THE EFFECTIVE AND ETHICAL DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

AN OPPORTUNITY TO IMPROVE OUR WELLBEING

EXTRACT

Full report available at
www.acola.org

HORIZON SCANNING

ACOLA
AUSTRALIAN COUNCIL OF LEARNED ACADEMIES

EXPERT WORKING GROUP

Professor Toby Walsh FAA (Co-chair)
Professor Neil Levy FAHA (Co-chair)
Professor Genevieve Bell FTSE
Professor Anthony Elliott FASSA
Professor James Maclaurin
Professor Iven Mareels FTSE
Professor Fiona Wood AM FAHMS

ISBN 978-0-6483303-2-5 (print)
ISBN 978-0-6483303-3-2 (digital)

DATE OF PUBLICATION

July 2019

PUBLISHER

Australian Council of Learned Academies
Level 6, 436 St Kilda Road
Melbourne Victoria 3004 Australia
Telephone: +61 (0)3 9864 0923
www.acola.org

SUGGESTED CITATION

Walsh, T., Levy, N., Bell, G., Elliott, A., Maclaurin, J., Mareels, I.M.Y., Wood, F.M., (2019) *The effective and ethical development of artificial intelligence: An opportunity to improve our wellbeing*. Report for the Australian Council of Learned Academies, www.acola.org.

ACKNOWLEDGEMENT OF COUNTRY

ACOLA acknowledges the Traditional Owners and custodians of the lands on which our company is located and where we conduct our business. We pay our respects to Elders past, present and emerging.

REPORT DESIGN

Lyrebird
jo@lyrebirddesign.com

THE EFFECTIVE AND ETHICAL DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

AN OPPORTUNITY TO IMPROVE OUR WELLBEING

AUTHORS

Professor Toby Walsh FAA
Professor Neil Levy FAHA
Professor Genevieve Bell FTSE
Professor Anthony Elliott FASSA
Professor James Maclaurin
Professor Iven Mareels FTSE
Professor Fiona Wood AM FAHMS

Supported by Dr Alexandra James, Dr Benjamin Nicoll, Dr Marc Rands, Michelle Steeper, Dr Lauren Palmer and the generous contributions of many experts throughout Australia, New Zealand and internationally as acknowledged throughout the report. A full list of contributors can be found in the written submissions section of the report.

PROJECT MANAGEMENT

Dr Lauren Palmer
Dr Angus Henderson

HORIZON SCANNING

ACOLA™
AUSTRALIAN COUNCIL OF LEARNED ACADEMIES

ACOLA
AUSTRALIAN COUNCIL OF LEARNED ACADEMIES

Working Together

The Australian Council of Learned Academies (ACOLA) combines the strengths of the four Australian Learned Academies

The Australian Academy of the Humanities (AAH) is the national body for the humanities in Australia, championing the contribution that humanities, arts and culture make to national life. It provides independent and authoritative advice, including to government, to ensure ethical, historical and cultural perspectives inform discussions regarding Australia's future challenges and opportunities. It promotes and recognises excellence in the disciplines that provide the nation's expertise in culture, history, languages, linguistics, philosophy and ethics, archaeology and heritage. The Academy plays a unique role in promoting international engagement and research collaboration, and investing in the next generation of humanities researchers.

www.humanities.org.au

Australian Academy of Science

The Australian Academy of Science (AAS) is a private organisation established by Royal Charter in 1954. It comprises more than 500 of Australia's leading scientists, elected for outstanding contributions to the life sciences and physical sciences. The Academy recognises and fosters science excellence through awards to established and early career researchers, provides evidence-based advice to assist public policy development, organises scientific conferences, and publishes scientific books and journals. The Academy represents Australian science internationally, through its National Committees for Science, and fosters international scientific relations through exchanges, events and meetings. The Academy promotes public awareness of science and its school education programs support and inspire primary and secondary teachers to bring inquiry-based science into classrooms around Australia.

www.science.org.au

ACOLA receives funding from the Australian Government Department of Education and Training.

By providing a forum that brings together great minds, broad perspectives and knowledge, ACOLA is the nexus for true interdisciplinary cooperation to develop integrated problem solving and cutting edge thinking on key issues for the benefit of Australia. www.acola.org

ACADEMY OF THE SOCIAL SCIENCES IN AUSTRALIA

The Academy (ASSA) promotes excellence in the social sciences and in their contribution to public policy.

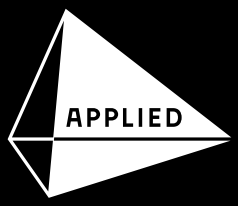
The social sciences are a group of like-minded academic disciplines that work on systematic development of logic and evidence to understand human behaviour in its social setting, including the nature of economic, political, and community activity and institutions.

ASSA is an independent, interdisciplinary body of over 650 Fellows, elected by their peers for their distinguished achievements and exceptional contributions across 18 disciplines.

ASSA coordinates the promotion of research, teaching and advice in the social sciences, promotes national and international scholarly cooperation across disciplines and sectors, comments on national needs and priorities in the social sciences and provides advice to government on issues of national importance.

