THE EFFECTIVE AND ETHICAL DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

AN OPPORTUNITY TO IMPROVE OUR WELLBEING

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

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

APPLIED

Australian Academy of
Technology & Engineering

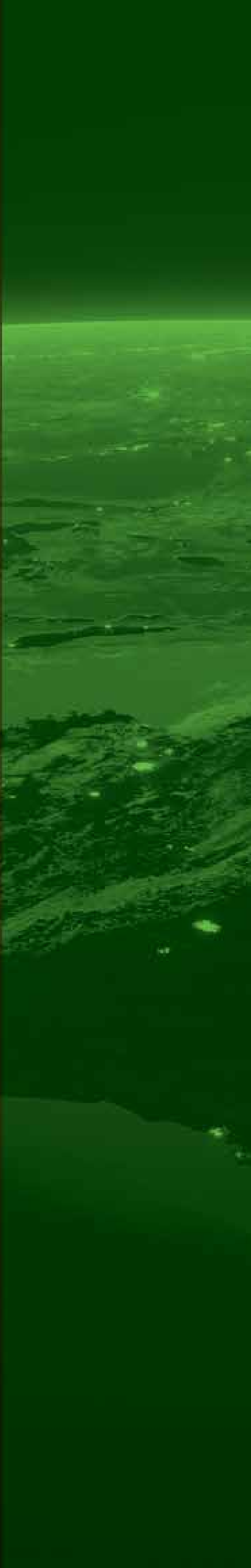
The Australian Academy of Technology and Engineering is an independent thinktank that helps Australians understand and use technology to solve complex problems.

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

FIGURES

TABLES

BOXES

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, this 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 this 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.

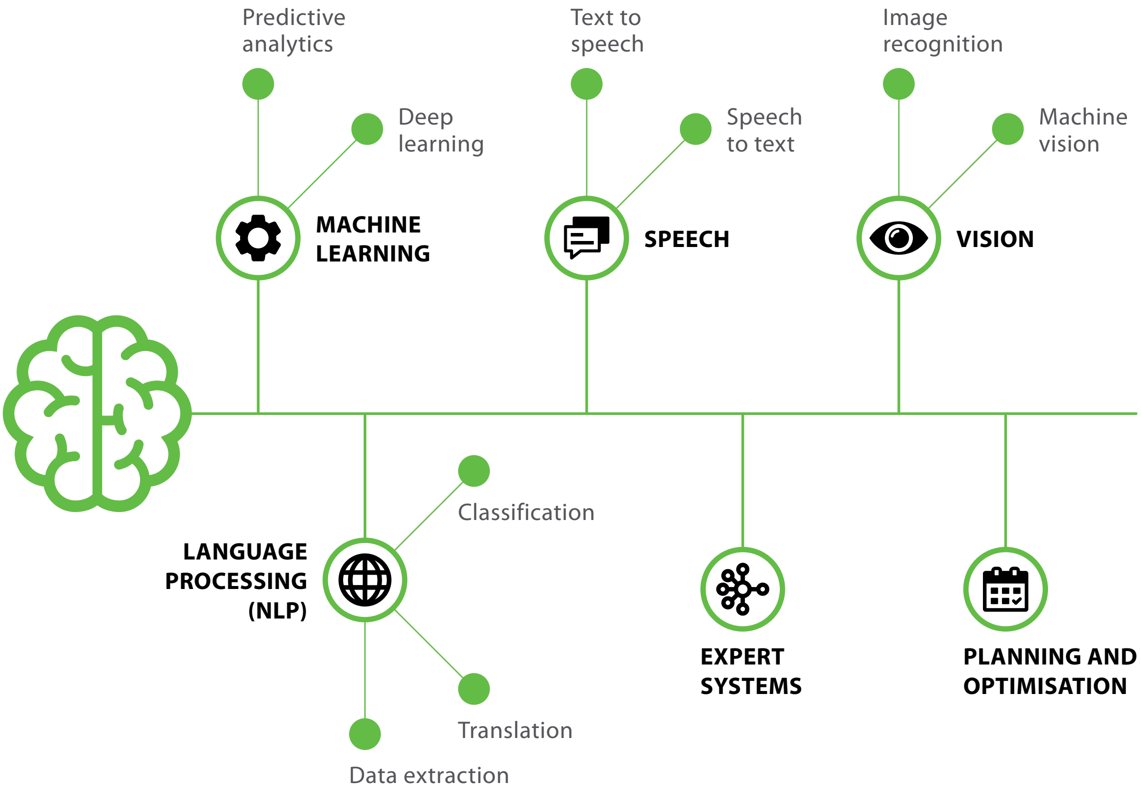


Figure 1: Components of AI

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 this 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.

This report will 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 will 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

This 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 this report complements and differs from others

This 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

This 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 this 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 this 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 this report and therefore has been excluded from consideration.

CHAPTER 1
A WORLD OF ARTIFICIAL INTELLIGENCE

This chapter is based on input papers prepared by the generous contributions of Ziyang Fan and Dr Susan Aaronson (AI and Trade); Professor Dr Andrea Renda (Global Governance); and Adjunct Professor Nicholas Davis and Dr Jean-Marc Rickli (Geopolitics). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

1.1 Overview of artificial intelligence

AI is already being used in many areas and will increasingly be the underlying technology that allows devices to run, communicate and analyse data. As AI becomes more advanced, its applications will become increasingly complex and will have widespread impact on our lives, workplaces, industries and the way we interact with each other. It offers opportunities for Australia and our neighbours for continued prosperity and global competitiveness. The way in which we interact with and adopt AI will fundamentally shape how it is developed in the future.

As mentioned in the introduction, AI is a collection of computational methods and techniques. AI applications touch all corners of the economy, including disaster management, the environment, logistics, health, education, manufacturing, warfare and government services. If pursued appropriately, the opportunities presented by AI may be as transformative as the industrial revolution.

However, without full consideration of the economic, ethical, social and cultural implications of implementation, significant issues such as social inequity, discrimination, breach of human rights, unemployment, loss of social cohesion, gaps in education, geopolitical tension, and poor public trust in governments, democracy and corporations, could come to bear and may prevent effective deployment of the technology and diminish the benefits.

At a broad level there are four types of AI:

- Narrow AI (or weak AI), is well established, available, and pervasive. It is usually designed to focus on a narrow task or application. Narrow AI ranges from the early instances of computers being programmed to beat humans at chess through to chatbots and digital assistants such as Apple's Siri. AI solutions in the market today are in this category, albeit with a very wide range of capability.

- Emerging and disruptive AI is under development and has emerging applications. Self-driving vehicles, drones, or advanced environments such as IBM's 'Project Debater' fall within this category of AI. AI of this type is characterised by a machine acting on what it sees based on either supervised or unsupervised learning (the latter of which is often referred to as machine learning, see Chapter 2).

- Generalised AI (or artificial general intelligence), is a machine that is of equal intelligence to an adult human. Unlike narrow AI, a general intelligence machine can theoretically make decisions irrespective of any previous training, instead relying on what it learns on its own. The basis for human consciousness is still unknown and therefore it is difficult to speculate when or even if a machine will be able to emulate it. As such, scientists are divided on how close we are to achieving artificial general intelligence.

- Superhuman AI (or artificial emergent intelligence) is the evolution of generalised AI and refers to a theoretical machine that has a far superior intellect in every field including creativity, social skills and general wisdom. In effect, this level intelligence would be representative of a machine that would be capable of constantly learning and improving itself.

As AI is developed, solutions will also move towards augmenting human intelligence. This will encompass systems that can learn from interactions with humans and the environment, and inform human decision making to select and weigh options. Augmented intelligence is a route whereby we can ensure that the human remains in the decision-making loop and that human capital is not rendered redundant by AI.

This report focuses on narrow and emerging AI and considers the potential opportunities and impacts of these emerging and disruptive technologies.

1.1.1 Promise

AI has been a field of interest and study for decades. However, more recently, increases in computing power, technology advancements, increases in data availability from the rise of social media, the digitisation of the global economy and the development of the Internet of Things (IoT) have led to its ascendance.

The accumulation, aggregation and manipulation of high volumes, high velocity (speed of data) and high variety (range of data types and sources) of data in real-time, provides increasingly accurate insights into the complexities of modern social life, which can enhance policy and service insights and enable better choice-making for consumers. For example, AI can be used to enable better resource management through the collective use of smart grids, which can provide detailed understanding of electricity usage at every stage in the grid. Other areas of AI, such as natural language processing, are strongly contributing to the automation and streamlining of various tasks including machine translation (e.g. Google Translate), dialogue systems (e.g. the back-end systems that underlie Apple's Siri and Amazon's Alexa) and automatic question answering (e.g. IBM Watson). Machine learning (ML) algorithms are also helping to automate a range of processes, from autonomous vehicles to medical diagnosis.

Recently, some AI systems have demonstrated the ability to outperform humans in forming inferences from large, complex datasets to solve problems such as classification, continuous estimation, clustering, anomaly detection, data generation and ranking (Chui et al., 2018). These techniques have resulted in advances in important aspects of AI such as computer vision, natural language processing, robotics, planning and intelligence analysis.

These advances in AI have the potential to transform economies and societies, in terms of innovation, effectiveness, process efficiency and resilience. In 2017, it was suggested that AI could contribute up to US$15.7 trillion to the global economy in 2030 (PwC, 2017), equating to more than the current output of China and India combined. It is estimated that of this amount, US$6.6 trillion would likely come from increased productivity alone (PwC, 2017).

While AI has the potential to advance society in new ways, much remains unclear about the future of AI and whether the promise of a radical transformation of economic and social life will be realised. Australia and New Zealand have an opportunity to develop policy frameworks for AI – frameworks that use data for national benefit and provide incentives for collaboration between industry, government, academics and everyday citizens.

1.1.2 Data

The internet has enabled rapid communication on a global scale, resulting in an unprecedented amount of data being produced, shared and recorded. Some analyses indicate that over 3.7 billion people use the internet, executing 5 billion searches, 720 million tweets and 432 million status updates every day (Marr, 2018). Additionally, smart or internet-enabled technologies and services in homes, workplaces, cities and governments, rely on the extraction and sharing of large volumes of data between individuals, organisations and governments. Such data is often personal and sensitive information about an individual. Governments can analyse data to better understand citizens' concerns and needs, while platform companies such as Twitter, Google and Facebook rely on this data to generate revenue in various ways. Much of this data is not *provided* by individuals per se, but rather *generated* through various internet-enabled technologies and services that produce continuous streams of data.

The availability of such large datasets is fundamental to the role of AI and underpins much of its development and use. Its collection, however, prompts questions about the legal, ethical and economic implications of data collection. For example, algorithmic decision-making tools raise concerns on potential bias and discrimination, while AI systems capable of deriving personal information from multiple datasets point to technical and legal challenges on tracing the 'provenance' of data.

1.1.3 International context

As AI continues to rapidly advance, many countries are responding with government and industry strategies and investments to take advantage of its potential benefits and opportunities (Figure 2). While not representing an exhaustive list of the initiatives underway, it does demonstrate the level of interest globally, particularly in the past two years.

The US is currently the world leader in AI research. Boosted by the world's largest private sector research and development (R&D) environment, with companies such Alphabet (Google), Amazon, Apple, Facebook, IBM and Microsoft, the US is leading the way in the adoption of AI in high-tech and telecommunications industries, and the automotive, financial and resource sectors. The US Defense Advanced Research Projects Agency (DARPA) has been crucial in supporting this agenda. In 2017, the value of AI in the US medical industry was estimated at US$369.25 million, with compound annual growth of 41 percent (Mordor Intelligence, 2017). In 2016, the United States National Science and Technology Council published a national AI strategy, *The National Artificial Intelligence Research and Development Strategic Plan* (NSTC, 2016). China, the UK, France and the EU followed suit, releasing national strategies that each demonstrate a different approach towards the uptake of AI-powered technologies. In 2018, the Australian Government joined in these efforts.

China, through technology companies such as Huawei, Baidu, Alibaba and Tencent, is already a key player in a wide variety of AI development activities, including autonomous vehicles, facial and voice recognition, targeted advertising and marketing, as well as policing. In 2017, China's State Council released the *Next Generation Artificial Intelligence Development Plan*, which lists several goals including: AI becoming a key source of growth; a primary driver of industrial advances and economic transformation; and for China to be the world's top AI innovation centre by 2030 (State Council, 2017). AI-driven facial recognition technologies developed by Chinese companies have been used in over 100 million smartphones (Bloomberg, 2018) and have been used to streamline boarding processes in large airports and draw attention to jaywalkers by projecting their image on large screens at intersections in cities like Shenzhen (Xu and Xiao, 2018). However, errors within these systems can also occur. A Chinese businesswoman was recently identified by the facial recognition system as having jaywalked when in fact the system had captured her image on a bus advertisement (Shen, 2018). The bus had been driving through an intersection when the facial recognition system incorrectly identified and displayed the infringement. Approximately 60,000 schools in China are participating in a program that uses AI software to grade the work of students, evaluate the structure of essays and incorporate notes from teachers (Chen, 2018).

In Europe, the European Commission has called for €20 billion in investment in AI R&D from public and private sources by 2020 (European Commission, 2018b, 2018c). The European Commission is also increasing its own investment to €1.5 billion via the Horizon 2020 fund, with that investment expected to spur an additional €2.5 billion in associated funding from public-private partnerships (European Commission, 2018b, 2018c). There are further initiatives such as the European fund for strategic investments that will help coordinate an additional €500 million in AI R&D investments by 2020 (European Commission, 2018b) and the Future Emerging Technologies programme. Citing the General Data Protection Regulation (GDPR), the European Commission's *Artificial Intelligence for Europe* report states that Europe is at the forefront of ensuring data serves humanity and suggests that the EU can lead in developing an approach to AI that 'benefits people and society as a whole' (European Commission, 2018c). The EU is hoping to find a competitive advantage in developing a more ethical approach that enhances privacy and trust and plans to release documents to support this in late 2018 (Rabesandratana, 2018b).

Building on a history of government investment in fundamental AI, the UK is home to some of the world's leading AI companies, including the headquarters of DeepMind, a British AI company acquired by Google in 2014 and considered perhaps the world's leading AI lab (Metz and Satariano, 2018). In October 2017, there were more than 200 start-ups and small-to-medium enterprises (SMEs) developing AI products in the UK (Hall & Pesenti, 2017). In April 2018, the UK Government released its national AI strategy as part of its broader industrial strategy and established several new bodies to support the development of AI: the AI Council, the Office for Artificial Intelligence, and the Centre for Data Ethics and Innovation

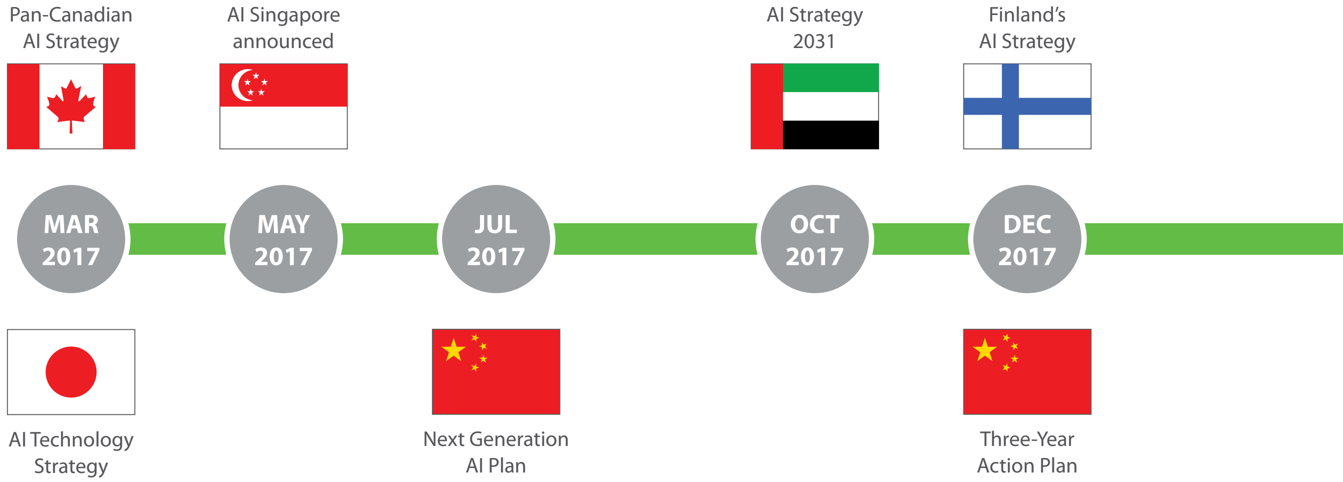


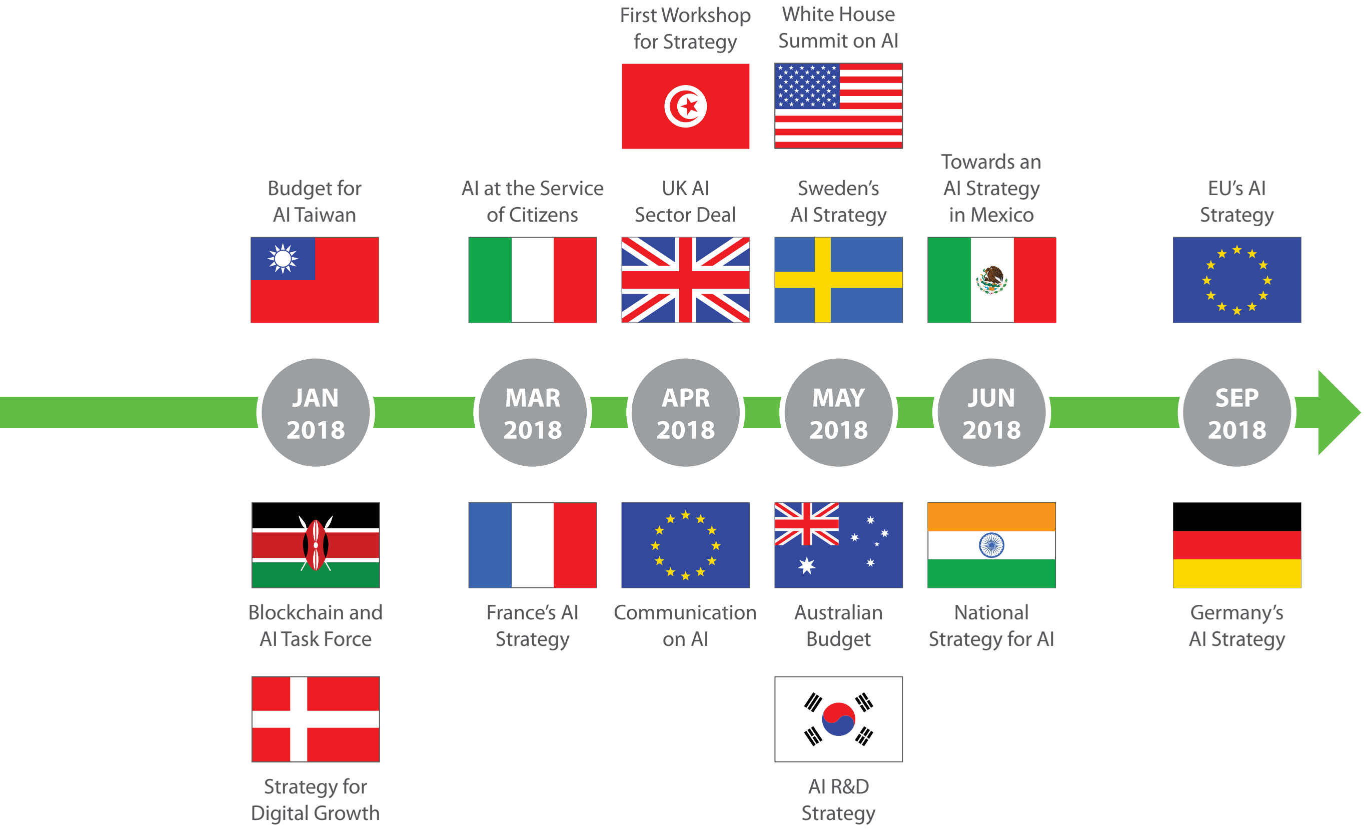
Figure 2: Global AI strategies and initiatives

Adapted from: Dutton, 2018.

(HM Government, 2017). Additionally, the Alan Turing Institute's remit was expanded as the UK's national research centre for AI.

Canada was the first country to release a national AI strategy. The strategy includes funding for centres of excellence in AI research and innovation. Canada is ranked third in the Government AI Readiness Index, indicating that the government is well placed to implement AI in its delivery of public services (Stirling, Miller and Martinho-Truswell, 2017). Canadian researchers and policy makers are producing strategies and principles to support the responsible development of AI. For example, the Université de Montréal is developing the Montréal Declaration for a Responsible Development of Artificial Intelligence (Université de Montréal, 2017), and Global Affairs Canada is leading a collaboration on AI and human rights with a number of Canadian universities (McKelvey and Gupta, 2018). Canada has also attracted international leaders in AI technology including Google, Uber, Facebook and Microsoft (Bernstein, 2018).

There are many other notable examples of national AI initiatives and programs.[1] France recently revealed plans to create a National Artificial Intelligence Program alongside the launch of a national AI strategy promising €1.5 billion for AI projects by 2022 (Dillet, 2018). Germany has announced €3 billion over six years for the Artificial Intelligence (AI) Made in Germany digital strategy with the aim to boost the country's AI capabilities. Matching funds are anticipated from industry, which will bring the total investment to €6 billion. The strategy outlines several goals,



[1] These are outlined in further detail in a supplementary report available on the ACOLA website (www.acola.org)

including the creation of 100 university chairs with a focus on AI, alongside other strategies to enhance research and translation (Buck, 2018). Estonia, a leader in digital governance initiatives, has already begun development of a legal framework for AI systems and is ranked sixth on the global Automation Readiness Index (Plantera and Di Stasi, 2017; The Economist Intelligence Unit, 2018a). The Indian Government recently released a report from the AI Task Force that outlines the key challenges for India in integrating AI into its economy and society, making a number of recommendations (NITI Aayog, 2018). Japan has produced a national AI strategy and with significant investment in R&D has been a major contributor to AI research (Japanese Government, 2017). Israel is host to a number of universities undertaking AI research, and Israeli start-ups have received overseas investment from large US companies (Solomon, 2017). Nigeria and South Africa are emerging as African leaders in the development of AI; the University of Lagos having opened an Artificial Intelligence Hub and the two countries hosting the majority of African industry start-ups (Ferrein and Meyer, 2012; Ndiomewese, 2018; The Guardian, 2018).

Globally, technology giants spent $20-30 billion on AI in 2016, with 90 percent spent on R&D and deployment and 10 percent on AI acquisitions. The majority of the funds (66 percent) were spent in the US, with China also receiving significant investment (17 percent). Indeed, investment in China is growing at a significant rate. Corporate investment examples include IBM's investment of US$240 million over 10 years in a partnership with Massachusetts Institute of Technology (MIT). The aim of this investment is to create an AI laboratory to conduct advanced research and to explore the implications of the technology on industries such as health care and cybersecurity as well as on society. However, in comparison to overall spending, IBM's investment is relatively small and indicative of the need for collaboration. In addition, Google has launched its AI-first strategy in 2016 and appointed a new research group dedicated to ML.

In contrast to the international initiatives, Australia is placed eighth in the Government AI Readiness index, sitting between Japan and New Zealand. However, Australia is considered to be at the forefront of AI development and experimentation in the Asia-Pacific region (FTI Consulting, 2018). In 2018, the Australian Government announced A$29.9 million investment in AI, including the creation of a technology roadmap, a standards framework and a national AI ethics framework. Further, the Australian Human Rights Commission is undertaking a major research project examining the impacts of technology on human rights, with a particular focus on AI technology (Australian Human Rights Commission, 2018a).

To keep pace with international advances and ensure a growing and strong economy, it is necessary to be inventive and capable in the adoption of AI. It is likely that Australia will initially be a receiver of internationally developed technology and data standards and constraints. The research and strategies being undertaken by leading Commonwealth countries provide opportunities to cooperate in this area at a Commonwealth level. Australia and New Zealand's AI capabilities and initiatives are further outlined in Appendix 1 and Appendix 2.

1.1.4 Emergence, impact and governance of AI

Global geopolitical changes are likely to occur as a result of developing AI technologies. While the US and China are global leaders in developing AI technology, other countries may develop AI expertise in niche areas. For example, Canada and Germany are emerging leaders in the development of ML and autonomous vehicles, respectively. To develop a globally competitive AI industry, Australia will require public and private investment, collection and sharing of large datasets, an appropriately skilled workforce and supportive regulation.

Given the unpredictability and the pervasiveness of the impacts of AI on every sector of society and every institution, it is not possible to be comprehensive in reviewing its implications for every aspect of international relations. What follows is a series of case studies, each focusing on an area of major importance in which thinking is advanced and policy development underway.

While AI technologies may present opportunities, advanced AI technologies and capabilities may also introduce new risks to human rights, the economy, the environment, democracy and social cohesion, from individuals, criminal enterprises, non-state actors and rogue states, who could use the technology in undesirable ways. Commercial technologies that have the capacity to be repurposed for surveillance or offensive uses are already being used for untoward purposes and it is likely that AI technologies will be no exception. Nationally, AI technology will cut across applications such as energy grids, internet pipelines, the food chain, banking networks, hospital infrastructure and transport logistics. It will therefore remain important to ensure data collection and storage systems are secure and protected from external intrusion and threats. It has been suggested that national stability may be affected by workforce disruptions resulting from job automation. However, new opportunities for employment are also likely to emerge from the adoption of AI technologies. To ensure national security, governments may need to consider the way in which technological research and associated datasets are shared and accessed.

1.1.4.1 Global governance of AI

While frequently associated with a slow and laborious process of negotiation, global governance has been successfully achieved in key areas of global concern including trade, human rights, security and the environment. The effects of AI are also likely to have far reaching, global impacts that would benefit from global management. The development of global governance in relation to AI was initiated on 12 July 2018 with the UN Secretary General's appointment of a High Level Panel on Digital Cooperation. The panel's ambition is to support 'cooperative and interdisciplinary approaches to ensure a safe inclusive digital future for all taking into account relevant human rights norms'. Rather than seeking to create new international treaties, which could struggle to find agreement, efforts on the global governance of AI should build on, and be derived from, existing relevant sets of global agreements, such as international human rights law, the law of armed conflict, trade agreements and Security Council resolutions.

The creation of an International Panel on AI was recently announced by Canada and France. The purpose of the panel is to promote human-centric AI, which is 'grounded in human rights, inclusion, diversity, innovation and economic growth' (*Mandate for the International Panel on Artifical Intelligence*, 2018). The panel's potential focus areas include, 'data collection and access; data control and privacy; trust in AI; acceptance and adoption of AI; future of work; governance, laws and justice; responsible AI and human rights; and equity, responsibility and public good' (*Mandate for the International Panel on Artifical Intelligence*, 2018). It is likely that other nations will join this panel, including members of the G7 and EU (Knight, 2018). Given the global context of AI development, international fora should encompass global participation and representation. Several countries beyond the G7 and EU members are also pursing the development of, and investment in, AI technologies and thus should be considered in this context.

In this context, opportunities exist for countries like Australia and New Zealand to adopt a leadership role in the development of global frameworks for AI use. For example, Australia and New Zealand have the opportunity to be at the forefront of discussions that reframe AI as a public good and to lead inclusive approaches to the development of safeguards associated with AI use. This would involve the increased engagement of a variety of actors, including private and start-up companies, governments, international organisations and the academic community. New Zealand is a member of the D7 group of digital nations, which includes in its charter the idea that member nations will 'lead by example and contribute to advancing digital government' (Digital Government New Zealand, 2018). Expansion of such panels and fora to include Australia could offer opportunities for the country to collaborate internationally and to lead by example.

1.1.4.2 Trade policy

AI technologies are powered by large quantities of data that are frequently sourced and exchanged across state borders. Data shared for the development of AI are therefore considered a commodity and subject to trade regulation. Trade guidelines are primarily established by the World Trade Organisation (WTO) and other bilateral, multilateral and regional trade agreements, such as the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP). The CPTPP contains rules specific to the trade of data.

However, several agreements, to which Australia and New Zealand are a member of, were established prior to the emergence of AI technologies. As a result, some agreements may require updating if international access to data is desired. There may be an intermediary step involved where existing regulations are applied to the new AI context prior to considering what amendments or new regulations may be required. Sourcing international data will be particularly important for states with smaller populations, such as Australia where there is limited capacity to access the large amounts of data needed for the development and application of AI technologies. For AI to be representative it requires representative data, otherwise models will be limited by incomplete data. WTO-plus trade agreements, which extend current WTO standards, could provide mechanisms for regulation around data obligations and trade.[2] For instance, the

[2] WTO-plus agreements are trade agreements wherein the contents and level of obligations exceed those required by WTO rules.

United States-Mexico-Canada Agreement (USMCA) includes a chapter on digital trade and the CPTPP, which was ratified by Australia in October 2018, also contains provisions for data localisation, cross-border data transfers and source code disclosure requirements.

International trade regulations will affect the operation of corporations that use internationally sourced data. For example, the EU's General Data Protection Regulation (GDPR) establishes specific rules for the use of data sourced from EU citizens (discussed further in Chapter 6).

1.2 Improving our wellbeing

While periods of rapid change have occurred throughout history, the persistence and acceleration of rapid change being experienced from new technologies is unprecedented. Individuals will respond to this change in different ways, with some faring better than others. Similarly, businesses will differ in their ability to keep pace. The technology will undoubtedly outpace policy responses. Therefore, countries should consider how to develop and implement technology in a way that allows it to flourish while protecting society and wellbeing.

As noted previously, AI presents both risks and opportunities in nearly every sector of society. It could lead to a dramatic rise in unemployment or a boom in new work opportunities, or both simultaneously. It could threaten democratic governance or make governments more responsive to a better informed populace. It could make many tasks much safer or make the world more dangerous. Only by managing its development in a way that places wellbeing at the centre can the social benefits be maximised.

There are several principles that should be considered in the development of AI that keep the wellbeing of society as the central consideration. These principles are: economic benefit, social benefit and sustainability. It will be necessary to determine what kind of society we want to have and how AI technologies might be able to uphold this vision.

Economic

In 2016, it was predicted that worldwide revenues from the adoption of AI systems across multiple industries will experience an increase from US$8 billion in 2016 to over US$47 billion in 2020 (International Data Corporation, 2016). Furthermore, as mentioned earlier, global GDP could increase by 14 percent, or US$15.7 trillion by 2030 because of AI, with US$1.2 trillion extra economic growth forecasted GDP gains in Oceania (PwC, 2017). By 2030, Australia could increase its national income by A$2 trillion from productivity gains afforded by increasing automation and AI technologies (AlphaBeta, 2017). The potential income gains of AI will, however, need to be set against the costs associated with its implementation, including the cost of archiving, curating, trading and protecting data and of reskilling workers.

AI should be implemented in a manner that limits economic disadvantage or exacerbation of inequalities, and instead generates broad positive economic benefit to society. It can do this by fundamentally reducing the economic cost of creating or producing goods and services, and by providing new goods and services that would have otherwise been impossible or not economically viable.

Societal

As a society, we will coexist with AI, form a variety of relationships and attachments to AI and will react to AI (both positively and negatively). Therefore, a discussion is needed about what kind of society we want to be, and embedded within that, the desired relationship with AI, and the boundaries that should be established and protected. Developments will require continual monitoring and agile responses, because they are certain to play out in ways that cannot be predicted, as different technologies and different developments interact with society. Governments, industry and society will need to ensure AI is developed and is implemented in a way that protects individual dignity, human rights and the capacity of people to advance.

Inclusive AI design can meet the needs of minority groups and create the possibility of better products and services for everyone. The implementation of AI must proceed with the aim of protecting and promoting human rights – including civil and political, as well as economic, social and cultural, rights – and enable more informed and objective decisions. For example, AI can limit direct and indirect discrimination by humans in decision making processes, who may act on their own prejudices and without empirical support. It can provide more accurate and targeted health diagnoses and treatment; improve emergency response planning; and enhance workplace health and safety. Further, AI algorithms – if rigorously and thoughtfully developed – can assist in identifying systemic bias and may present opportunities for more effective assessment of compliance with fundamental human rights. AI technologies should improve access to services and improve outcomes across a range of socio-economic indicators, through better systems or interventions in health and education, or for groups who experience vulnerability and disadvantage.

Environment and sustainability

AI can be used to create a more sustainable society. Environmental sustainability is a complex issue and requires geo-scale management and interaction with processes that are inherently poorly predictable. Dealing with complexity and improving predictability, sustainability depends on having enough data, using data that is available, and identifying where new data will make the biggest difference. AI can help deal with this complexity and help humanity make the best use of limited resources.

More specifically, AI can make a significant contribution to environmental management in a number of sectors. AI can, for example, reduce the environmental footprint of agriculture through better management of chemical use, soils, on-farm waste, and through improvements to animal welfare. AI can improve energy performance through enhanced data collection and analysis from smart meters and smart electrical grids, as well as ML algorithms in buildings to optimise energy consumption. Another area is precision of mining, where AI techniques can be applied to improve efficiency so that there is less waste, and less water and energy use. Blockchain technology can be used as a way to confirm ethical and sustainable production. AI can be used to create virtual scenarios minimising human impact on the environment. AI technologies can also provide opportunities to mitigate climate change and reduce pollution and can be used to optimise urban spaces, support individual and community use and ensure minimal environmental impact. AI and automation can also be used to assist in recycling processing. For example, Apple has developed Liam, a collection of autonomous machines that dismantle and sort iPhone components for recycling, and a Polish start-up has created a smart bin that is able to recognise and sort rubbish for the purposes of recycling and space management (Leswing, 2017; Best, 2018).

1.2.1 4Ds (dirty, dull, dangerous and difficult)

AI and automation have the capacity to free workers from dirty, dull, difficult and dangerous tasks. Some jobs include tasks that people do not want to do or should not be made to do. For example, robots powered by AI can undertake dirty tasks such as mine exploration, or inspecting, monitoring and fixing clogged sewer pipes. Dangerous tasks such as investigation of unstable structures, mining, disaster response and space exploration provide another avenue for AI use to minimise potential harm to workers. AI and applications in robotics are also being developed for difficult tasks that require a high level of detail with a low margin of error, such as surgery. However, AI may also threaten to automate interesting high-value tasks, rather than just unattractive tasks. This is discussed in further detail in Chapter 3.

1.2.2 Measuring success

Traditional measures of success, such as GDP and the Gini coefficient, will remain relevant in assessing the extent to which Australia and New Zealand are managing the transition to an economy and a society that takes advantage of AI opportunities. These measures can mask problems, however, and innovative measures of subjective well-being may be necessary. We may, for example, need to transition to measures such as the OECD Better Life Index, or other indicators such as the Digital Inclusion Index to better characterise the effect of AI on society. Measures like the triple bottom line, incorporating three dimensions of performance (social, environmental and financial), may need to be adapted to measure success in a way that makes the wellbeing of society central. Issues such as the knowledge gap, the digital divide, and economic and social stratification will need to be considered. New Zealand is moving away from GDP as a standalone measure of success – in 2019, the country will launch its 'Wellbeing Budget', which will draw on a range of measures to evaluate wellbeing. The indicators build on international best practice and are tailored to New Zealand citizens by incorporating culture and Maori perspectives. A similar approach could be applied in Australia.

