Responsible AI

Your questions answered

Australian Academy of Technological Sciences & Engineering

THE UNIVERSITY of ADELAIDE

Australian Institute for Machine Learning

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

Contents

This collection of short papers developed by the Australian Academy of Technological Sciences and Engineering (ATSE) and the Australian Institute for Machine Learning (AIML) at The University of Adelaide offers an insight into the world of responsible artificial intelligence and the opportunities this presents to Australia.

Foreword

Elanor Huntington FTSE
ATSE Board Member; Executive Director, Digital, National Facilities & Collections, CSIRO

ARTIFICIAL INTELLIGENCE (AI) is a powerful technology that has the potential to transform our lives for the better. A contributor to this volume, Stela Solar in her article identifies examples of how AI has been embedded into our daily routines, seamlessly enhancing accessibility, productivity and experiences. It is being used to improve healthcare, education, transportation, and many other areas. However, end-use AI product deployment in our daily lives also raises a number of ethical and social concerns. For example, how can we ensure that AI is used fairly and equitably? How can we prevent AI from being used to harm people? This needs to be addressed in the responsible and thoughtful design, development and application of AI technologies.

This collection of short papers provides key insights from experts on 'responsible AI' research. The papers explore some of the roadblocks and challenges ahead and importantly, highlight the need for Australia to capitalise on the promise of responsible AI.

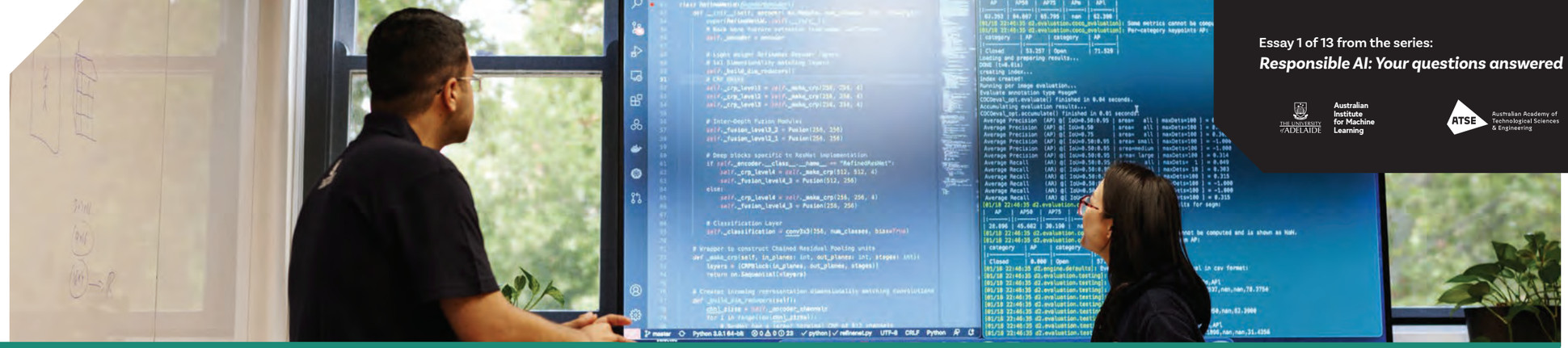
Australia needs to support the growth of a domestic AI ecosystem. As Professor Katrina Falkner FTSE explains in her article, this support must ensure our workforce has the necessary skills for this AI future. Australia must support schools in STEM education, and teachers must be one of our top priorities. This includes providing ongoing, sustained teacher professional learning to allow our teachers to improve their STEM expertise and build awareness of AI that they can use to train the future workforce.

Australia already has a number of pockets of success in AI applications. It is time for Australia to lean into the investment of our winners in AI and ensure that AI is used for good and that it benefits all Australians. For example, we have a strong track record in developing AI-powered solutions for healthcare, agriculture and mining. However, we need to do more to support the growth of our broader AI ecosystem and double down on our winners to create globally competitive products.

AI is a rapidly growing field, and we need to make sure that Australia is at the forefront of innovation. It has the potential to create significant economic benefits, and we need to ensure that Australia captures its fair share of these benefits. Australian-developed AI has the potential to address some of our most pressing social challenges, such as climate change and inequality. By investing in AI, we can ensure that Australia remains competitive in this globally competitive field and also achieve productivity gains that boost our economy, and improve the lives of all Australians.

This ATSE / AIML report is a great starting point to inform the importance of Responsible AI to better understand AI's growing role in our lives. I commend the authors and look forward to the amazing responsible AI developments to come that will, we must all hope, reshape the world for a more peaceful, stable and prosperous future.

***PROFESSOR ELANOR HUNTINGTON** FTSE is a board member of the Australian Academy of Technological Sciences & engineering (ATSE) and Executive Director, Digital, National Facilities & Collections, CSIRO. She leads the data-focused research, development and digital capability of CSIRO, and is a member of the CSIRO Executive. She has stewardship of a range of business lines including Data61, Space and Astronomy, and National Collections and Marine Infrastructure. Prof Huntington also leads major national infrastructure provided by CSIRO on behalf of the scientific community to assist with the delivery of research.*



Essay 1 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

What is responsible AI anyway?

Professor Jon Whittle
Director, CSIRO's Data61

THE FIELD OF 'responsible AI', which has become increasingly popular in recent months, is intended to put the development and adoption of AI on a sound footing and to ensure that AI has net positive benefits for society.

This is a broad definition, however. This article delves into the details and refines what responsible AI means in practice. In our forthcoming book on the topic,[1] my co-authors and I came up with the following definition:

> Responsible AI is the practice of developing and using AI systems in a way that provides benefits to individuals, groups and the wider society while minimising the risk of negative consequences.

There are many benefits arising from AI. CSIRO's *Everyday AI* podcast[2] presents many examples where AI is benefiting society: Tennis Australia is using AI to assist blind spectators to watch tennis more easily; CSIRO and Google are using AI to help manage invasive species in the Great Barrier Reef; conservationists are using AI to help track biodiversity; at St Vincent's Hospital in Melbourne, AI is reducing the time it takes for women to get the results of breast cancer scans; and AI is also being used as a tool by artists such as the Grammy Award–winning pop duo Yacht, who use AI to help write their music.

This article could go on for pages about all the benefits, but the focus here is on the potential negative consequences of AI and how to minimise them. What are these negative consequences? Unfortunately, there are many.

AI risks can be divided into three categories: risks associated with the **use of the technology**, risks that arise from **how the technology is developed**, and broader **societal risks**.

Risks from AI use

Let's take ChatGPT as an example. Putting aside consideration of how this was developed and its many benefits, two of the well-known issues with ChatGPT are bias and hallucinations:

Bias
It's been known for years that AI systems can suffer from bias. One of the earliest public examples of this was the COMPAS system in the US, which was used to predict reoffender rates at parole boards but was found to discriminate against black people.[3] In the case of ChatGPT, because it is trained on a large proportion of the text available on the internet, and because a lot of that text is biased and discriminatory, ChatGPT can also be biased and discriminatory.
In recent times, significant effort has been put towards trying to reduce bias in AI systems — for example, there are guardrails in ChatGPT that avoid much of the potential bias and discrimination. So one might ask, is ChatGPT an example of responsible AI? It's hard to say — the time and effort put into developing guardrails is evidence of a responsible approach. On the other hand, ChatGPT still suffers from some bias issues, which doesn't look like responsible AI. This example illustrates that responsibility in relation to AI isn't necessarily based on a binary between entirely responsible and irresponsible. Rather, responsibility lies on a spectrum, much like AI itself.

Factual inaccuracies (or hallucinations)
AI systems are never 100% accurate. This is the nature of the technology. Data-driven AI, in particular, applies statistics to look for patterns in data. But statistics, by definition, are not 100% accurate — it's all about probability. So a self-driving car will never perfectly identify obstacles in its path. And an AI-driven movie recommendation system can't always recommend the movie you want to see right now. Of course, AI systems don't need to be 100% accurate to be useful. However, some AI systems need higher degrees of accuracy than others. In the ChatGPT world, these factual inaccuracies have been called 'hallucinations'. Examples range from making up fake citations in academic papers, to providing wrong answers to simple math puzzles, and faking people's bios. So are AI systems that hallucinate irresponsible? Again, it depends on what the AI system is used for. Context matters when it comes to responsible AI.

Risks in AI development

Responsible AI isn't only about the end product. It's also about the *way* the product is designed and developed.

Indeed, one clear principle behind responsible AI is that responsibility must be considered at all stages of a product's lifecycle — from initial concept, all the way through design and implementation, to adoption and use. Responsible AI is like a chain — a weak link means the whole system fails the responsibility test.

In our forthcoming book,[1] we think about this issue in the context of three dimensions: governance, process and product. Each of these are characteristics of *how* an AI system is developed, and the care and rigour applied in each dimension will affect the extent to which the system can be considered responsible.

Governance
If you are an organisation developing or using AI, what governance do you have in place to ensure the AI is responsible? Governance in AI is a huge topic, which can't be fully covered here. But some of the questions organisations should be addressing are: Where is AI used in my organisation? What data is used to train the AI and do we have the right to use that data? Has the data been properly curated to ensure it isn't inherently biased? Have we considered what can go wrong and put in place mitigation strategies, or have we just assumed the AI will work as intended and everything will be okay?

Process
The best governance framework in the world won't save you unless your organisation has rigorous processes in place to monitor whether things are being done responsibly. Are the right kind of people involved in the development of your AI system (i.e., end users, and any relevant external stakeholders)? Have you defined what 'responsible AI' means in your particular context, and is this definition specific enough to be falsifiable? Have you introduced any relevant training for your workforce? Does your culture support responsible AI — is there psychological safety so that people feel free to speak up if needed? Is there a "human in the loop" for any critical decisions that an AI might make?

Product
The final dimension is the AI system — or product — itself. Important considerations here relate to the detailed design of an AI system and ensuring that best design practices are incorporated to ensure responsibility. This could involve, for example, having redundant systems so that a critical system doesn't rely on AI alone, quarantining new AI features until they have been shown to work responsibly in the field, or undertaking continuous real-time testing of an AI product so that responsibility can be monitored.

Broader societal risks

Our first two categories of responsible AI are concerned with AI systems and how they need to be managed at an organisational level. Arguably more important than this, however, is to understand the broader societal impacts of an AI system and whether there are unknown, unanticipated or unintended negative system-level consequences. As a good example, even if ChatGPT were implemented without any biases or hallucinations, the computing machinery needed to train and run ChatGPT for millions of users can have significant negative environmental impacts. Another example of negative consequences concerns those critical minerals that need to be dug out of the ground to build the data centres and mobile phones required to run AI. A third example is the use of low-paid workers in Africa to train ChatGPT, who were asked to label text containing violent, sexist and racist remarks so that ChatGPT could avoid generating such text.[4]

Broader societal risks are the hardest risks to manage. They are often hidden and not talked about. Systems thinking is one way to help understand and manage these hidden risks. Another way is to engage with experts outside the technology disciplines, such as lawyers, social scientists and anthropologists. These experts will typically bring a different lens that can enable hard questions to be asked that otherwise might go unnoticed.

While it is difficult, if not impossible, to guard against all negative unintended consequences, a truly responsible approach to AI development and deployment will involve a rigorous attempt to understand not just the AI system's risks but also the risks arising from how that AI system is used in a broader environmental and social context.

PROFESSOR JON WHITTLE is Director of CSIRO's Data61, the digital and data science arm of Australia's national science agency.

With around 1000 staff and affiliates, Data61 is one of the largest collections of R&D expertise in artificial intelligence and data science in the world. Data61 partners with over 200 industry and government organisations, over 30 universities, and works across vertical sectors in manufacturing, health, agriculture, and the environment.

Prior to joining Data61, Jon was Dean of the Faculty of Information Technology at Monash University.

Responsible AI extends beyond the realm of AI for good initiatives. To me, responsible AI is about the method of designing, deploying and governing AI systems to mitigate unintended consequences while creating value.

Three of the anchor points for responsible AI include ensuring AI systems are compliant with regulations and standards, that they're created and used in alignment with organisational and societal values, and that they're well governed and managed. Australia holds the title of the world's most AI-hesitant nation, and the responsibility to change this outlook lies with those creating and implementing the technology.

