tools designed to replace them.

These risks are not just theoretical. MEAA has already seen cases where performers' voices have been reproduced without their consent, while First Nations artists have had their styles replicated in AI-generated copycat works without permission or disclosure to buyers in the marketplace.

It's not just AI's impact on workers that are a concern. The Australian public needs to be able to trust that the news they read, the film and television they watch, and the music they listen to has been produced by artists and journalists whose work has not been compromised by AI.

For these reasons, MEAA holds that it is critical that Australia adopts a comprehensive, economy-wide AI Act. This AI Act must be underpinned by a specialised and well-funded independent regulator to ensure compliance.

Last, as a general comment, MEAA is concerned about the application of the guardrails to AI developers who have developed models abroad – outside of the Australian jurisdiction for the purposes of AI regulation. It is critical that the Government ensure that developers acting abroad are subject to the same requirements for transparency, the legal collection of data, and labelling as those operating inside Australia's borders.

Section 1: Proposed definition of high-risk AI

Developing a comprehensive and clear definition of high-risk AI is crucial for the success of the Departments' proposed mandatory guardrails approach. In general, MEAA endorses the definition of high-risk developed by the Department. However, we have a number of key concerns about the definition as it currently stands. First, MEAA is concerned that the stated principles do not provide sufficient clarity and certainty regarding the definition of high-risk in relation to a) work, and b) Australia's creative and media sector.

First, we endorse the ACTU's view that any AI that impacts the nature of work, the conditions of workers, or the means through which work is allocated and/or managed and distributed should be considered inherently high risk. This would remove ambiguities in the definition and make it explicit that the guardrails apply not only in the case of consumers and end-users but in the case of workers also.

Recommendation 1: That a further proposed principle be included: 'The risk of adverse effects on workers in regards to the employment relationship, the allocation of work (including through digital platforms) work processes and in all matters pertaining to rights at work and rights while conducting work.'

In addition, MEAA is concerned about the lack of attention given to Australia's cultural and media landscape in the proposed definition of high-risk. This is for two reasons. The first is the critical importance of these sectors for the health of Australia's democracy and for our unique cultural identity. Australia's storytellers, artists, actors, dancers, musicians, and technicians form the heart of our country's culture, and the health of our democracy and civil society is reliant on the commitment and tenacity of our journalists who deliver public interest journalism that scrutinises power, informs the public and provides a bulwark against the proliferation of disinformation and misinformation.

The second reason is that generative AI poses significant and particular risks to the creative and media sectors which fundamentally threaten the sustainability of these sectors. Generative AI has the potential to dislocate and devalue the work of artists, creators, and journalists. If left unchecked, it is conceivable that the increased use of AI tools in the media and creative industries could lead to a loss of jobs and the degradation of conditions in creative and journalistic work. Furthermore, while AI advocates spruik the technology's capacity to improve the productivity of workers and "democratise" the production of creative content, the reality is that businesses are utilising AI tools to replace their workforce with automation products. This means that the jobs of those working in the production of digital content – including actors, musicians, set and costume designers, and voice artists – are under significant threat. With that threat comes a broader threat to the production of distinctly Australian cultural work and journalism.

It is critical to note the impact of generative AI on First Nations art and culture. MEAA is aware that generative AI is being used to train AI software used to mimic the artistic style of Australia's First Nations artists and performers. There are now numerous reports of AI-generated 'Indigenous art' being commodified and sold online, including on Adobe, Etsy, and eBay.[1] These are adding to the competition Indigenous artists already face from the fake 'Indigenous art' market, threatening to further undermine their livelihoods and the sustainability of the sector.

Recommendation 2: That a further proposed principle be included: 'the risk of adverse impacts to Australia's cultural and media landscape.'

In addition, MEAA holds that there are several ways in which the proposed principles could better empower to First Nations people to protect culture, communities and country from exploitation. First, protocols for using Aboriginal and Torres Strait Islander peoples Intellectual and Cultural Property (ICIP) must be embedded into AI legislative protections and Australian copyright law. It is critical that these protocols must be incorporated into existing legislation to ensure proper enforcement.

These protocols must establish Aboriginal and Torres Strait Islander peoples' sovereign rights of ownership of data, lands, language, culture, art, stories, music, song, dance, voice, image, likeness, body, movement, and intellectual and creative output. In addition, if AI is engaged to depict the voice, image, movement or likeness of Aboriginal and Torres Strait Islander ancestors and/or deceased persons, all companies and entities must follow the applicable Aboriginal and Torres Strait Islander protocols and cultural Lore of those ancestors and/or deceased persons.

Recommendation 3: that Indigenous Cultural and Intellectual Property (ICIP) be embedded into law.

Our view is that all general-purpose AI (GPAI) models should be considered inherently high risk because they can be applied to a wide range of applications, some not possible to foresee, to a potentially very broad audience at very high speeds.

Recommendation 4: categorise all general-purpose AI (GPAI) as high risk.