CHAPTER 2
AI IN PRACTICE

This chapter is based on input papers prepared by the generous contributions of Professor Robert Williamson (Machine Learning); Professor Anton van den Hengel (Machine Learning); Professor Tim Baldwin and Professor Karin Verspoor (Natural Language Processing); Professor Lloyd Hollenberg (Quantum Machine Learning); Professor Dr Alberto Elfes, Dr Elliot Duff, Dr David Howard, Fred Pauling, Dr Navinda Kottege, Dr Paulo Borges and Dr Nicolas Hudson (Robotics and Manufacturing); Associate Professor Federico Girosi (Health and Aged Care); Professor Bruce MacDonald, Associate Professor Elizabeth Broadbent and Dr Ho Seok Ahn (Health and Aged Care); Chris Goodes, Adrian Pearce and Peter Scales (Mining); Sylvie Thiebaux (Energy); Professor John Quiggin (Environment); Professor Iven Mareels (Environment); Professor John Billingsley (Agriculture); Professor Salah Sukkarieh (Agriculture); Professor Mengjie Zhang (Agriculture); Dr Thomas Birtchnell (Arts and Culture); Professor Dr Sven Kesselring, Eriketti Servou, Dr Dennis Zuev (Transport and Mobility); Dr Malene Freudendal-Petersen and Robert Martin (Transport and Mobility); Associate Professor David Bissell (Transport and Mobility); Michael Cameron (Transport and Mobility); Professor Julian Webb, Associate Professor Jeannie Paterson, Annabel Tresise and Associate Professor Tim Miller (AI and Legal Services); Dr Adam Henschke (Defence, Security and Emergency Response); Professor Seumas Miller (Machine Learning and Autonomous Weapons); Dr Reuben Steff and Dr Joe Burton (Defence, Security and Emergency Response); 3A Institute led by Robert Hanson (Government); Westpac Technology (FinTech); Professor Mark Pickering and Dr Dimitrios Salampasis (FinTech); Koren O'Brien (FinTech); Dr Mark Lawrence (Finance); Dr Tiberio Caetano and Andrew Stead (SMEs and Start-ups). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

2.1 Introduction

The application of AI technologies within public and private sectors can provide national economic benefit and social value. AI represents the potential to address social problems, such as climate change, an ageing population and emergency response, as well as provide technologies and methods to enhance productivity. According to the Centre for Data Innovation (2016), the evolving nature of AI technologies and associated applications means that 'it is difficult to predict just how much value AI will generate'. Nevertheless, increases in global GDP by 2030 have been predicted as a result of the adoption of AI technologies, productivity gains from automation and augmentation of the existing workforce, and increased consumer demand of AI enhanced products and services (PwC, 2017). It is anticipated that the economies of Africa, Oceania and

Asia (other than China and developed Asia) will experience GDP gains of 5.6 percent as a result of AI adoption whilst China's GDP might grow by 26.1 percent. However, realising these gains will be largely dependent on the adoption and strategic deployment of AI technology by companies and industry (Bughin *et al.*, 2017). McKinsey states that gains from AI adoption are most likely to be experienced by developed economies with slower productivity growth. According to the McKinsey report, Australia is within the average threshold of global AI investment and research, but has higher than average potential to benefit from automation driven productivity gains (Bughin et al., 2018).

AI technologies have application across a wide range of industry sectors, however the pace of adoption and the value add will vary by sector. Sectors will need to apply AI techniques to specific areas of value that will most benefit from the use of AI technology and the potential to realise gains from the use of AI will be reliant on the availability of data and the applicability of algorithmic solutions. Those industries with complex business operations are reliant on forecasting and accurate decision making and are most likely to be at the forefront of AI adoption. McKinsey examined 19 industry sectors and identified the areas that are likely to experience the most incremental value from the use of AI technologies over other analytic techniques. For example, transport and logistics, supply chain manufacturing and oil and gas were among the areas discussed to have great potential to use and benefit from AI technologies. Industries which are expected to derive relatively smaller value from the use of AI when compared to other analytical techniques include insurance, advanced electronics and aerospace and defence (Chui et al., 2018; Peng, 2018). Given the rapid pace at which AI technologies and applications are anticipated to emerge, it will be necessary that industry are aware of, and are responsive to, these new developments (Callaghan Innovation, 2018).

The development of powerful mathematical tools and the increase in computer power has combined to make AI useful for a wide range of tasks. These tasks include: translating speech; enhanced computer vision and object tracking from video; enabling driverless cars; deep analysis of large datasets to find patterns and relationships; chatbots; and control of robotics in manufacturing and health and agricultural settings.

This chapter describes some of the key techniques of AI, the need for more powerful computing and explores how AI techniques can be applied across sectors of the Australian economy. Data considerations, including data governance, collection, storage and use are presented in further detail in Chapter 6.

2.2 AI Technology

As noted in the introduction, AI is not a specific technology, but rather a collection of computer methods and techniques (Figure 1). These techniques include machine learning (ML) and natural language processing (NLP) and 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 supervised learning, the AI learns a function from data labelled by humans. For example, an AI model used to distinguish human faces may derive its knowledge from a library of images where humans have labelled those that contain human faces. An unsupervised learning approach is where an AI model learns by improving its actions against a well-defined objective. For example, an AI model might learn to play chess with the aim of winning more games. Supervised learning can outperform humans (i.e. it can learn more data than any human can process). Further, there is the possibility for AI to learn in a semi-supervised mode using both labelled and unlabelled data.

2.2.1 Machine learning

ML is an important component of AI. It refers to making computers execute tasks through processes of 'learning' that derive inspiration from (but are not reducible to) human intelligence and decision making. Through analysis of large volumes of data and examples, ML algorithms can autonomously improve their learning over time. ML relies on algorithms ranging from basic decision trees through to artificial neural networks that classify information by mimicking the structure of the human brain. Predictive and anticipatory forms of ML are qualities that differentiate AI from previous forms of automation.

The use of ML techniques is based on large databases of information. This information is often collected and provided by a small number of powerful economic actors, such as Facebook, Google, Amazon, Alibaba and Uber, as well as institutions such as government, health providers and medical practitioners. However, this should not disparage the actual and potential benefits of using ML techniques across a wide range of sectors. Here as elsewhere, the question arises as to whether this new general purpose technology will be used to primarily meet important human needs (e.g. assist in the provision of analysed bulk data on which to base public policies to combat poverty, or the design of robots to assist people with a disability), as opposed to satisfying the economic imperatives of a small group of dominant market actors (e.g. profile based targeted advertising and marketing). The functions and needs of the public and private arenas often differ. As a result, AI may be designed in fundamentally different ways in order to meet varying applications of use.

ML is a highly effective pattern recognition system, however alongside the opportunities, there are questions about the benefits and technical limitations of ML and the distribution of these to disadvantaged groups under existing and emerging local, national and global institutional arrangements. Some of the issues facing the operation of ML include the following:

- It can be difficult – even for an expert – to understand how a ML system produces its results (the so-called 'black box' problem). There can be deficiencies and limitations of internal validity within the datasets that ML techniques rely on, such as false data and databases of information to which owners have not provided consent to access. This includes not only

Box 1: ML and privacy, national security and human rights concerns

In addition to the privacy, national security and human rights concerns associated with platform companies generating and providing data for ML applications, there is also concern that the widespread use of ML techniques by government to collect large volumes of data may undermine individual rights. In countries with liberal democracies, there are political, legal and other deliberative processes that are well underway to try to 'balance' security requirements against individual rights. However, these issues are far from resolved when it comes to the collection of bulk data comprised of personal information and the use of ML techniques.

The recent Cambridge Analytica scandal provides a case study of the moral problems and national security implications that can arise as a result of cooperation between state and market actors. Facebook and Cambridge Analytica are both market actors, yet bulk data stored by Facebook and ML techniques deployed by Cambridge Analytica played a central role in the targeting of 'vulnerable' voters in an attempt to undermine US democratic processes. Potential regulatory measures to deal with this problem include bans on micro-targeted political advertising, mandatory transparency by way of public registers of the source of any political messages being disseminated and, at a more general level, deeming Facebook and other social media platforms to be publishers or to have similar responsibilities and legal liabilities to those of publishers in respect of certain types of content communicated via their platforms.

personal data, but also data collected via 'background' technologies such as internet-enabled traffic lights, electric meters and sewerage systems

- Questionable or undesirable uses of ML techniques, such as Cambridge Analytica's use of ML techniques to intervene in electoral processes in the US and elsewhere.

International and domestic law and human rights standards provide a framework through which to assess and respond to these challenges. Rather than inventing new standards for AI, we should use the existing frameworks and, if there are gaps, explore how they can be filled. Indeed, existing frameworks may require expansion or addition in order to address new, AI specific challenges. In the same way that Europe has driven responsible regulation around the use of data, Australia has an opportunity to lead in the development of initiatives and frameworks governing the safe and ethical use of AI technologies.

2.2.2 Natural language processing

NLP is a core pillar of AI and encompasses all AI technologies related to the analysis, interpretation and generation (of text and speech based) natural language. NLP has prominent applications including machine translation (e.g. Google Translate), dialogue systems (e.g. the back-end systems that underlie Google's Assistant, Apple's Siri and Amazon's Alexa), and automatic question answering (e.g. IBM's Project Debater).

NLP has matured substantially in the past decade due to the unprecedented amount of language being produced, shared and recorded in electronic and spoken forms. This large volume of language information requires automated analysis and represents significant opportunities and challenges. NLP is strongly contributing to this automation as well as improving accuracy and tractability in production systems. Some of the elements of this maturation include better language models (meaning more reliable and fluent natural language outputs); a move towards character-based models rather than word-based models (leading to better handling of rare, misspelled, and otherwise, low-frequency words); and improvements in the ability to train models over multimodal inputs (e.g. text and images), vastly improving the accuracy of models at tasks such as image captioning. Many of these developments have been driven by 'deep learning' – a subset of ML that pervades modern-day NLP. However, NLP still has limitations as demonstrated by the Winograd Schema Challenge, a test of machine intelligence. The Winograd Schema tasks computer programs with answering carefully tailored questions that require common sense reasoning to solve. The results from the first annual Winograd Schema Challenge ranged from the low 30th percentile in answering correctly to the high 50s, suggesting that further research is required to develop systems that can handle such tests. Notably, human subjects were asked the same questions and scored much higher, with an overall average of approximately 90 percent (Ortiz, 2016).

Most of the advances in NLP over the past decade have been achieved with specific tasks and datasets, which are driven by ever larger datasets. However, NLP is only as good as the data set underlying it. If not appropriately trained, NLP models can accentuate bias in underlying datasets, leading to systems that work better for users who are overrepresented in the training data. Further, NLP is currently unable to distinguish between data or language that is irrelevant and damaging. This can create inherent inequities in the ability of different populations to benefit from AI; it can also actively disadvantage populations. To alleviate such biases, there generally needs to be explicit knowledge of the existence of the bias, with training data then used to mitigate the bias. This is discussed in further detail in Chapter 7.

There are advances in NLP capabilities that can be expected in the next decade, including models that can justify their outputs to humans, NLP with world knowledge and multilingual support. These are discussed in further detail in Box 2.

2.2.3 Computing hardware

Future AI and ML will require large amounts of data, and traditional computers are starting to reach the limits of possible data processing power. Classical computer architectures do not scale well with respect to the power required to process the information, and much of the complexity experienced in AI and ML scales poorly with increasing datasets and breadth of task attempted (such as language-independent or multi-lingual NLP). Quantum computers have the potential to alleviate this problem. The development of quantum computers, on both experimental and theoretical fronts, has accelerated in the past few years, due mainly to increased investment from industry and governments (Palmer, 2017).

The fundamental component in a quantum computer is the quantum bit, or qubit. Classical bits, which can take on binary values of either 1 or 0, act as strings of on/off switches that underlie computation. A qubit, on the other hand, can represent both 1 and 0 at the same time. This is referred to as 'superposition' of 0 and 1 states. In a quantum computer, binary strings can be encoded over multiple qubits, and the subsequent quantum register put into a superposition

Box 2: Advances in NLP capabilities

Models that can justify their outputs to humans

Trust and accountability will become increasingly important when it comes to tracing the provenance of a NLP output.

World knowledge

'World knowledge' refers to a model's ability to derive meaning from language by resolving ambiguities or picking up on subtle inferences based on background knowledge of the world.

Cross-domain and cross-task robustness

Significant advances are expected in general-purpose language processing through cross-training across multiple tasks and explicit domain debiasing. These advances will mean that off-the-shelf system components can be applied to novel tasks or domains with reasonable expectation of competitive performance.

Improved context modelling

Most NLP systems still operate at the sentence level. A NLP system will usually process a document by first partitioning it into its component sentences before processing the sentences one at a time independent of each other. Large advances are expected in the modelling of context, beyond simple document context to include social context (e.g. personalising the translation based on the identity of the author and their social network, the intended audience for the translation, or a particular viewpoint on the content) or author demographics (e.g. personalising the translation of a document or the output of a chatbot to a particular persona, in terms of age, gender and language background).

Multimodal processing

When humans learn language, they do so over a lifetime, in a rich, situated context using all their senses with a myriad of feedback mechanisms. There is an increasing body of work attempting to achieve similar outputs for NLP via multimodal AI – most notably by combining text and image analysis.

Improvements in task-oriented discourse processing

There have been many advances where hands-free language-based interaction with an intelligent agent enable more effective decision making. These include automobile navigation systems, and question-answer customer service bots, which are enabling more flexible interactions.

of states. However, the power of a quantum computer is not solely derived from the superposition over binary numbers. Rather, a quantum computer derives its power from its ability to correlate qubits with each other, thus enabling quantum logic through what is known as 'entanglement'.

While the long-term vision of a universal quantum computer is reasonably well understood theoretically in terms of the types of tasks that could be carried out, the experimental and engineering challenges involved in realising such a machine mean that it is likely to be decades away.

There are several opportunities for conventional ML to assist the development and deployment of quantum technology itself. For example, the design and optimisation of complex control sequences and analysis of quantum measurement data lends itself to a ML paradigm, and there are a number of examples of this application already (see for example Kalantre et al., 2018).

In Australia, research and development in quantum hardware is well supported through the Australian Research Council Centre of Excellence Scheme and the National Innovation and Science Agenda (e.g. the Centre for Quantum Computation & Communication Technology). Research and support in quantum software, specifically associated with quantum and ML, is more diffuse. One pathway for the future is for quantum software and hardware and ML communities to work together. Indeed, Australia's first quantum computing hardware company Silicon Quantum Computing, launched in 2017, is working to develop and commercialise a quantum computer by 2022 using intellectual property from the Australian Centre of Excellence for Quantum Computation and Communication Technology (CQC2T). The company is owned by the Australian Government, the Commonwealth Bank of Australia, Telstra, the University of NSW and the NSW State Government.

2.3 AI applications

Advances in ML, NLP and computing hardware will drive AI over the coming decades. Innovative researchers and companies are already establishing new AI-based products and services across a number of sectors.

This chapter aims not to provide a complete overview of AI applications, but rather to illustrate the breadth of applications that are being developed across sectors. Some applications will bring AI into our lives in overt ways, while others are likely to work in the background. Few areas of our lives will not be influenced by these advances and many industry sectors are expected to derive benefit from the use of AI technologies. However, given the evolving nature of the technology, which, in many instances, has unforeseen applications, there remains uncertainty as to the precise ways in which AI will transform and deliver benefit to industry. Therefore, to ensure growth throughout the economy, investment should not be restricted to singular sectors, but be wide ranging and explorative.

In parallel to industry innovation and investment, industry bodies have also started to develop and implement safety and ethical standards in relation to the development and application of AI. For example, at the request of its employees, Google pledged not to develop autonomous weapons. Additionally, the company also developed a set of principles and guidelines for the development of AI technologies, which includes a commitment to ensure the socially beneficial nature of AI (Mehta, 2018). To ensure both competitiveness and the safe implementation of AI systems, Australian business sectors should remain aware and responsive to the activities of global peers.

The applications of AI technologies discussed here include robotics, manufacturing, health and aged care, resources (mining and energy), environment, arts and culture, agriculture, transport and mobility, justice and law, defence and emergency response, government, financial services and infrastructure requirements. The opportunities highlighted are by no means exhaustive, rather they are illustrative of some of the recent advancements across different sectors. This broad range of applications is driving social change and, in some instances, influencing demand for further technological development. Inclusive design and equity of access will be crucial to ensure that everyone will be able to access the benefits and applications of AI.

2.3.1 Robotics

AI and robotics have a long history of interaction. In the 1950s, robotics arose from the fields of mechanical and electrical engineering, while AI arose out of the then-new field of computer science. Early researchers explored the nature of intelligence and the possibility of programming computers to solve problems that were traditionally seen as requiring human cognitive skills.

Robotics is often characterised as the intelligent connection of perception to action in engineered systems (Brady, 1984). Robotics include not only human-like robots, but any kind of technological system that uses sensors such as cameras, thermal imagers, tactile and sound sensors to collect data about the operational environment and build a 'world model'. A world model is an internal representation of the surrounding environment that enables a robot to interact with its surroundings. These interactions may include navigation tasks such as obstacle avoidance and trajectory planning, or manipulation and sensory tasks. Data from multiple sensors are combined using probabilistic sensor fusion techniques, and simultaneous localisation and mapping algorithms determine both the world model and the motion of the sensors or robot through the world. Application domains for robots include agriculture, mining, transport, defence, medical assistance and consumer services. The Australian Centre for Robotic Vision's Robotic Roadmap report (2018: p. 131) suggests that intelligent robotics and physical automation could provide a 'cost-effective way of addressing global maintenance and construction issues', especially in terms of 'building and maintaining (especially ageing) infrastructure, or difficult-to-access infrastructure over large geographic areas, while removing humans from dangerous working environments'.

Box 3: Strategic opportunities for AI-augmented robotics in Australia

There are several areas where substantially autonomous (AI-augmented) robotic systems are of significant strategic importance to Australia. Two examples are autonomous vehicles and unmanned aerial vehicles.

Autonomous unmanned ground vehicles that can operate in cross-country conditions are useful for applications such as long-distance transportation, mining, agriculture, biosecurity and biodiversity assessments, science surveys, and safety. Communication and technological advances are needed in the combination of localisation and mapping methods, motion planning, obstacle detection, obstacle avoidance and situation awareness, as well as in translating these technologies into operational and useful platforms that can increase productivity and safety across a wide range of applications (Kunze et al., 2018).

Unmanned aerial vehicles, popularly known as drones, are considered a core technology for a future digital society. They are especially important for Australia, which has low population density with only a small number of cities. The civilian market is booming, with the vehicles being primarily used to generate data at local scales. In the future, unmanned aerial vehicles may be extensively used for transport, delivery services, medical supply services, biosecurity assessments, agricultural surveys, and border surveillance on a continental scale. Indeed, a technology company is trialling drone deliveries of food and chemist supplies in Canberra with plans to make the service permanent and expand delivery locations (Jervis-Bardy, 2018).

For the next generation of robotics, priorities include the development of core technologies for highly autonomous, competent, and reliable robotic systems that can execute complex missions in Australia's unique environments. These robots also need to interact safely and seamlessly with humans and other dynamic agents and be deployed in a range of applications that are of strategic relevance to Australia and the world.

2.3.2 Manufacturing

Recent manufacturing trends are redefining business strategies across the sector (Lasi et al., 2014). The increasing adoption of sustainable practices, stronger demand for personalised products, blurred boundaries between manufacturing and the services sector and an interest from Australian producers in high-value activities across global supply chains, are all imposing opportunities and challenges for the domestic industry. Manufacturers are seeking alternative solutions to seize global opportunities.

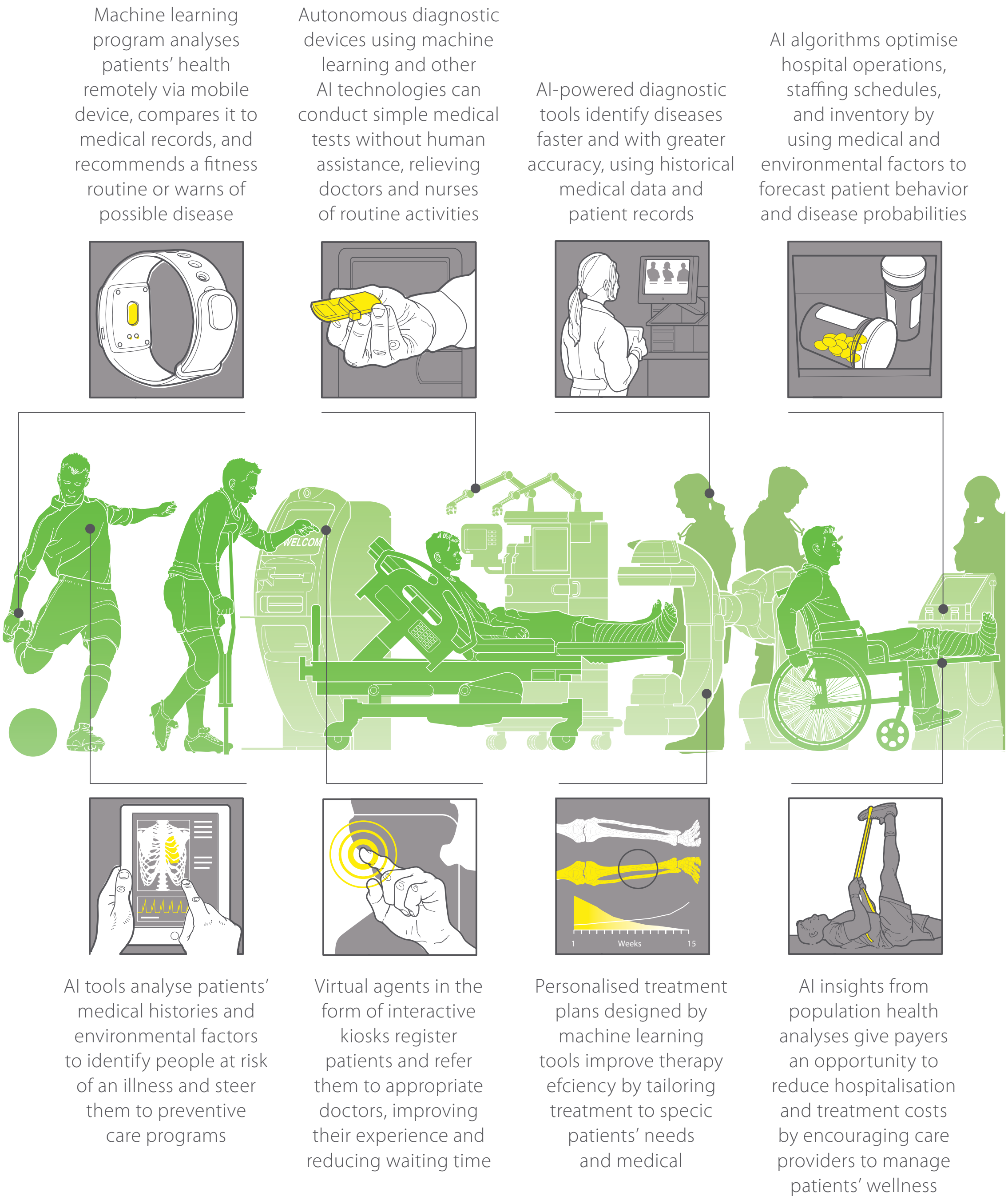
An area of opportunity is the development of AI assistive technologies to enhance manufacturing workers, rather than replace them through automation. This is an approach that has been suggested as having the potential to be an enabler for economic success (Brea et al., 2013). As AI-driven automation lowers production costs, there are opportunities for Australia to increase its competitiveness of manufacturing goods . In this context, AI may augment work processes, increase safety, or work with humans to increase productivity. However, if AI enhances productivity in these ways, it may result in companies employing fewer staff. Whether or not people are assisted or replaced in their workplaces may be considered a commercial and social issue rather than a technological one (see Chapter 3).

Manufacturing efficiencies can be gained through AI-driven automation and optimisation, and are essential for high cost economies, such as Australia, to remain competitive. AI and ML can be used within the manufacturing sector to optimise the manufacturing process. Across many areas of manufacturing and technology, there is a move from production lines of identical items to a new wave of increasingly flexible, adaptive, and customised products. Some of this is being driven by rapid prototyping and construction techniques. Diversity is provided by libraries of selectable 'base components' that are assembled into bespoke solutions. One area that is poised to benefit from this shift is robotics. Conventional automation, such as that used in automotive manufacturing, is driven by the need to automate specific mass manufacturing tasks. However, economic drivers demand less focus on large volume production, and more concentration on mass personalisation of products. This macroeconomic environment predicates a national quest for affordable assistive AI enabled automation that supports high-variety, low-volume production runs that are easy to implement, and highly flexible, adaptable operational processes.

These new technical capabilities, initially leveraged from advances in assistive mining technologies, are placing Australia at the forefront of a new AI assistive systems industry. The capabilities will also serve as a foundation for the next phase of the manufacturing evolution in Australia in association with future advances in AI and digitisation.

2.3.3 Health and aged care

There are numerous possibilities for the application of AI in health, many of which are already underway (Figure 3). The linkage of data provides opportunities for positive



Machine learning program analyses patients' health remotely via mobile device, compares it to medical records, and recommends a fitness routine or warns of possible disease

Autonomous diagnostic devices using machine learning and other AI technologies can conduct simple medical tests without human assistance, relieving doctors and nurses of routine activities

AI-powered diagnostic tools identify diseases faster and with greater accuracy, using historical medical data and patient records

AI algorithms optimise hospital operations, staffing schedules, and inventory by using medical and environmental factors to forecast patient behavior and disease probabilities

AI tools analyse patients' medical histories and environmental factors to identify people at risk of an illness and steer them to preventive care programs

Virtual agents in the form of interactive kiosks register patients and refer them to appropriate doctors, improving their experience and reducing waiting time

Personalised treatment plans designed by machine learning tools improve therapy efficiency by tailoring treatment to specific patients' needs and medical

AI insights from population health analyses give payers an opportunity to reduce hospitalisation and treatment costs by encouraging care providers to manage patients' wellness

Figure 3: AI in health care

Adapted from McKinsey Global Institute, 2017.

outcomes from understanding disease pathways, early recognition of disease, monitoring disease and end of life care. We will need systems of analysis to facilitate and optimise the use of health data to enhance medical understanding and improve care.

While still in the early stages of development, AI systems could be used to process large amounts of medical literature, clinical notes, and guidelines using NLP; interpret the results of diagnostic tests; design more personalised treatment plans based on data extracted from the analysis of large numbers of patients; detect fraud, waste, and error in health insurance; and collect, summarize, and relay information about a patient back to clinicians in a continuous loop. While the early developments in this area are promising, some aspects remain undeveloped. For example, there are currently limitations with this technology as highlighted by IBM's Watson, which has demonstrated difficulty in learning about the different forms of cancer (Ross and Swetlitz, 2017). Ensuring the systems are fit for purpose before integrating them into the healthcare system will be critical, as will establishing trust, transparency, and explainability in these AI systems and processes.

However, despite the limitations experienced with some technologies, there is also a role for AI to identify people at high risk of developing a chronic condition, experiencing an adverse event, or recovering poorly from an injury, by combining individual health data with expert opinion. AI-enabled systems could also allow clinicians to compare themselves to others, allowing them to make fair comparisons that consider the varying composition of patients they treat. AI solutions that are capable of recognizing and responding to human emotions have great potential to deliver computer-controlled assistants that can interact with humans, known as Intelligent Virtual Agents (Bedi et al., 2015; Luxton, 2015).

Further, AI can be used to tailor individual treatments or diagnosis. In Australia, a Queensland based collaboration has received federal government funding of A$2.6 million to help advance cancer patient treatment using AI. The collaboration will use AI to examine genetic information of an individual to tailor cancer treatment with the view to prescribe the most effective treatment for individuals (Crockford, 2018). In New Zealand, MedicMind has recently created a world-first AI medical platform for medical researchers and clinicians, that will eventually use AI to auto-diagnose a large range of diseases based on a single photograph (MedicMind, 2018).

Another example includes a recently announced collaboration between Microsoft and SRL Diagnostics in India, that aims to use AI to improve early detection cancerous cells. This collaboration will train Microsoft's AI system using more than one million pathology samples from SRL's records to make faster and more accurate assessments that support doctors to make a quick and accurate diagnosis. This will reduce the cost of treatment and better utilize the time of trained specialists to reach more patients (Microsoft, 2018b).

As the population ages, there is an increasing need for AI technologies that can help older people who wish to live independently or who suffer from chronic diseases. AI technologies and software applications can provide cognitive aids for monitoring health status, assistance in managing medication, therapies such as physical and breathing exercises, as well as tools for classifying activity patterns, detecting possible falls, entertainment, and social support for loneliness.

The government can play a crucial role in speeding up the adoption of AI in health by informing citizens about the benefits and implications of the personalisation of medicine. To achieve any level of personalised medicine, some personal data will need to be analysed, and an honest and informed discussion will need to occur between policy makers and citizens regarding the use of data.

Box 4: Case study: AI's application in health and aged care

In New Zealand, there have been field trials over the past decade to determine the acceptability, feasibility, and effectiveness of robots as cognitive aids to deliver aged care support. The results from the research have suggested that robot assistance is acceptable to the elderly and staff in aged care, that it is feasible to deploy such robots in people's homes and in aged care facilities, and that there can be cost benefits (Broadbent et al., 2016). MiCare in Australia has introduced autonomous mobile robots in aged care facilities to assist staff in the delivery of services, allowing for greater efficiencies and improvements in patient care (Stoyles, 2017). Additionally, studies of a companion pet robot suggested that robots can have psychological and physiological benefits for older people (Robinson, Broadbent and MacDonald, 2016).

Australian and New Zealand hospitals are also using smart technologies to manage rehabilitation for stroke sufferers and developing effective direct brain-computer communication to help people control prosthetics and communicate with technology that can provide assistance. Researchers are studying the feasibility of using data gathered from wearable devices such as accelerometers, to estimate people's activity patterns, to indicate conditions such as dementia, and to detect events such as falls.

The use of AI in healthcare is underway and will benefit from the development of ethical systems to facilitate implementation.

2.3.4 Arts and culture

The creative industries are typically at the forefront of technological change, embracing novelty, often at the sacrifice of employment stability (Threadgold, 2018). Despite references in the media to creative work being 'fun and free' (Brooke and Wissinger, 2017), many creative tasks involve menial and repetitive physical and mental routines that are appropriate for automation. AI could bolster numerous creative tasks across the range of different sectors from music to film, with examples including composing music and rendering digitally generated characters in a film to make them look more realistic.

There are many examples of applying AI in the creative industries. For example, AI is being used by platform companies such as Netflix and Spotify to determine gaps in creative content production and generate content recommendations. Software development tools such as Unity are providing videogame developers with ML techniques to promote innovative design. The Epoch Times, an Australian company, drew on an AI tool to gather insights into online readers and subscribers, to help the company provide tailored news and content. The computer generated imagery (CGI) used in films such as *The Hobbit* and *Lord of the Rings* relied on AI software to automate the individual movements and interactions of virtual soldiers so that they appeared lifelike and convincing, without recruiting hosts of humans in expensive and environmentally damaging real-world simulated conflicts. In the 2012 art instillation series, *Fifty Sisters*, academic and artist John McCormack electronically 'grew' plants using computer code. Using sensors, visitors to the gallery space found that their movements responded to various stimuli and influenced the growth of the flowers.

Box 5: Case study: AI in audio mastering

Various companies have started to use AI enabled audio mastering, web-based platform company LANDR being one example. LANDR and other AI-based audio mastering applications use AI to analyse audio mixes and then apply different mastering processes depending on what the AI engine determines as the audio mix needs. While not yet widely adopted, AI applications present audio mastering engineers with opportunities to reduce costs and decrease the potential introduction of error to the process. AI could subsequently increase opportunities for experts by creating a market for audio mastering among people who would not normally use the service or through promoting their own expertise through comparing their results to AI. Likewise, humans could offer mixtures of AI and their own signature sound through using AI to automate the routine or menial tasks in their work and to offer more cost-effective options. Humans could offer their critical listening skills to audit, or vet, the productions of AI masters to ensure acceptable standards and would therefore remain crucial as a safeguard against misjudgements by AI, yet be removed from the actual processes of labour. Finally, AI could be used in the toolchain alongside other analogue and digital technologies with differing degrees of automation. AI could be consigned to only menial tasks, such as error correction of metadata insertion into physical or digital media.

The application of AI to arts and culture can present mixed responses with experts in the field being uncertain about possible disruptions to their work (see for example, Ames, 2018). However, there are predictions that the music industry (Box 5) could enlist AI to 'create algorithms enabling the creation of customised songs for users', with the aim of helping 'sound creators to focus more on being creative', thereby boosting revenue (Naveed, Watanabe and Neittaanmäki, 2017).

2.3.5 Mining

Automated mining operations in Australia represent some of the largest industrial automation programs in the world. They are a combination of automated hauling and drilling, intelligent sensing, mine-wide asset and supply-chain optimisation, and remote tele-operation. Mine sites are less structured than many other application areas and autonomous haulage systems must operate in dusty and harsh environments as well as react to a range of unexpected events such as debris on the road. The application of AI technology in mining provides opportunities to reduce operational risk and increase competitiveness.

In a progression towards fully automated, intelligent mines, several companies (including Rio Tinto, BHP, Stanwell, Suncor, and Fortescue) have begun using autonomous haul trucks at their mines. Industrial vehicle manufacturers Komatsu, Caterpillar, and Hitachi have been developing these driverless haul trucks in close collaboration with mining operations, employing a combination of wireless communication, object-avoidance sensors, on-board computers, GPS systems, and AI software that enables the trucks to operate (semi-)autonomously where most trucks are supervised at a distance. Rio Tinto, which has employed a fleet of approximately 400 Komatsu haul trucks in its Pilbara mine, reports that their autonomous trucks have improved safety and cut costs by nearly 15 percent, partially due to the fact that the vehicles can be operated continuously. Further, Rio Tinto's autonomous trains are the first long distance, heavy haul trains and the world's largest and longest robots (RioTinto, 2018). The vehicles are controlled remotely from the Perth Operations Centre, 1,500 km away from the mine. Similarly, BHP is running remote operations centres to optimise mining, maintenance and logistic activities across the Pilbara.

Many other mining companies are using digital technology and machine automation to improve the productivity and safety of mining operations. Companies within Anglo American that have digitised their technical equipment have seen around a 30 percent improvement in their business (15 percent in productivity and 15 percent in cost savings) (Bloomberg New Energy Finance, 2018).

AI is being used in an effort to improve mineral exploration. In 2017, mining giant Goldcorp teamed with IBM Watson to comb through a vast quantity of geological information to find better gold deposits. Further, Australian gold miner, Resolute Mining, is building the world's first fully autonomous underground mine in Africa. As the level of automation increases it is likely that there will be a decrease in workers being flown in to work on the mines. In addition to autonomous vehicles, many mining companies are incorporating digital innovations into their operations to further enhance performance. This is ranging from remote monitoring and sensing, improving decision making through the use of real-time data, analytics and predictive tools, and block chain technology. AI and ML can help increase the efficiency of mining operations by improving the precision of mining, resulting in less consumption of water and energy, and less production of waste.

Through improving application of sensor technology, advanced analytics and process automation, mining sector digital advances have potential to add A$40-$80 billion in earnings before interest and taxes. Capturing this opportunity will require end-to-end integration for real-time performance monitoring, optimisation and control (McKinsey & Company, 2017).