This path begins by recognising responsible AI as an indispensable need that not only mitigates risk but also increases citizen trust — especially when co-designed with the customers and communities the system affects.

10 examples of AI that are here now and have been embraced by the general public

Stela Solar
Director, National Artificial Intelligence Centre, CSIRO

IN THE RAPIDLY evolving landscape of technology, artificial intelligence (AI) has quietly become an integral part of our daily lives. Once considered a futuristic technology, AI has been embedded into our daily routines for years, seamlessly enhancing accessibility and experiences.

Beyond daily conveniences, AI is playing a crucial role in solving some of the greatest challenges facing Australia. This includes adapting to a changing climate, protecting unique ecosystems, accelerating drug design, and helping cities and towns run as efficiently as possible.

Before we delve into these inspiring examples of everyday AI, let us remember that AI's potential to improve outcomes for business and community can only be realised through responsible and thoughtful design, development and application. It is the mission of the National Artificial Intelligence Centre (NAIC) to work with industry, academia, government and civil society to guide the creation and adoption of AI towards safe and responsible outcomes.

Below are some of the many ways AI has been embraced by Australians.

1. AI in your pocket

When we unlock our phones using facial recognition or fingerprint scanning, that's AI in action. These technologies use sophisticated algorithms to identify and authenticate users.

And it doesn't stop there — AI enables virtual assistants like Siri and Alexa to answer our questions and set reminders, navigation apps to plan the most efficient routes, and social media platforms to curate personalised content feeds based on what we like or dislike.

2. Let's talk chatbots

Chatbots are powered by AI technologies that allow them to engage in conversations with users. By leveraging natural language processing (NLP) and machine learning, these AI-powered assistants understand user inquiries and respond accordingly. They can answer common questions, are available 24/7, and provide personalised recommendations and solutions.

According to CSIRO's *Australia's AI ecosystem momentum report*, Australian businesses that deploy customer service bots experience an average incremental revenue benefit of $500,000 per project.[5]

3. Streamlined streaming

Video-streaming platforms are a great example of how AI can create a seamless and intuitive experience. Streaming platforms use machine learning algorithms to analyse vast datasets, identify patterns, make predictions, and continually refine their understanding of user preferences for personalised content suggestions. Some streaming services even use NLP algorithms to understand users' reactions to content to identify preferences.

By identifying which genres, themes and formats resonate with specific user segments, AI also enables studios and production houses to make data-driven decisions when developing new content.

4. Next-gen empowerment

Generative AI technologies have sparked our imaginations and engaged people at a scale never before seen with AI. The surge of new products and services underpinned by this technology is a testament to what human creativity can achieve in a short amount of time with new tools and technologies.

Australians are using generative AI to write content, create reports, analyse trends, build websites, generate ideas and create code. For sole traders and small businesses, generative AI is a revolutionary technology that can scale customer engagement and operations and transform time-intensive tasks into quick actions.

5. Game-changing

From testing its AI-powered predictive analytics platform on a handful of vending machines in Newcastle to counting Coca-Cola, Walmart and Red Bull as clients, HIVERY has successfully harnessed the power of AI.

Founded in 2015, this Sydney-based startup's AI solution generates insights from huge datasets to optimise product management and merchandising, empowering retailers to respond quickly to market trends.

SAM, HIVERY's Space Assortment Manager, can develop a planogram (a visual representation of a store's products or services on display) in several minutes, a process that normally requires 150 steps, numerous staff members and six months of work.

6. Preventing potholes

AI-equipped garbage trucks and council vehicles are identifying potholes, cracks and faded lines on our city streets before they become major problems. Since 2022, councils across Queensland, and now NSW, have been using AI-powered camera technology to conduct real-time street condition reporting before scheduling preventative maintenance work.[6]

Brisbane-based company Retina Vision developed the technology behind these initiatives. It aimed to automate the time-consuming and manual aspects of road maintenance, including gathering data, logging it in the system, and generating and closing out work orders.

This AI-powered approach enables maintenance teams to focus on improving roads, and in turn community safety, leading to significant cost reductions by identifying the need for the repairs early.

7. Keeping an eye out

Health in a Virtual Environment (HIVE) is a remote hospital patient monitoring system that has been assisting on-the-ground doctors and nurses since 2021. Powered by AI, HIVE continuously monitors patients who require close medical observation.[7]

Patients' vitals are analysed, including heart rate, blood pressure and oxygen levels. This data is relayed to a team of clinicians at HIVE's command centre at Royal Perth Hospital. If any anomalies arise, the HIVE team promptly alerts the attending healthcare staff to provide medical care, communicating through a two-way audio-visual unit.

This innovative use of AI technology facilitates swift medical intervention and helps ensure patients receive the best possible care.

8. Helping hands

LYRO Robotics, an Australian start-up, has developed an AI-powered solution to help farmers in regional Queensland sort and package fruits and vegetables.

The company's world-leading pattern-packing robot can lift delicate produce like avocados, sweet potatoes and pineapples from conveyor belts and neatly pack them into boxes.

The robot can be fitted into existing operations and installed in less than an hour, helping farmers optimise operating margins, reduce food wastage, increase efficiency and mitigate labour shortages.

9. Caring for Country

In the wetlands of Kakadu, rangers are using AI and Indigenous Knowledge to care for Country. As Kakadu is one of the largest national parks in Australia, it can be challenging to monitor the environmental health of this precious ecosystem and the wildlife.

But using drones to perform aerial surveillance, take thousands of photos and speedily analyse the images is a game changer for rangers, who would normally perform healthy country maintenance in 40°C heat and 60 per cent humidity.

All the drone monitoring, data analysis and reporting happens through the Healthy Country Dashboard developed by CSIRO, Kakadu Rangers and Microsoft.

This combination of Indigenous Knowledge and AI has already generated positive results. In 2019, rangers reported thousands of magpie geese, who are the markers of healthy wetlands, returning to areas that were once choked by weeds.

10. That's a wrap

Did you know AI is behind some of your favourite movie scenes? In 2021, Adelaide-based Rising Sun Pictures and the Australian Institute for Machine Learning (AIML) used AI to create visual effects for Marvel Studios' *Shang-Chi and the Legend of the Ten Rings*.

Using a 'deepfake' method, the team mapped stunt performers' faces with actors' faces, training each character's machine model with over 30,000 facial images of both stunt performers and actors. This enabled them to more efficiently swap the faces of the two, with 51 swaps needed for six key scenes.

Replacing the traditional use of 2D and 3D face-mapping tools, this new method saves time and money for high-intensity action scenes where stunt doubles are needed.

STELA SOLAR is the Director of the National Artificial Intelligence Centre, hosted by CSIRO's Data61. In this role, Stela is focused on building value for Australian people, businesses and the country, through use of artificial intelligence.

Over the past 15 years, Stela has cultivated expertise in capturing new revenue opportunities presented by emerging technologies and business model transformation.

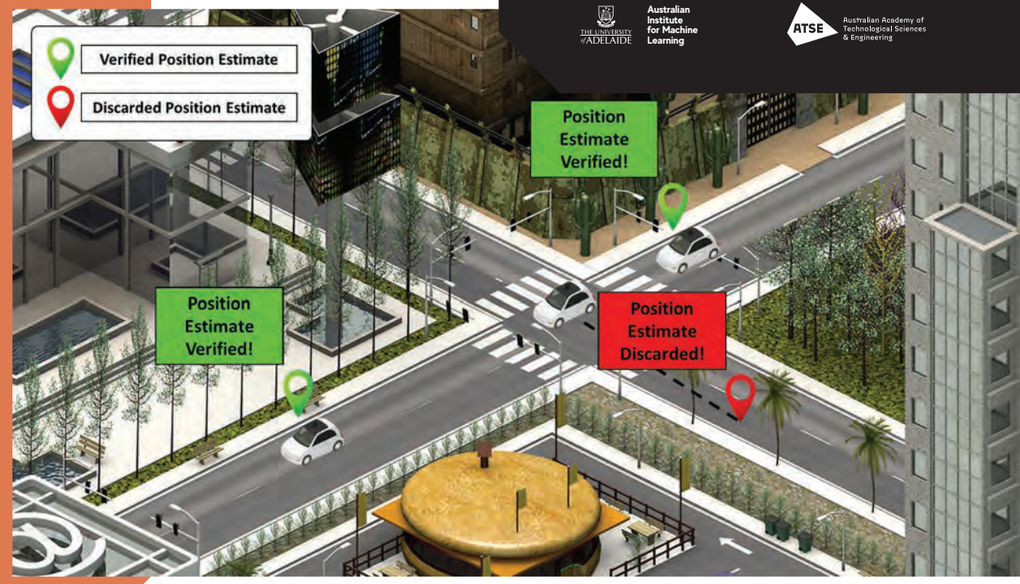
Stela is passionate about removing barriers to positive technology adoption and engagement. She leverages her broad experiences across business development, strategy, ecosystem development, marketing and product management to inform her insights surrounding cross-organisational factors affecting an organisation's ability to capture an advantage.

Essay 3 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

A unique opportunity for Australia: bridging the divide between fundamental AI research and usable, embodied AI

Professor Michael Milford FTSE
Joint Director, QUT Centre for Robotics, Queensland University of Technology



A responsible AI ecosystem is one where everyone — creators, regulators and end users — is accurately informed about the capabilities and limitations of the technology.

A responsible AI ecosystem means decision-makers can take into account the informed opinion of all stakeholders and make the best decisions about developing, deploying or retracting AI-related technologies.

In a responsible AI ecosystem, stakeholders acknowledge and manage the tension between the huge benefits and possible harms of any transformative technology like AI. Responsible AI practitioners do as much due diligence as possible to pre-emptively prevent harm, but also acknowledge, and prepare for, the inevitable need to remediate the unexpected impacts and consequences of even approved AI deployments.

RESEARCH IN AI and related fields like robotics, computer vision and machine learning is advancing at an unprecedented rate. While the popularisation of deep learning around a decade ago accelerated matters, that rate of change has stepped up again with the unlocking of generative AI capabilities, for example, through content generation and chat interfaces like ChatGPT. But amazing progress in domains like content generation masks an underlying grand challenge that will limit the expansion of useful applications of this technology unless major new research initiatives are forged to solve it. In this article, I describe the nature of the challenge and why its significance is only now being recognised, and I highlight how the substantial research needed to address this challenge represents an opportunity for the Australian research community in particular.

That underlying grand challenge is how to design a system that 'knows when it doesn't know'. To take a simple example, imagine an autonomous vehicle navigating a crowded pedestrian- and cyclist-filled city street. Part of its navigation system will rely on the vehicle understanding where it is located on that street: a process known as localisation in the research field, but what might more commonly be called positioning. The vehicle proceeds down the street safely until, at some critical juncture in time, perhaps when approaching a pedestrian crossing, the positioning system glitches and for a second or so, the vehicle imagines it is navigating a very different street, located on the other side of the city. The vehicle makes some navigational decisions as if it were suddenly driving down that other street, chaos ensues and a group of pedestrians end up with serious injuries.

Imagine instead that the algorithms and intelligence powering the positioning system had an introspection capability, so that the vehicle was able to analyse its performance and identify when it is not confident of what it is 'thinking', so to speak. Designed correctly, such a system would identify that when the vehicle positioning estimate jumps across the city to another street, the positioning system is not working as it should. Instead of ploughing on ahead blithely, the autonomous vehicle could engage a fail-safe mechanism— a sudden but safe braking manoeuvre to a stop and/or flagging to a remote operator that the car was in trouble. A potentially fatal situation could instead become a momentary cessation of operations while human operators work out what is happening — a far more desirable outcome from any perspective.

