

26/04/2024

Dear Select Committee on Adopting Artificial Intelligence

My name is Daniel Parris. I work as a sustainability finance professional, focusing particularly on responsible investment and ESG research. I am also the Chair of the Effective Altruism Canberra group.

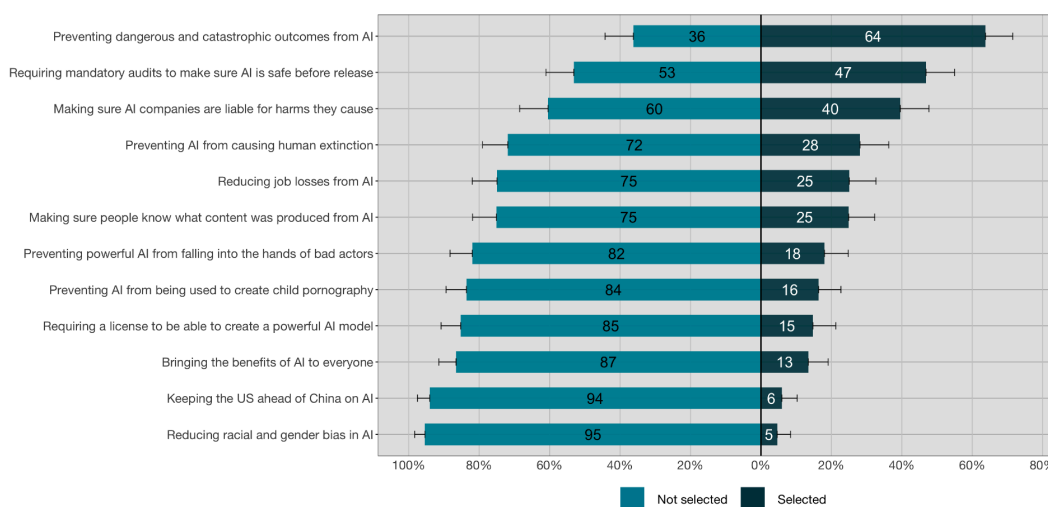
I am writing to you today about my concerns about Artificial Intelligence (AI). I believe AI should be treated as consequential, given its significant potential to shape our society, for better or worse. As one of the biggest risks to our collective future, unchecked advancements in AI could lead to unforeseen consequences, such as biased algorithms or disruptions in critical systems. It is one of the biggest risks we face together.

In personal conversations, friends and family have expressed concerns about the rapidity and safety of AI development. I believe these concerns are shared broadly, with many people apprehensive about the potential misuse of AI in severe cyber-attacks, creating novel biological weapons, or losing control of advanced autonomous AI systems not aligned with human values. World-leading experts, including Turing Prize winners Geoffrey Hinton and Yoshua Bengio, signed a statement that AI could pose catastrophic or existential risks if governments do not take safety seriously. And on average, AI researchers believe there is about a 20% chance of their research leading to catastrophic outcomes.

It's not just my circle of contacts, or international experts, who are worried about these issues. Research shows many Australians have concerns about AI safety. When asked about the topic in recent polling by Ready Research and The University of Queensland, the largest share of respondents emphasised concerns about the possibility of a lack of safety, trustworthiness, or alignment with human values in AI systems. When asked about the Australian Government's focus on different concerns related to AI, respondents suggested the government should focus on preventing dangerous and catastrophic outcomes from AI. Respondents also suggested the government should mandate audits of AI systems to ensure new AI models' safety and hold AI companies accountable for any harm they cause.

**Population estimates for Australian public's AI governance priorities**

What should the Australian Government focus on when it comes to Artificial Intelligence? (Select up to three actions)



Source: <https://www.aigovernance.org.au/survey>

26/04/2024

While the Government is working to drive AI adoption and address immediate problems with AI, I am concerned not enough is being done about larger, more consequential, risks. Indeed, I believe that Australia's current AI governance infrastructure has critical gaps. CSIRO's National AI Centre's primary mission is to drive the industry's adoption of AI in Australia. But, we need to go beyond this industry-focused approach to AI and ensure we are prepared for any significant global economic changes that rapid advancement in AI capabilities could bring about. We must not risk falling behind. So I urge this Committee to recommend that the Government prioritise AI Safety issues alongside its current focus on AI adoption.

I propose two additions that could help improve our current AI governance infrastructure: AI-specific regulation and an AI Safety Institute that is meaningfully separate from efforts focused on accelerating AI capability or adoption. While existing regulators have a role to play, they can only tackle some of the risks that AI presents in their current form. We need a new regulator to address risks at the point of AI creation, not just in its usage. Secondly, an AI Safety Institute should be established to evaluate advanced AI systems, drive foundational AI safety research, and partner nationally and internationally on AI Safety. Separation from commercial interests and government efforts to accelerate AI capability or the adoption of AI by businesses is crucial for maintaining integrity and trust. This would allow us to partner with international peers to advance AI Safety work. We cannot afford to wait. Such an institute should be responsible for evaluating advanced AI systems, driving foundational AI safety research, and partnering nationally and internationally on AI Safety.

In conclusion, AI's rapid and unpredictable progress could significantly harm our society. An AI Safety Institute, along with appropriate AI-specific regulation, could help protect our national interest and contribute to a positive vision of AI in society. I hope the Committee will consider my recommendations and act quickly.

Regards,

Daniel Parris