2.3.6 Energy

The Australian energy sector is currently undergoing profound changes triggered by high energy prices, the adoption of new technologies for renewable energy and storage (both residential and commercial), a growing consumer preference for increased control over energy usage, and the proliferation of smart energy tools such as smart meters. These changes are putting pressure on existing business models and creating the need for new regulations, policies and incentives. There is significant uncertainty concerning the transition to a cost-effective, secure, reliable and sustainable energy future. One example is in the planning of a carbon-free, cost-effective and secure future grid, including the ideal mix of generation and storage technologies over the next 30 years, the ideal locations of these generation and storage sources, and expansion of the network.

AI will play a crucial role in addressing these challenges and providing technological solutions. There is, for example, a need for AI systems that manage and optimise energy consumption, production and storage in residential and commercial buildings. This is a well-developed research area, and there are commercial systems in Australia that perform some of these functions. Examples include the BuildingIQ 5i platform for heating, ventilation and air-conditioning management in buildings, the Evergen Energy Management System and Reposit Gridcredits. Melbourne Water is also using AI to reduce electricity usage through the regulation and optimisation of pump speeds. The use of AI to determine and implement optimal pump settings is expected to reduce Melbourne Water's energy costs by 20 percent per annum (Melbourne Water Corporation, 2018). AI systems may play a role in coordinating distributed energy resources (such as residential energy management systems) to manage network constraints. AI systems that can handle automated planning and scheduling will play an important role, whether to schedule loads, coordinate distributed energy resources, plan renewable deployment and the expansion of the future grid, or to restore power systems following outages.

Many countries have public or private research institutes dedicated to future energy systems, which are well-versed in state of the art data analytics and AI and optimisation techniques (e.g. North America has the National Renewable Energy Laboratory and the Electric Power Research Institute; the French company Électricité de France has over 2,000 staff in six research centres across Europe; and Iran has the Niroo Research Institute). Such specialised R&D labs are lacking in Australia (not just in energy), which can make it more difficult for basic and applied research from universities and publicly funded research organisations to find its way into consumer products. Notably, France is starting to harness the power of AI by creating AI interdisciplinary institutes in selected public higher education and research establishments; each institute will focus on a specific application area.

Here as elsewhere, there are ethical concerns surrounding adoption of AI. For example, data analytics and smart meters might be used in potentially intrusive or compromising ways. Likewise, the use of AI systems to share, aggregate, and coordinate resources among participants in smart energy systems may exacerbate existing inequalities for people who cannot afford rooftop solar, batteries and energy management systems and AI systems may automate tasks performed by energy sector workers. Addressing these concerns will require a measured response from industry and government.

2.3.7 Environment

AI offers a new way by which to address environmental concerns and can make a significant contribution to the management of urban, rural, and natural environments. Climate change increases the complexity and uncertainty of managing the environment. Developing AI to augment human decision making and policy development will be important for ensuring environmental sustainability. According to the World Economic Forum, the use of AI technologies has immense potential to provide solutions for the Earth's greatest environmental challenges. This assertion has been echoed by the World Wide Fund for Nature (WWF) Australia. However, as with other electronic hardware, the development of AI technologies could also contribute to environmental degradation due to the extraction of materials to build equipment through to the disposal of superseded hardware. Therefore, an important consideration will be the sustainable development of AI technologies. Regulation or governance has been suggested as a way to ensure environmentally friendly AI systems (World Economic Forum, 2018a). Additionally, NGOs, social enterprises, academia and industry partners, could play a role in guiding the use of emerging technologies for public good. For example, WWF Australia has established Panda Labs, an innovation program, which, in conjunction with industry, start-up, and academic sectors, seeks to develop and advance emerging technologies for beneficial social and environmental impact.

Examples of AI used for environmental applications include the use of AI technologies and techniques to analyse data output from smart cities to improve the liveability and sustainability of cities or enhance conservation practices (World Wide Fund for Nature Australia, 2017). For instance, Deakin University has developed AI that has the ability to extensively record and track animals in national parks for the purposes of generating new park management processes (Deakin University, 2018). Given Australia's diverse natural environment, the country could present a testbed for the development and evaluation of AI solutions for environmental issues such as climate change and renewable resources. Currently, robots using AI powered software have been deployed to assist in the preservation of the Great Barrier Reef. Resulting from a collaboration between Queensland University of Technology, Google and the Great Barrier Reef Foundation, the RangerBot (previously known as COTSbot) uses computer vision to administer a lethal injection into the crown of thorn starfish that pose a significant threat to the reef. The robot is also capable of conducting underwater surveying and water sampling. This system is a world first and is significantly cheaper than traditional, acoustic, underwater systems (Gartry, 2018). Further, there is a trial underway in Western Australia to use AI to identify plants, vertebrates and insects that may pose a biosecurity threat to Australia.

Future developments in AI systems for environmental management may rely on data produced from the proliferation of internet-enabled devices, that is, the Internet of Things (IoT) (Mattern and Floerkemeier, 2010). IoT has the potential to allow for more efficient use of energy and resources by providing continuous streams of data that can be used by AI systems to model more effective responses to environmental issues. As an example, IoT devices could be set to use electricity only when there is an excess supply of renewable electricity. If electricity generated from coal-fired power stations late at night is cheap, such devices could exploit these sources.

Box 6: Case study: IBM's solution to China's air pollution problem

In 2014, IBM launched the 'Green Horizons' initiative in China, with the aim of alleviating air pollution in cities such as Beijing. The initiative draws on cognitive computing techniques in weather prediction and climate modelling to generate predictive models that indicate the source and likely effect of pollution. These systems enable city planners and officials to model scenarios and suggest potential actions to reduce the particulate load in the atmosphere. Since its inception, the initiative has reduced particulate load in the atmosphere by about 18 percent with minimal negative impact on the economy (because of source elimination).

2.3.8 Agriculture

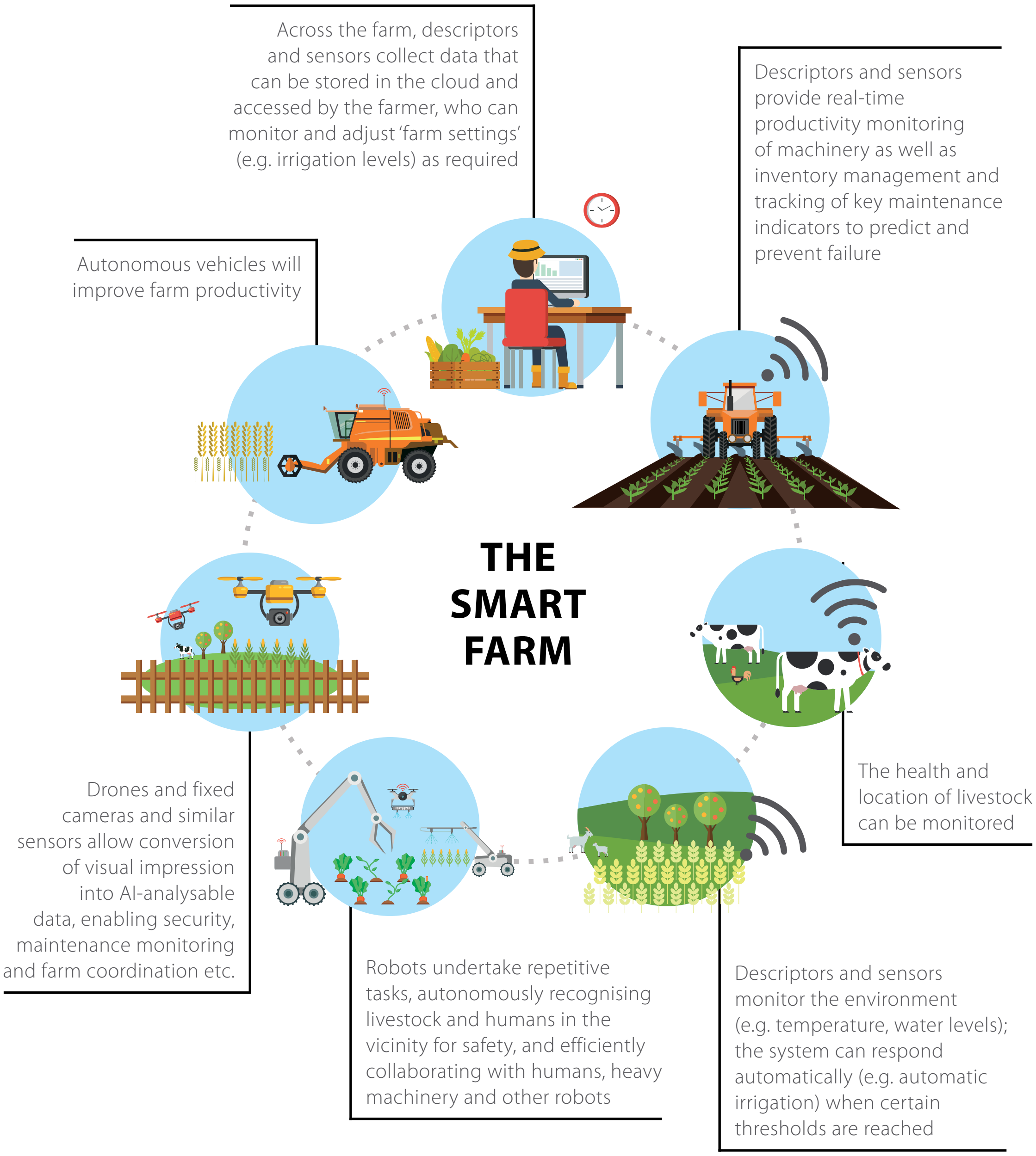
By 2050 there are likely to be close to 10 billion people on our planet, requiring a significant increase in food production. Most of this population growth will occur in Africa and Asia, where there will be an increased demand for higher quality and quantity of food, more protein-rich foods, fruit and vegetables, and an increasing vegetarian and vegan market. This will also result in a desire to reduce environmental footprint including chemical use, to improve soil management, to reduce on-farm waste and energy consumption, and to improve animal welfare. These increasing demands and desires will affect farmers who are seeking to meet demand, while dealing with issues such as climate change. Consumer demands for quality at low cost reduces growers' margins, further exacerbating the challenges. This has led to farmers around the world seeking new technologies that can help with their daily tasks on-farm, as well as provide a competitive economic edge.

AI systems and technologies are poised to have a major impact on-farm production of crops and livestock. A development that is already taking place is the design of autonomous farming machines, which can work throughout the night, collaborate with human and non-human peers, and request assistance if a condition arises that has not been programmed. There is increasing use of precision agriculture farming devices to collect and analyse data on, for example, crops and livestock, which can then be used to make informed farm management decisions (Figure 4). For instance, CSIRO is collaborating with partners in Queensland on the Digital Homestead project that aims to evaluate and demonstrate technologies that enable better decision making on farms, leading to improved productivity and profitability (CSIRO, 2015a). One of the technologies includes a solar-powered, wireless cattle collar that gathers information about the animal's location and behaviour. This provides information that can lead to better management decisions about grazing management, feed supply and when to muster. Additionally, researchers at the University of Sydney have developed an AI enabled robot that can identify weeds on an farm and autonomously apply herbicide in controlled amounts (Rural Industries Research and Development Corporation, 2016).

A component of AI, machine vision, can precisely locate a growing crop. This is superior to GPS guidance, which locates the vehicle relative to where the crop was thought to have been planted. Machine vision can assess the yield of fruit-bearing trees and, in the not-too-distant future, may lead to efficient selective harvesting.

While early applications are showing promise, key areas of advancement in this area include:

- Greater use of stochastic ML techniques that can capture and learn from semi-structured data, and that need to deal with very noisy and inconsistent data, such as changing light conditions, moving animals, plant variability and effects of different pests and diseases
- Developments in semi-supervised and unsupervised learning techniques to easily capture the great variety of food produced without the need for experts to train the algorithms for each type
- ML techniques for decision making, especially in automated crop and animal growth models, to assist in yield and quality prediction for each individual plant and animal
- Automated decision support tools that can identify what physical action needs to be undertaken to support the use of continuous on-farm robotic solutions. This includes automated mechanical weeding, targeted fertiliser applications, foreign object detection and removal, and eventually automated harvesting.



THE SMART FARM

Across the farm, descriptors and sensors collect data that can be stored in the cloud and accessed by the farmer, who can monitor and adjust 'farm settings' (e.g. irrigation levels) as required

Descriptors and sensors provide real-time productivity monitoring of machinery as well as inventory management and tracking of key maintenance indicators to predict and prevent failure

The health and location of livestock can be monitored

Descriptors and sensors monitor the environment (e.g. temperature, water levels); the system can respond automatically (e.g. automatic irrigation) when certain thresholds are reached

Robots undertake repetitive tasks, autonomously recognising livestock and humans in the vicinity for safety, and efficiently collaborating with humans, heavy machinery and other robots

Drones and fixed cameras and similar sensors allow conversion of visual impression into AI-analysable data, enabling security, maintenance monitoring and farm coordination etc.

Autonomous vehicles will improve farm productivity

Figure 4: AI and agriculture

Adapted from: Australian Computer Society, 2018a.

Box 7: Gaps in agriculture in Australia

The Australian agriculture market's relationship to technology is unique. There are many advances in agriculture technology that have been developed in Australia, or were initially tested here before going international. AI is expected to take the same pathway; however, for this to be realised, several gaps must be addressed.

The technical digital divide. Computer scientists and automation engineers lack practical agriculture knowledge and agricultural experts generally have little (or no) understanding of the complexities of technologies such as AI, ML and robotics. This is a common problem internationally, although in some countries where there has been a greater emphasis on digital technologies in food production (e.g. urban food production and large-scale greenhouses), multidisciplinary teams are being formed and there is additional training of engineers and computer scientists in agronomy or vice versa.

The spatial digital divide. Most AI courses, training programs, start-ups and AI communities occur within the city areas where there are large financial and engineering hubs. This generally means that any activity in AI for agriculture will gravitate to these areas at the expense of the rural areas where this knowledge is needed. This is a particularly significant challenge for Australian agriculture. A focus at the secondary school level, especially in rural schools, on greater ICT knowledge and the applications to food production would help facilitate bridging this divide.

Policy. Policies surrounding the implementation of AI technologies in agriculture in Australia are still in their infancy. Unlike other countries who are leading in the deployment and application of AI technologies, Australia has not implemented a national AI strategy. Industry could help rural development corporations develop a unified plan that is broad enough to deal with the various issues that AI can solve for the whole industry, while also dealing with the specific problems faced by commodities.

Data. There is great benefit in the collection and ownership of data across the value supply chain by those who write the algorithms. There is also benefit in the information being shared to support biosecurity concerns. However, the growers would be relinquishing a significant asset that could draw financial returns or could give up significant freedom of operation if the data were used improperly.

Telecommunication infrastructure. Telecommunication infrastructure is also another gap in Australia, with many developing countries having greater connectivity than Australia. This will affect how much data can be transferred, especially given that AI technologies are 'data hungry'. While it is currently difficult for ground-based networks to achieve 100 percent coverage, there is the potential that in the next 2-5 years use of LEO satellite constellations could provide pole-to-pole broadband coverage.

2.3.9 Transport and mobility

Autonomous vehicles (AVs) are among the most highly anticipated AI developments, and will have far-reaching impacts. AVs refer to a variety of transportation methods, including autonomous cars, planes, trains, trucks and ships. There has been international activity in the rollout of AVs, for instance, Rio Tinto is already using autonomous trucks and trains in Australian mining operations, France anticipates semi-autonomous trains to be deployed in 2020, with fully autonomous services being implemented by 2023, Rolls-Royce has partnered with Finnish universities to develop an autonomous unmanned ocean ship by 2035, and a Norwegian company is currently developing the world's first fully-autonomous, zero-emissions cargo ship (Rolls-Royce plc, 2016; Carlstrom, 2018; Railway Gazette, 2018). Given the expense associated with slow moving cargo ships, which remain a key component in global trade, the use of autonomous ships could significantly reduce costs associated with product transportation. Additionally, autonomous buses have been trialled at Australian cities and universities, with the initiatives demonstrating partnerships between academia, industry and government that have potential for widespread application (Monash Unviersity, 2018; Thomsen, 2018). The buses can operate on existing roadways without the need for additional infrastructure and can travel up to 45km per hour.

The potential benefits of road AVs are that they may decrease traffic accidents to almost zero, increase the efficiency of traffic control, decrease emissions, increase intermodality, improve accessibility for mobility impaired people, and increase social participation. However, these changes are anticipated to take effect over a prolonged period and we are likely to see semi-AVs long before fully-AVs. AI can also be applied to smarter road use for conventional vehicles through dynamic road pricing, optimising traffic management, and improved routing, utilising exiting roads more effectively.

The discourse around AVs promises more comfort and shorter travel times. However, the assumption that AI in transport will lead to any significant reductions in transport volumes has been challenged (Dennis and Urry, 2013; Dassen and Hajer, 2014; Rifkin, 2015; Maurer et al., 2016; Greenfield, 2017). Leading experts expect increases in vehicle use due to increased accessibility for people of all ages, including the aged, mobility impaired individuals and people who dislike driving (Dudenhöffer, 2008; Diez, 2017; Meyer and Shaheen, 2017). Forecasting has suggested that without the right planning and regulatory controls, there may be a danger of AVs creating more traffic, congestion, emissions and sprawl as a result of the increased uptake due to the convenience and comfort that may lead passengers to use AVs instead of mass public transport. Some also predict the increased use of AVs will lead to increasing social conflict and social inequalities, as early model AVs are likely to be expensive. The use of expensive technologies in these vehicles as well as the possible higher risks of accidents in a changed automated environment may result in insurance companies charging higher rates for driver-owners of cars. Additionally, truck, bus, taxi, and other transport drivers may find that their jobs are threatened by the uptake of AVs.

It is anticipated that it will take decades for AVs to replace all cars. Due to the slow rollout and varying costs associated with different levels of autonomy, it is likely that there will be AVs with different amounts of human involvement on the road simultaneously. Autonomous vehicles are likely to be shared, rather than individually owned (Deloitte University Press, 2016; McKinsey&Company, 2016). Moreover, the development of autonomous vehicles is likely to have significant implications for transport labour in Australia and New Zealand. This includes not only transport drivers, as discussed above, but also those associated with road maintenance and infrastructure. There is also the question of what people will do during AV transportation; work, rest, entertainment? Much will depend on public acceptance of new technologies and the trust people put into these highly complex systems (Fraedrich and Lenz, 2016). Carefully considered and implemented regulatory frameworks will provide a basis for public trust in AV and would signal that Australia and New Zealand are open for business for driverless technology. In 2018 the Australian Government announced that it would establish an Office of Future Transport Technologies to help prepare for the pending arrival of autonomous vehicles (Australian Government, 2018g). Approximately A$10 million in funding has been earmarked for the initiative which aims to improve transport and road safety outcomes.

Box 8: Impact of AI and AVs on social inequalities

Autonomous vehicle technologies have the potential to produce and perpetuate new and existing forms of social inequality. The design of autonomous vehicles, for example, is not necessarily value neutral. Research undertaken by Jensen (2007) highlights how the development of new mobility systems can intensify social segregation, leading to multi-tier services based on differential speed and comfort. For autonomous vehicles, the 'kinetic elite' (Elliott and Urry, 2010) may have greater access to autonomous vehicle services, allowing them to travel further and faster than others, and these privileged services may also provide higher levels of flexibility and comfort. Autonomous vehicles may also radically transform how car insurance operates, leading to new forms of inequality. This effect may be transient; however it will be important for industry and government to be aware of these potential inequalities and ensure equitable standards of design and implementation.

2.3.10 Justice and law

AI in the legal services focuses on the use of computer systems to perform or assist research, analysis and decision making normally performed by humans. Computers and automated services have assisted the legal profession for decades, using techniques such as Boolean keyword searching and simple hand-coded expert systems. In the legal profession today, AI is being used and developed to enable a range of automated solutions, including:

- Intelligent searching of primary sources of law and precedents
- Automated document review using predictive coding or statistical pattern analysis in, for example, contract analysis and e-discovery
- Smart forms that tailor legal information and advice to individual circumstances (e.g. to draft a will or settling financial arrangements following relationship breakdown or divorce)
- Legal data analytics for practice and judicial decision-support
- Online dispute resolution.

Box 9: AI in Australian judicial settings

Australia is more conservative than other countries in adopting AI technologies in judicial settings for criminal law, although there are some who advocate its use (Norton Rose Fulbright, 2018). The US, in contrast, has used AI-informed sentencing since the early 2000s (Dressel and Farid, 2018). Similarly, in China, 'robo-judges' have been used since 2016 to determine nearly 15,000 cases for criminal sentencing (Connor, 2017).

Criminal bail and sentencing are technically amenable to AI-informed decision making (Stobbs, Hunter and Bagaric, 2017: 261), but there remain critical questions about how such technology is to be used. The sentencing stage of trial requires the analysis of past sentencing decisions against the balancing of factors such as the maximum penalty, offence tariffs (if one exists for the offence in question), sentencing objectives and aggravating and mitigating considerations. Programs can build-in risk profiling and assessment factors that assist in determining whether a defendant is, for example, more likely to be a flight risk, or to re-offend (Stobbs, Hunter and Bagaric, 2017: 272). Supporters of an AI approach argue that once these factors are weighed, the result is quicker and more consistent than human decision making.

On the other hand, critics express concerns over using big data analytics predictively to create such individualised assessments. Debate over the use of COMPAS in the US specifically highlights design risks and uncertainties, and the negative consequences of (unintended) algorithmic bias in such high stakes decision making (Supreme Court of Wisconsin, 2016). COMPAS is a commercial tool used in the criminal-justice system that aids decisions about, among other things, parole. COMPAS scores, based on questionnaires completed by prisoners, are predictive of risk of reoffending, but a recent study in the US shows a strong correlation between COMPAS score and race (Larson et al., 2016).There is judicial recognition in the US that, at present, such tools should be no more than part of the material used in making a determination. Scholars have also stressed the importance of policymakers focusing on standards of 'fairness, accountability, and transparency' when deciding whether, and how to, deploy such tools (Kehl, Guo and Kessler, 2017).

It has been suggested that the use of AI in the legal sector could improve access to traditionally high cost legal services and thereby improve social equity. Applications in this area are already underway and include a chatbot that was developed by a Stanford University student to provide free legal advice. This service initially helped more than 160,000 people overturn parking tickets and has since been expanded to provide advice to individuals seeking asylum (World Wide Fund for Nature Australia, 2017).

Take-up and deployment of technology appears still to be slow and unevenly distributed within and across legal services markets. This is likely to reflect differences in market scale, patterns of both internal and external investment in the sector and restrictions on deployment flowing from legal services and regulation. There are no specific standards that regulate the use of AI in the Australian or New Zealand legal services market – beyond protections against the misuse of AI in the justice system, although the Lawyer and Conveyancers Act in New Zealand does specify that legal services must be delivered by a lawyer (interpreted as a person with a practicing certificate). Whether there should be standards is an important question, with significant ramifications for the development of the legal services market, and for access to justice (Bennet et al., 2018). These might include restriction on the use of automated legal information and advice; quality and competence of different automated advice systems; and transparency and explainability standards, which remain fundamental principles underpinning the justice system.

2.3.11 Defence, security and emergency response

AI will have implications for intelligence collection and analysis, logistics, cyberspace operations, command and control and emergency response. Internationally, it will have implications for military, information and economic superiority (Allen and Chan, 2017). These changes will require the development of skills to harness advances in AI, with training doctrine, recruitment and organisation structures having to adjust as a result (Nicholson, 2018).

AI could enhance the capacity of emergency services to react to, and prepare for, natural disasters and humanitarian emergencies and enable planners and responders to analyse population and physical environmental data, physical infrastructure schematics, risk assessments and response plans. This information could be merged with data from social media, and first responders' information, helping command and control personnel make effective decisions. By being continuously updated with new information, algorithms could provide a constant picture of changing needs and where resources should be prioritised during emergencies.

Deep cognitive imaging, an advanced form of pattern recognition, has been used in Australia to estimate the incidence of wildfires with respect to climate change (Dutta et al., 2013). The system was provided with a scenario based on Australia's climate between 2001-2010 and was able to replicate the real-world occurrence of fire hotspots with 90 percent accuracy. Further, CSIRO has also developed 'Spark', an AI powered system that has the capacity to predict the spread and location of bushfires and allow for better preparation, targeted deployment of resources, and to plan evacuation routes. Looking abroad, the Cincinnati Fire Department has developed an AI system that can classify the urgency of emergency calls and has effectively reduced delays in transporting patients to hospitals by 22 percent.

In Australia, drones have been employed at Australian beaches to detect sharks and other potential threats to swimmers. The Westpac Little Ripper Lifesaver and SharkSpotter uses ML techniques to analyse live video from a camera attached to the drone. It can identify sharks, issue alerts and can even conduct rescues by deploying a rescue pod. AI can also be used to better anticipate earthquakes through the use of neural networks and thus alert the public seconds to minutes before an event occurs (Kuma, 2018).

In the defence sector, robotic automation could augment or replace soldiers, freeing them from simple tasks and allowing them to focus on more cognitively complex work. Tasks undertaken by AI could be conducted faster, with greater precision and reliability, for durations that exceed human endurance and in dangerous environments, reducing the risk to soldiers in the field (Scharre and Horowitz, 2018).

A common public discussion about drones has been their use in support of military and covert actions for targeted killing and assassinations. Another significant proportion of drone activity is for surveillance and information gathering.

Box 10: Autonomous weapons

Autonomous weapons are AI systems that, once programmed and activated by a human, can identify, track and deliver lethal force without further intervention by a person. This weaponry includes those used in targeted and non-targeted killing, such as autonomous anti-aircraft weapons systems used against multiple attacking aircraft.

Autonomous weaponry that uses AI and ML can be categorised in the following way. These categorisations are quoted directly from a report prepared by Human Rights Watch (2012):

- Human-in-the-loop weapons: Robots that can select targets and deliver force only with a human command.
- Human-on-the-loop weapons: Robots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions.
- Human-out-of-the-loop weapons: Robots that are capable of selecting targets and delivering force without any human input or interaction.

Compared to countries that lead the world in the production of autonomous weapons, Australia's outputs are very limited, and so would be unlikely to have any significant impact on these technologies through technological leadership or innovation. However, even if Australia does not lead in the development of autonomous weapons, it is in our interest to ensure autonomous weapons are appropriately regulated, as we will have to deal with them in conflict zones. It is also in our interest to demonstrate ethical leadership in the use of new technologies like AI. Australia can have an impact through involvement in international dialogues and discussions to promote norms. For example, Australia has been active in United Nations discussions on the Convention on Certain Conventional Weapons. A parallel opportunity for influence at this international level is the United Nations Institute on Disarmament Research (UNIDR), one of the research arms of the UN that has been closely exploring and shaping debates about lethal autonomous weapons systems.

However, there are other applications, such as using AI in bomb or munitions disposal units (Singer, 2009) or for emergency response. The ethical, legal and social concerns with drone use vary depending on whether they are being used in defence, to gather information, or to support people.

Fully autonomous vehicles are already being deployed in the battlefield by states such as Israel (Gross, 2017). Militaries are working on capabilities to 'pair' older vehicles with newer ones, tasking them with conducting tasks to support manned systems (Hoadley and Lucas, 2018: 11). This could include carrying extra equipment and ammunition on the battlefield, reacting to electronic threats such as jamming, conducting reconnaissance, surveillance and removal of explosives. On-board sensors are being developed to alert users when repairs are required, allowing individually customised maintenance on an 'as needed' basis, lowering maintenance costs (Hoadley and Lucas, 2018: 9).

Box 11: Moral decision making?

AI cannot possess moral motivations, such as courage, moral innocence, moral responsibility, sympathy or justice, nor does it recognise moral ends. However, this does not that mean that AI cannot act in the interest of moral ends or principles. A robot can refrain from killing something because it is programmed not to kill things of that kind in the circumstances in question. If a robot is taught never to attack a vulnerable person (assuming the categories of vulnerability are readily identifiable for the robot), for example, it might be less likely to commit an immoral act than a soldier who has been subjected to the stresses of war and is presented with an opportunity to take revenge on a defenceless civilian (provided the robot understood and had a consistent framework for 'defenceless civilian'). The fact that the robot has no inherently moral motivation may not be critical, particularly if the moral motivation of the programmer is successfully imbued in the robot's decision-making system. However, moral decision making often requires the ability to infer the consequences of one's actions, which is something that narrow AI is particularly ill-equipped to do.

AI is likely to play an increasing role in decisions on military practices that are neither mission critical nor involved in the actual application or use of force. The command structures of militaries are likely to ensure that mission critical decisions and those relating to use of lethal force are likely to remain within the realm of humans.

The likely uses of AI in counter-terrorism, cyber warfare and network centric warfare include identifying 'abnormal' or 'antisocial behaviour' (Smith, 2014), facial recognition (Smith, B., 2018), moderation of illegal or offensive material online (Breland, 2018; Leetaru, 2018), recognition of foreign influence operations (Mueller, 2018), and, in the context of criminal law, sophisticated spear phishing (scam emails).

AI will be fundamental to harnessing and integrating ever-greater amounts of data across air, space, cyberspace, sea and land (Hoadley and Lucas, 2018: 11). This could transform command and control operations, enabling faster decisions and accelerating the speed of conflict (Hoadley and Lucas, 2018: 27). Additionally, identifying patterns across large datasets will allow improved image recognition, labelling and targeting. Better predictions of events such as terrorist attacks or civil unrest will also be possible (Scharre and Horowitz, 2018). Equally, there could be undesirable feedback loops ('flash wars').

A cybersecurity vulnerability for military and national security agencies is the security of their devices. This becomes increasingly important if the military use third party vendors to provide products from states whose strategic interests either do not align or are in direct conflict with Australia and New Zealand. Given that AI technologies used in this context are often quite opaque – technically and legally – they may present information security vulnerabilities. In a situation where an AI is used in an area of information security, it may be necessary for those involved in the procurement process to understand these opacities in order to determine that they cannot, and will not, lead to information security issues in the future.

2.3.12 Government

Many governments are employing AI technologies and systems for the purposes of, for example, managing access to, and delivery of, public services, health and aged-care, national security, employment, and making decisions based on legislation. It has many other applications, ranging from human resources, welfare, child support and services, assessing and providing advice on fines, homeland security, immigration and urban planning. The use of AI within the public sector therefore has the potential to deliver economic gains, increase productivity and efficiency, and deliver higher quality public service with the aim of increasing reliability, accuracy and accessibility (Capgemini Consulting, 2017). Given that the Australian government is custodian to large amounts of data, including aggregated and disaggregated personal data, there is great potential to adopt AI for various aspects of governance and public policy. New Zealand is also curating and using government-related data through its Integrated Data Infrastructure (IDI), which contains linked, deidentified data about people and households from government agencies, Stats NZ surveys, and non-government organisations, as well as its use of predictive risk modelling for policy development and implementation (Boston and Gill, 2018). However, care and consideration should be given to preparing people, organisations, functions and policy documents for this emerging landscape.

AI is being employed within government services for is its opportunities to alleviate administration processing. Administration processes are often tedious and repetitive and can delay governmental processes. This can be circumvented with the use of robotic process automation (RPA). RPA is a rule-based system, which is employed alongside ML, computer vision, speech recognition, and NLP to automate transactional, rules-based tasks by mimicking human interactions (Figure 5). RPA is most recognised in the form of a chatbot or virtual assistant and its incorporation into the workforce can provide employees with more time to perform complex decision-making tasks.

Globally, there have been several applications of RPA into government institutions to meet the increasing demands of paperwork processing and queries. Some examples include:

- DoNotPay Bot, a UK specific free app that assists people who have limited knowledge of the legal and welfare system with application filing. The app enables citizens to lodge applications and manage disputes over small legal matters such as parking fines, mail delivery as well as management of welfare, government housing, eviction and repossessions.
- The UK National Health Service has implemented the use of an AI assisted chatbot that can assess symptoms of urgent but non-life-threatening conditions of a patients to relieve pressure off emergency wards and helplines.

- The US Department of Homeland Security and Immigration use a computer-generated virtual assistant, Emma, that can to provide immediate responses to questions and direct users to where they may find more information regarding their matter.

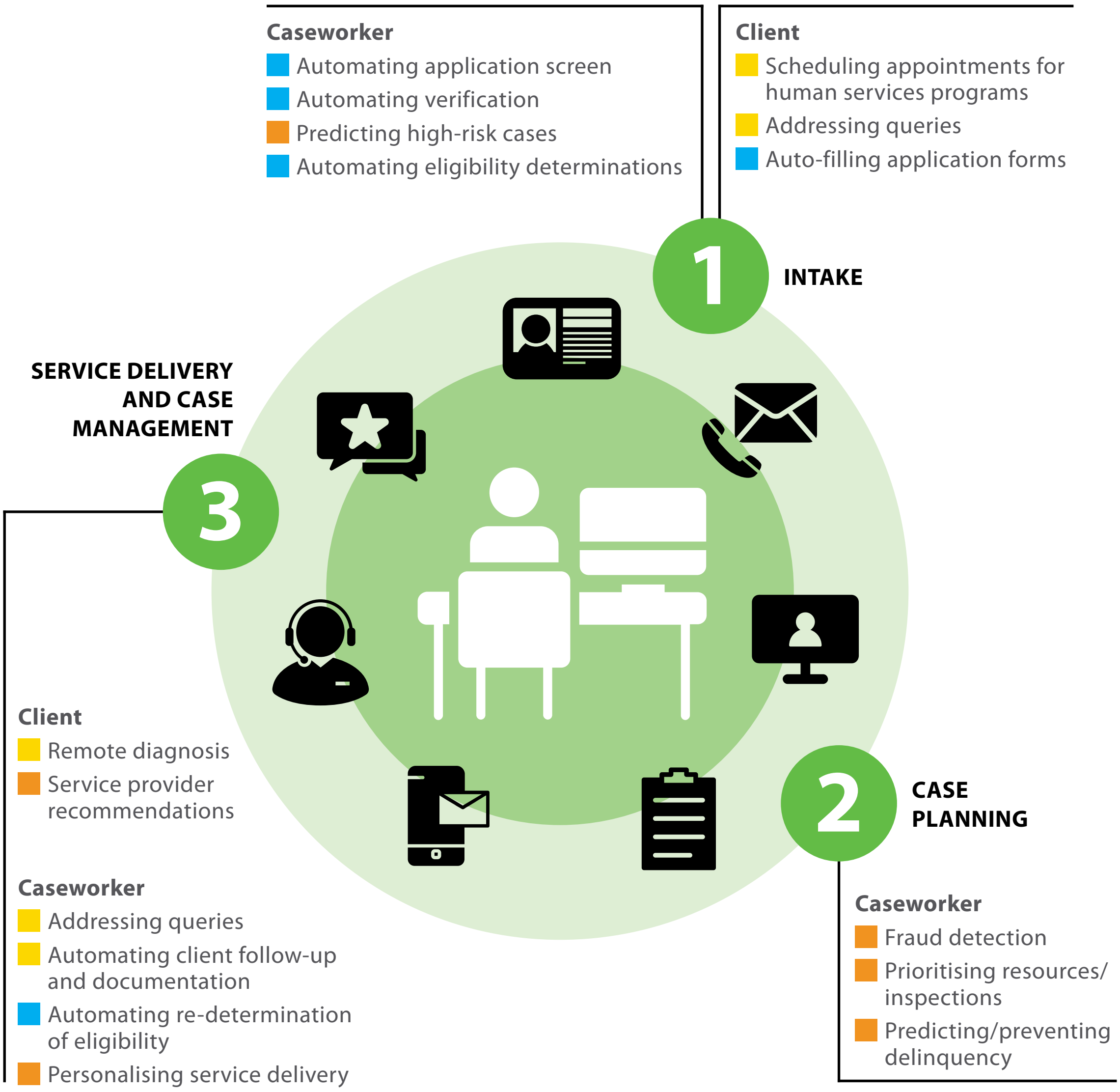
Australia has initiatives such as the Digital Transformation Strategy, which focuses on taking human services and business data to an online, central and accessible platform for both users. The strategy has three focus areas of development: 'Government that's easy to deal with; Government that's informed by citizens; and Government that's fit for the digital age' (Digital Transformation Agency, 2018). However, consistent improvement through the appropriate use of consumer or user data will be crucial to improving agility of AI tools in government applications. The Department of Human Services recently launched their virtual assistant, Roxy, to help process queries and minimise the time spent waiting for personnel to process consumer requests. The use of Roxy has successfully reduced workload and is able to respond to 78 percent of requests, however an expert is still required for more complex cases. Additionally, the taxation office uses a chatbot assistant named 'Alex' to assist with customer service. Alex exceeded industry benchmarks and achieved a first contact resolution rate of 80 percent (Capgemini Consulting, 2017).

