# What are we doing now to ensure that Australia is recognised as a global leader in responsible AI, and what else should we be doing now and into the future?

**Dr Ian Opperman** FTSE
**NSW Government's Chief Data Scientist, Department of Customer Service**

Responsible AI, ethical AI or trustworthy AI refers to the design, development, deployment and use of artificial intelligence tools and systems such that they align with ethical principles and values, respect human rights, and ensure fairness, transparency, accountability and safety throughout the AI lifecycle.

(A definition co-piloted by ChatGPT v3.5)

## Thoughts on responsible AI

Large language models (LLMs) and generative artificial intelligence (AI) have redrawn the frontier of what we thought AI could do. Ask any current generation of AI tools to whip up a short biography of your favourite artist, and you will get a succinct summary. Ask it to write a song in the style of this same artist, and you will get something impressive.

What has changed is the way AI works and the size of the datasets used to train it. Generative AI is trained to 'focus' and is training on datasets of unimaginable sizes to mere mortals, literally trillions of examples.

This unsupervised training occasionally leads to some surprises. While AI may provide a supposedly factual response to your query, some results may refer to 'real-world' sources that simply do not exist. Similarly, a request to generate an image from a verbal description may lead to something a little more Salvador Dali—like than you may have expected. This scaled-up version of the age-old adage of 'garbage in, garbage out' leads to the modern twist of 'garbage in, sometimes hallucination out'.

Nonetheless, the responses from the latest generation AI tools are pretty impressive, even if they need to be fact-checked.

So, what does this mean for people thinking of regulating AI or putting AI policies in place?

## AI is different from other technologies

Some of the concerns raised about AI are similar to those relevant to other technologies when first introduced. When addressing concerns with the use of AI, if you instead replaced 'AI' with 'quantum', 'laser', 'computer' or even 'calculator', some of the same concerns arise about appropriate use, safeguards, fairness, and contestability. Yet AI is different in that it allows systems, processes and decisions to operate or occur much faster and on a much grander scale. AI is thus an accelerant and an amplifier; it can also 'adapt', meaning that what we design at the beginning is not how it will operate over time. These three characteristics are referred to as the three A's.

Before developing new rules, existing regulation and policy should be tested to see if it stands up to the potential harms and challenges arising from those three A's. If your AI also 'generates', 'synthesises' or 'translates', then a few more stress tests are needed as this goes well beyond what you can expect from your desktop calculator.

## AI is no longer explainable

Except in the most trivial cases, the depth and complexity of the neural networks (number of layers and number of weights), alongside the incomprehensibly large training datasets, mean that we have little chance of describing and understanding how an output was derived, even if it were possible to unpick all of the levels and the impact of each training element. Any explanation would be largely meaningless.

For any decision that matters, there must always be an empowered, capable, responsible human in the loop ultimately making that decision. That 'human in the loop' cannot just be a rubber stamp extension of the AI-driven process.

Any regulation must not refer to the technology: There have been numerous calls to ban, 'pause' or regulate the use of AI. ChatGPT, one of the first user-friendly AI-powered chatbots built on an LLM, hit the scene in November 2022, arriving in our lives with a bang, and with the accelerator planted to the floor.

Every day new frontiers in AI capability seem to be announced. Buckle up when quantum supercharges AI. The orders of magnitude difference between the pace at which technology moves and that at which regulation adapts means the closer regulation gets to the technology, the sooner it is out of date. Regulation must stay principles-based and outcomes-focused. It must remain focused on preventing harm, enabling appropriate human-based judgement (even if AI-assisted), dealing with contestability, and remediation.

## Blanket bans will not work

Comprehensive banning of student use of generative AI has been announced by various departments of education around the world (including in Australia). The intention of these bans is to prevent students from using AI to generate responses to assignments or exams and then claiming it as their own work.

Such bans are extremely unlikely to be effective simply because those who have not been banned from using AI have a potential advantage (real or perceived) by accessing powerful tools or networks. The popularity of AI platforms also means that workarounds are likely to be actively explored, including the use of these platforms in environments outside the restrictions. The bans arguably address symptoms rather than root causes. In the case of education, rethinking how learning is assessed will be central to establishing the appropriate use of generative AI.

## We need to think long-term

AI is a technology that has been with us for a long time. It is suddenly renewed, and we are looking at it with little understanding of the long-term consequences. By analogy, electricity was the wonder of the 19th century. From an initial scientific curiosity, electricity is now embedded everywhere and has profoundly changed the world.

AI is likely to have as profound an impact as electricity. As AI becomes embedded in devices, tools and systems, it becomes increasingly invisible to us. Our expectations of these devices, tools and systems are that they are 'smarter': aligned to the tasks at hand; able to interpret what we mean rather than what we ask for; and able to improve over time. We do not expect to be manipulated or harmed by the tools we use.

## The NSW AI Assurance Framework version 1.0

The NSW Government developed an AI strategy and AI Ethics policy in 2020. The state government then developed, tested and mandated the use of an AI assurance framework.[46] The framework is NSW's attempt to connect the principles of its strategy and policy to the specific issues associated with the use of AI. The framework is a self-assessment tool supported by an expert AI review committee that is tasked to review AI projects with an estimated total cost of $5 million or those for which certain risk thresholds have been identified during the framework's self-assessment process.