MEAA also holds that certain kinds of AI applications should be banned, including that which is used for exploitative and deceptive purposes. MEAA endorses the ACTU's list of proposed unacceptably high-risk cases of AI use that should be, in our view, banned. These include cases where they involve automated decision making over the employment relationship that substantially affects the rights and interests of an employee, where they are applied into employment and work processes without the meaningful knowledge and consent of workers and where they are used for the purpose of intrusive monitoring or surveillance of workers without their knowledge or consent. MEAA would also add to this list any use of an AI system to create a digital replica using the biometric data of an individual without their informed consent.

Recommendation 5: that the government ban unacceptably high-risk AI.

Section 2: Proposed Guardrails

The proposed guardrails provide an important first step in managing the risks of AI in high-risk settings. In particular, MEAA strongly supports efforts to mandate transparency in the collection of training data, to ensure that datasets are obtained legally, and in efforts to ensure a system of labelling or watermarking for AI-generated outputs. These guardrails speak to key concerns raised by our members.

<u>Guardrail 2</u>

Workers must be involved in formulating the risk assessment and management process within an organisation, and it is also imperative that meaningful consultation and consent be obtained from workers prior to decisions about implementing AI systems in the workplace being made. We believe that there are circumstances where workers may identify or be exposed to risks from the use of AI in their work which may conflict with the priorities of their employer and that these can be avoided by ensuring meaningful consultation and engagement with workers.

The media and creative industries bring with them a range of unique challenges when considering and mitigating risk. Film and TV productions for example, are large undertakings that involve a diverse range of teams of experts responsible for aspects of the overall production. Therefore, there are many possible different uses of AI within a single production with the potential to expose workers to different kinds of risks. For instance, Generative AI tools could be used in post-production and editing, visual effects, photography, script writing, score writing, sound production and of course, in the performance of actors, dancers and musicians.

There are clear risks in the possible use of AI tools which have been developed using illegally obtained training data, and there are risks to workers (regardless of them being a performer, crew, technician or other member of staff) having their biometric data obtained with consent or compensation for the use in the production or used in other productions or even

commodified and on-sold. These are only two of the many possible risks associated with employing AI in film and TV productions.

Journalism is also exposed to a number of significant risks specific to the industry. Public interest journalism requires journalists to be able to verify and protect sources and interrogate information provided to them to ensure they are factual. Additionally, journalists who are members of the MEAA are bound by the Journalist Code of Ethics[2] which binds them to a set of professional standards in the production of public interest journalism.

The potential use of AI systems in newsrooms whether they are 'off the shelf' tools or 'fine-tuned' in-house AI systems present significant risks to the ability for journalists to uphold their professional standards. Risks to future work and appropriate compensation also arise when media companies wish to on-sell or license the work of journalists for use by AI developers.

Guardrail 3

MEAA holds that all AI models, not just those which are deemed high risk, should be subject to transparency requirements around the collection and use of training data. This is because copyright violation is an inherent risk of all AI models.

Second, MEAA holds that these datasets should be made publicly available – or at a minimum, that the copyright holder and original author are notified about any use of their work for the purposes of training. This is necessary because of the profound ways in which the theft of such work for the purposes of training is impacting creative workers, writers, and journalists.

Third, MEAA strongly believes that this guardrail should be applied retrospectively. We know that huge amounts of data have already been ingested for the purposes of AI training. For example, we know that the Books3 database (estimated to contain 196,640 books[3]) and the LAION-5B database (containing an estimated 5.85 billion images[4]) have been ingested by a range of major LLMs. Creators deserve to know whether their work has been used to build these models.

Last, we echo the ACTU's concerns about existing gaps in privacy and workplace laws, as well as the use of far-reaching exceptions which allow employers to use employees' data without restriction following the initial point of collection. We reiterate the ACTU's call to broaden this guardrail to further stipulate that this data must be legally obtained, with its use subject to informed consent. This should include clear avenues for redress where proper authorisation has not been obtained.

Guardrail 6

MEAA is concerned that a 'best efforts' approach will lead to sub-optimal outcomes when it comes to transparency, ensuring trust, and ease of use for end-users/consumers. Further to this, MEAA believes that, at a minimum, attempts should be made to standardise labelling

systems across industries and be incorporated into existing classification systems and practices where possible.

Furthermore, MEAA is strongly concerned about the narrow application of the requirement for labelling to high-risk AI only. We believe that labelling should apply to all content generated by AI, including those applications not designated 'high-risk'. This is necessary to ensure that, as a general rule, audiences can be informed about whether content – for example, a film, television show, media article, or musical composition – has been generated by AI.

Last, we echo the ACTU's concerns about employers potentially failing to recognise employees as end users – as well as the general public. We agree with the ACTU's recommendation that the government provide further clarification that employers have a responsibility to fulfil the obligations under this guardrail in relation to employees.

Section 3: Preferred Regulatory Options to Mandate Guardrails

In order to meet the community's expectations of government to ensure the safety and security of AI systems, it is necessary for there to be a comprehensive and coordinated approach across government and the economy that protects workers, citizens, and consumers, provides clarity and simplicity for users and businesses of all sizes, and is enforceable.

