Horizon Scanning Series

The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing

Data Integrity, Standards and Ethics

This input paper was prepared by Data61

Suggested Citation

Data61 (2018). Data Integrity, Standards and Ethics. Input paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing" on behalf of the Australian Council of Learned Academies, www.acola.org.

Abstract

Data is essential to artificial intelligence (AI). Machine learning algorithms require datasets to learn from and skilled practitioners of data analysis to develop the AI itself. Policies that affect data collection and sharing inevitably affect the development of AI.

As the costs associated with the collection, storage and analysis of data reduces, a rapid step change is occurring in the exploration and uptake of digital technologies, particularly in the area of AI.

Advances in core fields of data driven AI, including machine learning, image processing, predictive analytics and automation are seeing the complexity and capability of systems change at an exponential rate, with computers now able to more rapidly solve complex problems, often using self-generated strategies and with little instruction or guidance from human beings.

Artificial intelligence holds remarkable promise in its power to transform human existence and deliver improved quality of life. Every form of industry, and activity, is rapidly becoming a data business, and organisations are expected to continue to become more data-driven in the coming decades.

Very different approaches to data collection and incorporation into AI are emerging around the world. In the text of both the General Data Protection Regulation (GDPR) and the European Commission's AI development plans, The European Union has placed ethical AI with privacy and data collection restraints at the forefront of its strategy, with the aim of creating a competitive advantage out of ethical AI developed for the common social good.[i] A core part of this relates to the ethical collection of data.

Conversely, some investment analysts have pointed toward a lack of effective privacy controls in the collection of data in China as advantageous, giving the country an edge in the development of certain types of AI—though even in China, more privacy protections are anticipated.[ii]

In Australia, there are a number of concurrent processes occurring that will shape the development of data collection. In response to an inquiry into data availability and use, the Australian Government is developing reforms to allow greater use of data while ensuring the protection of individual rights to transparency and control over their own data[iii]. In addition, the launch of the data.gov.au website provides a platform to access over 25,000 government datasets and sets a precedent for data-resource sharing in Australia.

Australia is well positioned to be an early beneficiary or casualty of the rise of AI and with accessible data at its core it is increasingly important to invest in infrastructure to support innovation. A recent report by McKinsey identified the key sectors most likely to benefit from AI as early adopters including manufacturing, financial services, resources and health related companies and service providers, all significant economic and employment industries for Australia.[iv] AI holds the potential to unlock new opportunities for the development of personalised services, solutions and responses, targeted to the specific interests and needs of the individual. This includes areas such as healthcare, consumer products and government services.

Whilst these technologies have the power to help us in achieving remarkable new insight and outcomes, there will be significant legal, ethical and social challenges to contend with, as we adapt to an AI enabled world.

To keep pace with this change, the field of data science and informatics is continuing to explode, with growing demand for skilled data science experts, engineers and cybersecurity expertise at an all-time high.

Developments in our ability to rapidly collect, analyse and safely share data between individuals and organisations – without compromising individual privacy – will undoubtedly support the development of such AI enabled, targeted services, by enabling organisations the ability to understand our particular needs and characteristics, from observing our data.

For industries, the ability to access a broader base of information to support decision-making, understand patterns and anticipate needs will also enable new levels of efficiency, coordination and production, offering new economic opportunities and outcomes.

The impact of advances in data processing power is already evident in industry. Examples in the Australian context include the use of facial recognition technology to identify passengers moving through Australian airports and the use of virtual service agents, and voice authentication technology by a number of Australian banks – such as the National Australia Bank – enabling customers to authorize transactions by voice biometrics alone.[v]

The financial incentive for the development of AI is substantial. AlphaBeta estimates automation to be a $2.2 trillion market opportunity for Australia alone to 2030, through adoption of AI, robotics and IoT to unlock new levels of productivity support our industry base and workforce in transitioning[vi].

Realising the potential for AI in our industries and communities will require some significant advances in data processing, our approach to data sharing and interoperability, new levels of data security and privacy protection, and a deep exploration of the areas of AI ethics and the impact of bias.

This chapter explores some of the key opportunities and challenges facing AI in Australia, with an emphasis on issues related to data-collection and sharing.

Sections

Data is the feedstock of AI, and the quality, complexity, availability and origins of data will all have a critical impact on the accuracy and validity of the AI-based systems it powers.