AI can also be used for purposes such as fraud detection in massive datasets, easing congestion and optimising traffic management systems, optimising public spaces and generating public services that are transparent and accountable. However, these applications present both significant opportunities and challenges for governance. For example, facial and voice recognition technologies could be used to improve national security and delivery of public services – the Australian Taxation Office, for example, has already implemented a voice recognition system to authenticate callers – but these technologies are currently often inaccurate and may prevent people accessing essential services if deployed incorrectly. Likewise, the use of internet-enabled technologies and remote sensing devices for the collection of data may be useful in, for example, improving city planning, but may also compromise privacy. For the applications of AI to progress from assistive technology to an intelligent and integrative technology it will require continual development and refining of algorithms and policymakers will need to ensure that these (and other) AI developments comply with regulatory mechanisms and societal acceptance.

2.3.13 FinTech

Technological developments play an integrative role in the deployment of financial services. Intelligent financial service technologies (FinTech) are already being widely employed for a variety of tasks in financial services firms, including in Australia and New Zealand (Institute of International Finance, 2018). These algorithms and techniques have the potential to expand access to credit, better manage risk, reduce fraud, improve firms' compliance with laws and codes of conduct, influence the speed and correction of recovery in trading, and significantly expand industry revenues in the financial services sector.

FinTech may significantly disrupt the banking sector (PwC, 2017). The International Data Corporation predicts that worldwide revenues from the adoption of such cognitive systems across multiple industries will experience an increase from US$8 billion in 2016 to over US$47 billion in 2020 (International Data Corporation, 2016). Furthermore, by 2030,



Caseworker

- 🟦 Automating application screen
- 🟦 Automating verification
- 🟧 Predicting high-risk cases
- 🟦 Automating eligibility determinations

Client

- 🟨 Scheduling appointments for human services programs
- 🟨 Addressing queries
- 🟦 Auto-filling application forms

Caseworker

- 🟧 Fraud detection
- 🟧 Prioritising resources/ inspections
- 🟧 Predicting/preventing delinquency

Client

- 🟨 Remote diagnosis
- 🟧 Service provider recommendations

Caseworker

- 🟨 Addressing queries
- 🟨 Automating client follow-up and documentation
- 🟦 Automating re-determination of eligibility
- 🟧 Personalising service delivery

TECHNOLOGIES USED 🟨 Chatbot 🟦 RPA 🟧 Machine learning

Client problems alleviated

Tedious application process

Long wait times

Language issues for non-English speaking people

One-size fits all solutions

Caseworker problems alleviated

Following myriad program rules

Heavy administration burden

Not knowing which cases to prioritise

Figure 5: AI in government services

Adapted from: Deloitte, 2017.

Box 12: Case study: Adoption of AI within New Zealand banks

In New Zealand, AI adoption for improving customer experience is being developed by a number of banks:

- ANZ New Zealand launched 'Jamie', a digital banking assistant designed to interact via video or text to answer 30 of the bank's most frequently asked 'help' questions.
- New Zealand bank, ASB, announced a digital assistant named 'Josie', who helps people in the early stage of setting up a business. Josie is based at ASB's premises in Auckland and is available by appointment. ASB has also established AI-powered 'connected customer conversations', a multi-channel automated marketing program that aims to deliver timely and targeted customer conversations at scale.
- Westpac has released 'Wes', a text only chatbot accessible via its website.
- BNZ has created two chatbots – one for their internal helpdesk, and another built in Microsoft Azure, which is being trialled for KiwiSaver customers.

For the efficiency and accuracy of core business including risk management, BNZ has partnered with Intel, using the Saffron Anti-Money Laundering Advisor; Westpac has adopted ACI's Up Payments Risk Management Solution, which uses adaptive ML; and ANZ uses voice biometrics, powered by AI, to identify customers using the characteristics of their speech to improve security on mobile devices.

New ways of conducting lending and payments have also been created. Harmoney, a New Zealand FinTech that facilitates digital peer-to-peer lending, has created its own digital marketplace of 15,000 members while using AI to increase the accuracy of credit risk predictive models and to accelerate deployment of predictive models (CIO New Zealand, 2018). ANZ New Zealand and BNZ have addressed potential disruption in payments by forming enabling partnerships – ANZ New Zealand partnered with Apple Pay in 2016, and BNZ with Alipay in 2018.

global GDP could increase by 14 percent, or US$15.7 trillion because of AI with US$1.2 trillion extra economic growth forecasted GDP gains in Oceania (PwC, 2017). In addition, over US$1 trillion of today's financial services cost structure could be replaced by ML and AI by 2030 according to the 2018 Augmented Finance and Machine Intelligence report. Accenture estimates that AI will add US$1.2 trillion in value to the financial industry by 2035 (Purdy and Dougherty, 2017).

There are many opportunities for banks to explore these emerging technologies while rethinking corporate strategies, evaluating potential partnerships and paving the way towards a genuine transformation of the industry itself (Manning, 2018). FinTech may help banks improve customer experience by, for example, providing personalised customer interaction and advisory services through chatbots. The four major Australian banks are in the process of adopting AI tools in line with worldwide developments within the sector.

The Commonwealth Bank of Australia's 'Ceba' chatbot is able to assist customers with more than 200 banking tasks, including card activation, checking account balance, making payments and obtaining cardless cash. The National Bank of Australia has introduced the Digital Virtual Banker, which is able to answer approximately 200 customer service questions by drawing on data from countless customer service interactions. The ANZ Banking Group has created biometric voice capability with technology company Nuance to allow customers to bank by talking to the app (Eyers, 2018). In the same context, the Westpac Banking Corp is using AI to conduct data analytics and visualisation and provide personalised advice in managing financial matters.

These developments may lead to customers feeling more empowered to make choices that were previously accessible only to wealthy people who could afford experts such as financial advisors, researchers, coaches and consultants. FinTech can also improve the core efficiency, productivity, and accuracy of a bank, through AI systems that manage risk, security, transparency and accountability.

However, despite these opportunities, financial institutions are faced with numerous challenges in taking advantage of the benefits of AI in a timely manner (Capgemini, 2018). Such challenges include budgetary, regulatory, data quality and resource limitations in AI implementation; a lack of literacy and confidence with the technology among some consumers; as well as the potential for FinTech to exacerbate biases in areas of banking and insurance. AI systems deployed in these contexts are subject to cyber-crime and hacking and will require adequate protection. Policies that enable the public to understand clearly when AI is being used and whose interests AI is representing will be important to retaining the free market principles on which our economy has been successfully based.

2.3.13.1 Australia's response to the global financial crisis

The global financial crisis (GFC) revealed regulatory weaknesses within the banking sector. In response, risk models have come under greater scrutiny and regulation. The detailed standards and guidelines can be adapted and enhanced to accommodate AI techniques and ML algorithms, primarily in those parts of the business for which financial stability considerations are of high importance. A key lesson from the GFC is that business leaders must understand how the models at the heart of their businesses are designed, implemented, validated and used, and the limitations of those models, including their main assumptions and the nature of their reliance on historical data. It is therefore essential that business leaders take responsibility for the outcomes, decisions and actions that are created by, or a consequence of, the use of the models in the business, and to consider the use of AI techniques and ML algorithms in these models.

Issues such as these are of significant interest overseas where because of their exposure to the GFC, banks and regulators have invested deeply in risk management and developed strong regulatory requirements. For instance, the Federal Reserve's "Guidance on Model Risk Management" (SR Letter 11-7) recommends embedding critical analysis throughout the life cycle of an algorithm – from outlining assumptions in the underlying model through to the data used to train the algorithm (US Federal Reserve, 2011).

2.3.13.2 Resources and actions required to realise the potential of AI in the financial services sector

The emergence of FinTech has encouraged the experimentation and gradual adoption of numerous AI applications within the financial services industry, particularly in the areas of capital market, consumer banking, insurance and portfolio management. Some applications have already created a solid footprint; however, numerous areas remain undeveloped. Australia and New Zealand will need to continue to develop, or have access to, the technology and skills necessary to benefit from developments in FinTech.

Opportunities are likely to emerge for Australian companies and consumers to be at the forefront of FinTech, including:

- Being the international standard in the way AI-powered financial services are provided to the individual and society. This standard could see the ability for people to effectively allocate capital based on their individual circumstances and life goals. To achieve this goal, people will need to have access to the technology and confidence to use it
- Being an exporter of FinTech AI capabilities through local entrepreneurs using the technology to solve business problems in a way that creates international leadership in AI-powered FinTech
- Creating an engaged community with good domestic job opportunities and standard of living through an appropriate social and legal framework combined with well-developed education and retraining industries
- Building a prosperous and competitive economy through a world class financial services system that is able to mobilise and empower people and businesses to build and manage their wealth.

These benefits come with risks, including:

- The potential that advice and decisions issued through the 'black box' are not in the clients' interest or discriminate against customer groups. However, AI can be trained to verify if regulatory requirements are being met
- Procurement of AI and ML as well as deep customer related data become the main competitive factors in financial services. This could see the continued expansion of big technology companies such as Google, Facebook, Amazon and WeChat into financial services, potentially disrupting current institutions with implications for the whole sector
- A gap in workforce talent who possess the appropriate skills and experience to effectively develop, implement or work with AI systems

- A population that becomes distrustful of the technology as it misleads or breaches the trust of those who use it
- Misuse of AI systems to undertake fraudulent activities
- The risk that FinTech systems could be hacked or compromised.

Financial services regulators will need to develop a robust set of regulatory standards and detailed associated supervisory guidelines for model risk management and governance for banks and other financial services firms. These standards should be clearly and prominently articulated as part of a broader, coordinated, national strategy and approach for AI across all sectors of the economy. However, because of the unique and central role that banks play in our economy – and the very large risks that their failure may pose to the national economy and society – some very high quality, sector-specific controls are needed to ensure the safe and effective development and implementation of AI approaches in the financial services sector. Regulators may benefit from AI to implement oversight, and could make use of compliance AI bots to ensure that banking is conducted in accordance with established regulations.

2.3.14 SMEs and start-ups

Small-to-medium sized enterprise (SMEs) and start-ups will be a part of the shift towards an AI-enabled society, and it will be important to consider how they can benefit from, and contribute to, AI development.

Data and computers are critical resources to enable AI, yet often these resources are unevenly distributed. This may suggest that large corporations are destined to be in a more advantageous position when it comes to leveraging AI to grow their businesses. However, there are also benefits for SMEs and start-ups who can often adapt faster to technological change as they are usually not burdened by legacy IT systems or complex business structures. Further, AI has the potential to automate some human tasks that are expensive for small companies to maintain. This provides opportunities to free up time, money and human resources, making it possible for such businesses to more effectively compete with larger well-established organisations.

Smaller businesses can also take advantage of inexpensive application programming interfaces (APIs) and tools. Companies such as Google, Amazon, and Microsoft have developed APIs supporting the incorporation of AI functionalities, such as natural language and text processing, speech processing, image processing and computer vision. Some of these are open source, while others are inexpensive. These tools are enabling smaller companies to create novel AI applications without the need to hire software engineers.

2.3.14.1 'Off-the-shelf' AIs

There are numerous 'off-the-shelf' AI products that can add significant value to any business, and to SMEs and start-ups in particular. For example, SMEs may benefit from off-the-shelf tools that automate human resource management. Companies such as Tangowork provide chatbots with human resource functionalities, allowing employees to use natural language to ask questions or make requests about human-resource-related matters, such as 'I'd like to apply for leave from tomorrow until next Friday'. Over time the chatbot learns the patterns of interaction of each employee and effectively becomes a personal assistant, anticipating requests and providing notifications accordingly.

AI can also be used to coach salespeople to refine their conversational skills during a call to improve sales performance. For example, solutions such as 'Gong' record and transcribe calls, then correlate sales success with features of the call, such as choices of expressions, ratio of time talking versus listening and call duration.

Further, some AI tools can help with product-market fit, which is a major challenge for start-up companies. Natural language processing can be used to create sentiment analysis tools such as Keatext, compressing and interpreting vast amounts of textual data. This allows start-ups to screen different market niches for similarities with potential product offerings.

2.3.14.2 The impact of AI on SMEs and start-ups over the next 10 years

Leading nations are investing heavily to support AI in general and the AI business ecosystem in particular. The UK Government worked with over 50 technology companies to develop an AI sector deal worth over £1 billion, articulated in a policy paper released in April 2018 (UK Government, 2018). The French Government announced an investment of €1.5 billion in AI until 2022 (Rabesandratana, 2018a). In comparison, Australia's 2018-19 federal budget allocated approximately A$30 million to AI, a modest amount by comparison even when accounting for the differences in GDP and investment time horizons.

To support the growth of SME and start-up AI in Australia, there are several factors that will be key including incentives to support AI development, growing the AI talent pool and connecting entrepreneurs to AI talent.

Incentives for AI development – There are several mechanisms that can be used to support development of new technologies, such as providing both direct and indirect incentives for the private sector to invest in local AI developments and incentives for researchers to collaborate with industry (and vice versa).

Growing the AI talent pool **–** Without AI scientists and engineers, entrepreneurs cannot materialise their vision. Developing a strong AI skill base in Australia will be important as will retaining the existing talent in the country and attracting overseas talent. This could be done through several independent initiatives such as AI-specific postgraduate scholarships and programs. Additionally, Government programs such as the Australian Global Talent Scheme Pilot provide opportunities to attract skilled talent.

Connect entrepreneurs to AI talent – A wide range of mechanisms could help bring together entrepreneurs and AI talent, with the aim of spreading ideas, brainstorming solutions and seeding new SME and start-up teams. A range of activities, such as innovation precincts, AI events, technology meetups, and entrepreneurs-in-residence could provide opportunities for connections, communication and collaboration. For example, a recent initiative, the Victorian Innovation Hub, aims to connect entrepreneurs and start-up with mentors, investors and funding bodies such as LaunchVic that provide seed funding to support start-ups.

Given appropriate investment in research and development, by calling upon homegrown expertise and by attracting world-quality talent, we can play an important role in guiding the international development of AI.

2.4 Realising the potential

This chapter has highlighted some of the existing applications and emerging opportunities for AI across various sectors across the economy. To ensure these applications are effectively applied and to realise the economic and social potential associated with AI technology, a proactive approach in investment, leadership and coordination will be necessary. Government, civil society, and industry all have a role in establishing the future direction and adoption of AI technologies. The successful examples discussed demonstrate ways by which AI can be developed and harnessed for new purposes.

Whether the opportunities presented by AI are achieved is likely to depend on how both government and the sectors themselves address several fundamental risks and challenges. Some of these challenges include the following:

- technological unemployment and de-skilling and re-skilling
- a 'digital divide' of growing significance
- innovation risks
- business risks
- ethical risks
- unintended consequences of both the technology and its regulation
- vulnerability to cyber-attack
- misleading or biased AI
- expanding infrastructure requirements.

Managing the transition to AI-enabled sectors will require building on existing capabilities by promoting educational, interdisciplinary and inter-regional collaboration on AI development and literacy; establishing guidelines and advice for procurement, especially for public sector and small and medium enterprises; and reviewing regulatory mechanisms to address potential issues arising from the implementation of AI while also supporting local innovation and deployment. Australia's Industry Growth Centres could play a role to help drive AI innovation, productivity and competitiveness across different sectors.

Development of AI standards and guidelines that follow best practices in other jurisdictions will assist with the successful implementation of AI (data governance will be further discussed in Chapter 6). These regulations should apply to both human-created and machine-created models to ensure that current and future AI techniques are used appropriately and responsibly. Support through education, advice and community and government consultation will be important for enabling innovators to understand and comply with regulation.

AI specialists often do not have sector or industry specific knowledge in areas such as agriculture, energy, health and mining. Likewise, those in specific sectors do not necessarily have the technical knowledge to apply AI to their area. Education and training programs will help to develop, implement, work with, and harness emerging AI systems across sectors. Emerging university programs are seeking to address this gap by offering technology-related subjects. For example, the University of Technology, Sydney, now offers a major in 'legal futures and technology', and Melbourne Law School offers a small suite of legal technology subjects – law apps, new technology law and start-up law – as well as a legal research stream in this area.

Advanced economies have invested in research and development to take advantage of new waves of AI and automation. The technical skills and knowledge obtained gives these countries an edge over their international competitors. The more technical capabilities these countries can create, the more they are capable of putting together complex ideas and technologies to create higher value and complex goods (Hausmann and Hidalgo, 2010). An AI capability is essential to leverage current investments, maintain our high quality of life, and create an AI-enabled economy – a compelling ecosystem of high-tech businesses and highly productive workers in both private and public sectors.

None of this will be achieved without the appropriate infrastructure. AI is enabled by access to data and digital infrastructure that are secure, fast and accessible. However, the physical infrastructure to fully support AI is lacking. As internet and smart device use increases there will be an increase in the volumes of data being transmitted. This will require high quality connectivity to support the adoption of AI and associated digital technologies. Effective digital structures that help diffuse AI equitably – especially with ageing populations, people with disabilities and those living in remote and rural communities – will be required to ensure everyone can benefit. New or updated infrastructure may be required for the adoption of fully autonomous vehicles or the shift toward smart cities. In addition, soft infrastructure requirements for the adoption of AI include education, workforce and regulatory provisions. These requirements will be addressed in Chapter 3 and Chapter 5 respectively. Digital infrastructure requirements, such as data storage and cloud computing, should also be considered. Given the likely transnational nature of AI technologies and trade, global developments in these areas should be monitored. It will be important to keep pace with the global infrastructure requirements required for AI technologies.

In rural and regional areas, access to AI technologies supported by appropriate digital infrastructure could transform many sectors of the economy or alleviate social inequality by providing, for example, better access to healthcare, connectivity to social support services, education and employment opportunities. AI technologies are also poised to play a transformative role in agriculture. However, AI technologies cannot be successfully implemented in rural areas without the necessary infrastructure connectivity to support them. For example, an autonomous machine could work on crop management throughout the night, but to do this requires communication via satellite broadband or proprietary wireless networks, which are lacking in many areas of Australia. There will be a need to establish greater communication infrastructure in rural areas to speed the development and trialling of AI, with expansion to other areas in a staged manner.

To remain competitive, the deployment of AI will need to keep pace with international developments in telecommunications networks, capacity for data storage, cloud computing, computer infrastructure data at scale, and fast and secure connectivity. There will be a need to develop capacity to leverage data produced by IoT technologies and components to respond to growing and complex infrastructure demands. This will depend on a national implementation of up-to-date broadband and mobile connectivity infrastructure, fibre-optic backbone networks and data centres capable of storing and processing significant amounts of sensory data. These data could also be used to better manage urban populations and city planning, as evidenced by the increasing demand for smart city infrastructure projects and the use of 'open data' by city planners, entrepreneurs, businesses and citizens.

Recent government investment has provided support to certain sectors to advance data and computing capabilities. Examples of Australian initiatives to develop infrastructure to support AI include:

- The establishment of the Digital Transformation Agency to improve and increase the delivery of online government services
- Increased accessibility of open access data from the public sector in conjunction with fostered crowd innovation designed to use this data in a meaningful way through hackathons
- Investment for new supercomputers at the Pawsey Supercomputing Centre
- Investment into the Australian Digital Health Agency and rollout of the My Health Record; a platform for access, storage and integration of diverse data systems including genomic data.

These investments in certain areas of research and industry are a welcome addition to supporting AI developments and infrastructure, however further expansion will be needed across all industries.

AI techniques and technologies can present significant opportunities for development and can be used to achieve a more robust digital infrastructure. Intelligent robots can be deployed for infrastructure, such as using autonomous aerial vehicles to build or inspect (via machine vision) complex, precarious, or high-standing structures; below-ground infrastructure, such as using AI-enabled digging equipment to provide a legible map using radars, sensors, and sonars; and underwater infrastructure, such as using unmanned underwater vehicles to carry out inspections of cabling (Australian Centre for Robotic Vision, 2018: 132-3). AI developers will need to ensure these intelligent infrastructure systems are safe, able to accurately perceive their environments, have the necessary dexterity to perform tasks in and around complex structures, and can cooperate effectively and efficiently with human and non-human collaborators.

CHAPTER 3
SOCIETY AND THE INDIVIDUAL

This chapter is based on input papers prepared by the generous contributions of Professor Sharon Parker (AI and Work Design); Dr Ross Boyd (Employment and the Workforce); Professor Robert Holton (Employment); Alexander Lynch of behalf of Google Australia (Employment and the Workforce); Professor Greg Marston and Dr Juan Zhang (Economic and Social Inequality); Professor Rose Luckin (Education, Skills and Training); Professor Mark Reynolds (Training the Next Generation of AI Researchers); Professor Mike Innes (Psychological and Counselling Services); Professor Rafael Calvo, Dorian Peters and Professor Richard Ryan (Human Autonomy in AI Systems); Dr Eric Hsu and Dr Louis Everuss (Transformations of Identity); Professor Anthony Elliot (Transformations of Identity); Dr Oisín Deery and Katherine Bailey (Ethics). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

3.1 Introduction

The impact of automation and robots on society – particularly on employment – is fiercely debated. Some argue that the widespread adoption of AI technologies in workplaces will lead to massive job loss, disruption and demand for new skills (Ford, 2016; Turner, 2018), while others suggest that AI's impact on employment has been overstated, distracting us from other, more profound economic and social changes. Some credit the AI revolution with producing a world of comprehensive change.

There is much uncertainty about AI's impact on the future of work, society and the individual. The impact of AI and automation on the Australian workforce needs to be understood in a global context and this uncertainty is reflected in recent international reports on AI. For example, widely cited findings on employment in the USA by Frey and Osborne (2013, 2017), replicated by Haldane in the UK (Haldane, 2015) and Durrant-Whyte *et al.* in Australia (CEDA, 2015), suggest between 40-50 percent of jobs are vulnerable to replacement by new technology. Likewise, a 2016 report from the World Economic Forum estimates the net loss of over 5 million jobs across 15 developed countries by 2020 (World Economic Forum, 2016). A report, published by the International Labor Organization, predicts that over 137 million workers in the Philippines, Thailand, Vietnam, Indonesia and Cambodia are likely to be replaced by robots in the near future. However, a more recent report has suggested far fewer jobs vulnerable to AI replacement, claiming that approximately one-fifth of workers are in occupations that are likely to

shrink, with the authors stating that the figure is much lower than suggested by recent studies of automation (Bakhshi et al., 2017).

The research presented in this chapter shows that it is important to acknowledge the *potential* for AI to generate widespread economic and social change. It is important to consider the underlying social and economic forces that generate uncertainty about AI's impact on the future of work, and to acknowledge that this uncertainty may indirectly shape the way government and industry respond to the uptake of AI technologies.

Education should be considered in the context of an AI-enabled society – not only in terms of ensuring learners are equipped with the proper skills to develop AI systems and technologies, but also that AI techniques are effectively deployed in education. More broadly, people will need to be provided with the tools and support to have sufficient information to make informed decisions about how and when they interact with AI technologies in their lives.

3.1.1 AI and the future of work: An overview of key issues

The 2014 Australian Industry Report estimates that up to half a million people employed – many of them tertiary-educated – are at risk of their jobs being automated. However, the report notes that while innovation will inevitably lead to some job displacement in the short term, there is a lack of evidence to suggest this displacement is long term (Australian Government, 2014b). A 2015 CEDA report predicts nearly 40 percent of existing jobs are at risk in the next 15 years (CEDA, 2015). However, it is worth considering the historical and methodological factors that shape our thinking, as well as clarifying the short and medium-term effects of AI and reviewing other economic developments that may play a role in reshaping the future of work. A recent New Zealand report estimated that sectors that have a large labour force and high use of technology were most likely to benefit from AI, while sectors like agriculture, with relatively small labour pools and relatively low technology penetration, would expect less direct benefit from AI created labour efficiencies (The AI Forum of New Zealand, 2018).

Technological innovation from the Industrial Revolution onward suggests that major technological shifts create new forms of employment while simultaneously undermining others, often over prolonged periods. For example, while the introduction of computers in the latter half of the 20th century undermined much routine manual and clerical employment, computers helped to stimulate more complex cognitive, interpretive and abstract work (Borland and Coelli, 2017: 379). With respect to AI, there is the suggestion that machine-learning may also undermine many different forms of complex employment, a claim explored in later sections of this chapter. Further, it has been suggested that advancing AI technology might, for the first time in history, eliminate jobs faster than it can create new ones (Colvin, 2015).

Yet, AI technologies currently tend to affect tasks rather than whole occupations (this is true for narrow and broad AI but may change with general AI). Any given work role consists of several interconnected tasks. Machines are programmed to perform discrete tasks. The tasks that are easiest to codify are most likely to be automated. Where a *series of tasks* are involved, the occupation is far less likely to be completely automated. Humans may still be required to perform certain occupations (or oversee an AI performing certain tasks), even where a proportion of the tasks previously associated with them are automated. Moreover, changing the tasks involved in a job is likely to have significant effects on its value and desirability. Even so, some commentators argue that few if any, existing occupations will remain untouched by AI (Ford, 2016).

The methods and procedures used to measure the impact of AI and related digital technologies on employment are often based on subjective grounds. As such, there is a risk of exaggerating or inflating the scale of job vulnerability and loss of employment. Reports are often premised on subjective observations regarding the degree of routine, manual skill, social intelligence or creativity involved in any given occupation (see for example, Borland and Coelli, 2017). Yet, while there is undoubtedly a danger associated with overstating the impact of AI on the future of work, it is equally important not to simply assume that everything will be the same as it was before the introduction of AI.

One way of reconciling these conflicts is to distinguish between short-term and medium-term effects of AI. As discussed in the introduction to this report, short, medium and long term in this context is loosely considered to be within 5 years, approximately 10 to 15 years, and greater than 20 years, respectively. The short-term effects of AI will be associated with systems and technologies that can produce repetitive and predictable rule-based outputs. The tasks most under threat here include routine manual and cognitive work. Yet, while the proportion of jobs of this kind in Australia and New Zealand has fallen, this fall may be due to factors like globalisation as much as automation. In the medium-term, when advances in AI and ML render it possible for technologies to learn by actively interpreting and responding to data, higher-skill employment will also be at risk. Klaus Schwab (2017), founder of the World Economic Forum, has argued that the fourth industrial revolution is 'unlike anything humankind has experienced before'. According to Schwab, there will be multiple long-term impacts of the AI revolution on the economy, business, regions and cities, as well as geopolitics and global order.

Even if there remain enough jobs to retain full employment, the nature and frequency of occupations will change in an AI-enabled society, meaning that many people will have to transition between jobs during their working lives. There is therefore a danger that changes in the frequency of various types of work might effectively force significant numbers of workers into low value, low paid work, thereby exacerbating inequality (Turner, 2018).

An additional connection between AI and employment trends involves the platform economy and the growth of precarious casual employment in the 'gig economy'. Platforms are digital infrastructure that enable users to create, interact, and transact in diverse ways. For example, consumer goods platforms such as eBay, Amazon and Alibaba bring together buyers and sellers. Advertising platforms such as Facebook and Google aim to generate and extract data, which can then be packaged and sold to advertisers to match users to potential sellers. Other platforms such as Github, Job Rooster and Wannalo offer software tools for applications such as human resources. Such platforms typically transform types of employment that provide a space to mediate buyers and sellers. Many of these platforms make use of, or provide users with, AI APIs to, for example, translate or interpret large amounts of written text.

A key consideration is whether the growth of platforms undermines secure work and other employee benefits. Using mainly US evidence, Kenney and Zysman (2016) argue that such undermining may eventuate where the private governance structures of the platform economy escape public regulation. However, employment conditions vary depending on the platform and the kind of labour facilitated by the platform. Those directly employed by companies such as Google or Facebook, for example, generally retain traditional employment conditions. Those working in under-regulated areas such as taxi driving through Uber and Lyft, or those competing for episodic contracts to produce apps, are in a far less secure position. However, aspects of professional labour are also at considerable risk. Richard and Daniel Susskind (2015) show that new AI technology is reordering the professions and suggest that contemporary patterns of technological innovation are enabling intelligent machines and para-professionals to assume many traditional tasks once performed only by professionals.

Protection of worker rights will need to be considered as part of this workplace transformation and may involve consideration of civil society involvement, such as unions, throughout the process. How the transformation to automation is handled by management in relation to employees will be a key ingredient of a successful transition. Responses to digitally enabled employment changes have started to emerge in certain areas, including for-hire drivers working for Uber and Lyft (Fisher, 2017).

There is much uncertainty in knowing quite how AI will affect employment. There is also a broader question of how AI technologies can provide scope to revalue and reshape ideas on a meaningful recreational life. Australia and New Zealand will nonetheless need to prepare for the potential of widespread economic and social change. The most pressing need will be to focus on the types of employment most vulnerable to change in order to help facilitate retraining and income support. It may also be necessary to consider whether current work regulations need to be modified to anticipate and respond to challenges in the restructuring of employment tasks. These changes may also require a new workforce to create, maintain and monitor AI systems, as well as techniques for data curation and management. There may be a role for industry, governments and professional bodies to assist in this transition and collaborate to address potential issues associated with reskilling and upskilling of individuals as well as analyse the skills required for the future workforce. To this end, New Zealand has created a Future of Work Tripartite Forum and a New Zealand Digital Skills Forum (Digital Skills Forum, 2018; Robertson, 2018).

3.2 Employment and the workforce

3.2.1 Automation and the workforce

Contemporary debates about the transformative impacts of AI are often based on an arguably limited conception of autonomy, in which technologies are seen to be in conflict or harmony with human autonomy. This conception of autonomy informs not only the public response to AI, but also researchers in the field and those who design public policies and implement corporate strategies involving AI (Natale and Ballatore, 2017).

Recently, understanding of autonomy in relation to AI and robotics has shifted away from that of an independent agency towards a new model in which agency is understood as an emergent relationship. Ekbia, Nardi and Sabanovic (2015), for example, differentiate between automated and heteromated systems. Put simply, where automated systems are designed to shift some or many tasks performed by humans to machines, heteromated systems (e.g. Upwork, InnoCentive, Freelancer, Amazon's Mechanical Turk and other microwork or crowdsourcing applications) are designed to work by incorporating end users as indispensable system mediators.

By conceptualising autonomy as an emergent relationship, we can reframe the AI debate from one where humans are being displaced by robots to one where humans might play a more active role in moderating AI systems in their lives and work. Boyd and Holton (2017) stress the importance of critically evaluating the range of discourses about technological change because such discourses can constitute and direct or redirect change. Shifting the discursive frame of debate in this way enables AI and ML to be more closely aligned with ongoing processes of social learning and literacy regarding the adoption of new technologies (Stilgoe, 2018).

Recent economic modelling on the Queensland workforce suggests, based on conservative growth, that 250,000 more jobs will be created with a A$37.4 billion boost to the gross state product from the robotic and automation economy. Further, growth of the robotic and automation economy is predicted to generate three work categories: people who work for machines; people who work with machines; and people who work on machines (Synergies Economic Consulting, 2018).

Should the adoption of AI lead to a decrease in required working hours, this too could lead to benefits for both employers and employees. The four day work week has previously attracted attention with benefits including increased productivity, reduced worker stress, reduced strain on transport systems, a more equitable domestic division of labour and the potential to redistribute income across the economy (Jones, 2017). The four day work week has been associated with an emerging trend within the Netherlands, Germany and has been trialled in New Zealand.

Uncertainty remains as to the ultimate impact of AI on the workforce and, undoubtedly, many of the applications and associated effects of AI cannot be adequately foreseen at present. Should AI eventuate in workplace disruption that negatively impacts on certain populations, it will be necessary to investigate the ways in which these impacts may be ameliorated.

3.2.2 Productivity and changing employment

It is important to consider the implications of AI for ensuring employee satisfaction, autonomy and productivity in the workplace. As mentioned in 3.1.1, while the eradication of jobs and the associated need for people to reskill are important issues to consider, it is usually tasks that are automated, rather than whole occupations (Chui, Manyika and Miremadi, 2016). Tasks exist within a broader occupation or role, alongside many other duties that might escape automation. To take one example, precision medicine may be transformed by the application of ML to genomics, clinical imaging, and radiotherapy (Mesko, 2017). But this does not replace physicians, surgeons, medical scientists and researchers who are tasked with decisions involving interpretation, therapeutic intervention and professional responsibility.

This raises critical questions about how automated tasks fit within wider work roles, and within the whole system of work. For example, how might tasks be effectively shared between humans and machines? How do human workers interact with the technology and shape it to achieve their goals? How can people and machines best coordinate their activities, or work as a team, to achieve the overall goals and objectives of the workplace? These questions need to be considered alongside the wider implications of digital technologies that are transforming business models, where and when people work, the costs of production and many other aspects of work (Cascio and Montealegre, 2016).

A key concept through which to consider the impact of AI on employment is that of work design. Work design is concerned with the physical, cognitive, biomechanical, and social aspects of tasks involved in any work role (Parker, 2014; Safe Work Australia, 2015). Positive aspects of work design include providing employees with autonomy over work timing, methods, and decisions; a variety of tasks; the opportunity to make a difference or have an impact; job feedback; the chance for social contact and support; and moderate or reasonable levels of job demands (e.g. work load, emotional demands, and time pressure).

While work design research has long advocated for the need to consider human and technological issues together (see for example Clegg, 2000), scholars have called for renewed focus on how new technologies effect, and are effected by, work design (see for example Parker, Van den Broeck and Holman, 2017). For example, there is often a focus on replacing or automating human labour with new technologies, with 'leftover' tasks being allocated to people. Such an approach can result in poor work designs, with negative consequences for employee health, wellbeing, safety and productivity (Grote and Kochan, 2018). Rather than focusing on replacing human labour with AI and other automated technologies, it is important to consider how work systems operate as a whole, including the various tasks, responsibilities, and relationships that might elude automation. This means considering not only how existing skillsets need to change to fit new technologies, but also how new technologies can be designed, implemented, and managed to fit workers and organisational systems.

In preparing for an AI-enabled future of work, employers may need to consider which tasks and decisions should and should not be carried out by AI, as well as how to ensure optimum use of new technologies for existing skillsets.

Box 13: Australia's future workforce

In 2016, drawing on economic statistics from Australia, Google Australia commissioned AlphaBeta to provide an empirical view of the current state of automation in Australia and its effect on the workforce. The industry report suggests that from 2000-2015 the average Australian worker experienced two hours of automation for routine and predictable tasks, both physical and intellectual, across their working week.