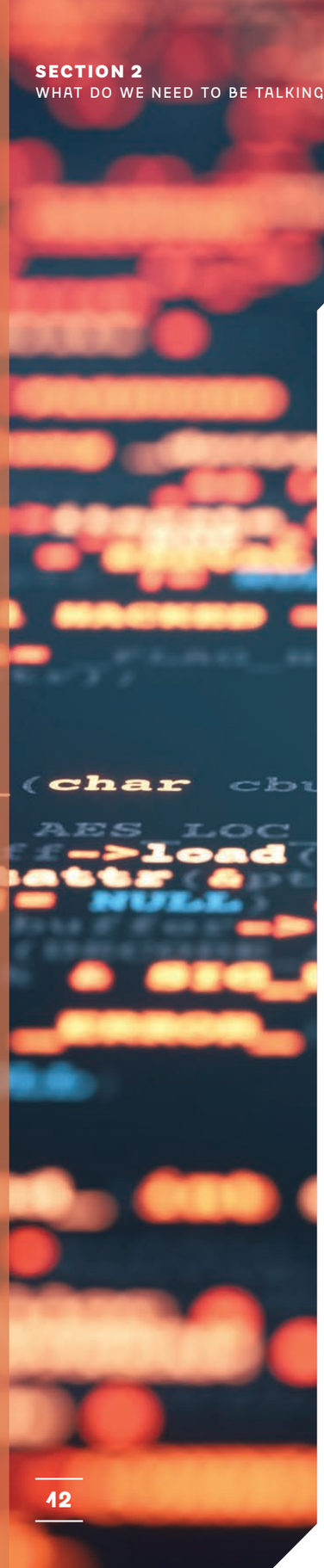
The scenario posed here relates to just one type of AI — positioning and navigation — but could just as easily apply to many other types of AI, such as the vehicle's pedestrian recognition system. Embodied deployments of AI are particularly useful for thinking about these issues because the consequences of the system not knowing when it is not functioning properly are usually self-evident: the autonomous car hits a pedestrian, the autonomous flying system crashes into a mountain, the surgical robot cuts an artery. The problem of AI knowing when it doesn't know how to respond has received relatively little attention in AI research to date, especially compared to the great deal of attention AI research has given to maximising percentage scores on research benchmarks for tasks like object recognition, where research careers are made and million-dollar salaries are obtained.

This imbalance in attention is no doubt also partly a result of how these AI systems have largely been deployed: in operationally tolerant, human-supervised tasks like image, text or music generation, where a human sees the output and can refine and iterate with the AI system until it is satisfactory. If the huge potential of these AI systems is to drive a large-scale transition into the embodied intelligence domain — robots, autonomous vehicles and automation — the systems themselves will likely need to have much better introspection capabilities. This need is especially present in safety or operationally critical domains where constant, vigilant supervision of the system is not feasible, and intervening after an incident is far too late.

Australia performs well in fundamental research per capita in many areas but has a poor track record in translating fundamental research into deployed technologies.[9] One applied technology area where Australia has demonstrated world-class capabilities is the field of deployed robotics: leading much of the development of some of the only robotic technologies in widespread, commercially critical operations, like the autonomous mining trucks that operate in both above- and underground mine sites around Australia[10] and the world. Australia has a number of major robotics research and development centres housed in universities and national agencies that have excellent links to the country's leading AI research powerhouses and to industry, where embodied intelligence in the form of robotics and more advanced automation could make their sectors more competitive than ever. Australia also has in many ways a particularly critical need for robotic and autonomous technologies, given its aging and expensive workforce, regular major economic losses from lack of access to labour in sectors like agriculture,[11] worsening financial outlook for the growing aged care sector[12] and the National Disability Insurance Scheme[13],and increased emphasis on sovereign manufacturing and supply chains[14] and defence capabilities[15].

When combined, these factors present Australia with a unique opportunity and need to lead the world in addressing a research challenge that will determine the success and reach of transformative AI technologies in the physical world. Our research capability and field robot experience provide us with a unique ability to tackle this challenge. Further investment at scale in Australian research will enable us to bridge the divide between basic, blue-sky AI research and introspective, trusted embodied intelligence in autonomous systems of all varieties. By doing so, we will be able to capitalise on this opportunity and take a leadership role in transitioning AI into the real, physical world.

PROFESSOR MICHAEL MILFORD FTSE conducts interdisciplinary research at the boundary between robotics, neuroscience, computer vision and machine learning, and is a multi-award winning educational entrepreneur. He is also one of Australia's most in demand experts in technologies including self-driving cars, robotics and artificial intelligence, and is a passionate science communicator. He currently holds the position of Joint Director of the QUT Centre for Robotics, Australian Research Council Laureate Fellow, Professor at the Queensland University of Technology, and is a Microsoft Research Faculty Fellow and Fellow of the Australian Academy of Technological Sciences and Engineering (ATSE).



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

Lana Tikhomirov
Australian Institute for Machine Learning (AIML), The University of Adelaide

Responsible AI reflects the values, needs and goals of humans by augmenting human lives and respecting human rights.

Associate Professor Carolyn Semmler

Responsible AI is AI that is developed with the purpose and understanding of the human system it seeks to serve.

Lana Tikhomirov

IF AI IS developed and implemented responsibly, it has the potential to be a positive force for Australians by reducing the strain caused by inefficiencies in our social systems. For instance, AI systems for health could minimise the millions of dollars lost yearly due to adverse events in the health system[16] by allowing us to invest in better care in different areas. However, this can only be achieved if AI is recognised as being one part of a complex network of working parts, known as a *socio-technical system*. Although these principles can be generalised to many applications of AI, here we focus on health, which is the field that is currently most influenced by AI and a major focus of the tech industry.

Socio-technical systems

All technology exists in a socio-technical system. In the context of the health sector, this system is made up of doctors, nurses, other healthcare professionals, the technical infrastructure, and the community they serve. However, a poor understanding of how AI fits into this system, combined with poor implementation, is increasing the potential for harm and error. To illustrate how this looks in the healthcare system, we might consider the example of doctors taking advice from AI when it conflicts with clinical best practice (over-reliance) or doctors choosing not to use AI when it could encourage adherence to clinical best practice (under-reliance). Further, recent research by the Australian Institute for Machine Learning has demonstrated that bias against minority populations can be transferred through AI systems, which can then negatively impact certain patients.[17]

Why there is no simple fit between AI and a complex socio-technical system like healthcare

Although doctors have specialised skills, they are privy to the same limitations as all humans, of which cognitive science can provide a basis for understanding. This begins by recognising how AI will never be a simple fit into a complex socio-technical system. First, the understanding of the tasks that AI can undertake is misguided. Too often, AI models are built without apprehension of the task that should be solved; and their development lacks input from the people who will use them, in this case, practitioners within the health system. Second, AI models have no ability to use context or meaning to inform their decisions. This is problematic because context critically determines the quality of outcomes for patients.

For example, AI algorithms to detect sepsis have previously missed a large proportion of cases by being unaware of the population characteristics in which they were deployed.[18,19] However, a doctor working in that same population will be able to draw on their knowledge (grounded in experience) of the different rates of sepsis among different populations to recognise the symptoms needed to accurately diagnose sepsis.[20] Furthermore, the datasets used to train AI are often not kept up to date to reflect the diversity of the population or the diseases they are trying to classify, significantly limiting the technology's adaptability and shelf life.

Designing AI to augment the human in the system

To overcome the problems we have identified, we need to take a radically different approach to the design of AI systems. This can be achieved by understanding how expert human decision-makers like doctors do their work, using the methods and knowledge of cognitive sciences. For example, cognitive scientists have developed a deep understanding of how radiologists can extract the features of a pathological condition from an image within milliseconds of seeing it. This understanding can help to guide when and where AI tools are needed to improve the skills and training of healthcare professionals. Doctors, unlike AI, have a responsibility to their patients and must maintain professional standards of care. Indeed, they are the pinnacle profession that needs to demonstrate responsibility. Therefore, ensuring the appropriate use of AI in their work represents a significant challenge. If AI is implemented poorly, it may add to their burden of responsibility and potentially expose doctors to the risk of poor decision-making. Alternatively, AI implemented with a responsible design informed by cognitive science will allow doctors to offload their cognitive tasks to the AI when appropriate and focus their attention on patients.

So what does this mean for Australians?

Responsible AI means giving all Australians, whose lives will be impacted by AI, information about its intention, data and decision-making processes. Further, responsible AI requires the development of legal frameworks to protect Australians from the potential harms arising from poorly developed AI and inappropriate deployment in socio-technical systems. Most importantly, Australians have the right to be informed about the limitations of AI to allow them to decide which aspects of their lives could benefit from it. AI could be a positive force; but only if our understanding of human cognition remains central to AI development.

ASSOCIATE PROFESSOR CAROLYN SEMMLER leads the Applied Cognition and Experimental Psychology (ACEP) research group at The University of Adelaide focusing on the application of theories and models of cognition, judgement and decision-making to legal and medical contexts. The research group uses experimental methods and modelling of cognitive processes to understand and improve human decision-making in important contexts, such as health, policing and national security. Technologies such as AI and social media have vastly changed human experience, and the group is working at the intersection between psychology and the many disciplinary areas that imagine, design and build new technology.

LANA TIKHOMIROV is a higher degree research candidate at the School of Psychology Faculty of Health and Medical Sciences, University of Adelaide. Her work resides in the nascent field that combines AI safety, cognitive science, and human factors research for the development of complementary and safe AI technologies. Specifically, she investigates the impact of medical AI systems in clinical radiology and is interested in ethical and safe AI development to ensure fairness and accountability practices. She is jointly supervised by Associate Professor Carolyn Semmler (School of Psychology) and Dr Lauren Oakden-Rayner (School of Medicine/ Australian Institute for Machine Learning).

Essay 5 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide
Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

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 AI is the development and use of AI systems in a way that upholds ethical principles of fairness, transparency and privacy. It involves understanding how AI technologies may have a negative impact on members of our society, the risks that are involved in leveraging AI technologies, and how we might counter those through the application of ethical principles.

It involves being transparent about how and when AI is being used, transparency in the provenance and sharing of data, fairness and equity in accessibility and benefit, and accountability for outcomes. It must be aware of and embrace diversity and inclusion to not perpetuate bias or discrimination, and considerate of different cultural, social and demographic backgrounds.

WHILE WE DON'T know what the future workplace is going to look like, we do know that artificial intelligence (AI) is already fundamentally changing it and the skills that our future workforce will need.

Enhancing disease diagnosis and predicting patient outcomes in our healthcare systems, creating personalised learning opportunities for students, freeing up human workers by taking on and automating repetitive tasks, and improving the safety of our workplaces through predictive maintenance and monitoring are just some of the examples of how AI can provide a significant opportunity for our society and workforce. But how do we set ourselves up to successfully engage with these changes, and to ensure that our workforce is prepared to lead in this space?

Key to our response is investment in STEM (science, technology, engineering and mathematics) education as a continued priority in schools and universities. We need to ensure we create the talent pool required to handle the challenges and opportunities presented by AI, starting with promoting core digital literacy and AI awareness from an early age, building awareness of basic AI concepts, tools and platforms along with broader digital literacy skills and an awareness of ethical considerations and potential societal impacts to shape the development and deployment of AI in a responsible manner.

Our investment in STEM education must have an emphasis on AI education and training. There is a need to develop AI-focused educational programs and training initiatives to support our schools in building the fundamentals and our universities and other educational institutions in shaping the next generation of STEM and AI experts, as well as supporting a culture of reskilling and continuous learning in our workforce. Furthermore, AI is changing rapidly — so this is not a once-off education need but must be an ongoing process in line with the rapid pace of technological change.

This investment must come from both the education sector and industry. Industry needs to partner with educational institutions to invest in the reskilling and upskilling of their teams, and to support workers transitioning from declining industries into AI-driven fields, mitigating job displacement and supporting economic growth.

How do we do this? Supporting our schools in STEM education, and our teachers, needs to be one of our highest priorities, ranging from offering ongoing, sustained teacher professional learning to allow our teachers to enhance their STEM expertise and build awareness of AI. This requires building dedicated AI-focused professional learning and curriculum, integrating real-world applications, and partnering with industry to align and reflect the latest industry trends and real-world examples of careers and impact.

But we are not starting from nothing — we can continue to support the many existing STEM-focused professional learning programs that are currently supporting so many of our teachers.

We have some fantastic programs across Australia, including Early Learning STEM Australia, Education Services Australia's Digital Technologies Hub, CSER STEM Professional Learning Program, and CSIRO's STEM Professionals in Schools program, providing a direct link from industry to our schools.

Beyond professional development, there is a wealth of STEM clubs and competitions calling out for industry and community support — encouraging students to actively participate in scientific experiments, engineering projects and coding exercises; fostering creativity; motivating students; and building a sense of community. Assisting these communities to support AI-specific opportunities and working with industry will enable rich AI-focused educational experiences that are relevant to the needs of both industry and community.