Established in 1971, replacing its parent body the Social Science Research Council of Australia, founded in 1942, ASSA is an autonomous, non-governmental organisation, devoted to the advancement of knowledge and research in the various social sciences.

www.assa.edu.au



Australian Academy of Technology & Engineering

The Australian Academy of Technology and Engineering is an independent thinktank that helps Australians understand and use technology to solve complex problems.

We bring together Australia's leading experts in technology, engineering and science to provide impartial, practical and evidence-based advice on how to achieve sustainable solutions and advance prosperity.

We champion STEM excellence and contribute robust and practical thinking to Australia's big debates. Like you, we're curious about the world and want to create a better future.

We are a Learned Academy made up of almost 900 Fellows elected by their peers.

www.applied.org.au

HORIZON SCANNING SERIES

We live in a time of rapid change; change that is driven by developments in science and technology and challenged by our capacity to adapt in the present and prepare for the future.

Commissioned by the Australian Government National Science and Technology Council and Australia's Chief Scientist, Horizon Scanning reports present independent and timely analyses to guide decision makers through the decade ahead.

Horizon Scanning reports by the Australian Council of Learned Academies (ACOLA) draw on the deep disciplinary expertise from within Australia's Learned Academies to analyse the future, navigate change and highlight opportunities for the nation. As interdisciplinary studies, ACOLA's reports include economic, social, cultural and environmental perspectives to provide well-considered findings that inform complete policy responses to significant scientific and technological change.

This project has been supported by the Australian Government through the Australian Research Council (project number CS170100008); the Department of Industry, Innovation and Science; and the Department of Prime Minister and Cabinet.

ACOLA collaborates with the Australian Academy of Health and Medical Sciences and the New Zealand Royal Society Te Apārangi to deliver the interdisciplinary Horizon Scanning reports to government.

Also in the Horizon Scanning Series

The role of energy storage in Australia's future energy supply mix
Published 2017

The future of precision medicine in Australia
Published 2018

Synthetic biology in Australia: An outlook to 2030
Published 2018

CONTENTS

PROJECT AIMS

1. Examine the transformative role that artificial intelligence may play in different sectors of the economy, including the opportunities, risks and challenges that advancement presents.
2. Examine the ethical, legal and social considerations and frameworks required to enable and support broad development and uptake of artificial intelligence.
3. Assess the future education, skills and infrastructure requirements to manage workforce transition and support thriving and internationally competitive artificial intelligence industries.

EXECUTIVE SUMMARY

Artificial Intelligence (AI) provides us with myriad new opportunities and potential on the one hand and presents global risks on the other. If responsibly developed, AI has the capacity to enhance wellbeing and provide benefits throughout society. There has been significant public and private investment globally, which has been directed toward the development, implementation and adoption of AI technologies. As a response to the advancements in AI, several countries have developed national strategies to guide competitive advantage and leadership in the development and regulation of AI technologies. The rapid advancement of AI technologies and investment has been popularly referred to as the 'AI race'.

Strategic investment in AI development is considered crucial to future national growth. As with other stages of technological advancement, such as the industrial revolution, developments are likely to be shared and adopted to the benefit of nations around the world.

The promise underpinning predications of the potential benefits associated with AI technologies may be equally juxtaposed with narratives that anticipate global risks. To a large extent, these divergent views exist as a result of the yet uncertain capacity, application, uptake and associated impact of AI technologies. However, the utility of extreme optimism or pessimism is limited in the capacity to address the wide ranging and, perhaps less obvious, impacts of AI. While discussions of AI inevitably occur within the context of these extreme narratives, the ACOLA report seeks to give a measured and balanced examination of the emergence of AI as informed by leading experts.

What is known is that the future role of AI will be ultimately determined by decisions taken today. To ensure that AI technologies provide equitable opportunities, foster social inclusion and distribute advantages throughout every sector of society, it will be necessary to develop AI in accordance with broader societal principles centred on improving prosperity, addressing inequity and continued betterment. Partnerships between government, industry and the community will be essential in determining and developing the values underpinning AI for enhanced wellbeing.

Artificial intelligence can be understood as a collection of interrelated technologies used to solve problems that would otherwise require human cognition. Artificial intelligence encompasses a number of methods, including machine learning (ML), natural language processing (NLP), speech recognition, computer vision and automated reasoning. Sufficient developments have already

occurred within the field of AI technology that have the capacity to impact Australia. Even if no further advancements are made within the field of AI, it will remain necessary to address aspects of economic, societal and environmental changes.

While AI may cause short-term to medium-term disruption, it has the potential to generate long-term growth and improvement in areas such as agriculture, mining, manufacturing and health, to name a few. Although some of the opportunities for AI remain on the distant horizon, this anticipated disruption will require a measured response from government and industry and our actions today will set a course towards or away from these opportunities and their associated risks.