The report estimates that automation is set to increase and by 2030 another two hours will be automated each working week (Figure 6). According to the report, this might allow workers to spend their time on higher-value activities (Figure 7) with an estimated boost to the Australian economy of A$1.2 trillion over 2015-30 (AlphaBeta, 2017). Tasks that have proved more resilient to automation include interpersonal interaction, decision making,

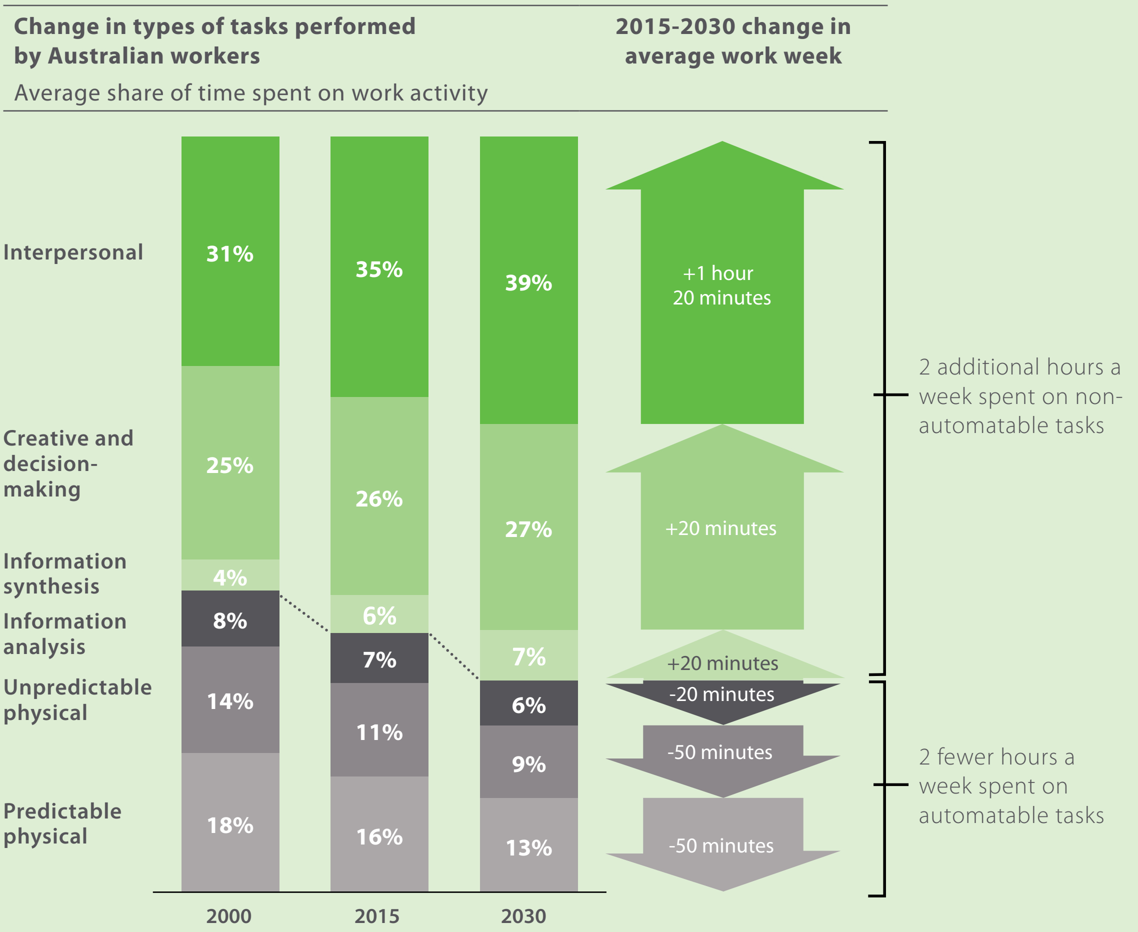


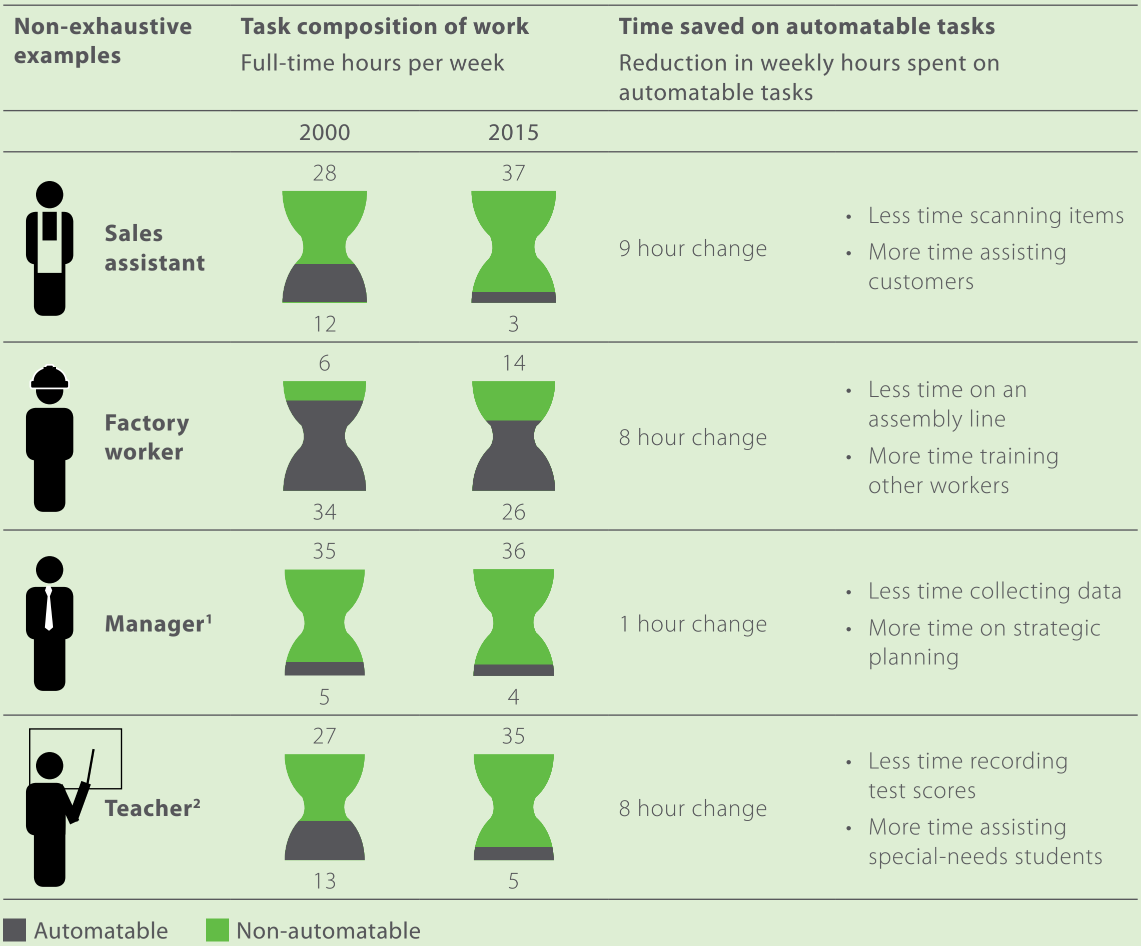
Figure 6: The effect of automation on work

Adapted from: AlphaBeta, 2017.

creativity, and synthesis of information from multiple sources and a degree of qualitative judgement.

The report estimates that 3.5 million Australian workers are at high-risk of being displaced by automation between 2015 and 2030. Workers who perform a large share of automatable tasks may need support to find new ways of working, either in the same jobs or in new ones. With an additional 6.2 million people projected to join the Australian workforce by 2030, the report suggests that Australia will need to adequately prepare its future workers for automation.

However, this domain is very difficult to model, and that the effects of increased automation are likely to be unevenly spread. Further insights in this area are required from academic experts and independent organisations.



Non-exhaustive examples	**Task composition of work** Full-time hours per week		**Time saved on automatable tasks** Reduction in weekly hours spent on automatable tasks	
	2000	2015		
Sales assistant	28 12	37 3	9 hour change	• Less time scanning items • More time assisting customers
Factory worker	6 34	14 26	8 hour change	• Less time on an assembly line • More time training other workers
Manager[1]	35 5	36 4	1 hour change	• Less time collecting data • More time on strategic planning
Teacher[2]	27 13	35 5	8 hour change	• Less time recording test scores • More time assisting special-needs students

■ Automatable 🟩 Non-automatable

Notes: Assumes a full-time worker works 40 hours per week, figures rounded to nearest hour.

1 Unweighted average of ANSZSCO 1 digit code used to estimate manager timeshares (excluding farmers and CEOs).

2 Example based on high-school teacher.

Source: ABS, O*NET, AlphaBeta analysis.

Figure 7: The effect of automation on the workforce

Adapted from: AlphaBeta, 2017.

3.2.2.1 Digital technologies and the quality of work design

There are many anecdotal examples of how AI can positively affect work, such as using chatbots to remove uninteresting and routine work (see Mesiter, 2018). Research on older technologies likewise supports the idea that technology can improve work design, such as electronic monitoring systems that enable people to monitor and improve their own productivity (Osman, 2010), and that this technology enhances job autonomy because greater information availability decentralises power and supports localised decision making (Davenport and Short, 1990).

However, there can also be negative consequences for work design, which affect the health, well-being, and performance of workers. New technologies can, for example, result in reduced work agency and deskilling. As an example, Eriksson-Zetterquist et al. (2009) describe how new global purchasing technologies radically altered the roles of purchasers in a Scandinavian automotive company. Where purchasers once had a high level of responsibility, autonomy, social contact, and a strong professional identity, the introduction of new technologies created an environment where purchasers mainly followed standard operating procedures with reduced need for skills, yet also experienced increased bureaucracy and workload. Likewise, electronic monitoring systems can result in excess surveillance, invasion of privacy and reduced job autonomy. As a consequence, employees can sometimes experience high levels of stress, fail to comply with organisational rules or engage in deviant behaviours (Alge and Hansen, 2014).

In theory, allocating tasks to AI technologies should leave operators free to do other tasks, but it can also create social, cognitive, or biomechanical problems. These might include decreased situation awareness; distrust of automation; misuse, abuse, or disuse of the machines; complacency; reduced vigilance; impaired performance; and erosion of skillsets (Redden, Elliott and Barnes, 2014; Grote and Kochan, 2018).

3.2.2.2 The effect of new technologies on work design and outcomes

There are many factors affecting the design of work such as national institutional regimes, employment policies, organisational culture and local leadership (Parker, Van den Broeck and Holman, 2017). These factors can also shape the impact of technology on work design. For example, in the specific area of computer-based monitoring, one study (Alge and Hansen, 2014) shows that the effects of electronic monitoring systems tend to be negative, resulting in reduced job autonomy, greater demands and higher stress. However, when the organisational culture is a highly supportive one, employees are more likely to be involved in the design of the monitoring system. This means that the focus is on fostering employee development, resulting in employees regarding the system as fairer and less stressful. Other factors, such as how the data are collected and their accuracy, also shape how employees perceive and react to monitoring systems.

3.2.2.3 Tasks and decisions carried out by AI or machines

AI technologies can assist workers in performing their tasks and in making decisions, but may be considered ineffective or inappropriate for certain tasks. For example, big data analysis can be used to simplify personnel selection, but it is unlikely to be a good substitute for leadership functions (such as inspiring employees), nor some of the other highly complex and cognitively demanding tasks in managerial jobs (Cascio and Montealegre, 2016). Moreover, AI may exacerbate existing human demographic biases in who performs which jobs, as seen with Amazon's recruitment tool that contained gendered bias (Dastin, 2018). However, if the algorithms are developed without bias and the underlying data is inclusive, AI holds the potential to be less biased than humans.

Likewise, there is the question of when algorithms should replace human judgement, and when they should not. For example, in the discussion of the use of algorithms in financial decision making, Bhidé (2010) argues that 'predictions of human activity based on statistical patterns are dangerous when used as a substitute for careful case-by-case judgment'. Bhidé describes how financial decisions are replaced by the 'robotization of credit', which can result in poor decisions (see also Alam and Kendall, 2017). In an ideal scenario, an AI system could provide data analysis, learn from its mistakes and provide feedback, while leaving the ultimate decision to a human agent.

3.2.2.4 Workers adapting, shaping and using AI technologies

Workers can use technologies in ways not anticipated by designers. One reason for this is that workers often do not trust the technology and hence do not use it effectively. If workers are to interact effectively with robots, for example, they need to trust the robots, communicate effectively with them and coordinate their actions with them. Research shows that the level of trust in AI is affected by factors such as the transparency of algorithms (Dietvorst, Simmons and Massey, 2016), having positive experiences with AI (Alan et al., 2014) and the responsiveness of the technology to humans (Bickmore, Pfeifer and Schulman, 2011). The degree to which workers have control over the technology can also shape their interactions with it. Attention needs to be given to how workers and machines coordinate their activities and work holistically as a team (Redden, Elliott and Barnes, 2014).

3.2.2.5 Summary – Productivity and changing employment

The adoption of AI will present both potential opportunities and risks to the workforce. On one hand, AI may assist with tasks and workplace shortages, and on the other hand it may replace tasks. The debate over AI and the future of work is not simply one of conflicting interpretations and arguments, it is one deeply immersed in the organisational structures and politics of our institutions. Employees have needed to adapt not only to the increasing presence of AI systems and technologies in workplaces, but also to the various power struggles within their organisations regarding the opportunities and challenges of AI. These power struggles are often determined by whether CEOs, directors and managers take optimistic, sceptical or balanced standpoints when it comes to adopting AI systems and technologies.

The question, then, is how businesses and industries will be led, organised and resourced in the age of AI, especially once it becomes clear that companies will need to adopt AI to remain competitive. Since no one organisational structure or management policy can accommodate the complexity of interpretations from optimists and sceptics, organisations will need guidance on how to make effective decisions about the automation of tasks and work roles. Individuals or groups who make decisions about work design such as, for example, managers, human resource personnel, consultants and IT staff, may benefit from specific education and training about effective work design.

Box 14: Changing employment. A case study on AI and the impact on psychologists

While the potential for change is often discussed within a variety of industries, occupations in human centred, health and helping sectors – particularly psychology – are frequently portrayed as immune to automation (see for example, Frey and Osborne, 2013; Susskind and Susskind, 2015; Reese, 2018). However, recent developments in automated therapy technologies are already changing the job characteristics of the psychologist.

A clinical psychologist is an *expert* in the analysis and understanding of the causes and consequences of human and animal behaviour. The job specification for a professional psychologist essentially specifies four tasks, whatever the area of specialisation (clinical, organisation, forensic, sport etc.). These are *to assess* the state of the client*;* to *formulate* hypotheses that account for causal relationships between observations and the behavioural, social, and economic outcomes that were the primary reason for the client contacting the professional; to propose an *intervention* in those causal relationships; and to *evaluate* the outcomes of said intervention.

Each of these areas – assessment, formulation, intervention, and evaluation – may be influenced or even replaced by automated procedures in the following ways:

Assessment. Over 60 years ago, Meehl (1954) demonstrated that statistical aggregation of assessment (tests or observations) was virtually always superior to aggregation by the clinician. This demonstration has been successively supported (see for example Dawes, 1994). The development of computer-aided tests has increasingly supplanted the provision of assessment by clinicians. The use of computerised assessment reduces the cost and time of assessment and may increase the accuracy of results (Kratochwill, Doll and Dickson, 1991; Nezami and Butcher, 2000). Computer based monitoring, including facial recognition, can be used to assess emotional changes in the client while being assessed, superior to many judgments made by clinicians. AI has been successfully developed to identify and recognise microexpressions – a task which otherwise involves advanced sensory and cognitive skills in addition to specialised training (Li et al., 2015).

Formulation. The tacit knowledge or *intuition* traditionally regarded as necessary in the development of hypotheses of cause and effect can be seen as resulting from training in uncontrolled environments wherein there are uncertain relationships between cues and decisions. These uncertain relationships can be identified and the clinician trained to make more predictable links (Kahneman and Klein, 2009). Machines can also generalise

to previously unseen cases and generate 'probably almost correct' responses to novel patterns, superior to the human operator. It is conceivable, therefore, that AI systems can automate intuition (see for example Morrison, Innes, & Morrison, 2017).

Intervention. With the increasing dominance of cognitive behaviour therapy in psychology, the relationship between the therapist and the client – previously regarded as important – has been downplayed. As a result, the *therapeutic technique* has become somewhat divorced from the therapist-patient relationship. Conceivably, this renders certain aspects of the psychologist's role – such as cognitive understanding required to identify and diagnose a psychological problem – susceptible to automation (see for example Innes & Bennett, 2010).

Evaluation. Evaluation of an intervention can be computer-based. This eliminates the unconscious biases often present when clinicians make judgments (see for example Lilienfeld et al., 2014).

The automation of assessment, formulation, intervention, and evaluation in clinical practice is in progress (Kamel Boulos et al., 2014; Innes and Morrison, 2017; Michie et al., 2017). Compared to human psychologists, AI systems can potentially perform these tasks with less bias, fewer computational and procedural errors and with no burn-out and fatigue.

However, if these aspects of psychology are subject to automation, many psychologists will still be required to develop psychological theory and methodology. Previously, the outcomes of computerised assessments were found to be useful only when utilised by a professional with adequate training (Nezami and Butcher, 2000). Psychologists of particular skill and insight may still be required. However, these will be a small proportion of those presently employed in Australia and New Zealand.

This also creates implications for the education system. Psychology is the second largest undergraduate program in Australian universities. While not all students studying psychology wish to become professional psychologists, many of them do. Therefore, the implications for the future training of psychologists are significant (Kennedy and Innes, 2005; Innes and Bennett, 2010), not only at postgraduate but at undergraduate levels.

There are other views of the factors that will affect the development of AI in forms that will affect the delivery of human services (see for example Aoun, 2017) but they do not address the fact that the model adopted in psychology is based upon the development of a technologically compatible structure that is liable for automation.

3.2.3 Ageing population

Australia and New Zealand's ageing population could present challenges to the overall supply of labour due to a decrease in workforce numbers (Brown and Guttmann, 2017). While migration and increased female workforce participation has thus far counteracted the reduction in labour supply encountered, the overall proportion of the ageing population is anticipated to significantly accelerate. By 2044, it is expected that one in four Australians will be aged 65 and over. The anticipated workforce reduction raises additional concerns including the capacity to provide public services and healthcare for an increasingly large portion of the population. However, AI technologies could potentially assist in replacing labour shortages. This is not a situation unique to Australia or New Zealand; the Japanese workforce has already encountered significant decline in workforce participation as result of a decreasing and ageing population (Schneider, Hong and Le, 2018). Japanese firms experiencing labour shortages, including the Japanese construction industry, have developed and adopted AI technologies with the view to counter these adverse effects. In this way, AI may provide opportunities to support declines in labour supply and productivity for some sectors.

3.2.4 Changing centres of employment

Developments in digital technology have, in some instances, resulted in increased urbanisation and reductions in the populations of smaller cities (see for example Porter, 2017). Transformations in the workplace enabled by AI may make remote working more attractive and more feasible. Physical co-location of workers may not be required for many employment roles. Given Australia's size, this fact may be particularly advantageous. People may choose to work far from the physical location (if any) of their workplace. Conversely, given technological developments and the use of AI systems, workplaces need not be located in the major cities in order to attract the most talented employees. This could have important implications for smaller centres and rural areas, leading to a renewed source of income and the reversal of population reductions. However, if these promises are to be realised, an appropriate high bandwidth infrastructure is required (see Chapter 2).

3.3 Education, skills and training

In January 2018, the Australian Broadcasting Corporation aired *The AI Race* (ABC, 2018) which presented data from a study into the risks to Australian jobs from AI-powered automation (AlphaBeta, 2017). Various jobs were explored from truck driving to law and few people felt well-prepared for the broad take-up of AI. Regardless of whether AI's transformative opportunities and impacts are perceived or actual, it is important that citizens feel prepared and informed to live, work, and interact with AI technologies. One of the challenges is to identify exactly what this process of education and training at all levels will involve in terms of skills, abilities, competencies, behaviours and knowledge. The following sections explore these challenges.

3.3.1 Agile and transferrable skills. What are the skills and knowledge we need to foster?

A recent UK publication provided a detailed analysis of how future employment is likely to change, and identified the implications of these changes for skills (Bakhshi et al., 2017). While the analysis focused on the US and the UK, the report provided insights of value to countries across the globe. The analysis suggested, for example, that in education, healthcare, and the wider public sector, roles and jobs are likely to increase in number and importance. However, some low-skilled jobs, such as those in construction and agriculture, are likely to be impacted.

The report identified the skills that are likely to be in greater demand in the future, including interpersonal skills, higher-order cognitive skills (such as originality and critical thinking) and learning strategies, namely the ability to set goals, ask appropriate questions, and take feedback into account as knowledge is applied meaningfully in a variety of contexts. The results confirmed the future importance of what are often referred to as 21st century skills, particularly interpersonal competencies – a finding that is consistent with those from other writers (see for example Tett, 2017) and reports (see for example the World Economic Forum and Boston Consulting Group's 2015 report on 21st century skills, and a similar report by Trilling and Fadel, 2012). Additionally, as AI technologies deliver new scope for prediction, there is increased need for human judgement to determine the best ways in which to action AI powered prediction (Agrawal, 2018; Agrawal, Gans and Goldfarb, 2018).

Some scholars suggest that a more future proof and appropriate approach to education and training – especially in terms of preparing for an AI-enabled future – is to focus on the notion of an 'interwoven intelligence' as the basis of an intelligence-based curriculum (see for example Luckin, 2018). Interwoven intelligence refers not only to academic and social intelligence, but also 'meta-intelligence', that is, the ability to develop an understanding of what knowledge is in different contexts. Currently, AI systems and technologies are limited to performing academic intelligence, but struggle when it comes to elements of meta-intelligence. As such, it may be useful to design and implement education systems based on a more 'interwoven' conception of intelligence.

3.3.2 The future of education

There are three core questions to consider in relation to AI's impact on the future of education and how Australia and New Zealand can ensure that students are well equipped with the sort of skills that will be valuable in an AI-enabled world: firstly, how can AI systems and technologies be used to augment education and learning; secondly, how can students receive adequate guidance and support when it comes to developing skills in AI development; and finally, how can we educate people about AI, so that they can make informed decisions about how and when to interact with AI systems and technologies?

There are several examples of AI systems that can teach well-defined subject areas, such as those that are routinely part of the science, technology, engineering and mathematics (STEM) curriculum. These systems can help learners build an understanding of the core principles of STEM education. Some AI systems provide individualised tutoring by continually assessing student progress. There are companies developing culture-learning and language-learning AI systems that specialise in experiential digital learning driven by virtual roleplay. Many of these technologies can be used by educators to augment and enhance a more traditional learning experience. If AI-based learning tools begin to displace some aspects of teaching, it will be important for teachers to focus on areas of knowledge acquisition and learning where AI is ineffective, such as meta-intelligence.

Box 15: AI for special educational needs

There are many potential benefits of applying AI to the education of special educational needs and disability students. For example, natural language processing to enable the development of voice activated interfaces can help students with physical disabilities who are restricted in their use of other input devices, such as keyboards.

The combination of AI and other technologies, such as virtual and augmented reality, can help students with physical and learning disabilities to engage with virtual environments and take part in activities that would be difficult for them in the real-world. Virtual reality becomes 'intelligent' when it is augmented with AI technology. AI might be used simply to enhance the virtual world, giving it the ability to interact with, and respond to, the user's actions in ways that feel more natural. Alternatively, AI might also be integrated into intelligent tutoring systems to provide intelligent support and guidance to ensure that the learner engages with the intended learning objectives without becoming confused or overwhelmed.

Virtual pedagogical agents might also be included, acting as teachers, learning facilitators, or student peers in collaborative learning quests. These agents might provide alternative perspectives, ask questions, and give individualised feedback. In addition, intelligent synthetic characters in virtual worlds can play roles in settings that are too dangerous or unpleasant for learners. For example, FearNot, a school-based intelligent virtual environment, presents bullying incidents in the form of a virtual drama. Learners, who have been victims of bullying, play the role of an invisible friend to a character in the drama who is bullied. The learner offers the character advice about how to behave between episodes in the drama and, in so doing, explores bullying issues and effective coping strategies.

AI can also help EdTech applications be more flexible through, for example, deployment online, meaning that they can be available on personal and portable devices within, and beyond, formal educational settings. The way that AI enables technology to be personalised to the needs of a learner can also make it beneficial for learners with special educational needs.

AI is being used by researchers at Athabasca University in Canada with students who have attention deficit hyperactivity disorder (ADHD). The goal of this work is to develop an AI-in-education system that detects ADHD earlier than current models, improves the quality of diagnosis of ADHD, educates instructors about methods that are effective for teaching students with ADHD, measures competency improvements and challenges of ADHD students and encourages ADHD students to study in an environment filled with anthropomorphic pedagogical agents.

The growth of AI development will depend on students with relevant post-secondary training in STEM, particularly mathematics. AI technologies may be used to support more engaging educational experiences by which to foster interest in STEM education (Rexford and Kirkland, 2018). Given that AI systems in the future may 'learn' to code parts of themselves, education in these areas may focus more on aspects of interaction design rather than writing computer code. STEM education across all levels could incorporate training on ethics, human rights and inclusive design to assist in the development of equitable AI technologies.

More broadly, if strengthening in AI development and the future workforce is to be bolstered, there will be a need to address clear gender disparities across STEM and humanities, arts and social sciences (HASS) education and participation. For example, women make up only 16 percent of Australia's STEM fields (Office of the Chief Scientist, 2016). Low rates of female participation in STEM fields could result in a deepening of gender inequalities as AI continues to emerge. Indeed, should STEM emerge as a preferenced knowledge and skill set, women may be prevented from accessing higher paid occupations requiring these skills. Initiatives aimed at encouraging and increasing gender diversity in STEM, such as Code like a Girl, Girls in STEM toolkit, Superstars of STEM and the Women in STEM decadal plan, are important advancements in this area and will play a crucial role in furthering the representation of women in AI development and application. Gender diversity is an important element of the design, development, and implementation of AI technologies to ensure inclusive design and equity of workforce representation. For example, a recently developed artificial heart was physically compatible for 86 percent of men but only 20 percent of women; while smaller versions will subsequently be designed for women, inclusive design from the outset would have helped mitigate the issue (Huet, 2013). Effective inclusion of women in the AI industry will require increased female participation in conjunction with efforts to combat workplace discrimination.

Additionally, there are similar patterns of gender inequality in AI-related HASS fields such as philosophy and ethics, and initiatives to increase female participation in HASS will be equally important. Alongside strengthening STEM training, there should be a focus on educating students in HASS subjects. HASS education equips students with expertise in 21st century skills such as communication, creativity and the social implications of technological developments. Graduates with expertise in gender and race, for example, may provide guidance on how AI researchers can develop equitable and non-biased AI techniques and interfaces. HASS graduates may also provide insight on how AI technologies affect recreational and leisure time, as well as their potential for uptake in arts and cultural industries, including important areas of the creative and cultural economy such as galleries and museums, entertainment (including screen, cinema and videogame development) and sport. To ensure a workforce that is cognisant of the implications of AI technologies, incorporating specific HASS subjects such as ethics and human rights into AI education and training programs should provide graduates with important all-round skills and expertise for AI development, application and use. Equally, HASS students should be exposed to digital and data literacy programs. Indeed, an increasing number of researchers in the humanities and social sciences are working at the interface of technology and culture. Future AI developers will need to communicate with, and develop proficiency in, legal, ethical, and philosophical frameworks that ensure AI systems are accessible, unbiased, and socially accounted for. As noted elsewhere in this report, this will not entail a wholesale reframing of existing regulatory systems; but rather a process of applying existing ethical and legal frameworks to AI and adjusting those frameworks where necessary.

The different roles and tasks that the future of AI will present may require a range of complimentary educational offerings, including interdisciplinary programs and programs that connect with industry.

In 2017, the Australian National University announced a 10-year plan to drive expansion of its program in engineering and computer science, including the establishment of the Autonomy, Agency and Assurance (3A) Institute. The 3A Institute has initiated a research agenda to build a new body of knowledge for that will apply scientific principles and interdisciplinary practices through a pragmatic lens to consider the full spectrum of benefits and harms presented by technology for the betterment of humanity. Further, programs such as Swinburne University of Technology's Factory of the Future provide examples of industry-based learning to practitioners, in this case, for the advanced manufacturing sector. Internationally, there has also been a recent surge in new ethics courses for AI and autonomous systems from academia and industry (Pretz, 2018; Singer, 2018).

AI specialists often do not necessarily possess sector or industry specific knowledge in areas where AI can be applied (such as agriculture, energy, health and mining). Similarly, those in specific sectors do not necessarily have the technical knowledge to apply AI to their area. Education and training programs may help to address this gap by offering technology-related subjects. MIT recently announced a US$1 billion investment in an AI college that equips students from a diverse range of disciplines such as chemistry, biology, physics, history, politics and linguistics to with the knowledge and expertise to apply AI and ML techniques to their disciplines. The college will also place an emphasis on ethical considerations relevant to AI (Vincent, 2018). As briefly mentioned in Section 2.4, university programs to bridge this gap are emerging, however, in order to be effective, there needs to be a cohesive and cogent approach across these initiatives. This is discussed in further detail as it applies to the research sector in Section 3.3.5.1.

To assist society in understanding AI technologies and applications, there will also be a need to develop and implement initiatives that provide all individuals with the opportunity to develop basic literacies in how AI systems and technologies function. This may involve establishing what AI can and cannot achieve, as well as what society can and should expect from AI. This will be important for ensuring not only that all citizens are able to take full advantage of the opportunities afforded by the broad take-up of AI, but also that they understand the potential implications and risks associated with using AI systems (e.g. consenting to data collection). One possible means through which to achieve this educational benchmark will be public communication initiatives and also through 'micro credentialing', which refers to mini-qualifications obtained online through tertiary and job training institutions.

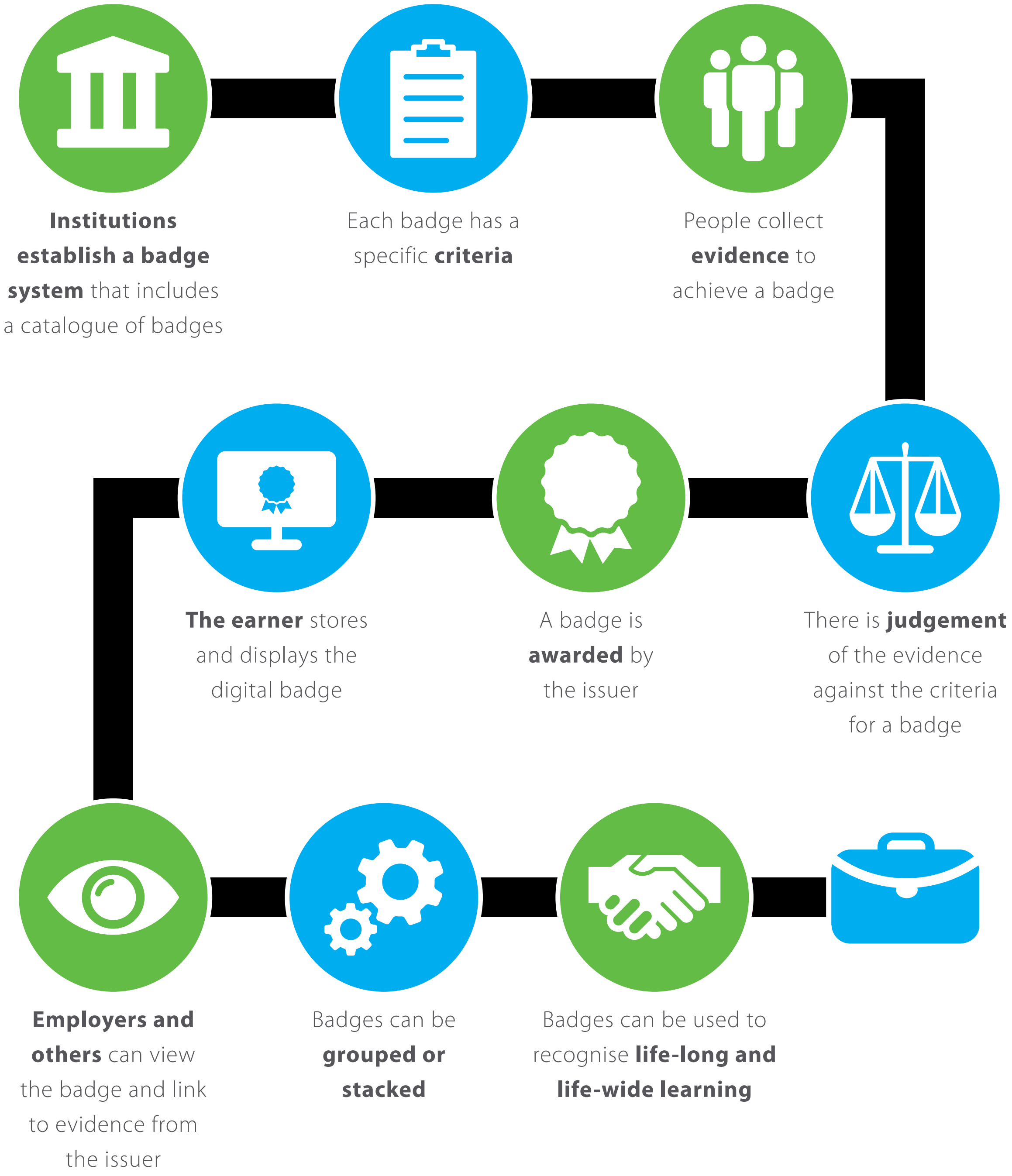
3.3.3 Micro credentials

Micro-credentialing is a way of certifying learning outcomes within an institution through online education platforms. Massive open online courses (MOOCs) such as Coursera often use micro-credentials to demonstrate a student's achievements with mini-qualifications in specific subject areas and capabilities. Given that micro-credentials are typically certified through online platforms, it is possible that AI systems and automation could play a role in their future. Moreover, as mentioned in the previous section, micro-credentials may be useful for people who require basic education and literacy in AI techniques and processes, so that they can effectively integrate AI technologies into their lives.

For micro-credentials to work well and maintain their value, there needs to be a healthy ecosystem of use, where there are employers and networks looking for competence-based skills, an easily updatable credentialing system, and students who want to gain the credentials. Without this ecosystem it is hard for the credentials to have an extended life. Blockchain technology offers support here. For example, Open Source University is using blockchain technology to enable the permanency of credentials and to ensure privacy and security. Future education platforms may provide virtual tokens as an incentive to increase learners' motivation and reduce digital course dropout rates.

Micro-credentialing is also a way of certifying personalised learning and peer review to achieve an educational certification. For example, Digital Promise in the US has launched a scheme of micro-credentialing for educators to provide competency-based recognition of the skills they learn throughout their careers. Once an educator has selected the micro-credentials they want to earn, they collate the evidence required and submit it via an online platform. An expert reviewer or educator who has already earned the related micro-credential will review the evidence, and if successful, the educator will be awarded the micro-credential in the form of a digital badge (Figure 8) (Digital Promise, 2018). AI can be used to identify skill gaps for people in the employment market (e.g. IBM's Watson Career Coach). Such AI systems can estimate the cost to retrain, overcome gaps and suggest learning pathways for individual development.

Digital badges are granular, verifiable records of achievement and can be thought of as a micro-credential. They offer a mechanism for valuing skills gained outside formal learning contexts. Learning management systems such as Blackboard, Moodle and Canvas have piloted the use of badges in many disciplines and levels. Open Badges go one step further, allowing skills, interests and achievements to be verified by attaching information to the badge image file, hard-coding metadata for future access and review.