We need to promote diversity and inclusion in our offerings. Encouraging our next STEM generation to see its place in the future STEM workforce will set us up as a nation ready to harness the diversity of thought needed for innovation and creativity, leading to robust and ground-breaking solutions that reflect the needs of our society. We also need to support and encourage representation of women and other minority groups in the future AI workforce.

As we grow to rely more and more on AI, we must be confident that our AI systems do not perpetuate bias or discrimination, and that our AI solutions are inclusive and considerate of different cultural, social and demographic backgrounds.

This is not intended to encourage all students to pursue a career in STEM. We also need to consider how students across all discipline (and workforce) areas will build their knowledge of AI, and of how to leverage it, how to work with it and when to trust it. Learning about how AI is being applied across different fields, including healthcare, finance and customer service, will open up new opportunities and career pathways for all graduates, and help all of our industry sectors embrace what AI can offer.

To support our industries to engage successfully with the changing world of AI requires investment in relevant infrastructure, ensuring that the necessary technological infrastructure, such as high-speed internet and cloud computing services, is widely accessible and that the digital divide in our society is mitigated. Programs that support and promote AI entrepreneurship and innovation, encouraging start-ups and small businesses to engage with AI, will be critical to building a workforce ready to take on the challenge of AI, and to foster and keep our future generations of STEM graduates.

PROFESSOR KATRINA FALKNER FTSE is a top 100 innovator completely transforming computer science education. Her work directly addresses inequities in access to technology, helping to build a fairer Australia. In 2013, 95% of teachers had no education in computer science. Katrina decided to turn this around. Her wildly successful computer science 'massive open online course' (MOOC) shows teachers how computer science works and highlights its importance for children's futures. Based on rigorous research, the MOOC blends computer science into any subject for an enjoyable learning experience. It has reached 45,000 Australian educators and more than two million students. Katrina's research shows that teachers are inspired and empowered by the course, moving rapidly from uncertainty to confidence in teaching computer science.

Essay 6 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

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

Responsible AI is a broad term that is intended to encapsulate a number of guardrails to protect against potential risks and harms that may stem from the inappropriate use of powerful emerging AI technologies.

These risks include but are not limited to those related to consent, privacy, accountability, fairness and accessibility. Responsible AI design, development and deployment necessarily require qualified experts, trusted data and sustainable infrastructure. Without widespread use and adoption of responsibly developed AI, there is a risk of irresponsible AI tech-nologies infiltrating business and society.

RECENT ADVANCES IN AI are bringing about both great excitement and significant concerns. It is widely recognised that the recent upswing of interest and success in AI technology is riding on the back of two significant advancements that previous booms did not have: access to cheap and elastic computational resources, including the cloud, high-speed networks and fast processors; and the availability of large volumes and a variety of digital information, or Big Data. Given the significant injection of both corporate and venture capital, the opportunities arising from the current AI boom are both promising and sustainable.

While the availability of large amounts of data is a critical factor behind the recent AI successes, it is also a cause for concern. A recent article on responsible data management[21] presents a mirror metaphor to explain pre-existing bias and the technical limitations of responsible data use:

'Data is a mirror reflection of the world. When we think about pre-existing bias in the data, we interrogate this reflection, which is often distorted. One possible reason is that the mirror (the measurement process) introduces distortions. It faithfully represents some portions of the world, while amplifying or diminishing others. Another possibility is that even a perfect mirror can only reflect a distorted world — a world such as it is, and not as it could or should be.'

Current AI systems are trained using very large amounts of data largely extracted from the web, the vast majority of which is typically publicly available (e.g., from Wikipedia). This raises several questions about quality, consent, ownership and privacy. Data is often collected and used without the awareness and explicit consent[22] of the people who created this online content. One of the most relevant pieces of legislation relating to this issue is copyright, which was developed to restrict how people re-use others' content but does not regulate the use of content to train AI.

The European Union's General Data Protection Regulation (GDPR) is an example of a regulation designed to tackle this issue of data copyright, particularly in terms of the data that users contribute to AI systems on the open web, such as through prompts to large language models like ChatGPT, which then eventually become part of the model. GDPR includes the right to be forgotten, which allows citizens to request that their data be erased from data products, including AI-assisted products. Given the current lack of regulation in most jurisdictions of the world on the use of public data to train AI systems, its usage may be regarded as legally compliant, but it may also be considered ethically questionable as it poses risks.

It is currently hard to guarantee that a generative AI model trained on certain data will not resurface pieces of content from the large amounts used to train it, thus potentially putting the privacy and safety of the content creators at risk. Note that the use of publicly available content without authors' consent is not a new issue.[23] It is important that the data used to train AI is either collected through informed consent processes and/or used in a way that provides a proportionate return for the content creators without putting them at risk.

Modern AI applications like chatbots, recommender systems and traffic-aware navigation commonly take user-generated real-time data as inputs and provide predictions or suggestions. To reduce the reliance on real-time user data, an emerging line of research is around the use of AI to generate synthetic data points[24,25] based on patterns extracted from real data. The application of synthetic data is already seen in several high-stake domains, such as medical analysis[26] and financial fraud detection.[27]

In the meantime, the computational cost of storing and analysing large-scale data is escalating. For example, the state-of-the-art computer vision algorithms for object classification are all trained with the ImageNet dataset, whose full version contains over 14 million images and is 1.3 terabytes in size.[28] Recent foundation models[29] and in particular large language models like GPT-4 have further demonstrated the tremendous cost of data collection, curation and storage as well as the computational and environmental cost of model training.

Recent advancements that provide distribution-level summarisation of real data points[30,31] with compression/condensation techniques offer promising solutions from both the data management and computational perspectives, wherein large and noisy data points can be substituted with a set of small but quality samples (or aggregated samples), to provide a data-efficient paradigm for training AI algorithms.

Advancements in model distillation (also known as knowledge distillation[32]) offer another promising avenue to help deal with computational concerns. Model distillation involves training a smaller model ('student') to reproduce the behaviour of a larger model ('teacher'), with the aim of creating a lighter, more efficient model to be executed in production, while maintaining a high level of effectiveness. This process is particularly useful for large language models that can have billions of parameters, making them computationally expensive to run. Distilling generic models, made available by third parties such as OpenAI's GPT-4 or Meta's Llama, into smaller ones makes them more viable for real-world applications where substantial computing power is lacking. Another advantage of model distillation is the student model's ability to assimilate a rich array of linguistic patterns from the teacher model, which may be trained on diverse, large and open-source data. The student model can then undergo further refinement via fine-tuning specific proprietary data, facilitating a degree of model sovereignty.

Data management will continue to underpin the development of AI technologies, and the responsibility with which we manage the data will determine whether we gain the benefits of these powerful technologies or are confronted by their risks and potential harms.

PROFESSOR SHAZIA SADIQ FTSE Is a Professor of Computer Science at the School of Information Technology and Electrical Engineering, The University of Queensland. Her research focusses on responsible data management and aims to reduce the socio-technical barriers to data driven transformation, by assisting organisations to create, protect and sustain agile data pipelines. She is the Centre Director for the ARC Industry Transformation Training Centre on Information Resilience 2020-2025, Chair of the National Committee on Information and Communication Sciences at the Australian Academy of Science 2019-2022, and member of The Australian Research Council College of Experts 2018-2021

PROFESSOR GIANLUCA DEMARTINI is an Associate Professor in Data Science at The University of Queensland, Australia. His main research interests in data science includeinformation retrieval, semantic web, and human computation. His research is currently funded by the Australian Research Council, the Swiss National Science Foundation, Meta, Google, and the Wikimedia Foundation. He received multiple best paper awards at artificial intelligence and information retrieval conferences. He has published more than 200 scientific papers at major computer science venues such as the ACM Web Conference, ACM SIGIR, VLDB Journal, ISWC, and ACM CHI. He is an ACM Senior Member, ACM Distinguished Speaker, and TEDx speaker.

DR ROCKY TONG CHEN received his PhD degree in computer science from The University of Queensland in 2020, and is currently a lecturer with the data science discipline, School of Electrical Engineering and Computer Science, The University of Queensland. He received the Discovery Early Career Researcher Award from the Australian Research Council in 2022. His research interests include data mining, recommender systems, user behavior modelling and predictive analytics.

PROFESSOR GUIDO ZUCCON is a Professorial Research Fellow at The University of Queensland, Electrical Engineering and Computer Science School, the AI DIrector for the Queensland Digital Health Centre (QDHeC), an Affiliate Professor at the UQ Centre for Health Services Research, Faculty of Medicine, and an Honorary Reader at Strathclyde University (UK). He leads the Information Engineering Lab (ielab), a research team working in information retrieval and health data science. He was an ARC DECRA Fellow (2018-2020). Guido's main research interests are information retrieval, health search, formal models of search and search literaction, and health data science.

Essay 7 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

Open the pod bay doors please, HAL

Andrew Dettmer
National President, Australian Manufacturing Workers Union

AI is a powerful tool. However, it must be limited — it is not simply a benign force for good. Humans are the only ones who should be able to make decisions about its creation and application.

We need to create a 'social licence' to ensure that those who AI would control have means of participation and recourse such that no system is applied without human beings at its centre. At all times, those who would benefit must be held to account for any of the decisions taken by AI systems. The safety of humans — mental, physical and social — must be incorporated into any AI system, and humans must be entitled to withdraw from any AI system and must be able to give informed consent to any limitation or control imposed by AI. Humans must retain, unimpeded, the right to our quiet enjoyment of life.

AMONG THE GREATEST fears we all have is a lack of control, of matters proceeding which involve us but over which we have no agency. Being a victim appeals to very few.

Australia has a belief in the 'fair go'. However, this is tempered by the reality of the experience of many, in the workplace and the community. In workplaces, seemingly benign bosses can subscribe to the principles of 'managerial prerogative'. At its worst, this entails bosses and supervisors conforming to the idea that they alone are the basis of all rational decision-making, with workers having the right to simply obey. In some ways, AI can be seen as an outgrowth of this prerogative, with even less human agency.

In Australia, rights in the workplace are governed by the awards of the Fair Work Commission (FWC). These instruments, having the force of statute, have been in existence for over a century. They form the basis of the rights and obligations enjoyed by workers. Formerly, awards were the result of a notional 'industrial dispute' between employers and unions. They were dynamic and responsive to the needs of workers and employers. Since John Howard's introduction of WorkChoices in 2006, however, they have been more akin to instruments of administrative law, such that unions and employers apply for variations to the FWC.

This is relevant to any consideration of AI given that rights in the workplace are determined through these instruments and enforced by the FWC and the courts. The only effective rights to a consultation that most workers have (and a notional defence against the depredations of AI in the workplace) are contained within the provisions known as the Termination, Change and Redundancy Case of 1984. The 'change' element is the relevant consideration in this discussion.

As one would imagine, the rights to consultation envisaged under that decision reflect the state of technology in 1984. At the time, workers and employers were grappling with early computerisation. Instruments such as program logic controllers and computer numeric control systems were then novel; it is worthwhile recalling that the most popular home computer (and the first for many), the Commodore 64, had only been introduced in 1982.

The introduction of change requires consultation. In the Manufacturing Award (MA10) — an award that covers approximately 900,000 Australian workers — Clause 41 requires an employer who has made a 'definite decision' to introduce 'major changes in production, program, organisation, structure or technology' to consult with their workforce.[33] In 1984, this was readily identifiable, and timeframes were created accordingly.

However, those rights appropriate to the time of the Commodore 64 are unlikely to provide significant rights where AI is concerned. When the notoriously anti-union Elon Musk complains about the scorched earth created by the unfettered application of AI, the rights of workers are at significant risk.

AI is 'artificial' only because the artifice involved does not need to be initiated by a human. But the directions, the process and the outcome of an AI-generated command all originate with a human.

AI is notorious for its capacity to descend into prejudice and abuse when an extended narrative is expected of it. But this, surely, is reflective of the prejudices of its originators and the data they provide. One of the many complaints about Robodebt, aside from the fundamental illegality at its core, was the repeated experience of many recipients of the computer-generated letters of demand not being able to engender a response from Centrelink. The mechanistic assertion of debt and the virtually instantaneous garnishing of wages and bank accounts were a source of great stress.