Development, implementation and collaboration

AI is enabled by data and thus also access to data. Data-driven experimental design, execution and analysis are spreading throughout the sciences, social sciences and industry sectors creating new breakthroughs in research and development. To support successful implementation of the advances of AI, there is a need for effective digital infrastructure to diffuse AI equitably, particularly through rural, remote and ageing populations. A framework for generating, sharing and using data in a way that is accessible, secure and trusted will be critical to support these advances. Data monopolies

are already occurring and there will be a need to consider enhanced legal frameworks around the ownership and sharing of data. Frameworks must include appropriate respect and protection for the full range of human rights that apply internationally, such as privacy, equality, indigenous data sovereignty and cultural values. If data considerations such as these are not considered carefully or appropriately, it could inhibit the development of AI and the benefits that may arise. With their strong legal frameworks for data security and intellectual property and their educated workforces, both Australia and New Zealand could make ideal testbeds for AI development.

New techniques of machine learning are spurring unprecedented developments in AI applications. Next-generation robotics promise to transform our manufacturing, infrastructure and agriculture sectors; advances in natural language processing are revolutionising the way clinicians interpret the results of diagnostic tests and treat patients; chatbots and automated assistants are ushering in a new world of communication, analytics and customer service; unmanned autonomous vehicles are changing our capacities for defence, security and emergency response; intelligent financial technologies are establishing a more accountable, transparent and risk-aware financial sector; and autonomous vehicles will revolutionise transport.

While it is important to embrace these applications and the opportunities they afford, it will also be necessary to recognise potential shortcomings in the way AI is developed and used. It is well known, for example, that smart facial recognition technologies have often been inaccurate and can replicate the underlying biases of the human-encoded data they rely upon; that AI relies on data that can and has been exploited for ethically dubious purposes, leading to social injustice and inequality; and that while the impact of AI is often described as 'revolutionary' and 'impending', there is no guarantee that AI technologies such as autonomous vehicles will have their intended effects, or even that their uptake in society will be inevitable or seamless. Equally, the shortcomings associated with current AI technological developments need not remain permanent limitations. In some cases, these are teething problems of a new technology like that seen of smart facial recognition technologies a few years ago compared to its current and predicted future accuracy. The nefarious and criminal use of AI technologies is also not unique to AI and is a risk associated with all technological developments. In such instances however, AI technologies could in fact be applied to oppose this misuse. For these reasons, there will be a need to be attuned to the economic and technological benefits of AI, and also to identify and address potential shortcomings and challenges.

Interdisciplinary collaboration between industry, academia and government will bolster the development of core AI science and technologies. National, regional and international effort is required across industry, academia and governments to realise the benefits promised by AI. Australia and New Zealand would be prudent to actively promote their interests and invest in their capabilities, lest they let our societies be shaped by decisions abroad. These efforts will need to draw on the skills not only of AI developers, but also legal experts, social scientists, economists, ethicists, industry stakeholders and many other groups.

Employment, education and access

While there is much uncertainty regarding the extent to which AI and automation will transform work, it is undeniable that AI will have an impact on most work roles, even those that, on the surface today, seem immune from disruption. As such, there will be a need to prepare for change, even if change does not arrive as rapidly or dramatically as is often forecast.

The excitement relating to the adoption and development of AI technologies has produced a surge in demand for workers in AI research and development. New roles are being created and existing roles augmented to support and extend the development of AI, but demand for skilled workers including data scientists is outstripping supply. Training and education for this sector are subsequently in high demand. Tertiary providers are rapidly growing AI research and learning capabilities. Platform companies such as Amazon (Web Services) and Google are investing heavily in tools for self-directed AI learning and reskilling. A robust framework for AI education – one that draws on the strengths of STEM and HASS perspectives, that cultivates an interest in AI from an early age and that places a premium on encouraging diversity in areas of IT and engineering – can foster a generation of creative and innovative AI designers, practitioners, consultants as well as an informed society. Students from a diverse range of disciplines such as chemistry, politics, history, physics and linguistics could be equipped with the knowledge and knowhow to apply AI techniques such as ML to their disciplines. A general, community-wide understanding of the basic principles of AI – how it operates; what are its main capabilities and limitations – will be necessary as AI becomes increasingly prevalent across all sectors. The demand for AI skills and expertise is leading to an international race to attract AI talent, and Australia and New Zealand can take advantage of this by positioning themselves as world leaders in AI research and development, through strategic investment as well as recognition of areas of AI application where the countries can, and currently do, excel.

Although AI research and development will become an increasingly important strategic national goal, a larger – and perhaps more significant – goal is to ensure that existing workforces feel prepared for the opportunities and challenges associated with the broad uptake of AI. This will mean ensuring workers are equipped with the skills and knowledge necessary to work with and alongside AI, and that their sense of autonomy, productivity and wellbeing in the workplace is not compromised in the process. Education should emphasise not only the technical competencies needed for the development of AI, but also the human skills such as emotional literacy that will become more important as AI becomes better at particular tasks. In the short to medium term, the implementation of AI may require the application of novel approaches. It will be important to ensure that workers are comfortable with this.