Institutions establish a badge system that includes a catalogue of badges

Each badge has a specific **criteria**

People collect **evidence** to achieve a badge

There is **judgement** of the evidence against the criteria for a badge

A badge is **awarded** by the issuer

The earner stores and displays the digital badge

Employers and others can view the badge and link to evidence from the issuer

Badges can be **grouped or stacked**

Badges can be used to recognise **life-long and life-wide learning**

Figure 8: An overview of the badge system process

Adapted from: West & Lockley, 2016.

3.3.4 Vocational training and lifelong learning

The sectors currently most impacted by AI are, perhaps unsurprisingly, those where significant revenue has been invested in AI technologies. Arguably, the sector most benefitting from AI is retail. Platform companies such as Alibaba and Amazon make extensive use of AI to better understand consumer preferences and habits, as a way of streamlining productivity and targeting consumer tendencies. These sectors are often premised on the need to process large amounts of data to identify patterns and relationships. Other examples include processing patient and treatment data for medical diagnosis, finding specific information from millions of documents in the legal profession, and recognising the identity of a person and their right to enter a country.

Vocational education is generally not well financed and has not seen significant investment in AI technology. However, the Industry 4.0 Higher Apprenticeship Program is an initiative supported by the Australian Government through the Skilling Australia's Fund, which seeks to train technicians to a higher skill level in the areas of the Internet of Things, automation and robotics, cloud computing, smart sensors and advanced algorithms. AI technology that is being developed for the school and university sector may in some cases be appropriate for vocational education. Examples include recommender systems such as 'Filtered', which is being used by companies for employee training to help make best use of existing company training materials, and specialist training such as that provided to the US armed forces by 'Alelo', which uses virtual roleplay simulations to teach aspects of culture, languages and interacting with locals (Alelo, 2018; Filtered, 2018).

Schools, the vocational education and training (VET) sector and universities should encourage broad-based skills and training beyond credentialing people for employment. Negotiation skills, creativity and critical thinking are human attributes that are resistant to automation and have the capacity to be cultivated. Education on ethics, social sciences and the humanities could be promoted with the view to strengthen democracy, develop ethical AI, and generate the new forms of governance that will be necessary to manage the impacts of automation on society, the economy and culture.

3.3.5 Next generation of AI researchers

A report commissioned by the Australian Computer Society suggests that there are shortages of workers in all areas of IT in Australia. The report states that 'demand for ICT workers is set to grow by almost 100,000 to 758,700 workers by 2023' but 'with fewer than 5,000 domestic ICT graduates a year, the only way we'll reach workforce targets is by importing labour, much as we've done for the past five years' (Deloitte Access Economics, 2018: 34; 3). Australian universities also export ICT education, with over 8,500 international ICT student completions in 2016.

AI technologies are being used primarily by banks and security companies to tackle risk and to improve fraud identification and by online companies to better match products with clients. However, over the next few years, AI technologies are expected to be more widely implemented in all industries, and in particular in manufacturing, retail and healthcare. There is likely to be an increasing demand for AI researchers and developers to support the development of applied AI technology across industry, government, and society.

The growth of AI technologies may also prompt researchers in areas such as IT and mathematics to engage with those in philosophy, law, and public policy. A key ethical question – and one that will be explored below – may involve working towards greater transparency of AI systems whose decision-making processes are currently obscured.

3.3.5.1 Research across disciplines

AI technologies may lead to significant changes in research *practice* in academic and industrial contexts. These changes will have varied effects depending on the discipline and area of research.

Many STEM researchers are using ML systems to perform repetitive identification tasks such as smoothing noisy astronomical pictures of galaxies, searching for the right sequence of reactions to synthesise small organic molecules (e.g. drug compounds), using AI systems in genetics research to predict how genes affect the functioning of nearby genes and using image processing algorithms to automate counting similar objects in a natural setting.

There are examples of AI techniques creating new research fields within a discipline. The data intensive field of precision medicine is creating a demand for research into AI systems for ML, advanced optimisation and searching techniques (Hodson, 2016). ACOLA's report, *The Future of Precision Medicine in Australia*, investigates this topic.

In humanities and social science research (and also some strands of STEM), natural language processing tools and methods such as analysis of text or language will enable researchers to sift through maintenance records, interview transcripts or meticulous records of human observations. This is likely to affect academic labour in these areas and create new avenues of research.

While artificial creativity may be a longer-term goal, it could be used in areas such as industrial design, architecture, engineering and art. There are claims that AI is poor at performing creative labour in these areas (see for example Rexford and Kirkland, 2018), but AI technologies are improving in some areas of creative work, such as sonnet writing, art and fashion design (see for example Downey,

Box 16: AI in health education

The lack of easy access to health data makes it difficult for university students to develop health specific skills in AI. Researchers at institutions with established research programs in areas such as linked health data may take advantage of existing projects and funding to gain access to valuable data, though this type of training remains inaccessible to most students. As a consequence, motivated students may move to other areas of AI application or graduate without adequate training. It is possible for students to gain experience using datasets from other countries, such as the US, which can be easily downloaded free or at modest cost. This leads to lost opportunities, as Australia and New Zealand would have gained if these students trained and published using local population datasets.

As noted in Section 3.3.2, there is also a potential lack of domain knowledge across sectors. Developing applications of AI in health requires a keen understanding of human health, population health, human behaviour and the regulatory environment. Higher education in AI and data science tends to be generic, with little intersection with health. Closing this gap may be possible by developing a range of degree programs that are more specific to health and by facilitating interactions between students and industry by, for example, taking advantage of institutions such as CSIRO, programs such as the Australian Cooperative Research Centres, Industrial Transformation Training Centres, and initiatives such as those supported through Innovation and Science Australia and, in New Zealand, by Callaghan Innovation, New Zealand's innovation agency. Industry partnerships could be incorporated into vocational, bachelor, masters and doctoral programs.

2016). Indeed, in 2018, the first AI generated painting sold at auction for US$432,500, almost 45 times more than its estimated value (Christie's, 2018).

AI technologies will have broader effects on basic research practices and processes, such as the use of AI systems to conduct literature searches. AI systems may be used to automate certain aspects of the publishing or peer-review process (see DeVoss, 2017).

3.3.5.2 Changing the training of researchers and developers

Future AI researchers and developers will need to respond to ethical, disciplinary, social and legal challenges in their research and development aims and outcomes. To investigate, design and build systems, AI researchers and developers will need to be equipped with the skills to work with other discipline experts from fields such as software engineering, human-computer interaction (or interaction design), information systems, business, psychology, economics, politics, industrial relations, human resource management, law, human rights and ethics.

- **Ethics.** Researchers and developers from all areas will need to be aware of ethical and human rights frameworks to ensure future AI technologies are transparent to inspection, predictable to those they govern, protect human rights, are robust against manipulation, and deployed in a context where we know who takes responsibility. Many IT degrees do not provide students with an education in ethical considerations.
- **Interdisciplinarity.** AI researchers and developers will need to work with other discipline experts to investigate, develop, design, and build effective AI systems. Software engineers, for example, may need to work with philosophers and law experts to familiarise themselves with the relevant ethical protocols, while HASS researchers may need to collaborate with AI developers to create effective tools and systems for their research needs.
- **Social impact.** AI researchers and developers will need to consider the social impact of their work – such as the implementation of AI technologies in certain areas of society or within certain social groups. Crucially, this will involve an acknowledgement of cultural diversity – in that AI technologies will need to be developed with a wide range of end users in mind. It will be important for AI research and development teams to develop strategies for attracting a diversity of experts and practitioners.
- **Legal compliance.** Researchers and developers from all areas will need to consider the various policies and legal frameworks that are being introduced to regulate the use and implementation of AI technologies. For example, the European Union's new General Data Protection Regulation legislation article 15 grants the data subject the right to 'meaningful information about the logic involved and the envisaged consequences of such processing for the data subject when automated decision making is used'. We will need to equip students with skills to detect poor, malicious or dubious AI systems.

Future research priorities for AI should include a broad and balanced interdisciplinary portfolio of projects that consider the full spectrum of opportunities and challenges likely to face society and the sector.

3.4 AI and transformations of identity

We live in a world where identities are technologically mediated to an unprecedented level and where the boundaries between digital and off-line worlds are becoming increasingly blurred. Today, identity intersects with life mediated by chatbots, softbots, touchscreens, virtual landscapes, location tagging and augmented realities. The impact of AI here must be considered in a broader context of interconnected technological, genetic and informational developments that are reconfiguring identity.

AI and related digital technologies are transforming what 'identity' and 'the body' mean. In addition to organ development technologies, 3D printers have been used in research to print living human embryonic stem cells (Heriot-Watt University, Edinburgh), blood vessels (German Fraunhofer Institute), human skin (Lothar Koch of the Laser Centre Hannover in Germany) and even sheets of cardiac tissue that can 'beat' like a real heart (Cabor Forgacs, University of Missouri in Columbia). In the light of these developments, some have claimed that growing bio-organs (by printing them) will eventually replace the need for donor organ transplants. Like 3D printing, developments in AI are part of this broader transformation of identity. Craig Venter, one of the leaders who mapped the first draft sequence of the human genome in 2000, has been at the forefront of the digitisation of synthetic life. In 2010, Venter and his team produced the first synthetic organism by transplanting synthetic DNA into a vacant bacterial cell. For Venter, developments in synthetic life are only in their infancy.

While such developments point to physiological crossovers between technologies and humans, AI technologies are augmenting our identities in more indirect ways as well, such as receiving Amazon recommendations, requesting Uber, getting information from virtual personal assistants and talking with chatbots. Furthermore, AI systems are increasingly able to mimic or perform human-like functions. Examples include the ability of AI systems to communicate and interact with humans in a quasi-human manner (Reeves and Nass, 1996; Zhao, 2006), to enter into states like 'sleep' to pair with and complement user schedules (Hsu et al., 2017), and to engage in sexual activities and relationships with human partners (Cheok, Levy and Karunanayaka, 2016). These technologies may rapidly enter into people's lives in some contexts, but encounter cultural resistance in others.

Additionally, the notion that AI may lead to a 'dispensing of the body' may carry political connotations for certain disadvantaged groups. A person with a disability, for example, may already have extensive experience dealing with technologies that augment their everyday bodily routines. As such, we need to avoid universalising 'the body' and 'identity' to consider a diversity of embodied experiences, including how race, disability and gender intersects with AI design.

3.4.1 Childhood development

The vulnerability of children requires consideration in the use and application of AI. The impact of AI-enabled technologies on children – including play, cognitive development, socialisation and identity formation – is the subject of much debate. As has been the case with major technological developments throughout history, this debate is often split between optimistic and sceptical positions.

An example of a sceptical position is Sherry Turkle's (2012) *Alone Together*, which compares play with traditional toys to play with digital pets such as Furbies and Tamagotchis. For Turkle, traditional forms of childhood play involve the child animating the toy – investing the object with imagination – to establish an emotional relationship. By contrast, robotic toys and AI devices appear to children as if already animated and full of intentions of their own. Children may be particularly vulnerable to these changes because, according to Turkle (2012), 'children need to be with other people to develop mutuality and empathy: interacting with a robot cannot teach these skills'. Moreover, many toys are now part of the Internet of Things, meaning that they can be used to collect and analyse data on children's play practices, thereby 'enrolling' children in data collection strategies without their informed consent or knowledge.

However, many researchers argue that these sceptical accounts fail to acknowledge the complexities of childhood play, socialisation and identity formation in a digitised society. For example, Seth Giddings (2014) observes that children's play often oscillates between online and offline game worlds; children may, for example, draw inspiration from an experience with an AI device for a scenario in a playground role-playing game, or vice versa. In this sense, children's play is often messy and unpredictable, meaning that it can exceed or confound the data collection strategies employed by designers and manufacturers of AI-enabled toys. Furthermore, much of the scepticism in this area rests on the assumption that children exist in – and should be encouraged to remain in – an 'innocent' reality removed from the digital society inhabited by adults. Yet, children are now born into a digital, data-driven and AI-enabled society, and are likely to express interest in actively participating in this society by, for example, wanting to explore aspects of their identities through play with AI-enabled toys. Therefore, it is important that adult mediators and, where appropriate, children are educated about what these AI-enabled devices are capable of, how they can be productively integrated into a child's life, and what measures can be implemented to safeguard children from potentially compromising scenarios, such as when an AI-enabled toys collect data without the child's consent.

3.4.2 The psychological impact of AI

The psychological impact of AI is treated as an increasingly important design imperative. In the past five years, researchers have developed design methods to support psychological wellbeing (Hassenzahl, 2010; Desmet and Pohlmeyer, 2013; Calvo and Peters, 2014). These design methods often build on existing psychological theories.

One such theory is self-determination theory (Ryan and Deci, 2000, 2017), which examines the factors that promote sustained motivation and wellbeing. The theory has gathered one of the largest bodies of empirical evidence in psychology and identifies a small set of basic psychological needs deemed essential to people's self-motivation and psychological wellbeing. Furthermore, it has shown how environments that neglect or frustrate these needs are associated with illness and distress. These basic needs are *autonomy* (feeling agency, acting in accordance with one's goals and values), *competence* (feeling able and effective), *and relatedness* (feeling connected to others; a sense of belonging).

While the concept of autonomy can be complicated (see the discussion earlier in this chapter), an autonomous person is often seen as one that has a sense of willingness, endorsement or choice in acting (Ryan & Deci, 2017). This is not the same as doing things independently or being in control; rather it means acting autonomously and in accordance with personal goals and values. Individuals often relinquish control or embrace interdependence on their own volition. Within AI development, the vast majority of research has focused on the design of autonomous *systems*, particularly robots and vehicles, rather than on supporting autonomous *humans* (Baldassarre et al., 2014). Recently, however, the Institute of Electrical and Electronics Engineers (IEEE) has developed a charter of ethical guidelines for the design of autonomous systems that privilege *human* autonomy and wellbeing (Chatila et al., 2017). As discussed earlier in this chapter, perhaps it will be necessary to move towards a model where the human-AI relationship is considered in more emergent terms.

Box 17: Evaluating discursive claims about AI

Attitudes to new technologies are often shaped by the way such technologies are *discursively* framed – that is, how they are presented in public discourses such as the news media. Consider, for example, the headline, 'Stanford's artificial intelligence is nearly as good as your dermatologist' (Mukherjee, 2017). The associated story was that researchers at Stanford University had claimed to have developed an AI system that could detect whether a skin lesion is cancerous or not (Esteva et al., 2017). But the AI system they developed is a statistical ML model – a model performing a supervised learning task, which is to classify images of lesions based on labelled images of lesions that it has previously seen. This is a remarkable development, but it is not accurate to claim that their system is 'nearly as good as your dermatologist'.

The Stanford system is weak AI, meaning that it is good at a specific task or range of tasks. The problem with headlines like the above is that they suggest we already have statistical models that have reached the benchmark of 'general AI' (defined in the introduction to this report).

The Stanford system was trained with images from three datasets, including the International Skin Imaging Collaboration Archive. In this dataset, there are only two or three images of lesions on tanned or darker skin. Because the Stanford model was mainly trained with images of lesions from Caucasian people, it is unable to reliably classify lesions in people of diverse ethnic background. This example highlights the risk of researchers inadvertently not noticing, or reporting, deficiencies in their training data. If the training was conducted on a broader cohort dataset, it may have resulted in a system that was more inclusive.

The Stanford example is only as good as the data that it is trained on. The potential weakness of the system may be overlooked by the medical profession and healthcare policymakers, particularly in light of overreported claims.

3.5 Changing social interactions

In 2018, it was widely reported that a family in Portland, Oregon, had received a phone call from an acquaintance advising them to disconnect their Amazon Alexa device. The device had recorded private conversations in the family home and forwarded these, apparently randomly, to a person in the family's contact list. Although the conversation recorded was mundane, commentators were quick to forecast the arrival of a dystopic world where chatbots spy on us and share our information without consent. This anecdote underscores the effects of AI technologies on social life – identity, relationships, communication and so on – as well as the relation between AI and privacy, explored further in Chapter 4. It also highlights the way in which public responses to AI can shape how it is received, used and developed, as discussed throughout this chapter in relation to optimistic and sceptical approaches to AI.

3.5.1 Spoken and text-based dialogue systems

Spoken and text-based dialogue systems are often embedded in mobile smart apps for wellness and personal health management, an area that has seen a significant increase in activities over the past few years. However, not every mobile health app necessarily qualifies as an application of AI, since many apps do not have an intelligent component (such as the ability to adapt to the user, to learn from past behaviours or datasets, to interact as a human or to perform some form of reasoning). In addition, the evidence that these apps provide any health benefits is often scant (Byambasuren et al., 2018).

Progress has been made, however, on the measurement of the quality of apps (such as the MARS scale in Australia (Stoyanov et al., 2015) or on their certification. In the latter area, the US has led the way by allowing the Food and Drug Administration to approve mobile apps the same way they approve drugs (hence the name 'prescription apps') (Bilbrough, 2014; Boulos et al., 2014), and the NHS in the UK has a library of trusted digital tools, some which may have an AI component.

The range of mobile apps in this area is broad; however, many of these apps may have only a small element of AI and often are mostly passive devices. Nevertheless, they are becoming increasingly intelligent and able to adapt to the needs of the consumers and interact in a more human form. Many solutions are devoted to helping consumers change behaviours and better manage chronic conditions, such as diabetes and heart disease. Apps often provide assistance or support to change dietary habits, to stop smoking or drinking and to increase physical activity. Many solutions include an element of monitoring and may include interaction with health providers, such as nurses, dieticians and mental health specialists (Palmier-Claus et al., 2012). An interesting opportunity offered by mobile apps is that consumers may be more likely to disclose information to an app rather than to a human (Lucas et al., 2014), opening the way for better informed services. Substantial innovation has taken place regarding applications of dialogue systems to mental health (Hoermann et al., 2017), which ranges from the design of virtual affective agents that can help patients with depression or autism (Luxton, 2015) to the delivery of interventions (Hoermann et al., 2017).

3.5.2 Digital-device-distraction syndrome

There is significant public and international scholarly debate surrounding the effects of AI technologies (such as automated assistants, chatbots and digital devices) on communication and social relationships. In Australia, New Zealand and in other parts of the Anglophone world, there is a tendency to regard such technologies as autonomous forces in society, which inevitably lead to certain social outcomes (Wajcman, 2002). Discussions about 'digital-device-distraction syndrome' commonly exhibit this tendency as well. Digital-device-distraction syndrome often presumes that the presence of digital devices will invariably cause users to become more distracted and less attentive to elements of their surrounding environments (see for example Fritz, 2016; Nixon, 2017).

This 'deterministic' way of understanding the role of technology in society has, however, been the subject of sustained criticism in the social sciences (Mackenzie and Wajcman, 1999; Guy and Shove, 2000). Scholars working in the field of Science and Technology Studies have shown the significance of cultural factors in developing the design, implementation and use of various technologies.

This culturally informed account of technology can help us avoid discussing AI's effect on social life in overly simplistic terms. For example, instead of attributing the phenomenon of distraction in the contemporary era to the use of digital devices alone, we should pay more attention to the various social forces that produce and privilege distracted modes of being in some social contexts (Hsu, 2014; Wajcman, 2008). Doing so will give us a better understanding of how people can moderate the amount of distraction in their lives, since distraction is as much of a social issue as it is a technical one.

3.5.3 The use of algorithms in the provision of social services

Algorithms have long been used in organisational life to allocate resources and services. But what makes their contemporary use in the digital world unique and potentially problematic is their ability to remain opaque and hidden. This partly refers to the difficulty of recognising when some algorithms in the digital world are switched on and active (Pasquale, 2015). It also refers to the ways in which the inner workings of digital algorithms are commonly difficult to understand (Beer, 2009). As with any opaque or hidden decision-making system, it is crucial to ensure that the user trusts the expertise that goes into said system. The Australian experience of automated debt recovery systems will be discussed in Chapter 4.

This feature of algorithms in the digital age has led some researchers to explore how algorithms can be made more transparent to promote fairness and democratic values (e.g. Diakopoulos, 2016). One emerging approach is to understand how the opaqueness of algorithms stems from different sources. According to the work of Burrell (2016), algorithms can be concealed on a number of levels: as a result of institutional secrecy, technical illiteracy or the sheer scale and complexity of their operation.

These three sources of algorithmic opaqueness need to be considered in the use of predictive risk modelling in the provision of social services. It may be useful to make the operations of certain algorithms accessible to the public, such as the algorithms underlying Facebook. Knowledge and training about the technical workings of algorithms could be expanded and collaborative. A careful 'supervised' ML approach, which some predictive risk modelling programs already employ (Gillingham, 2015), may be worth further pursuit.

Ultimately, the algorithmic underpinnings of predictive risk modelling need to be transparent to the parties concerned. Perhaps more importantly, they also need to be held socially accountable, which is a complex issue in need of further implementation and elaboration (Ananny and Crawford, 2018). With social accountability, there are a range of questions to consider: which society, which culture, who decides what is socially acceptable within those societies and cultures and how can AI be deployed in a way that aligns with these expectations and social norms.

3.6 Conclusion

In 2017, Brooks (2017) argued that discussions about AI's impact on employment, society and identity are often unhelpfully skewed between overestimates and underestimates. He states, 'AI has been overestimated again and again in the 1960s, in the 1980s, and I believe again now, but its prospects for the long term are also probably being underestimated. The question is: *How long is the long term*?' (Brooks, 2017: n.p.). While it is difficult to say with certainty exactly how disruptive AI will be – to what extent it will replace low-wage jobs, transform high-skill occupations, reconfigure identity and displace traditional means of social communication – what can be said with certainty is that these changes, while rapidly emerging, will not arrive immediately. The opportunity remains for Australia to steer the development, adoption and use of AI and its potential impacts on society. Through proactive planning and responses, and measures that include broad education, industry and workforce responses and interdisciplinary research Australia can be better prepared for the anticipated disruption.

CHAPTER 4
IMPROVING OUR WELLBEING AND EQUITY

This chapter is based on input papers prepared by the generous contributions of the Australian Human Rights Commission (Human Rights); Joy Liddicoat (Human Rights); Nik Dawson (Economic and Social Inequality); Professor Greg Marston and Dr Juan Zhang (Economic and Social Inequality); Associate Professor Ellie Rennie (AI and Indigenous Peoples); Professor Maggie Walter and Professor Tahu Kukutai (Indigenous Australians and Maori); Dr Manisha Amin and Georgia Reid (Inclusive Design); Dr Jane Bringolf (Universal Design); Dr Sean Murphy and Dr Scott Hollier (Disability); Professor Neil Levy (Fake News); Professor Mark Alfano (Public Communication); Joy Liddicoat and Vanessa Blackwood (Privacy and Surveillance); Professor Hussein Abbass (The Human-AI Relationship); Dr Oisín Deery and Katherine Bailey (Ethics). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

4.1 Introduction

The development of AI technologies provides Australia with the opportunity to ensure that the benefits derived are fairly distributed. AI has the potential to benefit Australian society and advance human rights, including social security, health, economic and cultural rights. However, it also poses societal challenges and new forms of human rights violations, including new forms of discrimination.

The State can ensure that everyone benefits from scientific advancement and its applications, but to do so means that governments must consider how to engage with the benefits of AI and also manage the related risks, including risks of increased inequality. Different groups are, and will be, affected by AI technologies differently.

Some groups are particularly vulnerable to human rights abuses – especially when they are affected by decisions that are made, or informed by, AI-powered systems.

This chapter examines issues of equity and human rights that arise from AI in relation to freedom from discrimination, the right to justice, the right to work and the right to security. The development of AI in keeping with the human rights framework is both in accordance with Australia's legal obligations under international human rights law and necessary for the responsible and safe implementation of AI technologies. Underpinning human rights, however, are broader considerations in relation to accessibility, inclusiveness and equity.

Such principles have long been integral to societal standards within Australian culture. These foundations provide additional considerations for the development and use of AI technologies which do not entrench inequalities and instead provide benefit to all Australians.

This chapter also considers what additional protections might be needed. It discusses the need for trust, inclusion and public communication. If AI is implemented and developed inclusively, with considerations of wellbeing and human rights at the centre, it can play a role in closing the gap in social and economic inequality. However, if developed and implemented poorly, AI could further widen this gap.

The following discussion contains concepts which are not without contestation. Even concepts that are generally perceived to have universal acceptance, such as the application of human rights, have the potential to attract some criticism. The many varied experiences of each individual leads to a society that contains a plethora of perspectives and options. Discussion of AI technologies typically invokes disparate responses due to the emotive and uncertain nature of disruptive technologies. This chapter seeks to provide a framework of overarching basic considerations for the development of AI technologies that foster an inclusive and equitable society. A human rights approach is similar, but not identical, to an ethical approach (Australian Human Rights Commission, 2018b: 17). The two approaches can be brought together, where human rights can provide the normative content that can be applied through an ethical framework to developing and deploying new technology while ensuring that any decisions on a human's rights, privileges entitlements can be linked back to an accountable human decision-maker. While notions of ethics, equity and inclusivity are subject to interpretation, these principles underpin the Australian ethos of a 'fair go' – everyone deserves freedom of opportunity.

4.2 Human rights

Human rights are fundamental rights that should be enjoyed by all people. These rights embody the idea that all humans are born free and equal in dignity and rights. These inherent and inalienable standards apply across all aspects of human life, including those affected by new and emerging technologies such as AI.

The international human rights framework is the foundation for assessing the human rights implications of AI (Figure 9). Since the advent of the Universal Declaration of Human Rights (UDHR) 70 years ago, this framework has proven robust and capable of adapting to changing circumstances and technologies.

International human rights treaties rarely refer expressly to a particular domain, such as new technologies. The task is to apply existing human rights principles to countries in which AI is increasingly prevalent.

While AI powered technologies present challenges in relation to human rights, AI may also be used to advance human rights and enhance accessibility, social inclusion, civic participation access to education and medical care. Beyond the more obvious applications of AI, even the right to intellectual property may be enhanced by AI systems such as plagiarism checkers.

4.2.1 The right to equality and freedom from discrimination

Equality and freedom from discrimination are fundamental human rights, designed to protect people from unfair treatment through either direct or indirect discrimination. Indirect discrimination includes any act or omission which may appear neutral but has the effect of producing inequity.

The use of AI to assist in decision making has potential to advance human rights by enabling more informed decisions. There is potential to minimise direct and indirect discrimination by humans, who may act on their own prejudices. Algorithms can assist with identifying systemic bias and may present opportunities for better assessment of compliance with fundamental human rights (Commissioner for Human Rights, 2018). Additionally, AI technologies may improve access to services and improve outcomes across a range of socio-economic indicators. Improvements may be through better systems or interventions in health or education, for example, or targeted programs and services for groups who experience vulnerability and disadvantage. Examples of AI assisted technology that is being developed or used to potentially minimise inequalities from around the globe include:

- Textio, an augmented writing platform, uses AI to analyse and monitor company position descriptions to provide alternative

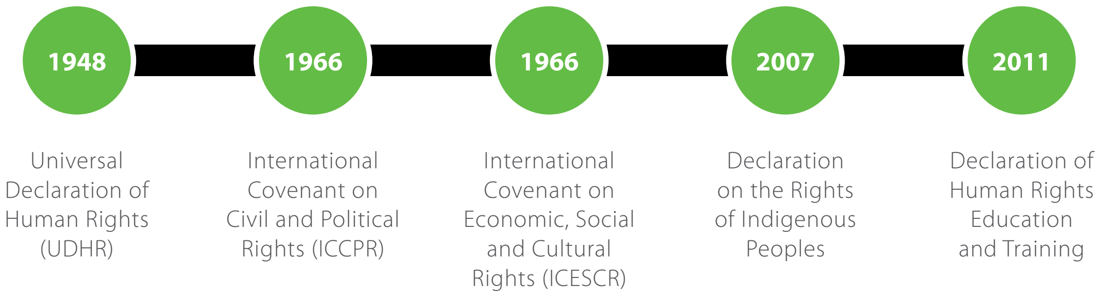

1948
1966
1966
2007
2011
Universal Declaration of Human Rights (UDHR)
International Covenant on Civil and Political Rights (ICCPR)
International Covenant on Economic, Social and Cultural Rights (ICESCR)
Declaration on the Rights of Indigenous Peoples
Declaration of Human Rights Education and Training

Figure 9: Overview of the international human rights framework

wording that engages passive candidates and eliminates unconscious gender bias language. One of the clients using this software has increased the proportion of hired female employees from 10 to 57 percent over two years (Halloran, 2017)

- Microsoft are developing innovative and ethical computational techniques that draw on deeper contexts of sociology, history, science and technology. The collaborative research projects address the need for fairness, accountability, transparency and ethics (FATE) in AI technologies (Microsoft, 2018a)
- IBM Research project Science for Social Good is aimed at leveraging the power of AI to address global inequalities and threats identified by the United Nations. Of particular interest is IBM's project investigating the automated identification and monitoring of hate speech online (IBM Research, 2018)
- Data for Black Lives is a diverse group of activists, organisers and mathematicians creating a network of data systems that seeks to provide data science supported solutions to black communities to 'fight bias, build progressive movements and promote civic engagement' (Data for Black Lives, 2018).

However, AI can also amplify discrimination, if the developed technology is misused– consciously or unconsciously. In addition, unequal access to new technologies, such as AI, may exacerbate inequalities, especially where access is affected by factors such as socio-economic status, disability, age or geographic location (O'Neil, 2016). Examples of applications of AI technologies with potentially discriminatory consequences include algorithms and tools that:

- target advertising of job opportunities on the basis of age, gender or some other characteristic such that, for example, people over a certain age never become aware of an employment opportunity (Angwin, Scheiber and Tobin, 2017)
- exclude applicants with mental illness (O'Neil, 2016: 4)
- lead police to target certain groups disproportionately, such as young people and people from culturally and linguistically diverse groups or minority groups (O'Mallon, 2017)
- entrench gender inequality, bias (Stern, 2017) and stereotyping (Steele, 2018)
- direct police to lower socio-economic areas, entrenching or even exacerbating the cycle of imprisonment and recidivism (O'Neil, 2016: 87).

From the technologies currently available and in use, it is indicative that the capabilities of AI are presently limited to assisting humans in performing tasks and functions. Additionally, AI applications are in their early stages and have the potential to include developmental flaws. Risk assessment tools that are employed in the administration of justice may use algorithms based on undisclosed criteria, or variables that result in algorithmic bias when applied to large datasets. This has been demonstrated in the NSW Police's risk assessment tool, 'Suspect Targeting Management Plan', which sought to target repeat offenders and people police considered likely to commit future crime (Sentas and Pandolfini, 2017). Analysis of those targeted by police revealed that young Aboriginal and Torres Strait Islander people were disproportionately targeted compared to other demographics (Sentas and Pandolfini, 2017: 1).

The transparency of the assessment criteria used in risk assessment tools will become increasingly important in determining the success of the aforementioned assistive AI technologies as will rigorous testing to minimise programmed bias.

Initiatives to minimise bias, increase transparency and fairness, and ensure privacy and human rights are not breached will need to be developed in parallel. The Australian Human Rights Commission is examining the challenges and opportunities for human rights of emerging technologies, and innovative ways to ensure human rights are prioritised in the design and governance of these technologies (Australian Human Rights Commission, 2018a). In New Zealand, the Ministry of Social Development's formation of the Privacy, Human Rights and Ethics framework (PHRaE) provides a set of tools that users of information can utilise to ensure initiatives don't breach clients' privacy or human rights (Sepuloni, 2018).

4.2.2 The right to equality before the law

AI is being used in criminal justice settings for a variety of purposes. For example, the COMPAS tool issued by some US court systems to help judges in determining questions about bail and sentencing. However, concerns have been raised about human rights implications of using AI in these settings. While human decisions also have the potential to contain bias, depending on how AI systems are developed and deployed, those tools can reduce, reflect or exacerbate bias (Angwin, Scheiber and Tobin, 2017). Researchers note that 'bias in automated decision systems can arise as much from human choices on how to design or train the system as it can from human errors in judgment when interpreting or acting on the outputs' (Reisman et al., 2018). While many researchers point to this risk of implicit or explicit bias in algorithmic decision making and machine learning (ML), humans can be equally biased as the AI replacing them. As a result, these structures of bias and discrimination can be replicated in training or programming the AI and in making final decisions involving human oversight. Biases founded on racial or other protected attributes in decision making may impinge on the right to a fair hearing.

4.2.3 The right to privacy

Article 17 of the International Covenant on Civil and Political Rights provides that no one shall be subjected to arbitrary or unlawful interference with their privacy, family, home or correspondence, nor to unlawful attacks on honour and reputation ('International Covenant on Civil and Political Rights', 1966: art 17).Technological developments present the opportunity to both enhance and challenge privacy and surveillance. For example, the use of AI in healthcare or AI enabled assistive technologies may provide vulnerable populations with greater autonomy, privacy and the reduced need for human intervention that could otherwise impact on their right to privacy. However, the right to privacy is becoming increasingly hard to protect due to the ease and power of collection, distribution and analysis of information – especially personal information – enabled by new AI technologies. In particular, AI offers new tools for surveillance technologies that may be deployed by government and non-government bodies. For example, AI-powered facial recognition technology is powerful technology that can have applications in identifying victims of child-sex trafficking, however if misused facial recognition can impinge on an individual's privacy as well as a range of other human rights (de Hert and Christianen, 2013; Cannataci, 2018). This can include the right to hold opinions ('International Covenant on Civil and Political Rights', 1966: art 19); peaceful assembly ('International Covenant on Civil and Political Rights', 1966: art 21); liberty and security of person, and protection from arbitrary arrest ('International Covenant on Civil and Political Rights', 1966: art 9); freedom of movement ('International Covenant on Civil and Political Rights', 1966: art 12); freedom of thought, conscience and religion ('International Covenant on Civil and Political Rights', 1966: art 18); and equality before the law ('International Covenant on Civil and Political Rights', 1966: arts 14, 26).

Broader concepts around AI and privacy and surveillance, beyond the human rights framework, are discussed in 4.5.

4.2.4 The right to freedom of expression

Everyone has the right to hold opinions without interference and the right to freedom of expression, including the freedom to seek, receive and impart information and ideas of all kinds ('International Covenant on Civil and Political Rights', 1966: art 19). AI tools may be used to influence or manipulate social media newsfeeds (see for example, Bastos and Mercea, 2017; Office of the Director of National Intelligence, 2017), advertising and search engine results (Birnbaum and Fung, 2017). Such interference can significantly impede the enjoyment of this right, as freedom of expression includes the free exchange of ideas and information.

4.2.5 The right to benefit from scientific progress

As stipulated by the International Covenant on Civil and Political Rights, all persons have the right to enjoy the benefits of scientific progress and its applications ('International Covenant on Civil and Political Rights', 1966: art 15b). While scientific progress has demonstrated the potential for harmful and negative impacts, technological progress frequently results in myriad of opportunities. AI can provide many benefits to people – for example, it can improve the enjoyment of the human right to life and access to health (Cohn, 2017; Lonstein, 2018). Human rights law in Australia requires States to take appropriate steps to ensure that all sectors of the community benefit from these applications of AI.

Some AI technologies provide significant benefits to people with disability. For instance, AI could advance the rights of people with a disability and foster greater accessibility and inclusion through AI powered assistive technologies such as speech to text technology. However other AI technologies are currently inaccessible for people in this cohort (Senate Community Affairs References Committee, 2017). Similarly, while children and young people face fewer difficulties using technology, they are particularly vulnerable to the potential harm of new technology, such as a breach of privacy or exploitation, made possible by the use of social media platforms (Australian Human Rights Commission, 2018b).