However, a key factor affecting the usability of data in analytics is its potential to be inconsistent. This includes inconsistency due to the poor quality of the data collected, the way the data has been recorded, or the potential for the data that is recorded to have been impacted by bias. This may be the bias of the contributor, whose data is captured, or bias of the collector. Our use of different data collection methods – ranging from verbal information, to paper documents, to sensors networks – inevitably results in a range in the level of quality and reliability of data.

In order for a dataset to be consistent and reliable in terms of its potential to contribute to an AI or machine learning based technology, we need to be able to trust in the integrity of the data, meaning we need to be sure that appropriate quality controls and processes – such as ethics and consent – are in place when it was collected and that the methodology of collection is well documented and available to users of the data.

Data integrity and standards

Data Integrity and Standards

Setting in place accepted and common standards to ensure the integrity of health data will act to accelerate the potential for data sharing and linkage, in turn offering opportunity for the development of new treatments, technologies and predictive systems, targeting individual and system-wide needs[vii].
The use of common metadata registries, such as those conforming with ISO11179, will facilitate the accurate capture and management of descriptive and structural health metadata (including assumptions and methodologies used in data capture) will aid more precise data combination and linkage, reuse of data and its governance.

The data integrity principles set in place by the Australian Bureau of Statistics (ABS) provide a succinct basis on which to base general Australian standards. The prerequisites of data integrity are "objectivity in the collection, compilation and dissemination of data to ensure unbiased statistics which are not subject to confidentiality breaches or premature release"[viii]. Adherence to these principles is largely supported by legislative frameworks.

Data Standards

In 2017, the World Economic Forum (WEF) released its Global Risks Report, which identified AI as a key risk in part due to the slow pace of the development and setting of globally accepted standards and norms for its use and application[ix]. In particular, the WEF pointed to risks around the development of machines that can substitute for humans in tasks associated with thinking, multitasking, and fine motor skills, generating new risks in decision- making, security and governance.

Organisations within Australia are keenly observing practice overseas, and by leading AI companies in identifying and setting potential best practice standards for the use of AI. Additionally, international standards will affect various Australian industries. The European Union General Data Protection Regulation will apply to all Australian organisations that have an establishment, trade with or collect information on the EU. It will be increasingly important for Australian industries working internationally to understand emerging requirements.

In its 2018 Budget, the Australian Government prioritised the development of a National AI Technology Roadmap and AI Ethics Framework, in order to identify emerging opportunities and challenges for Australia in the adoption of AI, and ensure adequate measures are in place to ensure its ethical use and application.

Achieving greater unification in global AI and data use standards across industry and government will help to minimize the potential risks from its use and adoption. In particular, establishing aligned settings for data privacy and confidentiality will greatly support organisations around the world in ensuring decisions made or tasks completed using AI do not result in unintended, or potentially irreversible, consequences.

Technologies enabling AI

Federating Data

Government and industry are significant data generators, however in most cases, the true power and potential that data could offer for insight into their operations, customers or constituents

remains untapped and underutilized due to challenges in data linkage – in particular the potential for beaches of privacy.

There is a tremendous opportunity for government and industry to share and leverage datasets across organisations, to build more powerful and insightful predictive models. Doing so has traditionally required co-locating all available data, or bringing a common format, which is often difficult and inefficient for legal, contractual and practical reasons.

Federated machine learning allows data owners to work together to build shared predictive models from data, without having to physically bring that data into one place. Instead they share information only about how the model performs on the data they own. This distributed optimisation approach means that data from multiple organisations can be drawn on and reflected in a single model that generates insight and makes predictions as if it has access to all the data.

Australia has a significant opportunity to establish an ecosystem of federated machine learning technologies across Government and industry, based on the use of open formats and application programming interfaces, that will encourage and support innovation in AI development and support new market development.

The principle of federated data has already been successfully demonstrated and is an emerging model in use globally, and by Australian Government agencies. Examples include ATO's Standard Business Reporting (SBR)[x] platform and Australian Government's NationalMap[xi] federated spatial visualisation platform.

CSIRO's Data61 is also currently working with the Department of Prime Minister and Cabinet on a project to improve the searchability, quality, indexing and discoverability of available datasets. The software developed - known as MAGDA (Making Australian Government Data Available) – supports better ways for locating and accessing data from across the country and this data can be used together with personal data for more targeted analytics.

1.1. Privacy Preserving Machine Learning

Traditional machine learning methods require data owners to expose or give away their confidential or potentially sensitive data to those building the models. This requirement generates serious privacy and competitive implications, as the data may contain trade secrets, or private information relating to individuals.