To ensure the benefits of AI are equitably dispersed throughout the community, principles of inclusion should underpin the design of AI technologies. Inclusive design and universal access are critical to the successful uptake of AI. Accessible design will facilitate the uptake and use of AI by all members of our community and provide scope to overcome existing societal inequalities. If programmed with inclusion as a major component, we can facilitate beneficial integration between humans and AI in decision making systems. To achieve this, the data used in AI systems must be inclusive. Much of society will need to develop basic literacies in AI systems and technologies

– which will involve understanding what AI is capable of, how AI uses data, the potential risks of AI and so on – in order to feel confident engaging in AI in their everyday lives. Massive Open Online Courses (MOOCs) and micro-credentials, as well as free resources provided by platform companies, could help achieve this educational outcome.

Regulation, governance and wellbeing

Effective regulation and governance of AI technologies will require involvement of, and work by, all thought-leaders and decision makers and will need to include the participation of the public, communities and stakeholders directly impacted by the changes. Political leaders are well placed to guide a national discussion about the future society envisioned for Australia. Policy initiatives must be coordinated in relation to existing domestic and international regulatory frameworks. An independently-led AI body drawing together stakeholders from government, industry and the public and private sectors could provide institutional leadership on the development and deployment of AI. For example, a similar body, the Australian Communications and Media Authority, regulates the communications sector with the view to maximise economic and social benefits for both the community and industry.

Traditional measures of success, such as GDP and the Gini coefficient (a measure of income inequality), will remain relevant in assessing the extent to which the nation is managing the transition to an economy and a society that takes advantage of the opportunities AI makes available. These measures can mask problems, however, and innovative measures of subjective wellbeing may be necessary to better characterise the effect of AI on society. Such measures could include the OECD Better Life Index or other indicators such as the Australian Digital Inclusion Index. Measures like the triple bottom line may need to be adapted to measure success in a way that makes the wellbeing of all citizens central.

Ensuring that AI continues to be developed safely and appropriately for the wellbeing of society will be dependent on a responsive regulatory system that encourages innovation and engenders confidence in its development. It is often argued that AI systems and technologies require a new set of legal frameworks and ethical guidelines. However, existing human rights frameworks, as well as national and international regulations on data security and privacy, can provide ample scope through which to regulate and govern much of the use and development of AI systems and technologies. Updated competition policies could account for emerging data monopolies. We should therefore apply existing frameworks to new ethical problems and make modifications only where necessary. Much like the debates occurring on AI's impact on employment, the governance and regulation of AI are subject to a high degree of uncertainty and disagreement. Our actions in these areas will shape the future of AI, so it is important that decisions made in these contexts are not only carefully considered, but that they align with the nation's vision for an AI-enabled future that is economically and socially sustainable, equitable and accessible for all, strategic in terms of government and industry interests, and places the wellbeing of society in the centre. The development of regulatory frameworks should facilitate industry-led growth and seek to foster innovation and economic wellbeing. Internationally-coordinated policy action will be necessary to ensure the authority and legitimacy of the emerging body of law governing AI.

A national framework

The safe, responsible and strategic implementation of AI will require a clear national framework or strategy that examines the range of ethical, legal and social barriers to, and risks associated with, AI; allows areas of major opportunity to be established; and directs development to maximise the economic and social benefits of AI. The national framework would articulate the interests of society, uphold safe implementation, be transparent and promote wellbeing. It should review the progress of similar international initiatives to determine potential outcomes from their investments to identify the potential opportunities and challenges on the horizon. Key actions could include:

1. Educational platforms and frameworks that are able to foster public understanding and awareness of AI
2. Guidelines and advice for procurement, especially for public sector and small and medium enterprises, which informs them of the importance of technological systems and how they interact with social systems and legal frameworks
3. Enhanced and responsive governance and regulatory mechanisms to deal with issues arising from cyber-physical systems and AI through existing arbiters and institutions
4. Integrated interdisciplinary design and development requirements for AI and cyber-physical systems that have positive social impacts
5. Investment in the core science of AI and translational research, as well as in AI skills.

An independent body could be established or tasked to provide leadership in relation to these actions and principles. This central body would support a critical mass of skills and could provide oversight in relation to the design, development and use of AI technologies, promote codes of practice, and foster innovation and collaboration.

KEY FINDINGS

1. **AI offers major opportunities to improve our economic, societal and environmental wellbeing, while also presenting potentially significant global risks, including technological unemployment and the use of lethal autonomous weapons. Further development of AI must be directed to allow well-considered implementation that supports our society in becoming what we would like it to be – one centred on improving prosperity, reducing inequity and achieving continued betterment.**

- AI offers opportunities across many areas including, for example, the potential to advance health treatments; transform government processes; improve the wellbeing of society; be used for emergency response and early detection of natural disasters such as earthquakes and bushfires; and be applied in dangerous occupations to improve health and safety.
- Change is inevitable and already underway; action and planning are critical; without assertive preparation for AI, we will be left behind and will be more reliant on importing AI technologies and expertise that may not be suitable for the local context.
- AI should be developed for the common good. The protection of human rights and fairness must be built in from the outset, to ensure that AI is implemented safely and sustainably, to benefit all of our citizens.
- Ensuring the safe, responsible and strategic development of AI would benefit from a national strategy that allows areas of major opportunity to be established while the range of social, ethical and legal challenges are embraced and held as core values for implementation.
- The national strategy would be complemented by an implementation framework that balances the need for social values, data-driven innovation and responsive regulation. The interplay between these pillars will determine the way that AI advances and the opportunities that we pursue.
- Meaningful dialogue between civil society, industry, academia and the highest levels of government is needed to shape the kind of society we want for future generations. For example, a national summit could be used to encourage advancement of AI and identify desired societal goals, as well as boundaries that ensure AI is developed within sustainable, ethical and socially responsible limits.