Further, women's economic and other opportunities may be compromised through the disparity in global access to technologies (United Nations Women, 2017).[3] To ensure that access to the benefits of AI technologies is universal and inclusive, tools and approaches need to be developed to address the issues new technologies raise for specific groups (Australian Human Rights Commission, 2018b).

[3] For example, the report notes that the global internet user gender gap has grown from 11 percent in 2013 to 12 percent in 2016. Over 1.7 billion women do not own mobile phones. Women are on average 14 percent less likely to own a mobile phone than men, which translates into 200 million fewer women than men owning mobile phones. When women do own mobile phones, they use them less frequently and intensively than men, especially mobile internet.

4.2.6 The right to life

Every human being has the right to life ('International Covenant on Civil and Political Rights', 1966: art 6). Individual AI-powered technologies can themselves both harm and promote the right to life. For example, unmanned aerial vehicles, also known as drones, can be used as lethal autonomous weapons, conversely they can also be used to transport vital medical supplies to hard-to-reach places (Cohn, 2017). Additionally, AI can promote the right to life through more accurate and targeted use of machinery, such as in medical diagnostics.

4.2.7 The right to work

All people have the right to work, which includes the right to the opportunity to gain their living by work which they freely choose or accept ('International Covenant on Economic, Social and Cultural Rights', 1966: art 6). The Australian robotics industry benefits our economy by employing almost 50,000 people and generating revenue of A$12 billion (Australian Centre for Robotic Vision, 2018). Conversely, AI automation technologies have the potential to displace an estimated 3.5 million workers in Australia in coming years (AlphaBeta, 2017). Estimates of the impacts of AI on workforces across the globe vary widely. As discussed in Chapter 3, many claim that AI will not lead to mass unemployment, but rather that it will transform the tasks involved in work, create new roles and establish new skill sets. Some job types and socio-economic groups are more likely to be adversely affected through increased automation of tasks. The consequences of widespread automation are likely to be different for women and men, with implications for socio-economic equality and the global gender gap (Schwab, 2017).

4.2.8 The right to security

Given the many decisions that governments must make, there are efficiency gains in using AI in decision-making processes. Some of these decisions may be in areas which particularly concern vulnerable people, such as in determining eligibility for, or compliance with, government assistance programs. This creates risk that some of the limitations of AI (or of poorly designed AI systems) – especially algorithmic bias – could lead to infringements of human rights.

For example, in the US in October 2013, the Michigan Unemployment Insurance Agency (UIA) launched an automated information system, *Michigan Integrated Data Automated System (MiDAS)*, to detect claimant fraud and pay unemployment insurance benefits to eligible claimants. The aim of MiDAS was to increase data accuracy, improve data security and privacy, reduce operating costs through automation, improve integration of organisational functions, and improve customer service (Office of the Auditor General, 2016).

Michigan state officials reported that during its 22 month operation, MiDAS had a 93 percent error rate when its attempts to identify unemployment insurance fraud weren't reviewed by humans. This resulted in over 20,000 claimants being falsely accused of fraud and subjected to fines between US$10,000 and US$50,000 (Claburn, 2017a; Claburn 2017b; Egan, 2017). There are also reports that the system was insecure – between October 2016 and January 2017, companies using MiDAS could reportedly see the names, social security numbers and wages of people whose payroll was managed by any of the 31 third-party vendors that worked with the UIA (Claburn, 2017b). Automated decision systems are a likely outcome of the implementation and application of AI. It will therefore be important that the application of AI in social and economic decision-making does not infringe the right to social security in the future.

The use of AI to create weapons introduces challenges to the right to security and the international humanitarian rules of war. Lethal autonomous weapons systems, such as drones and submarines or other weapons, can be programmed to act individually or in groups. These developments have raised serious concerns, leading to the establishment of a United Nations expert working group to consider the place of lethal autonomous weapons systems in the context of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be 'excessively injurious' or to have 'indiscriminate effects'. Similar concerns among non-governmental organisations has led to the establishment of a global coalition and the 'campaign to stop killer robots', which aims to ensure human control of weapons systems.

Discrimination and other human rights violations are not only unlawful – they also undermine public trust and can result in pre-emptory calls for regulation or reduced uptake of new technologies.

4.2.9 Protections needed – a human rights approach

Society must be aware of the both the risks and opportunities that AI and related technology pose to human rights. AI technologies present opportunities economically, socially and in the protection and fulfilment of human rights. Human rights protections can help build the community trust that will be needed to take these opportunities.

It is important to ensure that human rights are adequately protected and promoted in the context of new technologies. However, there are likely to be a number of acceptable ways to ensure that those developing and deploying new technologies, including those incorporating AI, do so in a manner that respects, protects, and fulfils the human rights of affected people.

4.2.10 Human rights in Australia

In order to adopt AI technologies in an equitable manner, it will be necessary to consider the values that are central to Australian society. Human rights provide a framework that could underpin the equitable implementation of AI, and Australia has ratified several major international human rights treaties.[4] As a party to these treaties, Australia has agreed to respect, protect and fulfil the human rights obligations contained in them. There are a number of mechanisms in Australia which to some degree protect and promote human rights.

Incorporation into domestic law

In order for international human rights law to have full legal effect in Australia, the relevant parliament or parliaments must incorporate the specific provisions of these laws into domestic law. Australia has incorporated some, but not all, of these international human rights treaty obligations into domestic legislation.

Federal law prohibits discrimination on the basis of race, disability, age, sex, sexual orientation, gender identity and some other grounds.[5] The *Privacy Act 1988* (Cth) protects primarily information privacy ('Privacy Act

[4] In addition to the UDHR, International Covenant on Civil and Political Rights (ICCPR) and International Covenant on Economic, Social and Cultural Rights (ICESCR) referred to above, these treaties include: *International Convention on the Elimination of All forms of Racial Discrimination*, opened for signature 21 December 1965, 660 UNTS 195 (Entry into force 4 January 1969); *Convention on the Elimination of all Forms of Discrimination Against Women*, opened for signature 18 December 1979, 189 UNTS 1249 (entered into force 3 September 1981); *Convention on the Rights of Persons with Disabilities*, opened for signature 13 December 2006, 2515 UNTS 3 (entered into force 3 May 2008); *United Nations Convention the Rights of the Child* UN GA Res 44/25 (20 November 1989).

[5] See, especially *Racial Discrimination Act 1975* (Cth); *Sex Discrimination Act 1984* (Cth); *Australian Human Rights Commission Act 1986* (Cth); *Disability Discrimination Act 1992* (Cth), *Age Discrimination Act, 2004* (Cth); *Fair Work Act 2009* (Cth).

1988 (Cth)'). There are also parallel state and territory laws that deal with discrimination and privacy. Two jurisdictions, Victoria and the Australian Capital Territory, have statutory bills of rights (Australian Capital Territory Legislative Assembly, 2004; Victorian Government, 2006).

Executive bodies

Australia has executive bodies that are responsible for promoting and protecting human rights. The Australian Human Rights Commission has primary responsibility in this area, including through a conciliation function (in respect of alleged breaches of federal human rights and anti-discrimination law), education and policy development (Australian Human Rights Commission, 1986 (Cth): s11). There are also specialist bodies with regulatory and broader functions in respect of specific rights. For example:

- the Office of the Australian Information Commissioner is responsible for privacy and freedom of information, and has regulatory functions regarding privacy (Australian Government, 2018h)
- the Office of the eSafety Commissioner is responsible for promoting online safety, with regulatory functions regarding cyberbullying and image-based abuse (Australian Government, 2018i)
- the Australian Competition and Consumer Commission (ACCC) is the national competition and consumer law regulator (Australian Government, 2018b) and is currently investigating digital platforms and their impact on Australian journalism (Australian Competition and Consumer Commission, 2018).

UN review processes

The Australian Government reports on the nation's compliance with human rights obligations through UN review processes. Some international bodies can hear complaints from a person in Australia alleging that the Australian Government is in breach of its obligations under one of its treaty commitments. In addition, the UN conducts its own investigations, and reports on human rights conditions in countries including Australia. These international processes generally make recommendations or findings that are not enforceable.

4.2.11 Opportunities for protection of human rights

The mechanisms discussed help guard against some of the potential adverse effects of AI technologies on the rights of citizens. However, there remain gaps in how this system promotes and protects human rights in the context of AI. For example, decisions that are wholly or partly informed by AI systems fall outside the scope of traditional forms of regulation for science and technology (Metcalf, Keller and Boyd, 2016; Conn, 2017; Vijayakumar, 2017).

As noted previously, there are numerous examples of how decisions arising from AI systems can lead to infringements of human rights. Where information about algorithms, datasets and resultant decisions is not available or comprehensible, it is difficult to ensure accountability for affected people (Australian Human Rights Commission, 2018b).

If a process of decision making is opaque, it can be difficult or even impossible to determine whether an impermissible consideration – such as one that is racially biased – has been taken into account (Australian Human Rights Commission, 2018b). Decision-making systems that rely on AI can be particularly susceptible to this problem. To ensure the protection of fundamental rights, the Australian Home Affairs Department has adopted a 'golden rule' process by which decisions that have adverse outcomes, such as visa refusals, will be determined by a human and not a machine.

Australia and other jurisdictions have started to grapple with these challenges, such as the recent announcements to develop a technology roadmap, standards framework and a national AI ethics framework to identify global opportunities and guide future investments (Australian Government, 2018c). The Australian Human Rights Commission is exploring the opportunities and challenges in relation to human rights and emerging technologies. Scheduled for release in 2019, the Australian Human Rights Commission report findings and recommendations seek to ensure the protection of human rights in Australia. In New Zealand, the Law Foundation's Information Law and Policy Project [ILAPP] will develop law and policy around IT, data, information, artificial intelligence and cyber-security. The project brings together experts to examine challenges and opportunities in areas like global information, cyber-security, data exploitation, and technology-driven social change (The Law Foundation New Zealand, 2018).

Other jurisdictions and technology companies have begun to approach some of the challenges posed by AI technologies through initiatives and forms of governance and codes of practice. For example:

- The EU's General Data Protection Regulation (GDPR) harmonises data protection laws across the EU and includes provisions relating to the transfer or export of personal data outside the EU – something that will influence how AI can be used on transnational datasets (European Commission, 2018a). The GDPR also imposes restrictions on how decisions based on automated processes may be made where they have significant effects on an individual
- New York City's Automated Decision making Task Force is examining the use of AI through the lens of equity, fairness and accountability and recommends redress options for people who are harmed by agency automated decisions (City of New York, 2018)
- The UK's proposed 'AI code', to be developed across the public and private sectors, including the Centre for Data Ethics and Innovation, the AI Council and the Alan Turing Institute, could provide the basis for future statutory regulation (Artificial Intelligence Committee – House of Lords, 2017)
- The European Commission's European Group on Ethics in Science and New Technologies has called for a common, international ethical and legal framework for the design, production, use and governance of AI, robotics and autonomous systems (European Group on Ethics in Science and New Technologies, 2018)
- Co-led initiatives by industry, NGO and academia to guide and frame AI ethical discussions include Open AI (Open AI, 2018) and the Partnership on AI (Partnership on AI, 2018).

Australia has the opportunity to develop codes of practice, guidelines and frameworks that reflect Australian values and incorporate standards of equity, inclusion, and human rights. However, global cooperation and input will also be required to accommodate the inherent international nature of technological development.

Co-regulatory and self-regulatory approaches

In addition to ordinary legislation, self-regulatory and co-regulatory approaches can promote and protect human rights in the context of new technologies. These approaches can include accreditation systems, professional codes of ethics and human rights-compliant design. These types of measures are generally led by industry participants and subject-matter experts. They may also influence the actions of manufacturers through the procurement process (Australian Human Rights Commission, 2018b).

An example of a self-regulatory approach is the proposed cross-sector ethical 'AI code' in the UK, which would require the establishment of ethical boards in companies or organisations that are developing or using AI (Artificial Intelligence Committee – House of Lords, 2017).

Responsible innovation organisation

Gaps in regulation of aspects of AI technologies, especially AI-informed decision making, are a cause for concern as automated and AI-informed decision-making systems become more widespread. Discrimination is both more likely and of greater consequence for those already marginalised (Eubanks, 2018b). Further, the often undisclosed algorithms employed in these systems (World Economic Forum, 2018b) are challenging the concepts of procedural fairness in decision making. It is essential that the public has trust in the systems and processes employed in the decisions that affect their lives. Discriminatory practices in AI may also prevent people from embracing the positive outcomes from AI-informed ML. An independent body could be established to provide leadership in the responsible development of AI technologies. The potential for an independent body is discussed later in the chapter.

4.3 Equity of access

The development and deployment of AI technologies in accordance with the human rights framework offers one way by which to ensure the safety of AI systems and guarantees Australia's adherence to international obligations. However, to ensure that AI delivers benefit to the entire Australian community, the implementation of AI should be underpinned by broader principles of equity and inclusion. In this way, AI could be designed to reflect Australian values, including that of a 'fair go'.

4.3.1 Economic and social inequality

As discussed in Chapter 3, AI technologies are likely to alter the nature of employment. While these changes were outlined in detail, the following section addresses potential inequality resulting from employment changes. Indeed, a primary concern associated with the emergence of AI is the potential for it to lead to rising inequality. Inequality broadly refers to unequal outcomes, rights or opportunities. Economic inequality concerns the unequal distribution of economic resources between, and within, groups of individuals, firms, industries and economies. Social inequality refers to the unequal distribution of resources and opportunities through norms of allocation that engender specific patterns of socially defined categories, such as race, gender or sexual orientation.

The World Economic Forum identified income and wealth inequality as the biggest global risk resulting from the adoption of AI technologies (World Economic Forum, 2017). While technological progress is one of several factors that affect inequality, concerns relate to the equitable distribution of benefits derived from the development of AI.[6] Given the uncertain nature of AI technologies and capabilities, there are divergent views on how AI will affect economic and social inequality. However, sharing the benefits of growth and equality of opportunity are important factors in ensuing social cohesion. This is as relevant for Australia and New Zealand as for all national economies.

4.3.1.1 State of inequality in Australia and New Zealand

The state of economic inequality in Australia is complicated and subject to interpretation. This reflects the complexities of inequality measurement and evaluation. From a national average perspective, income and wealth inequality have remained relatively constant over the past few decades (Australian Bureau of Statistics, 2017; Wilkins, 2017). According to the Gini coefficient, a common measure for income inequality, Australia ranks 24 and New Zealand 29 of 38 OECD countries (OECD, 2018a).

However, despite these averages, there are signs of rising economic inequality. For example, the top one and ten percent of income earners have commanded consistently higher shares of national income in Australia since 1980 (World Inequality Database - Australia, 2018). There is increasing wage growth for higher earners with higher levels of education, compared to middle and lower-income earners with lower levels of education (OECD, 2017a). Data reveals 13 percent of Australians and 14 percent of New Zealanders from 0-17 years of age live below the poverty line (OECD, 2018c). These levels of inequality are more likely to affect particular groups of the population, such as Indigenous peoples and people with a disability. The increasing adoption of AI is likely to affect these inequality distributions. If AI fulfils its projected economic impacts, it will certainly have structural effects on the Australian and New Zealand economies. The way in which the benefits are distributed will influence economic and social inequality outcomes.

4.3.1.2 Relationship between technology and economic inequality

Past experiences have demonstrated the disruptive force of technological and organisational innovations on the social and economic order (Schumpeter, 1975). Overall, technological progress has reduced inequality by lifting productivity, expanding the demand for labour, and increasing income, wealth and quality of life (Mokyr, Vickers and Ziebarth, 2015). However, this progress was not immediate and often required more than 50 years for economies to adjust and widely diffuse its applications (Jovanovic and Rousseau, 2005: 3-5). Therefore, the short-run disruptions of transformational technologies have caused profound structural changes to labour markets and economic activity. These initial decades have typically required significant labour transitions and have contributed to widening short-run inequalities (Bruckner, LaFleur and Pitterl, 2017). Similar to other general purpose technologies, such as electricity and personal computers, the impacts of AI are likely to be a continuation of this 'short-term pain for long-term gain' trend. These new technologies improve productivity for industries, populations and individuals to varying extents, which skews the distribution of benefits to those with the skills to make productive use of them (Milanovic, 2016). As a result, wage premiums are earned by those with the skills that complement these technological

[6] Other significant factors include economic performance, labour conditions and employment growth, education and training programs, minimum wage policies, taxation and redistribution policies, and trade and globalisation.

changes, which can cause or exacerbate economic inequality. Additionally, as the share of income shifts from labour to capital, tax collection also becomes more difficult for governments, which can strain public revenues (Abbott and Bogenschneider, 2017).

4.3.1.3 Risks of AI to economic inequality

AI represents a potential departure from other general purpose technologies due to the scope of capabilities, the speed of development and the scale of impact. Unlike traditional technologies, AI has the capacity to perform non-routine tasks that would otherwise require human cognition. Technological automation has traditionally occurred in areas of routine and manual labour because these tasks are relatively simple to codify. AI expands the scope of automation to include cognitive and non-routine tasks. The multi-use capabilities of AI techniques have developed rapidly over the past two decades, and development continues to accelerate (Shoham et al., 2017). This positions the economic impact of AI to be one of the most significant in the history of general purpose technologies (Brynjolfsson and McAfee, 2016). Therefore, the implications of AI on inequality should be examined according to the degree of structural changes in the economy. Among the most important is the impact that AI will have on labour demand.

Inequality from automation, the increasing skills divide and employment polarisation

There is the potential for new class division as a result of integrating AI into economic systems (Harari, 2017). Such a divide could emerge between people who are able to adapt to new techno-social transformations and those who might be left behind. New forms of exploitation, class disparities, and social exclusion could develop. Inequalities of wages emerge as the demand for labour skills that complement new technologies increase and attract wage premiums. For example, skills that are non-routine and cognitive, such as abstract thinking in ML development, benefit from advances in AI due to strong complementarities between routine and cognitive tasks (Autor, 2015). This raises the productivity and demand for workers with complementary skills to technology, thus driving up their wages.

However, typically these skills, and subsequent wage premiums, disproportionately favour the highly educated. This is problematic for inequality because jobs demanded by AI are likely to require higher skills and different mindsets, which could be difficult for many workers to develop. For example, it has been suggested that 36 percent of all jobs across all industries will require complex problem-solving skills by 2020, compared to 4 percent of jobs where basic physical abilities are a core requirement (World Economic Forum, 2016). Indeed, AI technologies are more likely to replace, rather than augment, routine tasks. Such tasks are disproportionately found in low-skilled to middle-skilled occupations with lower levels of education (Frey and Osborne, 2017). Therefore, these low-skilled and middle-skilled jobs, which are already missing out on the wage premium, are also more exposed to AI-enabled labour automation and shifts in skill demands (Bakhshi et al., 2017).

The effect of this increasing displacement of low-skilled and medium-skilled labour is referred to as 'employment polarisation'. This is where labour supply becomes concentrated at either ends of the skill spectrum, which can obstruct upward social mobility (Santos, 2016). If employment polarisation worsens, there will be fewer opportunities for people to climb the 'skill ladder', as the middle-skilled rung is weakened or shifted. Not all workers will have the training, skills or safety-nets to successfully transition into the new jobs created by AI. Additionally, in some instances, it will not be economically efficient to replace certain low skilled, low paid workers with AI given the costs associated with the technology. These factors can also result in the widening of income inequality. Strategies in response to workforce changes should therefore include helping displaced workers to train and acquire new skills. Upskilling will be necessary to ensure that some groups of people, lacking in the right education and special skill sets, are not disadvantaged by technological developments.

Young people from low socio-economic backgrounds, as well as those from rural and remote areas, are more likely to choose vocational courses rather than university education, and are more likely to end up performing routine low-skill jobs (Tomaszewski, Perales and Xiang, 2017). Even if enrolled in university, women and students from disadvantaged backgrounds are less likely to study STEM subjects. There is therefore a need for continued investment in providing equitable access to quality education to avoid the marginalisation of people from disadvantaged backgrounds in future labour markets.

In some occupations, it is likely that humans will work with smart machines, rather than be replaced by machines; automation of routine tasks will allow professionals to undertake more complex cognitive and creative tasks. Intelligent, creative, and emotional skills are still considered non-replaceable by machines, and therefore secure in the wake of the digital revolution. Non-routine, uniquely human skills that focus on care, creativity and human consciousness will continue to be essential within the workforce. By combining technical and interpersonal tasks in occupations, these people may become the 'new artisans' (Katz and Margo, 2014) of the new age. Future workers who are able to use these skills and deliver specialised services based on these skills are likely to remain competitive in the job market.

Opportunities for future work

Workforce issues do not pertain to a scarcity of work, but to the distribution of work among the population (Autor, 2015). In addition to labour market risks, there are opportunities to reduce economic and social inequality through the reimagining of the future of work. Paid work is becoming a less reliable and useful method for distributing wealth (Dunlop, 2016); a situation that will be further enhanced as technology continues to improve. AI could be used to redistribute work across the population, and therefore redress disparities in unemployment (Spencer, 2018). AI and automation afford new possibilities to extend creative activities in work. In this regard, less and better work can become a reality with the use of AI technologies (Srnicek and Williams, 2016).

Looking to the service sector as an example of future opportunities, in countries such as the US and Australia it has become the largest section of the labour market. The service sector includes banking, finance, tourism, hospitality, healthcare and social services. Technological changes that have generated new methods of service delivery and rising household incomes have aided this expansion. Some predictions indicate that the service sector will continue to play an important role in the economy and there is potential that, as with previous technological developments, the sector will have the capacity to absorb displaced workers as new occupations are created.

4.3.1.4 Increasing marginalisation

The use of models and algorithms has the potential to further contribute to the marginalisation of already vulnerable population groups. Unjust outcomes can be produced as a result of human decisions regarding both model design and implementation.

Bias can also occur in contexts wherein statistical models are used with the aim of reducing the influence of bias in human decision makers. For example, statistical recidivism models aim to reduce the influence of bias (whether explicit or implicit) of judges in the sentencing of crimes. However, these models often reproduce and disguise bias (see Angwin, Larson, Mattu, & Kirchner, 2016; Austin, 2006; Kehl et al., 2017; Labrecque, Smith, Lovins, & Latessa, 2014; Lum & Isaac, 2016; Prince & Butters, 2014; Vrieze & Grove, 2010). Data used in these models are typically obtained from questionnaires completed by perpetrators and often include details about upbringing, family, social connections, geographical location and proximity to other offenders. This is data that should be irrelevant to a perpetrator's sentencing. Nevertheless, these data are used to generate a recidivism score for sentencing decisions, which influences the outcome in a way it should not, and for minority groups often results in a high recidivism score. Statistical models can also be used for hiring to help reduce unconscious bias, however further marginalisation can occur when statistical models are used to inform hiring (and sometimes firing) decisions (Barocas & Selbst, 2016a; Hu & Chen, 2017). Statistical models are typically used for hiring and firing for lower paid jobs, whereas hiring (and firing decisions) are often made on the basis of personal judgment in higher paid jobs. This therefore generates the situation in which people from already marginalised groups in society can be further marginalised by statistical models.

Statistical models have been previously used to inform decisions in important arenas, such as the use of credit scores by financial institutions. However, the statistical models used by AI systems may differ as a result of widespread application areas, leading to limited capacity for explainability and reduced oversight.

4.3.1.5 Social implications of rising inequality

In Australia, it has been estimated that automation could add A$2.2 trillion to annual income by 2030 (Marin-Guzman and Bailey, 2017). However, if trends continue, the majority of these gains will not be returned in the form of increased wages and conditions. If economic inequality were to rise due to the effects of AI, the growth of AI could be inhibited and the risks of social fragmentation could increase. In scenarios where workers are displaced by AI, and they do not receive adequate transition support or subsistence compensation, those affected could oppose AI developments (Korinek and Stiglitz, 2018: 3). If a large part of the population does not economically benefit from the growth of AI, it is rational that they would defend their economic position. This rejection of modernity could compromise social and economic development. As a result, similar to the case with other disruptive technologies, AI could be less likely to be adopted and diffused throughout the economy, which would hamper economic growth and fuel political discontent.

Workers who are more likely to be adversely affected by AI are also more likely to experience inequality due to lower levels of education. It is therefore critical for the benefits of AI to be distributed equitably. Unless this is achieved, AI threatens to perpetuate entrenched disadvantages, which is harmful economically and socially.

4.3.1.6 Mitigating the rise of inequality

Public institutions play an important role in determining the market structures affecting economic distribution. This role requires that innovation is encouraged, while ensuring the equitable distribution of benefits. In the context of AI and inequality, policymakers have a range of mechanisms they can call upon, such as:

- Taxation and redistribution: Applying effective tax and redistribution systems to ensure that the surpluses earned by innovators and investors help to support those inadvertently impacted by AI. This is typically performed through progressive taxation and transfers, which provides workers with subsistence compensation during periods of employment transition
- Infrastructure: Effective digital infrastructure that helps to diffuse AI equitably, such as 5G mobile networks and standards that foster open-data sharing. Infrastructure, such as internet connectivity and access to digital devices, provides the backbone for the diffusion of AI. In a country as large and dispersed as Australia, ensuring equitable access to critical infrastructure affects the extent of benefits that AI provides, particularly for rural and remote populations
- Antitrust policies: Regulating anti-competitive behaviours by ensuring that companies do not stifle market competition and exhibit 'rent-seeking' behaviours that adversely affect innovation and the consumer
- Intellectual property rights: Creating incentives for companies to innovate by granting patents, but also ensuring that these exclusive rights do not unfairly block barriers to market entry
- Education and training: Investing in the development of high-demand skills for youth and targeted worker transition programs to assist people whose jobs have been displaced by AI. Recent work suggests that, while STEM skills will be important for those developing AI, HASS skills will be equally important for the larger group of people in other occupations in an AI-enhanced world (see for example Royal Bank of Canada, 2018: 12). As such, educational programs should ensure that students receive an adequate combination of training from both HASS and STEM disciplines
- Minimum wage: Helping to tackle poverty and alleviate precarity as a result of casual, part-time, and 'gig-based' employment
- Public research: In parallel with effective antitrust policies, public research can help reduce the scope for monopolies that capture large portions of innovation returns. Innovations that are funded by public expenditure can be owned by the State and achieve market returns that contribute to public revenue, such as the CSIRO Wi-Fi patent (CSIRO, 2015b).

Given the rapid development of AI and automation, public policy and administration will require mechanisms for responding to the new risks and opportunities associated with AI and automation. Within the public sector, there exists scope for decision making by data-driven AI systems which should be implemented with care with respect to algorithmic fairness. Across government and non-government agencies, large, multiple databases are matched and mined to produce new understanding of service users and activities (Gillingham and Graham, 2016: 135). The provision of social services driven by data-powered algorithmic decisions can pose challenges for already vulnerable populations (McClure, Sinclair and Aird, 2015: 128).

AI-powered technologies represent potential for significant widespread benefit. To capitalise on these benefits and ensure equitable outcomes, it is necessary to acknowledge the limitations and uncertainty associated with AI technologies (O'Neil, 2016: 208). The use of these technologies can reinforce prejudice and inequality resulting from the simplification of complex issues. Ensuring procedural fairness and digital inclusion will require that the economic progress enabled by AI developments is shared equally.

Universal basic income

Should AI result in the displacement of workers, a redistribution of the economic gains derived from AI technologies could be considered to ensure social equity. A universal basic income could provide this method of wealth redistribution. The universal basic income has been suggested as one means to provide economic security at a time of economic uncertainty and as a way of providing an economic floor as workers experiment with new forms of income generation in the so-called 'gig-economy' (Mays, Marston and Tomlinson, 2016). A universal basic income is an unconditional regular payment.

It has been suggested that a universal basic income could be implemented in a developmental process, which focuses on the 'basic' component of the income, rather than the 'universal' (Quiggin, 2017). This could be achieved through the initial introduction of a full universal basic income payment to selected, vulnerable populations. Subsequently, payment recipients could gradually increase until full universality is achieved. The estimated cost of everyone in Australia receiving a full basic income is around 5-10 percent of GDP (Quiggin, 2017). Integrating the universal basic income with the tax system may allow for a cheaper model of universal basic income. There are industries and countries experimenting with initiatives that reduce working hours but do not reduce income. These initiatives will serve as significant platforms for evaluation. Given that part-time and low-paid workers are predominantly female, a reduced working week with an adequate income could enable a more gender-equal distribution of wage work (Rubery, 2018).

4.3.1.7 Summary – equity of access

The use of AI technologies presents opportunities for future societal benefit. Developmental decisions will shape the way in which AI delivers these benefits. This is highlighted by Schwab (2017: 174):

> 'Neither technology nor the disruption that comes with it is an exogenous force over which humans have no control. All of us are responsible for guiding its evolution, in the decisions we make on a daily basis as citizens, consumers and investors. We should grasp the opportunity we have to shape the Fourth Industrial Revolution and direct it toward a future that reflects our common objectives and values.'

Societies have regularly adapted to industrial and labour transformations from previous general purpose technologies (Bresnahan and Trajtenberg, 1995). Therefore, consideration should instead be given to the types of new skills and jobs demanded by AI, how to equip people with these skills, and the implications on inequality if the labour market is slow, or fails, to transition to meet these new economic demands. A comprehensive and continual understanding of the changes taking place in the labour market, now and into the future, will be important for developing appropriate policy responses to deal with these changes. Public policies will play an important role for ensuring that the benefits of AI are not unreasonably concentrated or reinforce existing inequalities. While the developments of AI must be nurtured to help realise its potential, it should not be done by creating an unequal society.

The future direction of AI will be determined by human action within broader political and social frameworks. With collective vision, AI could facilitate a society that has increased leisure time in which 'familial, community, and creative development can flourish and replace our current society's incessant production and overwork' (Stubbs, 2017: 709).

4.3.2 Indigenous peoples

AI is, to an increasing extent, a part of the everyday lives of some Māori and Aboriginal and Torres Strait Islander peoples. In Australia, for example, Aboriginal technology entrepreneur Mikaela Jade is using augmented and mixed reality technologies to tell stories on country in Indigenous communities (Powell, 2018). In Aotearoa New Zealand, AI is being used for language revitalisation. Tribal radio stations Te Hiku Media are creating language tools that will enable speech recognition and natural language processing of Te Reo Māori (Collins, 2018).

It has been suggested that AI might be a less confronting notion from an Indigenous standpoint than it is from a western perspective (Black, 2018). Indigenous legal customs are determined by a sacred relationship between people and nature. This relationship shapes how people carry out their responsibilities and gain rights (Black, 2011). Therefore, the notion that there can exist a non-human decision-making system that knows us, possibly better than we know ourselves, is familiar to Indigenous peoples.

To realise the potential benefits of AI for Aboriginal and Torres Strait Islander peoples and Māori, it is necessary to consider the unique challenges and opportunities posed by AI systems across community groups. Aboriginal and Torres Strait Islander peoples and Māori are among the most disadvantaged in Australia and New Zealand, carrying the heaviest burden of disease, over-incarceration and broad-spectrum inequality. This is directly related to histories of colonisation and dispossession, as well as ongoing integrational impacts of social, cultural and political marginalisation.

AI decision-making systems have the potential to exacerbate these existing inequalities, if not developed with considerations of diverse Indigenous populations. For example, Aboriginal and Torres Strait Islander and Māori children and their families are disproportionately affected by the use of potentially biased algorithms in child protection. In New Zealand, more than half of children in state care are Māori even though they comprise only one-quarter of the child population (Office of the Children's Commissioner, 2015). In Australia, Aboriginal and Torres Strait Islander are nearly seven times as likely to be in state care as non-Indigenous children (Australian Government, 2017). The complex relationships between structural inequalities, ethnicity, patterns of system contact and system bias are still poorly understood (Keddell and Davie, 2018). Marked spatial differences in child protection substantiations relative to notifications suggests system bias is one of several explanatory factors at work. There are other signs of bias. For example, the overrepresentation of Māori children increases at each decision point within the child protection system, with 40 percent of children notified being Māori, increasing to 60 percent by the time decisions to remove children into foster care are made (Keddell and Davie, 2018). The following discussion considers the potential for AI technologies to contribute to inequality experienced by Indigenous people. Concerns specific to Indigenous people in relation to the collection and use of data will be discussed in Chapter 7.

AI technologies should be developed to safeguard against the entrenchment of such inequalities. Indeed, inclusive AI technologies may provide opportunities to address existing inequalities.

Harnessing the potential of AI for Indigenous peoples in Australia and Aotearoa New Zealand should be closely aligned with Indigenous leadership and Indigenous governance on the processes of how, when and in what circumstances these technologies are applied. It therefore follows that an important issue for Indigenous peoples is the extent to which AI impacts on their right to self-determination.

4.3.2.1 Self-determination

The United Nations Declaration on the Rights of Indigenous Peoples affirms the right to self-determination (article 3) and extends this right to self-government and autonomy in relation to internal and local affairs (article 4). In Australia, self-determination refers to inclusions in decision making for those affected by government decisions, and independent, territorial sovereignty (Ford, 2012). Self-determination with respect to AI may comprise Indigenous involvement in the design, use and implementation of AI technologies. Indigenous consultation is particularly important given that Indigenous peoples may have different priorities and needs associated with the use of AI. These requirements can be overlooked in systems and technologies that are solely focused on achieving efficient outcomes for the broader population.

This can be understood with reference to an example from New Zealand. Economists in New Zealand used large government datasets to develop a predictive algorithm for early intervention in child protection cases (Oak, 2016). An ethical review found that Māori people were disproportionally represented in the risk group. As a result, there was a risk that Māori people or communities might be subject to hyper-vigilance, including the removal of Māori children not at risk. Even if such a model were found to succeed in creating social benefits for the community (in this instance by mitigating child abuse), a Maori-centred approach should involve Māori at all stages 'from design to the follow-up of whānau [family/political unit] and the evaluation of the programme' (Blank et al., 2015: 10).

Where AI systems have not involved community input, some Indigenous people may prefer to opt out. However, the capacity to opt out may be limited when people could encounter disadvantages as result of this decision. For example, when AI systems are linked with public services associated with the delivery of healthcare or social and economic wellbeing, opting out of such a system may mean losing access to these services.

4.3.2.2 Digital inclusion

AI may be used in ways that may have benefit for Indigenous communities. For example, researchers are working with Google to build AI models that preserve Indigenous languages (Biggs, 2018). However, the extent to which people benefit from AI is dependent on the capacity to access digital technologies. Factors that could affect a person's capacity to use the digital systems and services underpinning AI include access limitations, costs associated with access and digital literacy abilities.

Measures of digital inclusion, such as the Australian Digital Inclusion Index, suggest that Aboriginal and Torres Strait Islander people access the internet less than the general population. While measures of digital inclusion have received criticism because, for example, a greater proportion of needs may be met through the examination of the way in which people use the internet rather than who uses the internet, such metrics continue to provide a basis for understanding population differences in access (Borg & Smith, 2018: 378). Although the number of Aboriginal and Torres Strait Islander people who access the internet (in non-remote areas) is increasing, differences in internet access and use remain; these differences have the potential to affect the extent to which AI services may be used. For example, Aboriginal and Torres Strait Islander people in non-remote areas are significantly more likely to use mobile-only internet services (Thomas et al., 2017). This restricts internet access to locations with mobile reception and requires the capacity to pay for mobile internet services. Regardless of available infrastructure, internet use can vary according to the social norms and choices of particular groups. In New Zealand, the Digital Economy and Digital Inclusion Ministerial Advisory Group seeks to reduce the digital divide. The 20/20 Trust is another New Zealand initiative which provides digital literacy programs.

