Australian Government
Chief Scientist

DR CATHY FOLEY AO PSM

SCIENCE AND TECHNOLOGY IN SOCIETY *FORUM*'S 12TH ACADEMY OF ENGINEERING PRESIDENTS MEETING

Building trust in AI

Monday 02 October 2023

Kyoto, Japan (virtual)

Thank you for inviting me to address this meeting again this year.

I'm pleased to see the focus on artificial intelligence in a forum such as this. As I'm sure we agree, AI presents immense opportunity to drive economic growth and productivity across our economies. As a scientist, I am a strong proponent of the power of AI to accelerate discovery, and to open new avenues for science.

Machine learning, combined with the coming revolution in quantum technologies, will transform medicine and diagnostics. We've already seen the potential in RNA science. AI will help us develop new materials for the energy transition. AI will become ever more integrated into the way we interrogate data and conduct research. It's a powerful tool, and these are capabilities that I look forward to seeing.

However, I'm sure we also agree on the need for guardrails. Misinformation and digital manipulation are a serious threat to social cohesion – even to our shared understanding of reality. Our estimation of the risk posed by the current generation of AI technologies – and the next generation – may vary from person to person, but all of us agree there are risks, and we must work together as an international community to mitigate them.

I value Japan's role in this work, including your efforts to bring leaders together to maintain momentum and focus. Important discussions will begin in Kyoto in a few days. The G7 nations are working together on the Hiroshima Process. These are welcome initiatives.

In Australia, close attention is being paid to the risks and opportunities of AI, to ensure we chart the best way forward. As you may be aware, the Australian Government released a Safe and Responsible AI discussion paper in June, and is now considering the responses from that consultation. It's clear from the sheer number of submissions that Australians are very engaged with this topic.

Advances in large language models, most obviously ChatGPT, have really concentrated people's minds on some of the ways that machine learning and algorithmic interactions are already impacting our lives. People everywhere are

concerned for the future of their jobs. There's concern about what AI means for the future of the arts. There's concern about whether our democratic systems are robust enough to withstand the misinformation and disinformation, deep fakes, and malign outcomes that are made possible by AI.

I'm a scientist and I couldn't be a more enthusiastic advocate for the power of science to change our lives. Nevertheless, I see the arts as our most cherished human capability, and humanity's gift to the future. So I share the concerns, but I'm also convinced of the benefits, and of the importance of following the science.

As I said, AI, machine learning, and the digital revolution that is underway now will, together with quantum technologies, transform capability. They have the potential to solve some of the world's most challenging problems, and to accelerate discovery. The challenge for us is that these benefits will only be realised if we can build public trust, so that digital technologies can be used to best effect – whether it's in science and research, in business and industry, or in government.

In my view, the keys to building trust are ensuring transparency, protecting privacy, and embedding diversity. Diversity is about ensuring the digital world reflects who we really are. For reasons that are clear, skewed datasets and biased algorithms are not the basis for good outcomes. Privacy is important because it is built into the way we learn. A private life allows us to make mistakes and to reflect, to learn and grow over a lifetime of experience. This must be protected in an era characterised by indelible digital footprints. And transparency drives accountability. It allows us to see below the surface, understand the inputs and understand the processes that are driving machine intelligence.

No one knows how ChatGPT works, but the inputs are known, and the instructions are known. With knowledge about the algorithmic inputs, the system design, we can then contextualise outcomes and use judgement in the actions we take. If we can get the guardrails right around those questions of diversity, privacy, and transparency, we will go a long way towards building trust and confidence. In Australia, we're well placed for this challenge, given our strengths in standards, governance and inclusivity.

I want to finish with a reflection on the purpose of science – and this is a comment that applies equally to all endeavours in science and technology. The rapid development of advanced technologies is not an end in itself. We're not pursuing new technology for technology's sake, or science for the sake of science.

The purpose is to advance transformative technologies for the good of the planet and society. Keeping our eyes on this prize, will give us the best chance at success.