2. **Proactive engagement, consultation and ongoing communication with the public about the changes and effects of AI will be essential for building community awareness. Earning public trust will be critical to enable acceptance and uptake of the technology.**

- AI presents opportunities to make society more inclusive, to improve living standards for people with a disability and those experiencing disadvantage, and increase representation of minority groups. To maximise these benefits, there is a need to ensure that advancement is inclusive, protects human rights and is well communicated to align with social values that are openly accepted.
- Increased focus on accessibility and inclusive AI design can minimise possible harm to society by reducing prejudice and bias introduced by AI systems. This includes access to digital infrastructure that supports, enables and diffuses AI systems; designing AI systems for diverse needs rather than adopting a 'one-size-fits-all' approach; and working to increase representation of marginalised groups in the development of AI technologies. There are opportunities for us to lead in this area.
- Ensuring the protection of human rights may involve, for example, extending existing legal concepts such as liability to encompass decisions made by AI and protections for employees; or establishing ethical standards that will help to leverage the benefits of AI while also managing associated risks.
- There is a need for initiatives that promote and provide broader digital literacy and understanding within society to support the transition to an AI future without marginalising sections of the community.
- Community education initiatives should promote general knowledge and understanding of the principles of AI; how data are used; what it can and cannot achieve; and what we can and should expect from it. Explaining AI in such a manner will be critical to ensuring that people can make informed decisions about AI and how they use it in their everyday life.
- Education should also encompass the risks and opportunities of AI. The public should be aware which risks are realistic and should understand that risks can be managed through adaptation or intelligent policy.

3. **The application of AI is growing rapidly. Ensuring its continued safe and appropriate development will be dependent on strong governance and a responsive regulatory system that encourages innovation. It will also be important to engender public confidence that the goods and services driven by AI are at, or above, benchmark standards and preserve the values that society seeks.**

- Regulatory systems must engender public trust and limit adverse outcomes. Gaps in regulation, for example in automated decision-making technologies, raise significant human rights implications, especially regarding discrimination, implicit bias and undisclosed decision-making processes. It is therefore essential to identify where there are gaps in our regulatory frameworks for AI technologies in order to address such gaps.
- While greater regulation will be required for the application of AI within industry sectors, industry should take proactive steps to ensure safe implementation and readiness for AI systems. In doing so, industry should continue to explore and refine the use of AI and monitor the actions of global peers, competitors and activities in the research sector.
- An ethical certificate and privacy labelling system could be created for low-risk consumer technologies such as smartphones or home assistant technologies. Such a system could be maintained by experts and consumer and industry groups and reviewed by an independent auditor.
- Transparency and explainability are important for establishing public trust in emerging technologies. To establish public confidence, it will be necessary to provide the public with an explanation and introduction to AI throughout the initial adoption stage.

4. **AI is enabled by access to data. To support successful implementation of AI, there is a need for effective digital infrastructure, including data centres and structures for data sharing, that makes AI secure, trusted and accessible, particularly for rural and remote populations. If such essential infrastructure is not carefully and appropriately developed, the advancement of AI and the immense benefits it offers will be diminished.**

- AI technologies rely on digital infrastructure that is accessible, secure and fast. However, the lack of adequate infrastructure will inhibit the broad uptake of AI and will reduce the benefits it offers, particularly for remote and rural communities.
- To be competitive in the AI sector, infrastructure development will need to expand and should keep pace with international progress in telecommunications networks, cloud computing, data at scale, and fast and secure connectivity.
- AI will require high quality and comprehensive datasets that are accessible and useable for learning algorithms. The use of AI technologies to bolster data accumulation and aggregation can lead to positive societal benefits, particularly in healthcare. However, there are also potential negative impacts associated with data collection, including AI's ability to derive personal information from aggregated datasets, and related considerations of consent, privacy and sharing. Transparent and fair data collection policies and procedures will be essential to building trust in how data are collected, accessed and used, and ensuring existing privacy provisions are not bypassed.