Development of the Privacy Act (1998) may be required to include provisions in their principles for the use of Australian data in developing AI as they have done for use of information classified as sensitive such as biometric data and criminal records.

Technologies will be required to enable access and analysis of data, where it is stored, rather than requiring it to be shared across organisational boundaries. The development of tools and technologies able to send in machine learning algorithms to where the data is stored, or enabling machine learning to be carried out on encrypted data in a central location, provide strong potential to overcome these barriers. These technologies will also further enhance the applicability, robustness and attractiveness of federated machine learning.

1.1. Deep Learning

The rapid acceleration in exploration and adoption of AI is largely driven by advances in "deep-learning" technology, that is the ability to process large volumes of information and rapidly recognize patterns, with little or no human involvement. Leading companies, such as IBM, Apple, Facebook and Google heavily utilise deep learning in their service and product offering.

For example, Apple's Siri and Google Assistant both utilise deep learning to train on, recognise and respond to our voice commands.

The development of these types of technologies is heavily dependent on the use of datasets and the quality of the products will be determined by the ability to access representative and unbiased datasets.

Additionally, as the cost of computing continues to reduce, companies of every size will be able to access deep-learning powered software and the ability to explore patterns in their own data, identifying new patterns they'd previously been unable to spot through human review alone.

1.1. Causal Inference and Data Driven Policy

The challenge in deploying AI technologies in service delivery is that predictive algorithms are making predictions with an unfixed and ever changing degree of certainty. Before those algorithms can be trusted to endorse an action that will impact a citizen (e.g. withhold welfare payments) causality with a high degree of confidence is required.

Causal inference is a vital area of research in machine learning that goes beyond prediction to try and understand why an event occurs, and how to influence it. As such, it is critical for government and industry to move from making predictions, to utilising the information and insight it gains from analysing available data as an evidence base in the development of future government service delivery.

Taking full advantage of causal inference will require new expertise and ways of thinking for decision makers, and will ultimately change how industry and government plans, acts and reviews its strategies and policies.

Ethics and AI

The opportunities for AI machine learning to support rapid and personalised predictions will continue to grow as more data becomes available. These technologies will enable us to better understand how markets and industries perform, how government services are best delivered, and ultimately lead to an improved understanding, at the individual and population level, of the factors that drive our economy and society.

Despite the enormous potential for benefit, the ethical use of AI requires robust frameworks and guidelines to ensure that we are protected from infringements to our privacy, wellbeing and human rights.

Ethical Use of data

With machine learning playing an increasing role in our everyday lives, one of the major concerns is the use of data – including personal information – in the development of predictive models. This may include the use of data for profiling, without consideration of historical biases affecting such data and the use of algorithms that may unfairly generalise, based on attributes such as age, gender and race.

There remains much research to be done on ways to design suitable objectives into machine learning approaches which will consider the (often conflicting) ethical imperatives, such as reducing racial, gender or ideological bias, valuing privacy and ensuring reliability.

A 2018 report released by the Australian Government attempts to address ethical issues associated with the use of data across government, community and industry. The main goals of

the reforms are to ensure that the necessary frameworks are in place to protect the privacy of Australians, to establish the best use of our collective data and to develop government oversight on the way that all sectors use data. These reforms are intended to provide greater access and use of Australian data generate and promote innovation while adhering to best practice ethical use. The government has committed to the following[III]:

1. A Consumer Data Right as a new competition and consumer measure to allow consumers to harness and have greater control over their data.

2. A National Data Commissioner to support a new data sharing and release framework and oversee the integrity of data sharing and release activities of Commonwealth agencies.

3. A legislative package will streamline data sharing and release, subject to strict data privacy and confidentiality provisions.

There are also existing frameworks developed for government departments that utilise automated decisions. These frameworks will by design need to deal with the ethical sharing of data and privacy concerns, as well as accountability for improper use of data.

A number of laws stipulate that relevant ministers are accountable for decisions made by automated systems.[xii] In addition, a 2004 report prepared for the Attorney General's office outlines 27 principles for government departments that utilize automated decision-making processes.[xiii] These include drawing a clear distinction between decisions that require discretion which should not be automated, and situations which require a large volume of decisions and the facts are already established, with the latter being suitable for automation. An inter-agency report from 2007 also covers this issue in detail, and highlights the need for external agencies to be involved in shaping automated decision policies and being able to review the data involved.[xiv]

Preserving Privacy and Confidentiality

Techniques are emerging which hold promise to enable confidential data to be accessed for insight, without putting its content at risk. For example, new federated, privacy-preserving analytics methods developed by Data61 – known as Confidential Computing – enable encrypted queries and responses to be sent and received by third party datasets, without requiring raw data to be shared or exposed.

