Submission
No 17

INQUIRY INTO ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES

Name: Mr Christopher Leong

Date Received: 17 October 2023

NSW Enquiry Submission

I'm pleased to see that NSW is taking forward-leaning approach to AI regulation and striving to remain at the forefront of this rapidly evolving and impactful technology.

Often, discussions in Australia might not fully capture the monumental implications of AI – including just how bad it could be if we don't keep our eyes on the ball. For this reason, I think our governments have a real responsibility here to be aware of any possible risk AI may bring and take action to mitigate these risks.

I'd like to bring your attention to a recent survey by Roy Morgan. This found that one-in-five Australians believe AI could present a risk of human extinction in the next 20 years, and 57% thought that AI would create more problems than it solves. Survey participants weren't just worried about job losses, but also felt that there was a need for regulation, were worried about how it could be misused and the potential for unknown unknowns.

I've been struck by immense number of AI experts who have expressed concern about these risks, including through the Center for AI Safety's Statement on AI Risk and the Future of Life Insitute's open letter calling for a Pause on Giant AI experiments. In a survey of experts in the field, 48% of respondents gave at least a 10% chance of an extremely bad outcome from AI.

I beleieve NSW's first priority should be doing everything that we can to mitigate the risk of the worst possible harms of AI. The sooner that we're in a position where we can feel confident that we've avoided the worst possible outcomes, the earlier we can get started on maximising the ways that things could go well.

Biosecurity:

The Combating Terrorism Centre at West Point has sounded the alarm over the potential hazards of growing public access to sophisticated AI and biotechnology:

> *It is likely that terrorist organizations are monitoring these developments closely and that the probability of a biological attack with an engineered agent is steadily increasing.*

Soon, the means to design, produce, and disseminate harmful and unprecedented pathogens might become commonplace, potentially within the grasp of the general populace. In 2021, Professor Brian Schmidt AC, Vice-Chancellor of the Australian National University, expressed his gravest concerns regarding this widespread access to biotechnology:

> *"[The ANU] is one of the first places to be able to do CRISPR… in the next 5 to 10 years there's every reason to believe that you're going to be able to use literal mass-market printers to do what you want, and it won't be just hijacking an existing disease, it will be the ability to create new diseases... [T]hat is what really scares me. That is my number one fear."*

Following Professor Schmidt's observations, the demand for synthetic DNA, specialized reagents, and AI tools that enable the general public to harness this technology has been on the rise. If things go as they are, then by 2025, the tools required to engineer a severe bioweapon might be "become accessible to school students and others with low skills and resources," notes MIT's Professor Kevin Esvelt.

Australia needs to address these issues on both the bio front and the AI front.

One major component fueling bioterrorism possibilities is the production and importation of synthetic DNA. Australia has already implemented a system to oversee this. It's imperative for the NSW Government to promptly liaise with the Commonwealth, urging them to mandate screening protocols for all DNA export orders to Australia. Concurrently, the NSW Government should mandate that its affiliated research entities only collaborate with labs that enforce thorough screening measures for every order.

While private entities like the International Gene Synthesis Consortium (IGSC) have initiated self-regulation with standard protocols, still, one out of every five orders goes unchecked. Firms such as IBBIS and secureDNA provide the technology for screenings, emphasizing the pressing need for governmental intervention.

While regulating these pivotal components is crucial and can provide a buffer, advanced AI might still empower malicious entities to circumvent such rules, unless the AI itself is safeguarded. Bearing this in mind, it's crucial for NSW regulators to establish and uphold stringent safety benchmarks for cutting-edge AI implementations within NSW. Clear guidelines should be laid out for AI developers and users: AIs possessing the potential for dual-use applications that could usher in massive threats shouldn't be tolerated. To gain entry to our economic landscape, AI products must first address and resolve AI safety concerns, ensuring their harmlessness before being introduced to the market.

Governance:

NSW deserves recognition for initiating a transparent AI assurance framework. However, there's a pressing need for this to be developed further.

One of my main concerns, listed on page 13 of the assurance framework, is that “the key factor that determines risk is how the AI system is used”. I don’t believe that this is true and when it comes to open source software, it could be dangerous.

While the system's application is undoubtedly a consideration, NSW should promptly shift its focus to also evaluate inherent risks within the AI system itself. Specifically, the NSW Risk Assessment should identify features that can affect an AI's risk profile. For instance, more opaque "black box" systems or those with a higher propensity for errors, increased autonomy, misalignment with human values, or that are at the frontier of AI development should be deemed higher risk. Conversely, systems characterized by transparency, interpretability, evaluated by leading AI safety labs, and consistent, controllable performance should be viewed as less of risk. Open sourcing frontier models should be considered especially risky because dangerous capabilities can’t be fixed after release.

Today's unpredictable systems demonstrate that they can be harmful, even in seemingly benign settings like entertainment. As AI technologies advance and operate with greater autonomy, the inherent safety of the AI itself becomes paramount, surpassing the intentions behind its use.

On that note, I'd like to bring to NSW's notice the Responsible Scaling Policy recently introduced by Anthropic. This policy outlines "AI Safety Levels (ASLs)", categorizing models from ASL-1 through to ASL-4+. This concept mirrors the Biosafety Levels (BSL) in place for labs handling infectious diseases, a practice familiar to NSW. By integrating this approach into its existing risk assessment framework, NSW could substantially improve its risk assessment.

Conclusion:

In summary, AI technology is swiftly advancing and reshaping our society. I trust that NSW will consistently reassess its AI strategies, adjust based on new findings, and inspire other Australian governments to follow suit. While the future trajectory of AI's impact remains uncertain—whether profoundly positive or negative—we do recognize the imperative for alert governments that monitor these developments and stand prepared to respond.