5. **Successful development and implementation of AI will require a broad range of new skills and enhanced capabilities that span the humanities, arts and social sciences (HASS) and science, technology, engineering and mathematics (STEM) disciplines. Building a talent base and establishing an adaptable and skilled workforce for the future will need education programs that start in early childhood and continue throughout working life and a supportive immigration policy.**

- Governments should prepare and commit to long-term initiatives that prepare workers, business and the economy for technological change. This would include developing policy and legislation to ensure the benefits brought by technology are shared equally.
- Education curricula at all levels of schooling, particularly higher education, must evolve for students to develop the skills and capabilities required for changing occupations and tasks. Human skills will become increasingly important for AI and subsequently for the education and training of AI specialists. There is a place for education systems to focus on elements of human intelligence and how to protect basic human rights, dignity and identity. Ethics should be at the core of education for the people who are developing AI technology.
- Specific education and training programs will be essential for developing an appropriately skilled AI workforce. Specialist training will often need to augment established domain knowledge in fields such as health, energy, mining and transport and should be driven by deeper interactions between industry and the university sector. There also needs to be effort invested in ensuring diversity in AI training programs.
- AI technologies tend to impact on tasks and processes rather than whole occupations. While the full extent of displacement of workers is uncertain, skills and role types are evolving, new jobs are appearing and there will be a need to respond to these changing workforce needs by upskilling affected workers. Consideration should be given to not only upskilling and reskilling workers specifically in AI, but also across other unrelated industries and roles.
- There may be a need to rethink the context of work itself. People will need to be meaningfully engaged in activities and roles independently of work. Income support could be considered for those displaced if they cannot be appropriately reskilled.
- Skilled working visa programs aimed at transferring experience and capability from overseas would benefit the advancement and uptake of AI and help the nation stay abreast of global development. The Australian Global Talent Scheme Pilot is a welcome approach to attracting skilled talent.

6. **An independently led AI body that brings stakeholders together from government, academia and the public and private sectors would provide a critical mass of skills and institutional leadership to develop AI technologies, as well as promote engagement with international initiatives and to develop appropriate ethical frameworks.**

- Through collaboration, there is an opportunity for us to compete on the international stage, become international role models and provide trusted environments for AI development. This would be stimulated by a robust, harmonised regulatory environment that is designed to support local innovation, help start-up companies to commercialise AI technologies and foster economic development. Sandbox opportunities include prominent industry areas such as healthcare, agriculture, mining and advanced manufacturing. Once demonstrated, established AI technologies can be exported internationally.
- International cooperation and coordination in AI, data, privacy and security issues could be nurtured through increased participation in international fora. Cooperation between governments, corporations and researchers would support increased measures of global governance for AI.
- An independent body that considers the full spectrum of interdisciplinary aspects of AI and allows stakeholders to connect, collaborate, exchange and train staff and share resources would provide significant value to the advancement and uptake of AI. Whether a new institute or an existing body with an enlarged remit, the institute could bring together researchers, developers and policy experts from both HASS and STEM disciplines to undertake long-term projects on issues spanning human rights, psychology, regulation, industrial relations and business. Such an institute could conduct integrated interdisciplinary design, facilitate stakeholder collaboration, develop cyberphysical systems, inform broader policy standards and allow for the full remit of AI to be explored in a holistic manner.
- Basic and translational research in areas of identified priority must be supported to ensure that we are among the most innovative AI nations.

INTRODUCTION

Artificial Intelligence (AI) is not a specific technology, but rather a collection of computational methods and techniques. There is no single AI and there is a lack of consensus among AI researchers on a universal definition. This is because AI means different things to different people and can be used in conjunction with a variety of other technologies, such as the Internet of Things and robotics. However, in the report we define Artificial Intelligence as: a collection of interrelated technologies used to solve problems and perform tasks that, when humans do them, requires thinking.