The framework assists project teams using AI to analyse and document a project's specific AI risks. It also helps teams to implement risk mitigation strategies and establish clear governance and accountability measures. The framework will get a boost with version 2.0 planned for release in late 2023.

## Summing up

For AI to be used responsibly, much more is required than the application of simple checklists. It requires oversight and that we remain vigilant to the negative consequences of AI use, individually, for our society, and for the environment.

Our focus must be on ensuring a safe and level playing field for users of AI as it continues to amplify, accelerate and adapt. That focus also has to stand the test of time.

*DR IAN OPPERMANN is the NSW Government's Chief Data Scientist working within the Department of Customer Service, and Industry Professor at University of Technology Sydney (UTS). Ian has 30 years experience in the ICT sector and has led organizations with more than 300 people, delivering products and outcomes that have impacted hundreds of millions of people globally. He has held senior management roles in Europe and Australia as Director for Radio Access Performance at Nokia, Global Head of Sales Partnering (network software) at Nokia Siemens Networks, and then Divisional Chief and Flagship Director at CSIRO.*

# Essays

## SECTION 1: INTRODUCTION

**What is responsible AI anyway?**
Professor Jon Whittle – Director, CSIRO's Data61

**10 examples of AI that are here now and have been embraced by the general public**
Stela Solar – Director, National Artificial Intelligence Centre

## SECTION 2: WHAT DO WE NEED TO BE TALKING ABOUT?

**A unique opportunity for Australia: bridging the divide between fundamental AI research and usable, embodied AI**
Professor Michael Milford FTSE – ARC Laureate Fellow, Joint Director QUT Centre for Robotics

**Responsible AI means keeping humans in the loop: what are other social implications of the mainstream adoption of this technology?**
Associate Professor Carolyn Semmler School of Psychology, Faculty of Health and Medical Sciences, The University of Adelaide and Lana Tikhomirov – Australian Institute for Machine Learning (AIML), The University of Adelaide

**AI is changing the way people work: how do we skill our future workforce to ensure these new jobs stay on shore?**
Professor Katrina Falkner FTSE – Executive Dean of the Faculty of Sciences, Engineering and Technology, The University of Adelaide

**Responsible data management: a precursor to responsible AI**
Dr Rocky Chen, Associate Professor Gianluca Demartini, Professor Guido Zuccon, and Professor Shazia Sadiq FTSE – School of Computer Science and Electrical Engineering, The University of Queensland

**Open the pod bay doors please, HAL**
Andrew Dettmer – National President, Australian Manufacturing Workers Union

**Innovation needs to create value: how do we tool universities to remain relevant to industry needs?**
Professor Simon Lucey – Director, Australian Institute for Machine Learning, The University of Adelaide

**An AI-literate community will be essential for the continuity of social democracy**
Kylie Walker – Chief Executive Officer, Australian Academy of Technological Sciences and Engineering

## SECTION 3: WHAT ARE THE NEXT STEPS?

**What are the limits of current AI, and what opportunities does this create for Australian research?**
Professor Anton van den Hengel FTSE – Director, Centre for Augmented Reasoning, Australian Institute for Machine Learning, The University of Adelaide

**Australia's unfair advantage in the new global wave of AI innovation**
Professor Mary-Anne Williams FTSE – Michael J Crouch, Chair for Innovation, UNSW Business School

**The $1 billion dollar question: What should Australia's responsible AI future look like?**
Kingston AI Group

**What are we doing now to ensure that Australia is recognised as a global leader in responsible AI, and what else should we be doing now and into the future?**
Dr Ian Opperman FTSE – NSW Government's Chief Data Scientist, Department of Customer Service

For acronyms, abbreviations and endnotes please see the composite document with all the essays.

# Responsible AI

## Your questions answered

ACKNOWLEDGEMENTS
The Australian Academy of Technological Sciences and Engineering (ATSE) and the Australian Institute for Machine Learning (AIML) acknowledge the Traditional Owners of the lands on which we meet and work and we pay our respects to Elders past and present. We recognise the deep knowledge and practices embedded in the oldest continuous culture on the planet. Australia's history of engineering, technology and applied science spans more than 60,000 years.

This artefact is produced by ATSE in partnership with AIML. We would like to thank all the experts for their contributions to this edition.

PROJECT TEAM
Eddie Major, Dr Kathy Nicholson, Peter Derbyshire and Suryodeep Mondal

DESIGN AND PRODUCTION
Elizabeth Geddes, Edwyn Shiell and Alexandra Horvat

SUGGESTED CITATION
Responsible AI: Your questions answered. Australian Academy of Technological Sciences and Engineering (ATSE), and the Australian Institute for Machine Learning (AIML) at The University of Adelaide. Canberra, Adelaide 2023

*Cover image: An artist's illustration of artificial intelligence (AI). This image represents the boundaries set in place to secure safe, accountable biotechnology. It was created by artist Khyati Trehan as part of the Visualising AI project launched by Google DeepMind. Source: unsplash*

# Responsible AI

## Your questions answered

Australian Academy of Technological Sciences & Engineering

THE UNIVERSITY of ADELAIDE

Australian Institute for Machine Learning