

**Submission  
No 7**

## **INQUIRY INTO ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES**

**Name:** Mr Yanni Kyriacos

**Date Received:** 5 October 2023

---

Thank you for the opportunity to make a submission and share my views about how AI might impact NSW, including the risks and challenges it presents.

As a member of the NSW community, I've been thinking a lot about AI and how it is already affecting my work and my community, and the impacts it could have in the future.

When I read NSW's AI policy and assurance framework, I appreciated that the NSW government has been ahead of other jurisdictions in understanding the pace of change of AI and its transformative nature. I want government to make the right calls in a timely way, and stay ahead of emerging issues and risks from AI. The recommendations I provide below hope to contribute to that.

Terms of reference 1.(k) asked after measures other jurisdictions, both international and domestic, are adopting in regard to the adaption to and regulation of AI.

One of the global developments that is most promising is the creation of "national laboratories" to enable technical tests on AI models, provide technical reports and provide ongoing monitoring and assurance. Singapore has established the AI Verify Foundation, the EU has created a Centre for Algorithmic Transparency, the UK has a Foundation Model Taskforce and the Tony Blair Institute for Global Change has proposed that the UK create "Sentinel" with a similar goal.

Without a similar lab in Australia or in the region, deploying trusted and safe AI in Australia might become impossible as capability and capacity increases.

NSW is well positioned to collaborate with other jurisdictions to create or support a national laboratory for AI safety, modelled on international best practices.

This approach is exciting because there's international best practice to follow, and NSW could use the laboratory to ensure the AI products it uses are safe and can be subject to effective ongoing monitoring and assurance.

Perhaps even more importantly, we are already seeing various kinds of dangerous and risky AIs. If we had a trusted national laboratory, it could assess AI products before they go to market. In the same way we don't let cars on our roads without them going through safety tests, a lab like this would allow us to block AIs until they've passed appropriate safety tests.

NSW's AI assurance framework (page 23) includes consideration of possible harms of an AI system, and frames that in terms of the residual consequence after mitigations are applied. NSW should be commended for going further and considering secondary or cumulative harms (page 24). This is often neglected, and second-order effects can often be much more significant than primary effects.

That said, there is room for NSW to improve what it considers a secondary harm. Specifically, the bulk of the future risk of AI systems could turn on the values and priorities of the developers of frontier models - including factors like how committed they are to safe and ethical AI systems and how much they are investing in AI safety research. A future where commercial incentives encourage AI developers to set aside safety considerations is a much worse future than one in which they prioritise it.

Phrased another way, if NSW signs contracts with AI developers who do not take AI safety seriously, the secondary harm could be very significant.

In light of that, NSW should update its guidance regarding secondary harms to include the implications of engaging any particular AI developer – including the reputational benefit for that developer and the implications of it receiving further funding. If NSW is reaching an agreement with a frontier model developer, its main considerations should relate to the demonstrated commitment of that developer to long-term AI safety. NSW should only deal with AI developers with strong commitments to AI ethics and AI safety – including demonstrated investments in and commitments of computing resources to longer-term AI safety considerations.

I appreciate the chance to share my views about how NSW could continue to refine its approaches to AI. I think a growing expert consensus is that AI is the most important issue of our time. It could go really well and usher in a new era of abundance. It could go really badly and lead to the destruction of the things we care about most. With the stakes so high, we need to take this issue seriously at all levels of government and society.