To the Select Committee on Adopting Artificial Intelligence (AI),

My name is Damien Granet. I'm a current student at the Australian National University, concerned with governance, institutional stability, and catastrophic risks – including those presented by AI. Over the past year, I have also been organised multiple events from AI researchers and ethicists as part of my role as President of the Effective Altruism club at the university. All this to say that I'm writing as a concerned individual, but also as someone who sees how concerned my community is about AI, and their frustration with the lack of government response.

AI is undoubtedly one of the major players poised to shape our collective future. Amongst both technical experts (and their political science and philosophy colleagues), there is a remarkable consensus that unregulated advancements in AI could lead to numerous unforeseen consequences, ranging from biased algorithms to disruptions in critical systems. As such, it is essential that AI safety is properly considered. I also think that this is research that is not internalised by the market, and should be supported by the government.

First and foremost, I would like to advocate for the establishment of a National AI Safety Institute in Australia. This is a model that has already been adopted in the US, UK, Canada, and Japan. The Australian Government has indicated an interest in regulating AI. However, given the current pace of AI development, waiting for regulation may not be the best course of action. Instead, creating an Australian AI Safety Institute would allow us to begin addressing safety concerns immediately.

To be effective, this safety institute would need to fill a few key roles:

- It would need to be able to evaluate advanced AI systems. This means a sophisticated program of red-teaming, analysing characteristics and capabilities, and considering their potential implications. We need an early warnings system that allows us to stop dangerous AIs becoming available to an entire population. This isn't just because we don't trust humans to use AIs safely (though this is of course a concern). AIs are capable of doing dangerous things by accident, particularly if they are poorly designed.
- It would also need to spearhead technical AI research (which basically means figuring out how to make machine learning algorithms which don't tend towards dangerous, inhuman or discriminatory outputs). Importantly, it would need to provide support for safety research: how do we make these models safer. It should not support capabilities (making AI more powerful) research because this will create a distinct conflict of interest. Companies will (and have) been neglectful about public safety in the pursuit of profits.
- Finally, an AI Safety Institute would need to take part in global efforts (e.g. the UK and US AI safety labs have recently made it clear that they want to collaborate with other AI safety labs). AI Safety will require international cooperation, and Australia

should capitalise on its position as a world research leader to further global progress in this space.

Furthermore, the institute would also help prepare Australia for any future regulatory regime, and the technical capability and capacity required to administer that legislation. Overall, the establishment of such an institute aligns with our current approaches to AI governance and serves as a crucial step towards safeguarding our national interest.

As a conclusion to this submission, I want to cover two risks which I think are particularly urgent.

The first major risk that is critical to address are biosecurity threats posed by more advanced AI systems. Recent studies have shown that AI can be used to create lethal molecules and bioweapons. In a famous recent example (which I'm sure you'll hear about a lot) Collaborations Pharmaceuticals reported that an AI designed 40,000 lethal molecules in less than six hours. Similarly, a study on Large Language Models (LLMs) revealed that these could be used to manufacture synthetic DNA for bioweapons.

In response to these risks, President Biden issued an Executive Order in 2023, addressing the biosecurity risks posed by frontier AI models. However, Australia has yet to follow the US's lead in this regard. Therefore, I urge the Senate Inquiry to seek evidence from various departments to understand their awareness of and action plans for these biosecurity risks.

The second risk is that cybersecurity risks may become significantly more sophisticated and *easy* with AI. A Google experiment found that ChatGPT, a language model, was able to pass an interview for a high-paying engineering position. This level of AI capability, if unregulated, could result in an increase in cyber attacks, which already cost Australia $29 billion per year.

Recent studies have shown that LLMs can autonomously hack websites and generate malware (you no longer need to actually understand what you're doing), highlighting the need for improved safety measures in AI systems. If left unaddressed, we risk living in a world where all services accessed remotely become highly vulnerable, or we find ourselves in a perpetual cyber arms race.

I think its pretty clear that the risks associated with unregulated AI advancements are too high to be left to chance. I would urge the Australian government to prioritise the establishment of a National AI Safety Institute and to address the biosecurity and cyber threats posed by AI. By doing so, we can ensure that AI serves to improve our future, rather than jeopardise it.

Regards,

Damien Granet