HAL was the computer at the heart of Stanley Kubrick's film, *2001: A Space Odyssey* (which is based on the novel of the same title by Arthur C. Clarke). HAL was the killer of many astronauts, none of whom could control it. The most important scene occurs when astronaut Dave Bowman asks HAL to allow him to re-enter the spaceship. HAL refuses to cooperate. This is the great fear that is at the heart of every human interaction with AI — that it will not obey a human command.

AI has at its heart the involuntary gathering of human data. There should be no question that the right of individuals to ownership of their data is absolute. For promoters of AI, however, no such limitation appears to apply.

In the workplace, AI can be used in multifarious ways. If it is in control, whether work is decent — or safe — is of no concern to a computer; workers are simply units of production. Whether they work seven hours or 20 is of no consequence. In the dystopian vision of HAL, all humans are disposable.

The rights of workers to protect their data must be extended to the community. End-use AI companies gather personal data, without restraint. We need to impose a social license on the generators and users of AI. In particular, workers need to have the right to refuse their data being collected.

We need to ensure that workers' rights to consultation are enhanced. Just as importantly, such rights need to ensure that humans are not controlled by and do not respond to machines. The calm voice of HAL refusing to open the pod bay doors cannot be replaced by the soothing electronic tones of Jeff Bezos telling Amazon workers when they can go to the toilet.

ANDREW DETTMER is the National President of the Australian Manufacturing Workers Union and is the ACTU representative to SafeWork Australia. Andrew sits on the boards of the Industry Capability Network; the Australian People for Health, Education and Development Abroad (APHEDA) and TAFE Queensland. He co-chairs the Future of Work, Education and Training/Test Laboratories stream of the AiG Industry 4.0 Forum.

Essay 8 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

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

For me, the term responsible AI represents the inflection point we've now reached with artificial intelligence. AI has transitioned from a laboratory curiosity into a deployable commodity for governments and companies to drive transformative change across industry and society. This means AI has reached a certain maturity level where we should start asking questions such as "How is this technology affecting our society?" and "How is it used in a responsible way?"

Australia really needs to be at the forefront of responsible AI, because if we can get ahead of the rest of the world, we can buy into the opportunities being created.

ARTIFICIAL INTELLIGENCE IS at an interesting inflection point. The technology is now rapidly transitioning from a perception as a laboratorial, theoretical curiosity, to something tangible that's really transforming global business and making a big impact in people's lives.

However, while we consider where we fit in an AI-enabled world, the Australian industry needs a bit of a wakeup call.

We enjoy an excellent standard of living, but for a country that pitches itself as an advanced economy, we have a dangerous lack of economic complexity, ranking 79th in the world — behind Chile and Kazakhstan[34]. The world's most economically complex countries are some of our closest allies and trading partners: Japan, Singapore, and the United States.

Our industries have incredible potential but have demonstrated a long-held aversion to risk and a lack of interest in serious research and development. The average Australian Securities Exchange (ASX) 200 company spends just 3 per cent of its revenue on research and development[35], half that of the Organisation for Economic Co-operation and Development (OECD) average and not nearly enough to spark sufficient innovation on our own soil.

Instead, Australia's prosperity hinges on a 'dig and ship' mentality, where our economy is propped up by exporting precious resources to the world, leaving us vulnerable to the volatility in commodity markets.

When it comes to the critical technology we need to complexify our economy and remain globally competitive, I'm genuinely concerned that Australia risks becoming too comfortable with becoming AI adopters, and not AI creators. Australian biggest companies cannot afford to sit tight and wait for AI technology to be developed abroad and buy it 'off the shelf' when they feel ready.

So how do we start to turn things around?

Australia has impressive universities that are engines of innovation. They undertake about 40 per cent of our R&D[36], and the industry would be foolish to not leverage this wealth of knowledge on campuses across the country.

One of the greatest international examples is Stanford University, which has been an exemplar of successful university technology transfer and commercialisation for decades. Its industrial affiliates programs bring multiple companies together with faculty and students to explore research ideas in a pre-competitive environment. For a small membership fee, companies get direct contact with skilled researchers, industry-focused research presentations, and access to a student talent pool for internships and graduate recruitment. Where's the interest from Australia's top companies for these kinds of opportunities?

Traditionally, universities have conducted low technology readiness level (TRL) research on initial ideas before they are spun out into stand-alone companies, where they then mature and climb the TRL ladder. But AI offers us — and requires us to develop — new kinds of university-industry partnerships for the future.

AI is a lightweight technology. It climbs the TRL ladder more rapidly because it doesn't require heavy physical infrastructure, algorithms can be prototyped and tested rapidly and cloud-based services offer lower barriers to market entry. AI relies on datasets, making it ideal to roll out across existing industries, where it can be integrated into existing systems to dramatically augment capability.

Accompanying AI's rapid development is a growing global demand for the main source of AI capability: talented people. Universities need strategies to ramp up and meet global demand. Hybrid appointments — where AI researchers split their time between academic research supervision and leading a company's applied AI lab — are one way that Australian universities can form deeper engagements with local and global industry.

Increasingly common in the US but relatively new to Australian universities, hybrid roles are particularly useful in niche fields such as AI, where specialist skills are in high demand and top researchers command salaries that public universities can't match.

Rather than headhunting research talent outright, tech companies understand the strategic value of building an ongoing connection with the latest research developments and drawing from a growing talent pool of PhDs and graduates.

Students benefit immensely from working on real-world problems with professors who are connected with the best graduate employers in their field.

The federal government also has a role to play in sparking innovation and helping our universities and industries work better together. While grants and piecemeal funding are beneficial, there needs to be fundamental change if we want to support the next era of innovation through start-ups, small- to medium enterprises, and broader industry.

We can also reimagine the government's role as an AI *customer* and require governments — both state and federal — to purchase a certain percentage of their AI product requirements domestically. It's a great way to build confidence in Australia's tech ecosystem.

The idea is hardly new. In the early 1980s, the California state government implemented novel tax credit arrangements that saw Apple put computers into 9,000 public schools. This allowed them to get a strong foothold in the education market and revolutionise personal computing through the 1990's.

For startups, scaleups, and tech companies seeking to do new things, having the government as a customer is vital in building a brand and name recognition. After all, it's these start-ups and small- to medium-sized tech companies that have been at the forefront of AI innovations through the past decade.

Australia has countless success stories in AI research, and we have many expats doing amazing things abroad, we just don't often hear about them. High school students need to see role models in AI who are championing innovation if they're to forge careers in STEM — kids might actually want to do their maths homework if they could see where it can lead them.

AI has arrived and very soon the idea of 'AI' and 'not AI' is going to be outdated. When the internet first came to homes in the mid-1990s, it was only accessible via a dial-up modem attached to a desktop computer. Now it's effortlessly connecting every aspect of our daily lives that we don't really think about it anymore.

In an age where AI is quickly becoming the facilitator of global innovation, Australia stands at a pivotal juncture. Our industries and academic institutions must recognise the synergy that could fuel our growth in this new landscape. We should aim to be at the forefront of AI technology creation and implementation. By fostering a culture of research, risk-taking, and close university-industry relationships, we can diversify our economy, bolster our global standing, and create a fertile ground for a new generation of Australian tech innovators.

PROFESSOR SIMON LUCEY is the director of the Australian Institute for Machine Learning (AIML) and a professor in The University of Adelaide's School of Computer and Mathematical Sciences. Prior to joining AIML in 2020 he was an associate research professor at Carnegie Mellon University's Robotics Institute, in Pittsburgh, Pa. Professor Lucey has received various career awards, including an ARC Future Fellowship (2009-2013). His research interests span computer vision, machine learning, and robotics. He enjoys drawing inspiration from AI researchers of the past to attempt to unlock computational and mathematical models that underlie the processes of visual perception.

Essay 9 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

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

Responsible artificial intelligence (AI) requires the development and use of AI to be aligned with the values of social democracy, such as freedom, respect, fairness and equality of opportunity while being transparent and accountable to the public.

People should be able to broadly understand how AI systems work, and how and when they are being used. AI should be deliberately crafted to be inclusive and have regard for marginalised members of society, and it should be applied for the benefit of all, rather than a select few. It should be used in a way that does not discriminate against or disadvantage any group of people.

IN CATALYSING THE industrial revolution, the steam engine fundamentally changed the way people lived, worked, defined and organised themselves. It brought widespread access to education, the rise of professional specialisation, cheaper goods, medical advances and longer life expectancy, and the promise of social mobility. It also brought exploitation, overcrowded cities, climate change, unmanageable volumes of waste, dislocation and social exclusion.

AI is the steam engine of today. It can, and almost certainly will, fundamentally change the way we live, learn, care, play and work. And it will change the way we define being human: when machines can responsively apply language and mathematics, and create art as well as, or better than, people, what does it mean to be human?

The opportunity to meaningfully shape the direction of this shift is still open — but it won't be for long. An AI-literate community is essential if we are to ride this revolutionary wave in a way that enables social democracy to thrive. AI will permeate every part of our lives; this means it is untenable for AI know-how to be the exclusive purview of engineers, scientists and technologists. To shape an inclusive and socially responsible future, we must prioritise AI literacy for those who care for the vulnerable, those who engage with social media, those who teach and learn, and those who work in every economic sector. In a word, we must prioritise AI literacy for everyone.

AI technologies replicate human-like intelligence, achieved by training machines and computer systems to do tasks that simulate some of what the human brain can do. They learn from and draw on large datasets that serve as their knowledge base and are trained to predict specific outcomes based on the patterns and structures found in these large collections of information.

It's tempting to think of datasets as neutral sources of information, but they are not. The people and places from which they're drawn, the questions they're tasked with responding to, and the context in which they're analysed all impact on the quality and focus of the information, and therefore drive the outcomes. For example, most large voice-based datasets are overwhelmingly male voices. As a result, voice-activated software responds much more readily to lower registers than to higher-pitched (i.e., female) voices. These systems can improve over time, becoming more complex and accurate as they take in more information. Understanding this can help users to better grasp how, when and why these potent AI technologies are used to inform the decision-making of political, social and economic organisations — and the potential inadvertent outcomes they can thereby produce.

AI literacy is even more important among decision-makers and their advisers. It is also essential that human rights and rule of law principles be incorporated into the development of AI, particularly in contexts where AI systems assist or replace human decision-making. Australian democracy is built on the core values of freedom, respect, fairness and equality of opportunity:[37] these values are central to our community to promote a secure, inclusive, prosperous and peaceful place to live. They should therefore be the guiding principles for developing and using AI-powered technologies. The Australian Government's AI Ethics Principles framework[38] provides a starting point (so long as it is consciously and actively applied). Further binding and non-binding instruments, regarding specific aspects of AI and its use in specific sectors and contexts, will be needed to comprehensively address the use of this complex and rapidly evolving technology.

The relative merits of using AI in different applications is highly context-dependent, and the strengths and limitations of the technology must be considered and reconsidered as it evolves. AI technologies are already being applied in deciding who gets a job interview or a mortgage, as well as which movie a streaming service recommends to individual users. An AI-related mistake or built-in bias in the first two instances has much more serious consequences than in the latter. Literacy for AI users (such as loan assessors and human resource officers) would support greater awareness of the technology's limitations, and hopefully a more equitable outcome.

AI systems should also be required to be safe and secure throughout their operational lifetime, and it should be easy to verify their security where applicable and feasible. Therefore, policies and regulations that apply to organisations and industries require appropriate tailoring, particularly in alignment with the potential risks and harms resulting from AI technologies.

If used appropriately, AI technologies could progress commitments to the United Nations Sustainable Development Goals, including Goal 4 (which aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all). Applied by a sophisticated teacher, AI can be used to personalise a student's learning journey by analysing a student's learning history and suggesting improvements. It could also reduce teachers' administrative workloads, freeing them up to devote more time to their students.

The development of AI tools, and the algorithms and inputs they use, is often quite opaque in nature, naturally leading to concerns about how these tools are being applied. It is therefore crucial to increase transparency in the processes involved in AI development and deployment where possible.

To support agency and independence, it's also important to communicate with those impacted about how an AI solution was created and developed, why it was put into use, how it is maintained and updated, and the circumstances under which it may be retired.

