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.

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