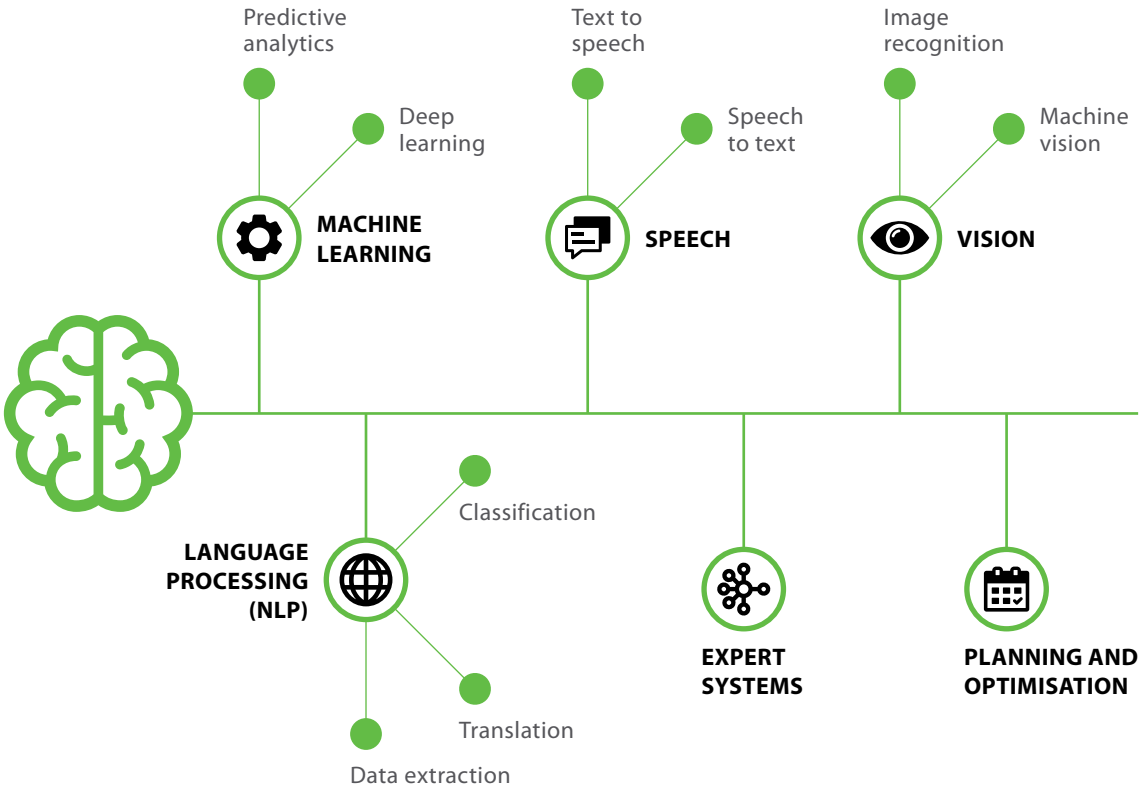


Figure 1: Components of AI

Adapted from: G2 Crowd, 2018.

AI is sometimes equated with machine learning (ML), an often data intensive process in which a computer program 'learns' to do a task from examples. However, ML is only one part of AI, just as learning is only one part of human intelligence. AI also includes: natural language processing (NLP) to enable computers to understand and manipulate language; speech recognition to enable computers to understand speech; computer vision to enable computers to perceive the world; and automated reasoning techniques such as planning, scheduling and optimisation, which enable computers to reason about and solve complex goals. AI is used within a number of areas like robotics and intelligent user interfaces (Figure 1).

AI can be distinguished from simpler software technologies in its ability to handle problems involving complex features such as ambiguity, multiple and sometimes conflicting objectives, and uncertainty. AI software often, but not always, incorporates an ability to learn and improve over time. AI techniques can lead to computers learning through the extraction of information from data and optimising techniques such as self-improvement (unsupervised learning) or by being taught by a developer (supervised learning). In this way, AI is enabled by access to data and depends on existing digital infrastructure. Minsky, a founder within the field of AI described AI as computer systems that are able to perform searches, pattern recognition, learning, planning and inductive reasoning. For the purposes of the ACOLA report, we discuss narrow AI, which are relatively simple systems limited to narrow problem domains.

AI techniques may solve problems in a different manner to how humans solve the same problems. However, AI is currently limited in its ability to solve many problems. For example, while ML is effective at finding patterns in high dimensional data sets, it also has technical limitations. ML systems will often break in strange ways, do not provide meaningful explanations, and struggle to transfer to a new domain. AI systems currently have only a narrow focus and this will likely be the case for many years. AlphaZero, for example, learnt to play two-person complete information games like Go and Chess at above

the level of humans. However, AlphaZero cannot learn to play a game of chance like poker, translate English into Mandarin, or read x-rays.

The ACOLA report does not consider Artificial General Intelligence (AGI), the attempt to build programs that match the full breadth of ability of humans. This is a very ambitious goal, that may not succeed, and is expected to take many decades or even centuries if it does. We focus instead on the application of AI to narrow specialised problems where progress has already been made.

However, despite the limitations described, there have been recent advances in certain areas of AI and it is emerging as transformative technologies that promise to significantly alter our environment. AI is involved in many technologies and applications that already have an influence on our lives. As PwC stated in a 2017 report (PwC, 2017: 3):

> 'What comes through strongly … is just how big a game changer AI is likely to be, and how much value potential is up for grabs. AI could contribute up to [US]$15.7 trillion to the global economy in 2030, more than the current output of China and India combined.'

AI development is a truly global enterprise. It is being pursued by countries around the world because of the perceived benefits it has to offer and is likely to underpin economic competitiveness for both businesses and countries in the foreseeable future. For example, AI can advance health treatments to improve the wellbeing of society; be used for emergency response and early detection of natural disasters such as earthquakes and bushfires; and be used in dangerous occupations to improve workplace health and safety. Yet, as with most endeavours, AI also carries risks for both individuals and societies and it is likely that the changes will shift the prosperity and competitiveness of nations.

AI has deep implications for our lives, including the protection of human rights, quality of life, employment prospects, geopolitics, social inequality, trust in governments and corporations, education, ethics and law, the meaning of democracy, and identity and social relationships. It may be too early to say whether AI will be as transformative as the Industrial Revolution in the 18th and 19th century. However, what can be said with confidence is that it is moving at a far greater pace and is immediately global in a way that the Industrial Revolution was not.

It is therefore important that the development and implementation of AI is managed such that society can enjoy the benefits and opportunities presented without being harmed by the risks it can pose. With increasing development of AI, it is timely to consider what kind of society we want to be, what we would like to accomplish with machines and why. This consideration is important because the short-term choices we make in this field will have long term impacts. The pace of technological change demands agile and responsive policy responses to ensure that people feel prepared for the opportunities and challenges associated with the broad uptake of AI.