All these scenarios rely on a level of AI literacy to support informed consent and decision-making, to mitigate the risks of AI being used by bad actors, and to create the conditions for AI to benefit all members of society.

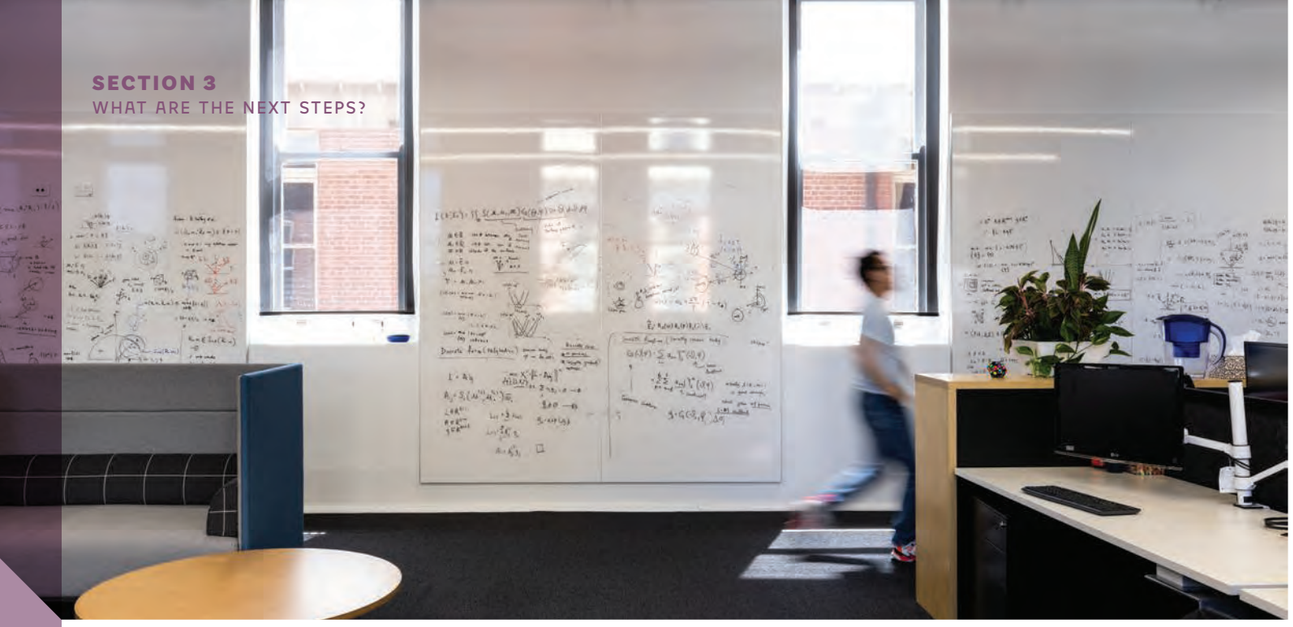
AI literacy includes the ability to understand the basics of AI, how it works, and its potential impacts. It further requires a whole-of-life approach to learning for every member of the community.

An AI-literate society is one in which everyone can exercise agency and discretion as we engage with AI technology. It's a society in which we understand how to set the boundaries for this rapidly evolving technology and direct it towards an inclusive, kind, well, productive and happy society.

An AI-literate society is one in which these powerful new tools can be used for the benefit of all people.

KYLIE WALKER is the Chief Executive Officer of the Australian Academy of Technological Sciences and Engineering. She works with Australia's leaders in applied science, technology and engineering to advise decision-makers, lead crucial national conversations to solve complex challenges, and support Australia's technology-powered, human-driven future.

She is a recognised leader in equity, diversity and inclusion in STEM.



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

THIS LAST YEAR has seen incredible growth in public awareness of artificial intelligence, but perhaps not as much public understanding. Despite ChatGPT and a raft of other consumer AI software releases, the public perception of AI still hovers excitedly around visions of sentient physical beings: loyal and attentive robots at our side, diligently doing the jobs we humans find so annoying. The truth is far less Hollywood.

While related, AI and robots are not the same thing. The vast majority of what people call "AI" today is machine learning. It's math and computer code. It's software that's able to analyse and interpret vast amounts of information, and make accurate predictions, far more efficiently than any human. And while AI is now driving a technological revolution and powering the world's largest companies, it's not about to cook you dinner and do the dishes.

The capabilities of even the most advanced contemporary robots are far more modest than the public imagines. The truth is the robot vacuum cleaner in your home is one of the smartest pieces of robotic technology you can buy. Most robots deployed in the industry today lack any form of true AI, rendering them essentially elaborate machines for the basic automation of repetitive tasks. They can't deal with complexity, and they stop working if they encounter even the slightest unexpected change in their surrounding environment. They're not intelligent, by the broadest of possible definitions.

This prevailing public misconception tells us a lot about the kinds of opportunities Australian AI research could be well positioned to pursue.

Right now, AI has trouble operating in the real world and interacting with the environment. *Embodied* AI tries to solve that problem.

Embodied AI operates inside smart devices like robots and drones and allows them to perceive, navigate and understand the real world in all its rich complexity. Perhaps the most publicly well-known example of robots with some basic attributes of embodied AI are iRobot's Roomba range of vacuum cleaners; and they can trace part of their origins to Australian robotics and AI research.

Alongside two of his MIT students, Australian roboticist Professor Rodney Brooks founded the iRobot company in 1990. Thirty years and 30 million robot vacuum cleaners later, he's internationally lauded for challenging the traditional AI approaches of the time and pioneering the commercially successful development of behavior-based robots. Modern Roombas are equipped with advanced visual navigation systems so they don't get lost in your living room; and that is a downstream result of landmark research by The University of Adelaide's Professor Ian Reid, who co-invented the computer vision AI technology that effectively transforms an inexpensive digital camera into a powerful geometric sensing and mapping tool.

Embodied AI holds the potential to radically change our economy. Consider Australia's vast landscape: we have plentiful land to cultivate and resources to manage, but manual labor at scale is both inefficient and expensive. Robots guided by advanced machine learning algorithms could potentially be deployed to perform some of these tasks autonomously, capturing significant economic value while doing the jobs Australians don't want to do themselves. This technology could open the door to advanced manufacturing industries previously unviable in Australia due to our high labor costs. Future robots will not be restricted to controlled factory conditions but will operate in open, dynamic environments, executing complex tasks.

The public dream of what AI technology should be—robots that listen to us and carry out our natural-language instructions—has been with us long before *The Jetsons* first appeared on TV screens sixty years ago; and while it's still a way off, recent AI advances are encouraging.

ChatGPT's great mainstream adoption is prompting people to now ask why they can't have similar interactive experiences with other machines. Enter *vision-and-language* AI, a burgeoning field at the intersection of computer vision and natural language processing techniques. It's an area where Australia has a very strong research talent, and an opportunity we should pursue.

The next generation of robots will possess natural language capabilities, allowing for more seamless human-machine interactions, while also interpreting and navigating the physical world in real time. Imagine asking your robot to "clean up that mess in the kitchen," and it not only understands you, but is able to effortlessly avoid obstacles, find the mess in the kitchen and maybe even empty the bin when it's done.

So, what's the broader implication? Australia is well positioned to be at the forefront of AI research in these emerging fields, but it requires investment. While the rest of the world is also accelerating in AI capabilities, Australia has a unique set of assets that make it viable for leadership in AI research.

The gap between current AI capabilities and the public's expectation isn't a drawback, it's an opportunity for Australia to invest in building technology that will significantly alter our economic landscape and daily lives. The question remains: will Australia capitalise on this fertile ground for AI research, or will we let another opportunity slip through our fingers?

Essay 10 of 13 from the series:
Responsible AI: Your questions answered

THE UNIVERSITY of ADELAIDE
Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

PROFESSOR ANTON VAN DEN HENGEL FTSE is a leading researcher in computer vision, the founding director of the Australian Institute for Machine Learning (AIML) and currently director of its Centre for Augmented Reasoning.

His research expertise is in the complex way that information is encoded, especially in images, and methods for recovering it. He is a powerful advocate for the importance of machine learning as a critical technology for Australia's future economic growth and social success.

SECTION 3
WHAT ARE THE NEXT STEPS?

Australia's unfair advantage in the new global wave of AI innovation

Professor Mary-Anne Williams FTSE
Michael J Crouch Chair in Innovation, UNSW Business School

Responsible AI is the commitment to designing, developing, deploying and overseeing AI systems by aligning with ethical principles to prioritise the safety, rights, values, and interests of all stakeholders.

Responsible AI adheres to best practice guidelines, standards, regulations and laws and considers immediate and long-term ecological, social and economic impacts.

It aims to provide technical robustness, making AI systems reliable, secure, resilient and sustainable, and it mandates continuous monitoring, learning and adaptation, proactively addressing biases, adversarial threats and evolving societal needs.

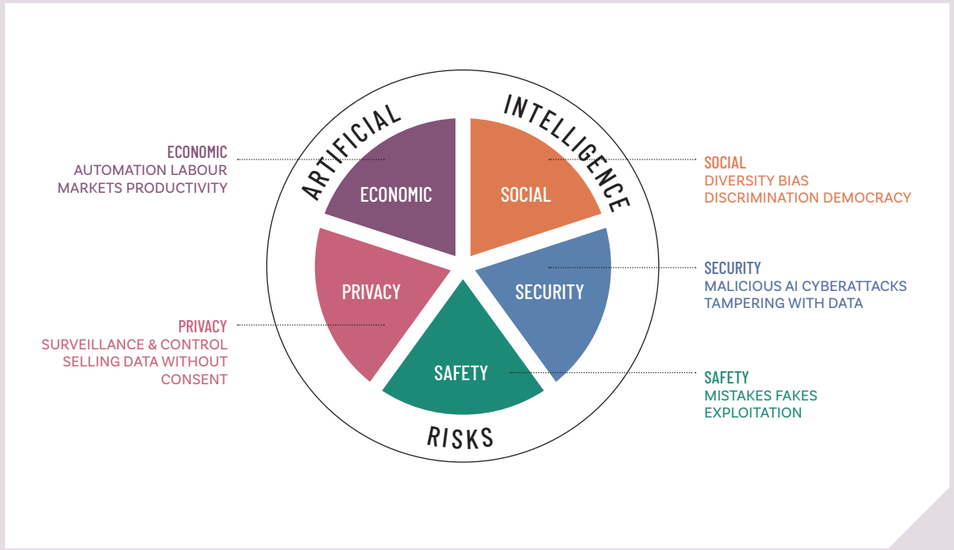


Figure 1: Risks associated with AI tools and technology. *Source: Mary-Anne Williams*

PROFESSOR MARY-ANNE WILLIAMS FTSE is an international authority in artificial intelligence (AI) and human-robot interaction. Her research has changed design paradigms in intelligent systems and significantly advanced decision-making under risk and uncertainty in open, complex and dynamic environments. The results of her work have been adopted by industry, including IBM, Infosys, Boeing and Visual Risk, and leading international research groups at CNRS France, Stanford University and Carnegie-Mellon University. Her high standing both in academia and industry is acknowledged through numerous awards, invitations, and distinguished appointments.

AS ARTIFICIAL INTELLIGENCE (AI) shapes the future, nations worldwide are grappling with significant challenges in seeking to harness its potential and doing so safely and responsibly. Australia has been a leader in AI since the 'expert system revolution' in the 1980s, when banks started to invest in developing smarter financial computer systems. But as other countries have scaled their efforts, Australia has lacked leadership in both the public and private sectors and has fallen way behind.

The recent and continuous release of game-changing generative AI technologies by US companies highlights Australia's lack of capability and capacity to contribute to a world dominated by these technologies. Without a game plan, Australia is at risk of being left even further behind.

Some have argued that Australia is currently in a do-or-die situation and that its lack of focus on the many challenges we need to face is wasting precious time and squandering future opportunities. Focusing on the translation of other nations' AI technologies and 'safe and responsible' AI is not enough; it is critical that we also develop homegrown AI expertise and technologies.

Australia has been lagging in government policy and funding, university research and teaching, business innovation and investment, and industry adoption over the last two decades. This exposes us to problems that can impact our future prosperity making Australia uncompetitive in national and major international markets.