Our approach is guided by these criteria:

- That the need for regulation is urgent,
- That the regulatory response must ensure the strongest possible protections for workers citizens and consumers,
- That the regulatory response must be comprehensive and work seamlessly with other economy-wide regimes such as workplace, copyright and privacy laws,
- That the regulatory regime must be clear and simple,
- That the regulatory regime must be enforceable, and
- That the regulatory regime must allow scope for alignment with similar regimes from international jurisdictions.

On this basis, we endorse Option 3: an economy-wide AI Act. This option is the only one that will deliver on the ambition to deliver safe and responsible AI in Australia in line with the criteria mentioned above. Much like the implicit rationale for other comprehensive regulatory regimes in Australia, the case for a comprehensive approach to regulating AI through an Act is clear. AI-enabled applications are already installed on devices sitting in most homes and workplaces. AI systems are already being used across business, academia and government, and plans are being put in place to use AI tools to replace thousands of jobs in the coming years.[5] We also know that AI is turbocharging the spread of disinformation and propaganda contributing to the decline of trust in democratic institutions and undermining the work of journalists.[6]

We are concerned that options one and two would force government to take a piecemeal approach to amending laws and regulations and this would greatly inhibit the government in implementing regulations with the urgency required to match the speed of the rollout of AI systems in Australia. It is our view that government must act with urgency because of the rate at which AI is developing and setting new norms in the absence of comprehensive regulation. Not only would these approaches be slower to enact than a comprehensive AI Act, but it would also make Australia's AI regulatory regime complex, difficult to understand and prone to unnecessary duplication. This would hinder the successful implementation of AI regulation, it would leave workers, citizens and consumers unsure if their rights and it would be a disincentive for small and medium sized enterprises to positively embrace regulation.

It is also our view that options one and two contain inherent limitations which would impair the ability for government to bring the guardrails into effect. We concur with the analysis of DISR in the proposals paper[7] that both options one and two will retain gaps across regimes, will leave industries siloed, will limit regulation to the scope and powers of current regulatory arrangements and will be largely unenforceable. Option three presents the best opportunity for a coordinated approach to regulation. Some of the proposed guardrails go hand in hand and would be much less effective if implemented in isolation or with different effect industry by industry.

Option three also presents the best environment for ensuring that all related regulatory regimes such as copyright and privacy are linked together as seamlessly as possible. As we have stated in other sections of this submission, we believe that there is a gap in the proposed guardrails when it comes to ensuring protections for workers. We endorse the position of the ACTU on the need for a comprehensive review of workplace laws, particularly the Fair Work Act, to ensure they remain fit for purpose in the context of the rapid adoption of AI systems, and the potential risks of AI.

Such piecemeal approaches would also add complexity to Australia's regulatory landscape, encouraging litigation and adding to the costs of doing business. It would enable well-resourced companies to game the system to suit their interests over the interests of the community. Such scenarios have precedents overseas where big tech companies – who also count amongst the most significant developers and deployers of AI – have been accused of flouting competition laws and other regulations in the European Union and United States[8].

Examples like this highlight the need to ensure the enforceability of these regulations, and only a comprehensive AI act can achieve that. It is our view that an AI regulator must be instituted and properly funded to oversee this regime and to make sure that there is sufficient oversight. Self-regulation cannot to be depended on to ensure that comprehensive safeguards are enacted and upheld. Considering the enormity of the scale of change that AI is predicted to have on our society, we cannot afford to implement a regulatory regime that is limited unambitious.

Recommendation 6: The government adopt Option 3: an economy-wide AI Act as the preferred option for mandating the guardrails.

References

[1] Cam Wilson (2024) 'AI is producing 'fake' Indigenous art trained on real artists' work without permission' Crikey https://www.crikey.com.au/2024/01/19/artificial-intelligence-fake-indigenous-art-stock-images
[2] *MEAA Journalist Code of Ethics* https://www.meaa.org/meaa-media/code-of-ethics/
[3] AIAAIC (2024) *Books3 AI training dataset,* https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/books3-ai-training-dataset
[4] LAION (2022) *LAION-5B: A new era of open large-scale multi-modal datasets*, https://laion.ai/blog/laion-5b/
[5] Paul Smith (2024), 'CBA explores replacing local call centre staff with AI', *AFR* https://www.afr.com/technology/cba-explores-replacing-local-call-centre-staff-with-ai-20240913-p5kafb
[6] Tate Ryan-Mosley (2023), 'How generative AI is boosting the spread of disinformation and propaganda', *MIT Technology* Review https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/
[7] *Safe and responsible AI in Australia: Proposals paper for introducing mandatory guardrails for AI in high-risk settings*, DISR, pp 48-50
[8] Anna Cooban (2024) 'EU could hit Apple with a huge fine after accusing it of breaking new tech rules', *CNN*, https://edition.cnn.com/2024/06/24/business/apple-eu-competition-rules/index.html