Good Ancestors Policy

Select Committee on Adopting Artificial Intelligence (AI)

Adopting AI

Friday, 16 August 2024

Canberra

Questions on notice for Mr Sadler (Good Ancestors Policy) and Mr Pour (Harmony Intelligence)

Question: The Chair asked Mr Sadler and Mr Pour to think on notice about the concept of incorporating workplace participation in the overarching regulatory model they discussed.

Answer: Good Ancestors Policy and Harmony Intelligence are aware of efforts to bring together governments, employers and workers' organisations in the context of best managing the growing involvement of AI in workplaces and the challenges and opportunities of the transition period.

While this is an important period, government also need to lay the groundwork for what happens after the transition. Future AI models are likely to be highly capable and could meet or exceed human capabilities in many domains. The chart below, from *Artificial Intelligence Governance & Safety Canada*, shows how adopting generative AI could lead to significant job displacement in the immediate term, while Agentic AI or Artificial General Intelligence would cause mass job losses and associated disruptions. The diagram shows that **governments might have a limited window to prepare for rapid change**.

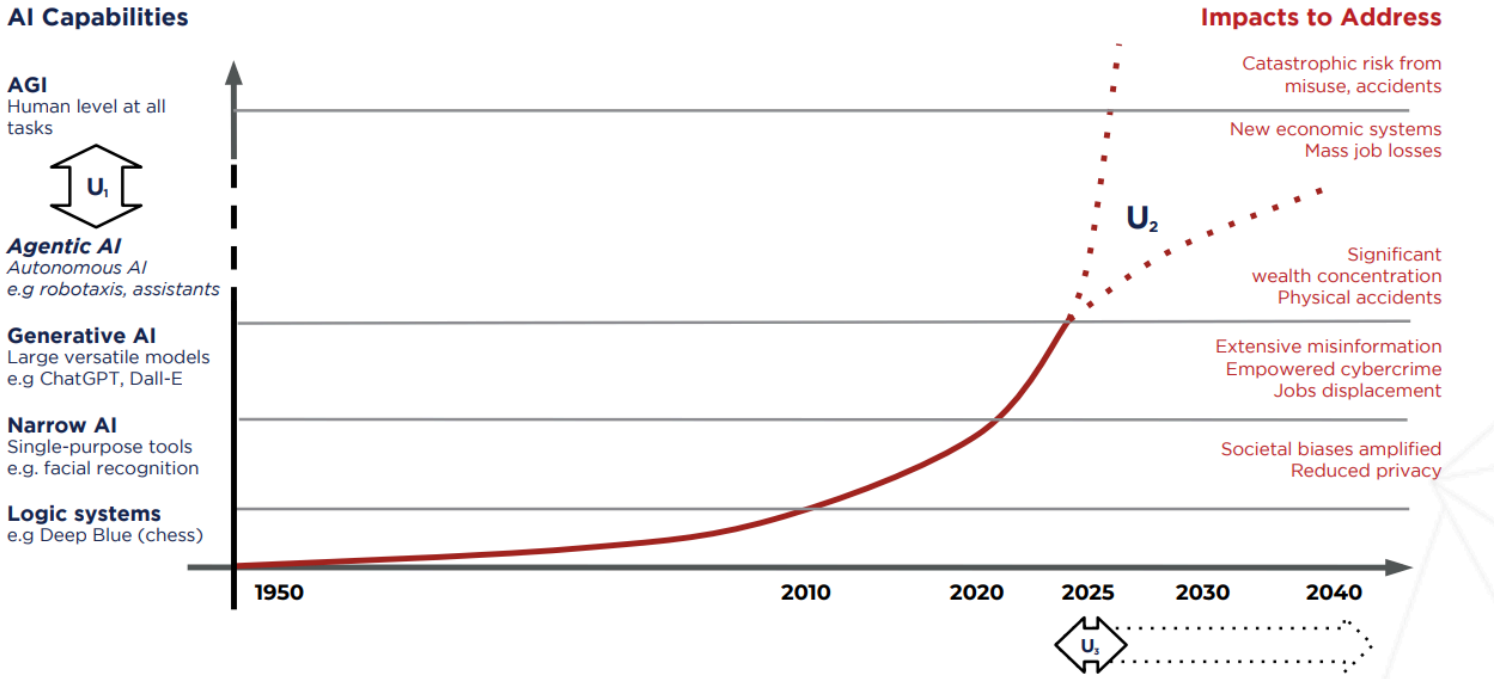

Chart 1: High-level sketch of AI capabilities and impacts over time

Key areas of uncertainty are U_1: how many technical breakthroughs are needed to reach AGI, and U_2: whether the current exponential rate of progress will continue. This determines U_3: how much time governments have to prepare for the potential catastrophic risks and major social upheaval.

One approach that has been broadly discussed, including in the contexts of ethical and responsible AI, is the importance of "humans in the loop" or other kinds of "effective oversight and control" of AI systems. This is seen as attractive, including because it connects AI systems to the risks and opportunities we are more familiar with managing (we feel more comfortable dealing with human-bias than AI-bias) and it seemingly creates a pathway for jobs linked with AI systems (humans who oversee AI systems). Policymakers might imagine this creates an enduring solution for human jobs alongside AI.

While this may seem intuitive, it faces practical challenges. Currently, humans provide oversight in situations where they have value to add to a process. This might be an experienced employee overseeing a junior employee or a worker "overseeing" a process. These relationships involve the person doing the oversight function having skills or capability that the person or process being oversighted benefits from. These can be fulfilling jobs and create career pathways.

However, instead, imagine a job overseeing an artificial intelligence that works hundreds of times faster than you, ingests orders of magnitude more data than you, and never sleeps. **A job "being in the loop" with such an AI would not be meaningful work.**

To ensure a future with meaningful work and human wellbeing, governments need to invest in substantive research into how to interpret AI models so we have meaningful alternatives to being "in the loop". We need to work to make models transparent and explainable. While trust and accountability are critical - we need to generate them in realistic ways that avoid creating a new class of meaningless work (i.e. jobs purporting to oversee superhuman AI models).

To achieve this, we need to increase funding for research into issues like AI interpretability and start building effective governance institutions that will allow for meaningful work alongside the AI models of tomorrow.