Even worse, Australia's investment in digital transformation — a prerequisite to AI adoption — has paved the way for other countries with AI solutions ready for deployment to swoop in and reap significant rewards. This has been aided and abetted by a vacuum of enabling government policy for domestic AI research and innovation and a lack of university and industry collaboration and investment in deep AI research, innovation and training.

However, Australia is the lucky country, the game is not over, and we can still change our trajectory. Technology adoption takes time; and despite AI's 'magical' powers, its widespread adoption will be no different, largely because of the significant risks AI brings to business, government, individuals and society (see Figure 1).

Australia needs a coordinated and integrated approach that leverages its unfair advantage.

As a nation, we have critically important unfair advantages in the AI space that we could leverage, and if we do it quickly, we can get back in the game and lead.

We need to rapidly identify the gaps in capability and capacity, consolidate our societal values, and develop an ambitious vision and strategies to translate into action.

Australia's unfair advantages relevant to building AI capability

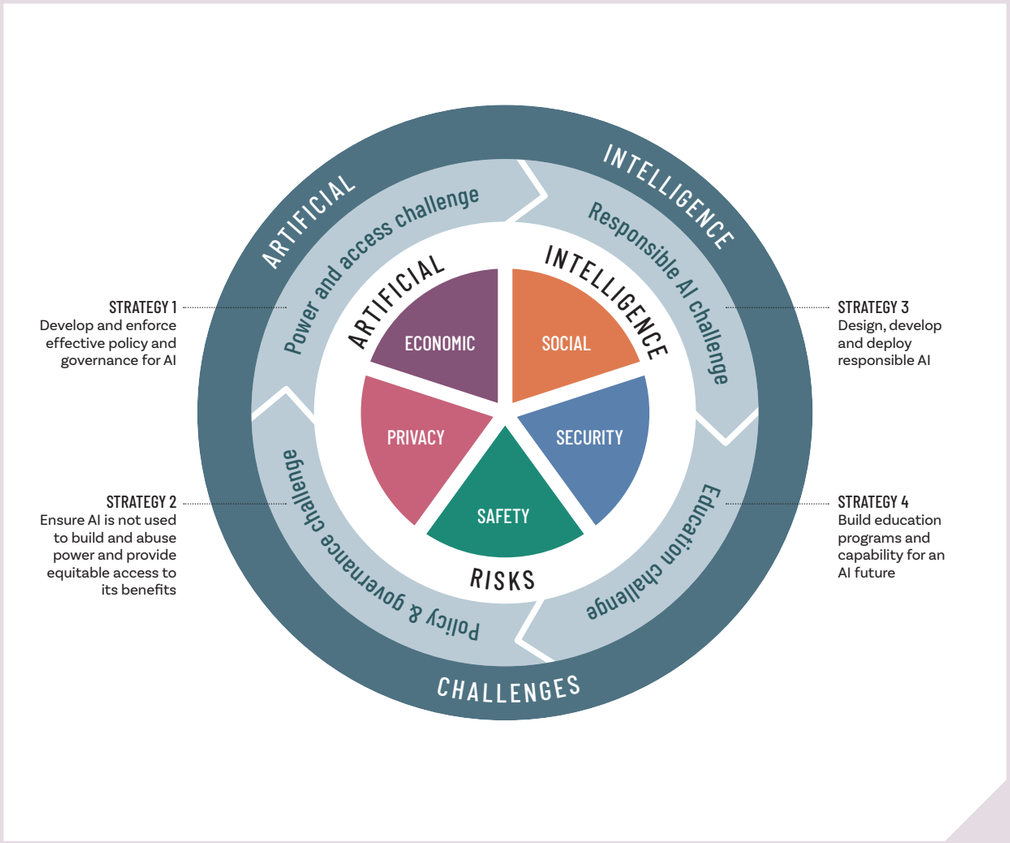


Figure 2: Risks, challenges and strategies to put Australia at the forefront of developing and using AI tools and technology.
Source: Mary-Anne Williams

and capacity lie in our rule of law; laws and governance structures; democratic principles, fairness, equity and societal diversity and inclusion; rapidly growing innovation ecosystem; cultural inclination towards technology adoption; and investment and progress in change and digital transformation. These advantages mean that Australia is well-positioned to embrace and drive AI-led and -enabled innovation. We need to develop integrated strategies that leverage our advantages to address all the risks and challenges simultaneously.

Australia's strong democratic principles represent an unfair advantage when confronting the challenges of governance and policy in relation to AI. Democracies are fundamentally about balancing interests and providing everyone with a voice, making Australia's democratic heritage a solid foundation for navigating policy debates around AI, including questions of privacy, equity and access. Unlike nations, where politics is extremely polarised, or policy may be swayed by a single entity or monopolistic businesses, the Australian approach embodies a wider set of societal perspectives, offering a model for inclusive AI regulation that reflects a broad consensus, which in turn enables rapid innovation.

Australia's existing legal, economic and political infrastructure constitutes an unfair advantage. Australia enjoys a stable political environment and a strong rules-based legal system, both of which are crucial to enforcing AI regulations effectively and maintaining public trust. Our robust economic structure, characterised by a high level of digitalisation and a strong service sector, makes the economy resilient and adaptable to the integration of AI. This strong institutional framework can attract both domestic and international AI stakeholders, providing them with the security and predictability necessary to innovate and invest.

The cultural inclination of Australians towards technology adoption also serves as a critical unfair advantage. Australians are known for their openness to new technologies, with high levels of smartphone usage, internet penetration and digital services adoption. This readiness for technological adoption can speed up the integration of AI across different sectors, making Australia a potentially attractive testing ground for new AI applications. Furthermore, this cultural trait ensures that the development of AI in this country is driven by a population that is both understanding of AI's potential and aware of its ethical implications.

Australia has an exceptional education system, which can be leveraged to develop a new generation of AI capabilities. Australian universities are breeding grounds for innovation, offering the potential to uplift the national workforce to embrace and exploit AI-led innovation. The existing strong ties between academia and industry further facilitate the practical application of AI research, turning theory and insights into tangible benefits. This relationship requires more investment to scale. There also needs to be more experimentation to learn how AI can be used to generate value for business and society.

Australia leads in critical industries where we expect AI to have the biggest impact, in particular, financial services, medicine and health. Australia's rich diversity gives it an unfair advantage in the development of safe and inclusive AI. As one of the most multicultural and harmonious societies globally, Australia can ensure that AI algorithms are trained on diverse datasets, thereby reducing algorithmic bias and improving the fairness of AI systems.

Australia's democratic principles, robust infrastructure, technological readiness, high-quality education system and multiculturalism are not just assets but provide us with significant 'unfair' advantages. Leveraged effectively, these distinctive strengths can enable Australia to overcome the challenges related to AI, and to carve out a global leadership role in AI. This is a path that relies not on matching other countries' stride for stride, but on leveraging Australia's unique strengths to create its way in a world increasingly shaped and influenced by AI.

By leveraging its unfair advantages, Australia can not only gain a leadership position in AI but also hone and strengthen these advantages to create a continuous virtuous cycle that sustains our differentiation and competitiveness.

For example, to help build a brighter future for all, Australia can improve the inclusion of more diverse societal groups like women, people with a disability and First Nations peoples. Similarly, there is considerable scope to improve our education system, not just in computing but also in many other disciplines. Since AI is a transformative technology, every industry will be affected and will require new workforce capabilities that enable people to utilise AI tools to create business and societal value and benefit.

Australia has distinctive strengths that position it uniquely to address the risks and challenges of AI and to lead in the pursuit of opportunity and national prosperity.

Essay 12 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

The $1 billion dollar question

What should Australia's responsible AI future look like?

Kingston AI Group
Australian Professors of Artificial Intelligence

Contributors to this article:
Professor Joanna Batstone FTSE
Professor Stephen Gould
Professor Anton van den Hengel FTSE
Adjunct Professor Sue Keay FTSE
Professor Dana Kulić
Professor Jie Lu
Professor Simon Lucey
Professor Michael Milford FTSE
Professor Ben Rubinstein
Professor Svetha Venkatesh FTSE
Professor Toby Walsh

If Australia invested $1 billion in AI research and development, where should the money be spent and what impact would we see?

ARTIFICIAL INTELLIGENCE (AI) IS now starting to impact every facet of human work and endeavour. AI will likely transform our world far more comprehensively and rapidly than digital technology predecessors like the internet.

Globally, AI is projected to contribute $15.7 trillion to the world's economy by 2030[39] and in Australia, CSIRO's Data61 forecasts that we could boost our economy by $315 billion from digital technologies, including AI, as early as 2028[40].

Public investments in AI are expected to play critical roles in protecting and furthering Australia's economy as well as its defence capabilities; natural resources and the environment; health, ageing, and disability sectors; and cities, towns, and infrastructure.

But Australia, despite its high-quality AI research outputs and capabilities, has been both slow to invest in AI and has invested far less than countries of similar size and wealth. As part of the 2023-24 Budget, the Australian Government announced only around $100 million to support businesses to integrate quantum and AI technologies into their operations[41].

In comparison, in 2021, Canada invested more than A$500 million to support AI commercialisation, standards, talent and research[42]; and the government of Singapore has invested about A$565 million in artificial intelligence research and development over the last five years[43].

Given current strategic investments in AI abroad, it's timely to investigate how a similar investment in Australia would look.

If Australia invested $1 billion in AI research and development, where should the money be spent and what impact would we see?

A national initiative for Australian AI research and innovation

Australia has globally recognised AI research expertise in its universities, but despite this, the country lags behind the rest of the world in building its domestic AI capability.

Many of our top AI researchers leave Australia because of the lack of research funding here. University funding is a catalyst for education and innovation excellence.

Kingston AI Group
Australian Professors of Artificial Intelligence

Top researchers supervise numerous PhD students, and design and deliver high-quality curricula to educate thousands of undergraduate students who will form our future workforce. They build smart teams, compete with the best in the world, and do great ambitious things that benefit all of society.

Investment in Australian research is really an investment in people, and we already know it delivers economic results; for every A$1 the Australian Research Council puts into research, it returns A$3.30 in economic output back to the Australian community[44].

So, how do we turbocharge Australian AI research to deliver for the country?

We could start by funding a national AI initiative focused on unleashing the value of our university AI research and ensuring it supports Australia in capturing the benefits of the AI revolution. The centrepiece of this initiative would be a coordinated network of AI research and innovation centres across the country, co-located with our leading AI universities in:

New South Wales / Australian Capital Territory — a centre would be integrated with the Australian National University (ANU), UNSW Sydney, and University of Technology Sydney

Victoria — with The University of Melbourne, Deakin University and Monash University

South Australia — with the Australian Institute for Machine Learning (AIML) at The University of Adelaide

Queensland — with Griffith University, Queensland University of Technology, and University of Queensland

And possibly other states.

'De-risking' AI and developing Australia's AI ecosystems

Universities are key to driving innovation and business development in all high-tech sectors, with knowledge spillover being a vital element of innovation precincts internationally. Many startup unicorns (companies that are valued at US$1 billion or more in a new funding round) have a campus origin.

To extract maximum value from our university research assets, Australia needs more AI entrepreneurs; but we have to de-risk funding tech innovation and create and fund prime innovation systems. The best way to de-risk a technology is for governments to invest strongly and make it as robust as possible. It's what happened in Silicon Valley, in Israel and in other nations, but we've talked ourselves out of doing it here.

Australia's AI venture capital market is smaller than that of other nations. Globally, billions of dollars in venture funding are pouring into AI startups and scaleups with the International Data Corporation expecting funding on AI technologies to increase to $110 billion per annum by 2024, double what it was in 2020[45]. In contrast, Australia has only a few AI-specialised venture funds whose offerings are typically significantly lower than those provided internationally.

Creating ecosystems between universities, startups, and industry that allow for AI innovation should be another goal of this $1 billion dollar investment. Setting up localised ecosystems that have substantive budgets that are committed for at least five to 10-year periods, will help attract and retain the world's most talented people.

These ecosystems will enable leading AI researchers to pursue somewhat unorthodox programs of talent recruitment and training, and to provide medium-term career development options and trajectories that would not be feasible in the current university system.

Creating partnerships and sovereign datasets for critical sectors

Australia's comprehensive public health service is the envy of much of the world, but our centralised health data assets are what can make Australia a leader in medical AI.

If we can look at partnering investments in core research technology with other investments, then we can start to lay out a translational journey of core technology investment in universities that has an impact on the quality of delivery of health care or on our ability to reach disadvantaged communities with digital health tools.

Investing in pooled/shared datasets and dedicated research computing resources will be necessary to drive the ethical, curated development of Australian AI in health, defence, and other priority areas.

These sovereign datasets are integral to the stability of Australia's AI future.

Investing in education and public engagement

AI isn't a discrete standalone technology restricted to a specialised workforce. It's a general-purpose technology that's starting to roll out across all industries and sectors.

Therefore, investment in education and AI literacy from the primary school to university levels is also key to creating the homegrown, sovereign AI capabilities needed for Australia to keep up in an increasingly competitive, AI-impacted global economy.

But it's not just formal education; there is also a need for extensive continuing professional development programs so our current workforce can upskill, reskill and remain occupationally nimble as AI technology disrupts workplaces and industries nation-wide. We are certain that Australia will benefit the most when it is a nation of expert AI creators, not just consumers.

A broad, systematic campaign that targets every sector, with extra depth for priority areas— such as government and primary / secondary education—is paramount to instilling enthusiasm toward AI among the broader public and stakeholders.

Given AI's transformative power to affect an endless array of sectors and industries, everyone must be involved in the conversation around its growth and impact here in Australia.

Essay 13 of 13 from the series:
Responsible AI: Your questions answered

The University of Adelaide | Australian Institute for Machine Learning
ATSE Australian Academy of Technological Sciences & Engineering

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.

Acronyms & abbreviations

AI	Artificial intelligence
AIML	Australian Institute for Machine Learning
ATSE	Australian Academy of Technological Sciences and Engineering
FWC	Fair Work Commission
GDPR	General Data Protection Regulation
HIVE	Health in a Virtual Environment
LLMs	Large Language Models
NAIC	National Artificial Intelligence Centre
NLP	Natural Language Processing
STEM	Science, Technology, Engineering and Mathematics
TRL	Technology Readiness Level

Endnotes

1 Qinghua Lu, Liming Zhu, Jon Whittle, Sherry Xu. *Responsible AI: Best Practices for Creating Trustworthy AI Systems*. Addison Wesley , 2023.

2 Jon Whittle. CSIRO Presents: Everyday AI, podcast hosted by Professor Jon Whittle. CSIRO . 2023. https://www.csiro.au/en/news/podcasts/everyday-ai-podcast (accessed 14 Sep2023).

3 Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner. Machine Bias. *ProPublica* 2016.https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 14 Sep2023).

4 Billy Perrigo. OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. *TIME Magazine* 2023.https://time.com/6247678/openai-chatgpt-kenya-workers/ (accessed 14 Sep2023).

5 CSIRO. Australia's AI ecosystem momentum report. 2023https://www.csiro.au/en/work-with-us/industries/technology/national-ai-centre/ai-ecosystem-report (accessed 11 Sep2023).

6 Lucy Stone. How your local garbage truck is helping fix south-east Queensland's pothole problems. ABC News. 2022.https://www.abc.net.au/news/2022-05-18/queensland-potholes-fixed-with-garbage-trucks-and-ai/101076146 (accessed 11 Sep2023).

7 East Metropolitan Health Service. HIVE (Health in a Virtual Environment) . 2021.https://www.facebook.com/EastMetropolitanHealthService/videos/686127998735106/ (accessed 11 Sep2023).

8 CSIRO. AI transforms Kakadu management. 2019.https://www.csiro.au/en/news/all/news/2019/november/ai-transforms-kakadu-management (accessed 11 Sep2023).

9 WIPO. Global Innovation Index Report. 2021wipo.int/edocs/pubdocs/en/wipo_pub_gii_2021.pdf (accessed 11 Sep2023).

10 Mining Technology. Australia continues to dominate the use of autonomous haul trucks. 2021. https://www.mining-technology.com/comment/australia-autonomous-haul-trucks/ (accessed 11 Sep2023).

11 Natasha May. Labour shortages may cause farmers to lose crops and consumers to face food price hikes. Guardian. 2021.https://www.theguardian.com/australia-news/2021/oct/21/labour-shortages-may-cause-farmers-to-lose-crops-and-consumers-to-face-food-price-hikes (accessed 11 Sep2023).

12 Lin Hatfield Dodds. Budget 2023: The shot in the arm Australia's care economy badly needed. Mandarin. 2023.https://www.themandarin.com.au/219934-budget-is-the-shot-in-the-arm-australias-care-economy-badly-needs/ (accessed 11 Sep2023).

13 Michael Read. NDIS on track for $5.7b budget blowout. Australian Financial Review. 2023.https://www.afr.com/politics/ndis-on-track-for-5-7b-budget-blowout-20230322-p5cuaq (accessed 11 Sep2023).

14 AuManufacturing. Hopes for the federal budget to revive Australia's manufacturing sector. 2023. https://www.aumanufacturing.com.au/hopes-for-the-federal-budget-to-revive-australias-manufacturing-sector (accessed 11 Sep2023).

15 Australian Government. National Defence: Defence Strategic Review 2023. 2023https://www.defence.gov.au/about/reviews-inquiries/defence-strategic-review (accessed 11 Sep2023).

16 Australian Commission on Safety and Quality in Health Care. The state of patient safety and quality in Australian hospitals 2019. https://www.safetyandquality.gov.au/publications-and-resources/resource-library/state-patient-safety-and-quality-australian-hospitals-2019 (accessed 11 Sep2023).

17 Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen L-C et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health 2022; 4: e406—e414.

18 AI Incident Database. Artificial Intelligence Can Improve Health Care—but Not Without Human Oversight. 2021.https://incidentdatabase.ai/cite/123/#r2705 (accessed 11 Sep2023).

19 Tim Simonite. An Algorithm That Predicts Deadly Infections Is Often Flawed. WIRED. 2021. https://www.wired.com/story/algorithm-predicts-deadly-infections-often-flawed/ (accessed 11 Sep2023).

20 Scott MC. Defining and Diagnosing Sepsis. *Emerg Med Clin North Am* 2017; 35: 1—9.

21 Stoyanovich J, Abiteboul S, Howe B, Jagadish H V., Schelter S. Responsible data management. *Commun* ACM 2022; 65: 64—74.

22 Uri Gal. ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned. The Conversation. 2023.https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283 (accessed 11 Sep2023).

23 Kaphle S, Kornhaber R, Hunt S, Watson R, Cleary M. When is public private? Tweets, privacy and consent in health research. *Nurse Educ Pra*ct 2022; 63: 103396.

24 Cai Z, Xiong Z, Xu H, Wang P, Li W, Pan Y. Generative Adversarial Networks. *ACM Comput Surv* 2022; 54: 1—38.

25 Borisov V and SK and LT and PM and KG. Language Models are Realistic Tabular Data Generators . arXiv preprint arXiv:221006280 2022.https://openreview.net/pdf?id=cEygmQNOeI (accessed 11 Sep2023).

26 Yan C, Yan Y, Wan Z, Zhang Z, Omberg L, Guinney J et al. A Multifaceted benchmarking of synthetic electronic health record generation models. *Nat Commun* 2022; 13: 7609.

27 Jayoung Kim, Chaejeong Lee, Noseong Park. STaSy: Score-based Tabular data Synthesis. 2023. https://openreview.net/pdf?id=1mNssCWt_v (accessed 11 Sep2023).

28 Tal Ridnik, Emanuel Ben-Baruch, Asaf Noy, Lihi Zelnik-Manor. ImageNet-21K Pretraining for the Masses. 2023.https://openreview.net/pdf?id=Zkj_VcZ6oI (accessed 11 Sep2023).

29 Rishi Bommasani et.al. On the Opportunities and Risks of Foundation Models. 2022.https://arxiv.org/abs/2108.07258 (accessed 11 Sep2023).

30 Timothy Nguyen, Roman Novak, Lechao Xiao, Jaehoon Lee. Dataset Distillation with Infinitely Wide Convolutional Networks. 2021.https://proceedings.neurips.cc/paper/2021/file/299a23a2291e2126b91d54f3601ec162-Paper.pdf (accessed 11 Sep2023).

31 He L, Wang X, Wang D, Zou H, Yin H, Xu G. Simplifying Graph-based Collaborative Filtering for Recommendation. In: *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. ACM: New York, NY, USA, 2023, pp 60—68.

32 Gou J, Yu B, Maybank SJ, Tao D. Knowledge Distillation: A Survey. *Int J Comput Vis* 2021; 129: 1789—1819.

33 Australian Government. Manufacturing and Associated Industries and Occupations Award 2020. Fair Work Ombudsman. 2020.https://library.fairwork.gov.au/award/?krn=MA000010#_Toc141266026 (accessed 14 Sep2023).

34 Goran Roos. For Want Of Industry Policy, Our Living Standards Are Set To Fall. Committee for Economic Development of Australia. 2014.https://www.ceda.com.au/NewsAndResources/Opinion/Economy/For-want-of-industry-policy,-our-living-standards (accessed 28 Sep2023).

35 Paul Smith. Why Australian businesses must spend more on innovation to survive. *The Australian Financial Review*. 2023.https://www.afr.com/technology/why-australian-businesses-must-spend-more-on-innovation-to-survive-20230324-p5cv47 (accessed 28 Sep2023).

36 Universities Australia. Research and Innovation. 2021.https://universitiesaustralia.edu.au/policy-submissions/research-innovations/ (accessed 28 Sep2023).

37 Australian Government. Australian values. Department of Home Affairs. 2010.https://www.homeaffairs.gov.au/about-us/our-portfolios/social-cohesion/australian-values (accessed 11 Sep2023).

38 Australian Government. Australia's AI Ethics Principles. Department of Industry, Science and Resources. 2023.https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles (accessed 11 Sep2023).

39 PWC. *Sizing the prize PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution* What's the real value of AI for your business and how can you capitalise? 2017.https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html (accessed 6 Oct2023).

40 Misha Schubert. Why must Australia invest in AI research and development? 2022.https://www.adelaide.edu.au/aiml/news/list/2022/08/03/why-must-australia-invest-in-ai-research-and-development (accessed 6 Oct2023).

41 Australian Government. Investments to grow Australia's critical technologies industries. Department of Industry, Sciences and Resources. 2023.https://www.industry.gov.au/news/investments-grow-australias-critical-technologies-industries (accessed 6 Oct2023).

42 Government of Canada. Government of Canada invests in responsible artificial intelligence research at the Université de Montréal. Innovation, Science and Economic Development Canada. 2023.https://www.canada.ca/en/innovation-science-economic-development/news/2023/06/government-of-canada-invests-in-responsible-artificial-intelligence-research-at-the-universite-de-montreal.html (accessed 6 Oct2023).

43 Wu Xinyi. About S$500 million invested in AI innovation in last 5 years: Josephine Teo. *Singapore Business Times*. 2023.https://www.businesstimes.com.sg/startups-tech/about-s500-million-invested-ai-innovation-last-5-years-josephine-teo (accessed 6 Oct2023).

44 Australian Government. New report shows ARC-funded research is delivering exceptional outcomes for Australia. Australian Research Council. 2023.https://www.arc.gov.au/news-publications/media/media-releases/new-report-shows-arc-funded-research-delivering-exceptional-outcomes-australia (accessed 6 Oct2023).

45 Slonim A, Nicholson K. Special Report: Artificial Intelligence — Your Questions Answered. 2022https://www.aspi.org.au/report/artificial-intelligence-your-questions-answered (accessed 6 Oct2023).

46 NSW Government. NSW Artificial Intelligence Assurance Framework. 2020.https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework (accessed 11 Sep2023).

Printed on ecoStar+ made with 100% recycled post-consumer waste. ecoStar+ is an environmentally responsible carbon neutral paper and is manufactured in a process chlorine free environment under the ISO 14001 environmental management system. Printed using vegetable-based low VOC (Volatile Organic Compounds) inks.

RECYCLED Paper made from recycled material FSC® C010983

ATSE Australian Academy of Technological Sciences & Engineering

THE UNIVERSITY of ADELAIDE

Australian Institute for Machine Learning