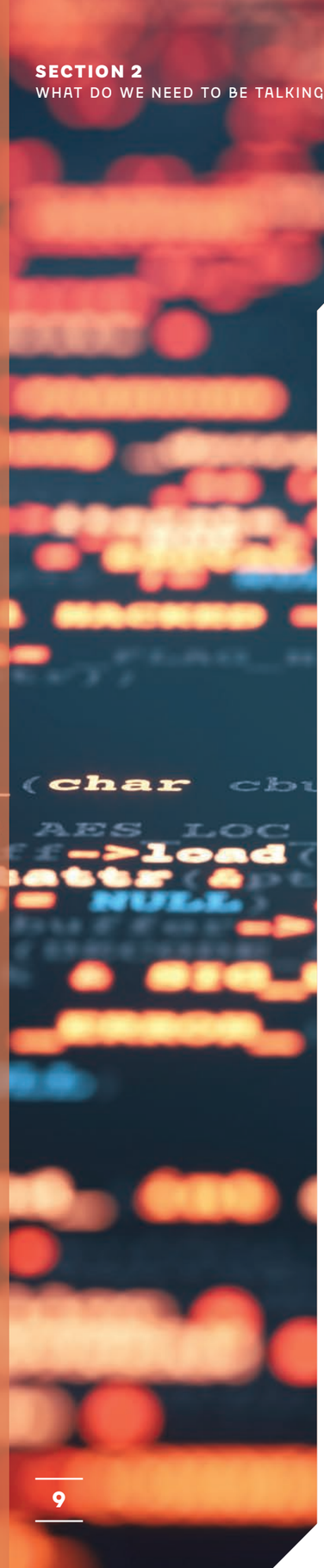

Essay 4 of 13 from the series:
Responsible AI: Your questions answered

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

Responsible 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).

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