The structure of the report

The report considers a range of AI technologies and applications across sectors that permeate or will permeate our society. It places wellbeing at the forefront of AI development and implementation and considers what governments, industry, education institutions and society will need to anticipate in the coming years. While no time horizon is formally specified, the use of short, medium and long term is loosely considered to be within 5 years, approximately 10 to 15 years, and greater than 20 years, respectively. The

huge uncertainty that is inherent in the rapidly evolving technological, social and economic contexts prevents specific prediction.

Chapter 1 provides an overview of AI, its promise and implications for international relations. The chapter discusses AI in relation to international treaties, global governance and geopolitics.

Chapter 2 describes the scope of AI technologies and considers AI applications and infrastructure requirements. An overview of some of the various sectors impacted by AI is presented. While this overview cannot be comprehensive, it aims to illustrate some of the uses for AI technology.

Chapter 3 discusses the future education, skills and workforce needs in a world of AI. It considers the potential impact of AI on these key areas and examines issues on the transformation of the Australian community, from the individual through to the workforce.

Chapter 4 examines the equitable development and implementation of AI technology in Australia. It considers the potential for inequality to be either exacerbated or reduced as a result of AI technologies and explores issues of human rights, public communication and inclusive design. Key considerations and principles for the equitable adoption of AI are also outlined.

Chapter 5 details some of the regulatory and legal implications surrounding AI, including liability for AI decisions, the ability to appeal an AI decision, and the effects of the EU's General Data Protection Regulation. It provides suggestions for regulatory considerations and explores the potential for an independent body to provide oversight and governance in relation to AI technologies.

Chapter 6 outlines the significance of data to the development and implementation of AI and describes the technical and legal components to data usage, including data collection and consent, data governance, data management and storage.

Chapter 7 examines data with respect to social and ethical considerations. Trust, accessibility, indigenous data sovereignty and the potential for discrimination and bias are discussed.

Chapter 8 provides an overview of the report and details the possibilities for AI.

How the report complements and differs from others

The report places society at the core of AI development and explores issues specific to Australia and New Zealand such as our workforce, our education system, cultural considerations and our regulatory environment. It identifies areas of importance to Australia and New Zealand. Enlisting expertise from Fellows of Australia's Learned Academies, the Australian Academy of Health and Medical Sciences (AAHMS) and the Royal Society Te Apārangi (New Zealand), the ACOLA report provides a comprehensive interdisciplinary study to map and establish a detailed understanding of the opportunities, benefits and risks presented by AI, including examinations of:

- **Technological enablers and barriers**, spanning trends in uptake
- **Standards and operating protocols** to support interoperability, accessibility for users, innovation and technology advancement
- **Employment and the workforce**, including displacement and skill change, labour standards, the changing geographic distribution of workers and the career long interaction between education and work.

- **Education** to ensure the effectiveness of education initiatives, support equity of access and increase public understanding and provision of appropriately skilled human capital
- **Social implications** and establishing frameworks to manage the array of potential issues spanning ethics, public trust, safety, productivity, employment, health and inequality
- **Cultural impact** and supporting positive public attitudes to technology uptake and change
- **Industry and research capabilities** and identifying niche areas of opportunity where Australia and New Zealand have a strategic advantage and can develop, adopt and lead.

While Australia does not yet have a formal plan or strategy for AI, there are several national initiatives underway or completed. In 2018, the Australian Government launched *Australia's Tech Future* (a digital economy strategy), the Australian Centre for Robotics Vision released a report *A Robotics Roadmap for Australia 2018*, and the Australian Government announced A$29.9 million in funding over four years for CSIRO's Data61 to develop a national roadmap for AI including a national ethics framework and to strengthen Australia's capability in AI and Machine Learning. Further, the Australian Human Rights Commission is undertaking a three-year project on the relationship between human rights and technology (Australian Human Rights Commission, 2018a).

It is anticipated that the ACOLA report will provide a broad interdisciplinary framework to support policy makers in Australia.

Exclusions from scope

The ACOLA report builds on a number of existing national and international reports on AI. ACOLA and the working group have engaged with concurrent Australian, New Zealand and international initiatives to ensure the reports are not developed in isolation. It is hoped that the findings of this report can contribute to the effective and ethical development of AI as an opportunity to improve societal wellbeing.

While application of AI to cybersecurity is important, it is not directly addressed in the ACOLA report. Cybersecurity is strongly addressed by current Australian Government policy and program initiatives and therefore is acknowledged rather than analysed in the instances where it underpins other applications or implications of AI development.

The Internet of Things will be addressed by a parallel ACOLA Horizon Scanning report and is similarly not considered in the ACOLA report, except when it underpins other applications or implications of AI development.

Artificial general intelligence (machines that match the full breadth and depth of human expertise) is also excluded from the report. General AI is likely to be pursued in the coming decades, but its development will require a longer time horizon than the issues presented in the ACOLA report and therefore has been excluded from consideration.

HORIZON SCANNING

ACOLA
AUSTRALIAN COUNCIL OF LEARNED ACADEMIES