The technique combines three underlying technologies - distributed machine learning, homomorphic encryption and secure multiparty computing - to provide a platform to draw insight across organisations, or from individuals, without requiring direct access to raw data.

As an example of the potential applications for this capability to precision medicine, there is potential that confidential computing could enable an individual - holding all of their own health information on their own computer or smartphone (such as their genome or personal health record) - to securely map their data against a third party database or health provider's system, and receive personalized results back, without their private data ever being disclosed to another party.

Building on our Strengths

Australia has a significant opportunity to be a leading technology developer and adopter of data-driven AI systems and technologies.

Australia has established, and globally recognised strengths in some of the key, data-driven capability areas core to AI, including data sharing or federation, trustworthy systems, machine learning, image analytics, natural language processing and automation. In addition to this

Australian is culturally diverse and serves as a desirable population in which to gather robust datasets, a core requirement for un-biased and effective AI.

We also have deep research capability and industry strength in some of the primary sectors expected to be impacted by AI, including the energy, manufacturing, agriculture and health sectors.

Realising this opportunity for Australia will require investing in a focused and coordinated effort, linking Australia's national AI capabilities and domain knowledge to the particular challenges and use cases for AI identify by industry and government.

This will only be achieved by actively seeding and nurturing a deep partnership between government, Australia's leading AI and digital researchers and industry, aimed at identifying and driving opportunities for rapid technology experimentation and adoption, as a national priority.

The Australian Government has already taken significant steps towards adopting a more federated approach to data sharing and management, enabling coordination and accessibility, but where control of the raw data continues to reside with its custodian organisation.

Supported through the National Innovation and Science Agenda (NISA), initiatives such as Platforms for Open Data (PFOD) are enabling Australian Government agencies to work with CSIRO's Data61 to test and validate techniques for allowing trusted access to high-value, Government datasets, whilst preserving the data's confidentiality and integrity

In order for Australia to take full advantage of the opportunities presented by data driven AI, governments, businesses and the community will need to strengthen their levels of awareness, adoption and acceptance of AI's use, and that will require a deeper level of trust in the integrity of AI based systems. There is a role for researchers, companies and governments to ensure appropriate safeguards are in place in the development and deployment of AI, to ensure opportunities are maximised, without that important trust being compromised.

.

References

[i] European Commission, 2018 "*Artificial Intelligence for Europe*", https://ec.europa.eu/digital-single-market/en/news/factsheet-artificial-intelligence-europe

[ii] Nee Lee, Yen. 2018. "*China will win the A.I. race, according to Credit Suisse."* CNBC. https://www.cnbc.com/2018/03/22/credit-suisse-china-will-win-the-ai-race-due-to-lack-of-serious-laws-on-data-protection.html

[iii] Commonwealth of Australia, Department of the Prime Minister and Cabinet 2018, The Australian Government's response to the Productivity Commission Data Availability and Use Inquiry

[iv]McKinsey and Company 2017, "*Artificial Intelligence – The Next Digital Frontier",* https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

[v] Australian Financial Review, 2017 "The four industries making best use of artificial intelligence", December 4 2017: https://www.afr.com/leadership/the-four-industries-making-best-use-of-artificial-intelligence-20171130-gzw0br

[vii] Commonwealth Scientific and Industrial Research Organisation 2017, *Medical Technology and Pharmaceuticals: A Roadmap for unlocking future growth opportunities for Australia*, https://www.csiro.au/en/Do-business/Futures/Reports/Medical-Technologies-and-Pharmaceuticals-Roadmap, pg 16

[viii] The Australian Bureau of Statistics, 2007, Information Paper: Quality Dimensions of the Australian National Accounts

[ix]World Economic Forum 2017, "*The Global Risks Report 2017"* 12th Edition, http://www3.weforum.org/docs/GRR17_Report_web.pdf

[x] http://www.sbr.gov.au

[xi] https://nationalmap.gov.au

[xii] http://www.abc.net.au/news/2017-07-21/algorithms-can-make-decisions-on-behalf-of-federal-ministers/8704858

[xiii] Administrative Review Council. Automated Assistance in Administrative Decision Making. 2004. https://www.arc.ag.gov.au/Documents/AAADMreportPDF.pdf

[xiv] https://www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf