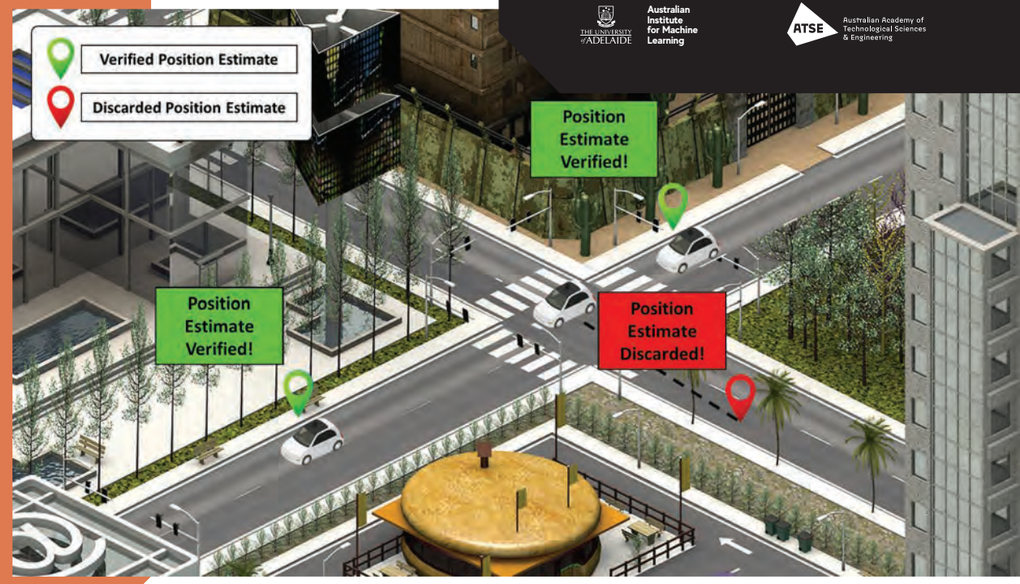
Essay 3 of 13 from the series:
Responsible AI: Your questions answered

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).

Essays

SECTION 1: INTRODUCTION

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

SECTION 3: WHAT ARE THE NEXT STEPS?

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

Responsible AI

Your questions answered

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

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

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

DESIGN AND PRODUCTION
Elizabeth Geddes, Edwyn Shiell and Alexandra Horvat

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

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

Responsible AI

Your questions answered

Australian Academy of Technological Sciences & Engineering

THE UNIVERSITY of ADELAIDE

Australian Institute for Machine Learning