It is likely that services will increasingly be provided online as more people access the internet. Those who remain without internet access (or with intermittent access) will experience difficulties as face-to-face services are removed or reduced. People excluded from accessing online services are most likely to be vulnerable and in need of social support services. However, the application and use of AI technologies have the potential to assist some people in accessing online services. Opportunities exist for AI to resolve access barriers related to digital skills, language or disability. For example, chatbots may be used in ways that overcome barriers associated with digital skills and abilities.

4.3.2.3 Summary – Indigenous peoples

As AI is developed, it is important to consider how Indigenous knowledge systems might inform their deployment, as well as how the governmental and philosophical implications are conceived. AI technologies may enable more appropriate services for Indigenous peoples, including services in language, or which accommodate group needs and norms in ways that those designed for the majority cannot. The decisions made as a result of ML may impact on individuals' agency. AI might be responsive to group norms in ways that existing technologies are not or generate supra-state governance through their decision-making abilities (Bratton, 2015). Much of the debate on the social and ethical implications of AI has been concerned with the quality of data and design. These issues will be further discussed in Chapter 6.

4.3.3 AI and inclusion

Emerging technologies can provide opportunities for greater inclusion and demonstrate potential to enhance the lives of people with a disability, older people, children, and others who experience social disadvantage. For example, technology may be able to replace the use of guide dogs, and development is underway for autonomous wheelchairs (Scudellari, 2017). To ensure a fair Australian society, everyone should be provided with opportunities for access and inclusion in addition to the freedom to opt-out of instances of social inclusion. However, there are challenges associated with AI technologies. Consideration should be given to the regulatory framework governing AI, its implications for the rule of law, and the inclusion of subconscious biases in data. While AI holds promises of life enhancing technology for those who might be considered disadvantaged, algorithmic bias could also reinforce certain disadvantages. The design of AI should be shaped by decisions associated with our desires for the future of society. AI systems

Box 18: Universal design in practice

Can AI be racist? This is a question asked in a Microsoft inclusive design team blog post (Chou, Murillo and Ibars, 2017). Microsoft and other software developers have been following inclusive design principles in the development of their software for some time. They found that by designing for the broadest possible number of users, they have created more accessible, convenient and useable programs and apps.

Microsoft states that its first inclusive design principle is to recognise exclusion and identify bias. They describe five biases: association, dataset, interaction, automation and confirmation bias. Microsoft has recognised that by identifying who is excluded rather than trying to include everyone has made the task easier. A similar approach was developed some ten years ago by the Inclusive Design Team at the Engineering Design Centre, University of Cambridge in the UK.

The inclusive design team, through the development of their inclusive design toolkit and their exclusion calculator, used population demographics and other factors to ascertain how many people will be left out of a design based on a particular level of ability such as seeing, hearing, lifting or grasping (Cardoso et al., 2007). For example, making something useable for people with poor grip strength (e.g. a lever handle) makes it easier for everyone – it does not exclude people with good grip strength.

should be developed to incorporate diverse human factors to ensure the benefits of AI are equitably distributed.

Inclusive design seeks to accommodate and involve those experiencing difference, disability or disadvantage (Center for Inclusive Design and Environmental Access, 2011). Although inclusive design is most commonly considered in the built environment, it is also considered in ICT, teaching and learning, service provision, written documents and in policy development (Centre for Excellence in Universal Design, 2018). Incorporation of inclusive design considerations during the development of relevant policies and AI technologies could address issues on trust, ethics and regulation. Public consultation in the development of inclusive AI design may help foster trust in the technology.

4.3.3.1 Defining AI and social disability

The social model of disability is the internationally recognised way to view and address disability. The United Nations Convention on the Rights of Persons with Disabilities sets the standard for approaches to disability. People with disability should not be seen as objects of charity, medical treatment and social protection. Rather, they should be regarded as subjects with agency, rights and obligations, capable of claiming those rights and autonomy, and navigate and participate in the world based on free and informed consent. The responsibility for inclusion does not lie with the individual with a disability but relates to the design of the environment.

4.3.3.2 Inclusive inputs

Big data underpins the functioning of AI systems. To date, concerns regarding big data have focused on the risks of inclusion – the threats arising from the collection, analysis and use of personal information. However, there is also the risk of an individual's data not being collected, or if collected, that it is dismissed as an outlier. The elderly, people with disability, or those who experience disadvantage, may be prevented from owning or engaging with the technology responsible for producing such data. This technology may be inaccessible or have prohibitive costs. Costs associated with technology are significant given the high poverty and unemployment rates affecting, for instance, people with disability (OECD, 2009; Australian Government, 2011). The subsequent outputs produced by AI systems using this incomplete data may have limited applicability to those users not represented within the initial dataset. This may be of particular importance for AI outputs that support medical or legal decisions.

People with disability are less likely to access the internet than people without disability. Only 60 percent of people with disability have home internet access, and they are 20 percent less likely to own smart-devices, home broadband, and a range of technology that is essential to the creation of data and the use of AI technologies (Australian Government, 2014a). Elderly Australians use the internet 50 percent less frequently than their younger counterparts and 98 percent of this internet use is within the home, creating less useable data (Anderson, 2015b, 2015a; Australian Government, 2016). An estimated one million Australians over 65 have never accessed the internet (Anderson, 2015b, 2015a; Australian Government, 2016). This means that the vast array of individual data is not collected.

This has significant economic consequences for the use of AI and big data in targeted advertisements, and trade and hiring decisions. There are also potential political problems resulting from the exclusion of minority representation in data, particularly when government uses data in political decision making (Australian Government, 2018j). Concerns about representation and exclusion are as inherent in technology as they are in traditional conceptions of political participation.

The reverse is also true; when data are collected, people with disabilities, culturally and linguistically diverse groups, women, and people who identify as LGBQTI, are at risk of being discriminated against (Danks and London, 2017; Knight, 2017). Recent exposure of the bias in COMPAS and PredPol demonstrates this. The proliferation of artificially intelligent female personal assistants entrenches gender bias (Stern, 2017), facial recognition threatens culturally and linguistically diverse groups (Bowles, 2016; Lohr, 2018; Shah, 2018), and chatbots can learn antisemitism, racism and misogyny in a single day (Mason, 2016). Some researchers caution that it may be impossible to create fairness and equality in algorithms (Miconi, 2017). As human creations, they are inevitably biased. There is the risk of 'automating the exact same biases these programs are supposed to eliminate' (Lum, 2017). This bias has resulted in the refusal of parole and disproportionate prison sentencing of culturally and linguistically diverse groups, the over-policing of neighbourhoods with large populations of culturally and linguistically diverse groups, the arrest of a Palestinian man over an incorrect Facebook auto-translation and the overrepresentation of Indigenous people on the NSW suspect targeting management plan (O'Mallon, 2017).

Research demonstrates the opportunity to include data outliers through the training of AI with messy data. The initial outcomes are encouraging, with AI taking longer in the initial processing phase but producing richer and more varied results. Additional work is being undertaken to change the shape of the bell curve to allow AI programs to read and understand the outliers as part of the dataset.

Box 19: AI and disability

In 2015, the Australian Bureau of Statistics reported that 18 percent of the Australian population identify as having a disability. It is generally agreed upon that the global rate of people with a disability is increasing due to age, new diseases and conflicts. AI presents unique opportunities and challenges for people with a disability. The development of appropriate policy and legislative frameworks could help to deliver benefits and protections to people with a disability. In conjunction with policy frameworks, the integration of AI technology could serve to decrease the unemployment rate for people with a disability; develop a more inclusive education system; promote access to existing and new media content, information and print publications; increase accessibility to consumer goods, computer and telecommunication technologies for all; and result in a more inclusive society.

Different disability groups present distinct needs and requirements with accessibility of information and technology. For example, people with a disability, particularly those who are blind or vision impaired, are not able to adopt computing and internet-related technologies at the same rate as the able-bodied population (Hollier, 2007). However, some AI-powered technologies, such as image recognition, are adopted at a faster rate by people with a disability. Over the next decade, it is likely that AI will provide significant opportunities for engagement and independence, particularly in the areas of mobility, home automation and information access. However, the design of user interfaces must support the relevant assistive technology used by people with a disability.

Often, commercial technologies can unintentionally include barriers to use. This has largely resulted from a lack of awareness, limited regard for accessibility and concerns about additional costs. In the consumer sector, touch screen technology can be a major barrier for people with a disability. However, mobile technology provides an example of the successful incorporation of inclusive design for disability requirements. Apple, Microsoft and Google have included accessibility and assistive technology into their operating systems.

AI-powered technologies that deliver data and respond to commands may constitute a form of assistive technology. AI provides support in a similar way to popular assistive technology software. For example, a screen reader may provide content to a person who is blind (Hennig, 2016). However, unlike assistive technology solutions, AI provides always-on real-time connectivity, which ensures that people can quickly and easily obtain assistance and support. Use of AI may improve quality of life and facilitate social and economic participation (Domingo, 2012). For example, self-driving car technology was used in conjunction with eye tracking and brain electrical activity sensors to develop self-driving wheelchairs and via the cognitive assistance project for visual impairment, vision-impaired users can see what is around them in context (Baker, 2014; IBM, 2015; Malewar, 2018).

While AI may provide benefits to people with disabilities, there are also possible risks of exclusion. These concerns pertain to interoperability, accessibility support, identification and configuration, privacy, and security and safety (Hollier and Abou-Zahra, 2018). For example, AI has been largely developed without consideration for the needs of people who are blind or vision impaired (Maguire, 2018). While there is design of AI technology specifically for vision impaired users, mainstream AI applications are largely developed separately (Maguire, 2018). As such, design considerations pertain to the use of systems by those with accessibility issues. An example of this includes the use of facial recognition and related biometric algorithms in airports, which may exclude blind people as their eyes may not be visible or not able to focus on facial technology. Accessible tools should be provided within industries responsible for the development of AI technologies. This will ensure that people with a disability will have equal opportunities in both the development and use of AI technology. Accessibility must be incorporated at each stage of the product life-cycle, not just at the end of the process. As a component of this, it is important to provide training and education on accessibility in education programs targeted at product management, design, development and marketing.

Accessibility is often considered to only benefit those with a disability. However, in practice, accessibility benefits a large fraction the community. For example, increasing text size on web pages up to 400 percent benefits people who have difficulty with small text and are not legally blind. Multimedia captions help those with English as a second language who find it easier to understand the written word.

For people with a disability to enjoy the benefits provided by AI devices along with protections relating to privacy and security, it is necessary to have effective legislative support and technical standards. Article 9 of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) states: 'To enable persons with disabilities to live independently and participate fully in all aspects of life, States Parties shall take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, to the physical environment, to transportation, to information and communications, including information and communications technologies and systems, and to other facilities and services open or provided to the public, both in urban and in rural areas' (United Nations, 2006). As a signatory to the UNCRPD, Australia has some policies but lacks specific disability-based legislation addressing ICT requirements for people with a disability. Issues of privacy, security or interoperability relating to AI present specific concerns to people with a disability. Australian policy or legislative framework does not account for these concerns. For example, technologies may be used by human support agents to obtain information on an individual with a disability without their knowledge. This could include financial information, personal habits, and other information that the user accidentally revealed.

Policies and legislative frameworks should be amended to provide greater support for people with a disability. Additionally, federal and state governments could provide private sector incentives to encourage accessible design of products and services. This approach could particularly incentivise small organisations to incorporate accessibility in products and services.

4.3.3.3 Inclusive design

Every design decision has the potential to include or exclude people. Understanding user diversity will result in maximised inclusion. Users vary in capabilities, needs and aspirations with differences in ability, language, culture, gender, age and other forms of human difference. While the concepts of accessibility, inclusive design and universal design are often intertwined, the goal is always the same – that is the human right to universal access. While the underlying principles of universal and inclusive design are virtually identical, the difference is a matter of perspective and source (May, 2018). Inclusive design seeks to expand the range and diversity of end users recognising that one size doesn't fit all. This notion is particularly suited to technological advancement.

Fundamental principles in inclusive design include:

- recognising diversity and uniqueness
- inclusive processes and tools
- broader beneficial impact.

It is important to incorporate a diversity of insights and voices in the design process. To achieve this participation, design and development tools must be accessible. Inclusive accommodation throughout the design process will ensure that the entire spectrum of users reap the benefits of the technology, and the inputs, processes, outputs and governance are inclusive and universal. In addition, attention to inclusive design will provide autonomy and dignity for people with disability and those who may otherwise experience exclusion.

4.3.3.4 Inclusive outputs

To ensure that AI outputs accommodate a wide variety of users and more users derive benefits, diversity should be recognised in both datasets and AI design. Recent debate over the emergence of My Health Record has shown that not only does the system risk marginalisation of people with disability, drug users, and sex workers, but the output itself – the self-managed health record – is largely inaccessible to people with disabilities, who perhaps could benefit most from the technology (Inclusion Australia, 2018).

There are many successful examples of inclusive AI; and many more examples where inclusively-designed AI challenges preconceived notions of violations of privacy and perpetuated prejudice. For example, AI can provide people with disabilities greater privacy.

- the use of AI in the design of an intelligent cognitive orthosis for people with dementia and Alzheimer's disease minimises the encroachment of a full-time care team into the personal life of a person with these conditions. This would provide dignity, autonomy and privacy to people with dementia and Alzheimer's (Mihailidis, Fernie and Barbenel, 2001; Mihailidis, Barbenel and Fernie, 2004)
- prototypes that provide for the monitoring of the onset of psychosis in people with schizophrenia promise greater autonomy and reduced intrusion into an individual's private life (Corcoran et al., 2018)
- speech recognition programs that identify non-standard speech allow for people with a spectrum of disability to access applications and technologies previously unavailable (Kewley-Port, 1999; Breakstone, 2017)
- Google's DeepMind creates closed captions and audio descriptions with greater accuracy than a person employed to do those tasks, at a markedly lower cost, increasing the access people with hearing and vision impairments have to media (Vinyals et al., 2017)

- Facebook is collecting data from its disabled-identifying users to address the issue of cleaned, absorbable datasets
- Microsoft has invested US$25 million in the development of accessible AI.

While these examples provide immediate benefits to people with disability, AI that is capable of lip-reading, non-standard speech recognition, and self-driving cars will benefit all businesses and people. While simple text on a webpage, lifts at train stations or an electric toothbrush may have been designed for disability, they benefit us all.

To ensure inclusive practice, it is necessary to address:

1. industry benchmarks and modelling
2. funding support for incubation, testing and piloting the resources and methodology in Australia and New Zealand, given our specific challenges including geographical reach, diverse populations and the Indigenous experience
3. a governance framework that combines technical understanding, community and industry

Box 20: Connected and automated vehicles

Automated vehicles present opportunities for inclusive design. Trials of driverless vehicles are being conducted with the aim of eliminating human drivers. Some of the technology is already available in newer model cars, such as automatic braking and assisted parking. It is claimed that road accidents will reduce significantly with the removal of human error. Autonomous vehicles have the potential to provide point-to-point transportation for people unable to hold a driver's licence. However, there must be attention to design to ensure the inclusive potential of autonomous vehicles. For example, visually impaired people may not be able to use touch screens. Voice activation is a problem for people who are non-verbal or hearing impaired. There is the question of addressing use of wheelchairs and baby strollers. Inclusive design principles can help with development of autonomous vehicles.

4.3.3.5 Governance and regulation

AI that is not inclusive, accessible, or universal poses potential threats to the democratic system and the rule of law. Democracy is predicated on a citizen's right to be freely informed and make political choices. Biased AI can produce social fragmentation, and the issue of reduced privacy, increased surveillance and targeted advertising all threaten this democratic principle of free and informed choice. The use of big data and various other inputs that exclude, marginalise, or pathologise minorities has significant legal ramifications. The use of big data underpinning AI is problematic with regards to the right to privacy and risks perpetuating imbalanced power structures. Consideration could be given to ethical and inclusive use of AI and data within the Age, Sex, Disability and Racial Discrimination Acts. Such regulatory considerations may include the erasure of data generated by people with social disabilities.

Achieving inclusiveness in AI is contingent on the establishment of governance and a regulatory framework. Inclusive design principles with oversight to ensure compliance will address many of the issues raised for those experiencing disadvantage and disability and will benefit the entire community. Methods of inclusive design can be refined by using technology to design for one extreme experience at a time and then including the next. It's an *and* rather than *or* model. While a number of inclusive design tool kits exist, none is unique to the Australasian situation, where geographic and population demographics, including our Indigenous populations, pose unique challenges and opportunities.

4.3.3.6 Summary – AI and inclusion

AI presents both challenges and opportunities. AI holds the promise of life-enhancing technologies that have the potential to broadly benefit the community. If programmed with inclusion, decisions made by AI systems could contain less bias than human decisions. The implications of exclusivity in the inputs, design, outputs, and regulatory framework have significant implications for the ability of people to use

Box 21: Case study: Using AI technologies to predict criminality

Research claims to have found evidence that criminality can be predicted from facial features. Xiaolin Wu and Xi Zhang (2016) describe how they trained a model to be able to distinguish photos of criminals from photos of non-criminals with a high level of accuracy.

However, Wu and Zhang's results can be interpreted differently depending on what assumptions are presented and what questions are posed. The authors make the assumption, contrary to overwhelming evidence (see for example Bobo & Thompson, 2006), that there is no bias in the criminal justice system. Consequently, Wu and Zhang assume that the criminals whose photos they used as training data are a representative sample of criminals in the wider population (including those who have never been caught or convicted for their crimes). The question posed by Wu and Zhang is whether there is a correlation between facial features and criminality. Given their assumption, the results suggest that there is such a correlation.

However, if the initial assumption was that there is no relationship between facial features and any putative criminality trait, then in place of this question, one might instead be interested in whether there is bias in the criminal justice system. In that case, Wu and Zhang's results could be presented as evidence that there is indeed such bias–that is, the criminal justice system is biased against people with certain facial features. This hypothesis would also explain the difference between the photos of convicted criminals and the photos of people from the general population. The authors did not consider this alternative possibility. Indeed, they appear to be saying that while humans may be prone to bias, ML systems are not.

However, it is clear that the data on which this system was trained had ample scope for human bias to enter at many points, from the initial arrest to the conviction of each person whose photograph appears in the dataset. Deploying a system like this in the real world could have detrimental consequences.

Human biases can infect the data on which 'neutral' statistical models train. This results in the model being biased and, in this example, potentially amplify the biases already present in the criminal justice system. Such false positives could have ethically unacceptable results, such as unwarranted scrutiny of people who have done nothing wrong, or worse, arrests of innocent people.

and benefit from this technology. To ensure collective and public benefit from AI technology, it is necessary to prioritise the potential for public good. Consideration should be given to issues of fairness, equity of access and broadly distributed economic wellbeing. Consulting widely is a key aspect of the inclusive design process. Some specialised technological functions have the potential to be broadly applied, providing wide benefit.

4.3.4 Profiling

AI technologies collect and use large amounts of data to generate predictions and results. Known as profiling, this use of AI enables online interfaces such as Google to individually tailor search suggestions (using predictive analytics) and answers to search queries (using prescriptive analytics). The use of profiling could enhance and accelerate decision-making processes through the use of aggregated datasets; such datasets are too large and complicated to be processed via traditional methods. However, profiling has important implications in relation to ethics, human rights, and discrimination. As discussed with reference to COMPAS, the use of profiling has the potential to adversely affect people, particularly those in minority groups. Design decisions with respect to the use of data and profiled information should be carefully considered.

The wellbeing of people should be prioritised in the consideration of profiling design consequences. Organisations responsible for the introduction of new profiling systems powered by AI technologies should evaluate the impact of these systems and make informed production decisions.

4.4 Informed citizens

To ensure that all Australians can equally participate in public life and engage with AI technologies on a consenting basis, it is essential that Australians are informed about its uses and capabilities. Equipping the community with an appreciation of AI systems is necessary in order to avoid individual exploitation. This is particularly pertinent given the power and knowledge imbalances between those who develop AI and those who use AI. While it may be difficult to explain the complete workings of AI systems, transparency could involve informing people about the use of AI and its applications.

The exercise of individual freedoms and participation in public life are predicated on privacy and freedom from surveillance. Important considerations exist in relation to AI and collected data, discrimination and consent.

The advent of the internet has made information more freely accessible, providing opportunities to increase knowledge, public communication and engagement. The scope and content of public discourse will be further affected by AI-powered technologies. AI can be used to help people find information, make friends, navigate cities, determine whom to hire and fire, predict epidemics, diagnose medical conditions and identify and track criminals. Until recently, decision making in these domains was the exclusive purview of humans. Our epistemic, ethical, and political capacities enable us to engage in such activities and – in the ideal case – explain our decisions to the people they affect.

The human capacity to make and explain decisions is a critical component of democracies. This is only possible in a social and political environment in which people have adequate access to the reasons that bear on their choices. In addition, one of the presuppositions of democratic deliberation is that citizens have access to enough of the same information and truths that they share common ground on which to debate policies, institutions and other arrangements.

Given the potential for AI to increasingly determine how people acquire and circulate information, it is necessary to consider the way in which these systems work, the capacity for these systems to be explained, and how these systems can or have been misused. The increasing use of online media has brought into focus the problems of 'filter bubbles' (Pariser, 2011), 'echo chambers' (Nguyen, 2018) and group polarisation (Sunstein, 2017).

4.4.1 Explainability in AI

The algorithms underlying AI are sophisticated, but their workings are often difficult to decipher and explain to people without a strong mathematical background. The workings of recent developments, such as Google's TensorFlow, are opaque even to their designers (Lewis-Kraus, 2016). This type of AI is built on artificial neural networks that respond holistically to a very large number of variables based on very large training datasets.

AI can embed human biases and systematic errors in the algorithms and data trained with it (Caliskan, Bryson and Narayanan, 2017). When training data are not made publicly accessible, it can be difficult to understand or explain how errors arise. An example of incorrect outputs generated by AI-powered TensorFlow is demonstrated by Google Translate. In this case, repeated instances of the word 'dog' from several input languages including Hawaiian, Maori, and Yoruba was translated into English as: 'Doomsday Clock is three minutes at twelve We are experiencing characters and a dramatic developments in the world, which indicate that we are increasingly approaching the end times and Jesus' return' (Christian, 2018). In instances of AI unsupervised learning it is, in principle, impossible to assess outputs for accuracy or reliability (Hastie, Tibshirani and Friedman, 2008). However, we often apply a double standard, requiring a much higher level of transparency for AI than for human decision makers.

While the underlying algorithmic process may not always be explainable, the ethical adoption of AI requires consideration of transparency. For instance, it may be important to notify the public in instances where they are interacting with AI and also be informed in an accessible manner when their data is being collected and how it will be used. Notifying individuals about the use of AI systems is important to ensure their capacity to appeal in instances of grievance. In addition to allowing for recourse, providing the public with an understandable and accessible introduction to AI will be especially important in establishing trust in AI during the initial adoption stages. AI technologies are likely to advance and change over time, however, establishing public confidence and knowledge around initial AI systems will aid in the continued support for future systems. Explainability in AI with respect to regulatory systems is discussed in Chapter 5.

4.4.2 Public communication and dissemination of information

Establishing public trust in AI technologies is not only equitable, but will ensure greater uptake of technologies that have the potential to deliver significant benefits. To generate public trust in AI technologies, a multi-faceted approach will be required and will need to involve developing standards that are informed by ethical principles, and which subsequently underpin regulation and policy (Winfield, 2016).

Public concern with the emergence of AI technologies has been well documented (Winfield, 2016). In some cases, these concerns have been exacerbated by sensationalised publications in the news and popular media. However, legitimate considerations exist in relation to the impact of AI technologies on employment and privacy (Winfield, 2016). Given that AI technologies are likely to have broad-ranging societal impact, it is important to both address these concerns and develop frameworks that reflect community values.

In Europe, there has been a recent decline in positive public perceptions towards autonomous systems (Winfield, 2016). Community engagement and consultation on the adoption of AI-powered technologies is likely to foster greater public support for this emerging technology. For example, increased exposure to, and engagement with, autonomous systems tends to increase favourable attitudes towards this technology (Winfield, 2016).

A process of public consultation was undertaken during development of Europe's digital single market strategy (European Commission, 2017). The aim of this strategy is to ensure that society and the economy benefits from the use of digital data. Stakeholders involved in the consultation process included industry representatives, self-employed people and members of the public (European Commission, 2017).

Furthering the concept of public consultation, processes of deliberative democracy may provide a useful framework for navigating the diverse community interests and values associated with AI technologies. Deliberative democracy has been proposed in instances where there are divergent ethical perspectives and competing public interests (Molster et al., 2012). Rather than a top down approach of community engagement, deliberative processes 'enable more informed citizens to collectively decide their shared values and acceptable trade-offs in public interests through a process of fair, inclusive and respectful reasoning with each other' (Molster et al., 2012: 83). This approach has been used in Australia to successfully underpin public policy development on biobanking and associated research in Western Australia (Molster et al., 2012: 88).

Inclusive community engagement is an important consideration in the framing of consultation and deliberation on AI. People with disability, culturally and linguistically diverse groups, women, and people who identify as LGBQTI, experience limited representation within public discourse. To ensure these people are not further excluded from participation in the public sphere, deliberative processes should involve inclusive representation.

In general, technology that is subject to regulatory frameworks helps build public trust (Winfield, 2016). The development of these frameworks should include community input and be reflective of people's values and interests. Failure to incorporate community views may result in lessened support for the use of AI technologies.

4.4.3 Democracy, information and AI

AI increasingly determines how people acquire information. Many get their news from Twitter and Facebook, both of which are underpinned by algorithms. In addition, people search for information and translate texts using Google's tools, which also relies on AI infrastructure and algorithms.

The technologies described have the potential to produce negative consequences as a result of a combination of negligence and malicious interference. While public misinformation campaigns are not new, a significant proportion of the population could be misinformed or disinformed as a result of AI systems, and it would be very difficult to trace, track and address the causes. For example, platforms could be hijacked and websites, social media accounts and links could be created and inserted. Indeed, there is evidence that this has already happened in connection with the Brexit referendum (Booth et al., 2017; Sabbagh, 2018), the 2016 US Presidential election (Smith, D., 2018) and other high-stakes processes.

These concerns are exacerbated when the training data and code these platforms use are not released for inspection and correction. However, even if training data and code were to be released, the personalisation of newsfeeds and search results makes it difficult to reproduce the processes that led to the information outcome (Alfano, Carter and Cheong, 2018). This in turn means that it is difficult or even impossible to diagnose and correct these processes.

For example, Google creates suggestions either by aggregating other users' data or by personalising for each user based on their location, search history or other data. In addition to the individual's own record of engagement, others' records can be used to profile that individual. Engagement, in this context, refers to all recorded aspects of a user's individual online behaviour. To the extent that a record of engagement – even in depersonalised aggregated form – is more similar to that of one set of users than to another, an individual is likely to be profiled among the former. For example, predictive analytics suggest, based on a user's profile and the initial text string they enter, which query they might want to run. The same predictive searches conducted in another place and at another time by an account with a different history and social graph will yield different results.[7] For example, a search for the query 'cafe' returns results for cafés nearest to the user; the top results will be different in Amsterdam from in Abuja.

In cases like this, Google suggests questions and then answers to those very questions, thereby closing the loop on the first stage of human reasoning. If reasoning is the process of asking and answering questions, then the interaction between predictive and prescriptive analytics can largely bypass the individual's contribution to reasoning, supplying both a question *and* its answer.

YouTube uses AI-powered predictive modelling to find patterns in individual and group preferences, then recommends clips (Newton, 2017). The vast majority of video selection on YouTube is driven by algorithmic recommendations rather than search or linking. Predictive modelling risks providing people with only select information. Some platforms may tailor this information in line with specific viewpoints, biased information or even bizarre and violent content (Lewis, 2018). Indeed, suggested queries that contain bias or produce discriminatory results can further entrench prejudicial beliefs.

Even if only a portion of the population is influenced in the ways described, our democratic institutions may be adversely impacted. People may find themselves in disagreement about what should be common knowledge. Each side will be able to point to their own sources of information as supporting evidence. Determining

[7] Depending on a user's profile, the content of search results can be subject to change, as in the case of Google's personalised search, which can 'customize search results for you based upon 180 days of search activity linked to an anonymous cookie in your browser' (Google, 2009).

which sources are problematic will be difficult or impossible, both because the AI that recommends the sources is difficult or impossible to explain and because the training data and code are treated as confidential.

A combination of several approaches, including potentially regulating microtargeting, may help in remedying these concerns. Legislative frameworks could be developed with the view to require search engine platforms and social media corporations to reveal both their datasets and the AI algorithms and infrastructures. A Google initiative takes steps in this direction. As with traditional print media, online platforms could be made liable for dissemination of information, news and content. Intra-industry and government research should be undertaken with respect to the explainability gap in AI. Australia and New Zealand could consider following the EU in upholding a legal right to explanation. Indeed, the opportunity exists to go further than the EU in enforcing this right.

4.4.4 Fake news

AI technologies may be used to circulate false information and news reports via the internet. Referred to as fake news, it is designed to influence public opinion and behaviours. Fake news is often presented in a format that mimics the authority of legitimate and trusted news sites. While AI technologies are used to generate and spread fake news, they can also be used to combat fake news stories.

While the full extent of the influence of fake news is ill-understood, it appears to be attributable to a number of independent and interacting factors, including echo chamber effects, biased assimilation of information and confirmation bias. Echo chambers describe the way some people consume information largely from sources that support particular viewpoints. Indeed, over 10 percent of US information consumers receive information only, or largely, from sources that promote fake news (Guess, Nyhan and Reifler, 2018). Confirmation bias may also explain the selective sourcing of information (Nickerson, 1998). Confirmation bias refers to the positive predisposition towards information that supports our beliefs, and a disinclination towards information that is contrary or undermines these beliefs.

Fake news in Australia

In 2017, the Senate established the Select Committee on the Future of Public Interest Journalism with the view to better understanding the challenges and opportunities associated with journalism in a digital society (Commonwealth of Australia, 2018). The Committee reviewed the significance of fake news with respect to contemporary media and journalism. It examined the roles of prominent online platforms, such as Google and Facebook, in facilitating the spread of fake news. However, aware of the importance of reliable and trustworthy news, these platforms are currently undertaking initiatives to combat the spread of fake news. For example, Google is using algorithms to identify reliable or unreliable content, as well as to pinpoint misleading advertising content. Likewise, Facebook is using algorithms to reduce fake news and fake user profiles. Algorithms are not partially well suited to these applications, however, and it is unfeasible for large companies like Facebook to have people fact-checking significant portions of information. As a result of international concerns about the impact of fake news on democratic processes, legislation will be introduced into Australian parliament in order to prevent this occurrence.

4.4.5 Nudges

'Nudges' – a concept derived from behavioural economics – can be designed to make people more receptive to testimony. Nudges are designed to use positive reinforcement and suggestions to influence the context in which people form beliefs, make informed decisions and act in certain ways (Thaler and Sunstein, 2009). For example, people are more receptive to the testimony of others who are perceived to have similar values (Levy, 2017a). Nudges can be designed to take this into account and may involve, for example, ensuring that messages are promoted by people across the political spectrum. There is evidence to indicate that such nudges are effective. For example, corrections of false claims are effective when they come from sources that share the ideology of the receiver (Nyhan and Reifler, 2013). Exposing people to a wider array of opinions may also make them more receptive to alternative information.

Nudging is controversial because it can be seen to circumvent our individual reasoning and thereby affect autonomy (see Levy, 2017b for an overview of these concerns and a response to them). Regardless of whether nudges *do* respect autonomy, they may be *perceived* to disrespect it or be otherwise unacceptably manipulative. Therefore, any attempt to increase public trust in, or acceptance of, AI must consider a possible perverse effect: there is a risk that people will perceive the measures designed to increase trust as themselves untrustworthy. To avoid a possible backfire, any such measures must be designed transparently, in ways that are sensitive to public attitudes.

4.5 Privacy and surveillance

4.5.1 Privacy and AI

In general, Australians and New Zealanders have good internet and related technology uptake. To enable this for AI use, trust is a key issue: for AI to succeed in the private sphere, people 'need to know that their privacy is respected and maintained' (Kelly, 2018). A survey by the Office of the Australian Information Commissioner in 2017 found Australians are concerned about their privacy. The results demonstrate that 69 percent of Australians were more concerned about their privacy than five years ago. Further, 83 percent believe there are more privacy risks dealing with an online organisation than an offline one, 79 percent are uncomfortable with a business sharing their personal information and 58 percent have decided not to share with an organisation because of privacy concerns (Office of the Australian Information Commissioner, 2016a). A survey commissioned by the New Zealand Privacy Commissioner had similar findings (UMR Research, 2018). Nevertheless, a recent survey by Samsung found that despite concerns over data security, 38 percent of New Zealand respondents agreed they would feel more secure if they used smart technology to monitor their home.

The New Zealand Privacy Commissioner has identified two significant privacy risks from data analytics related to AI and automated decision making: lack of transparency and meaningful accountability. The Commissioner notes that:

> '… systems may appear objective and yet be subject to in-built bias leading to discrimination. Many algorithmic assessment tools operate as 'black boxes' without transparency. This lack of transparency is compounded when private commercial interests claim trade secrecy over proprietary algorithms so that even the agencies using the tools may have little understanding over how they operate.'

Accountability for decisions made using AI raises complexities as some decision-making techniques are more amenable to explanation than others. The result is an emerging field of 'explainable AI', where methods for explanation capability are being developed (The AI Forum of New Zealand, 2018).

4.5.2 Surveillance

Increased perception of surveillance might affect people's behaviour; people alter the way they think and act even when faced with only the possibility of being under surveillance. This can include people avoiding talking or writing about sensitive or controversial issues, which not only has a 'corrosive effect on intellectual curiosity and free speech' but inhibits the kind of democratic discussion necessary for a free society (Munn, 2016).

Box 22: Intelligence law reforms in New Zealand

In the 2018 New Zealand survey on individual privacy, 62 percent of New Zealanders said they trust government organisations with their personal information, while only around a third of New Zealanders trusted private companies with that same information. Public discourse on privacy and security has led to significant reforms of intelligence laws in New Zealand. The *Intelligence and Security Act 2017* contained the most significant reforms of intelligence agencies in New Zealand's history including increased transparency of surveillance practices and the operation of intelligence agencies. The reforms may in part explain the greater levels of public comfort with government surveillance and the shift in public discourse from scrutiny of government actions to scrutiny of corporate information collection and surveillance.

The rise of corporate data surveillance, including embedded tracking in computing and smart devices, raises new privacy and surveillance issues. In 2016, the Office of the Australian Information Commissioner and 24 other privacy enforcement authorities across the world evaluated 'Internet of Things' devices, finding that 71 percent of devices did not provide a privacy policy that adequately explained how personal information was being collected and managed (Office of the Australian Information Commissioner, 2016b). Devices that allow or facilitate the pervasive collection of personal information mean that companies can increasingly use aggregate surveillance data to profile, predict and manipulate customer behaviour. AI which supports this predictive analysis will increase the scope and availability of tools to evaluate

and 'correct' individuals into their preferred course of action, which may be to increase profit and for the benefit of corporate interests rather than for a societal 'good'.

Private sector predictive data analytics increasingly provide support for government agency functions, including law enforcement, healthcare and public policy. In these situations, personal information collected with the surveillance power of the state can be used to inform those privately developed analytical tools. Privacy experts warn that these new practices need to be monitored closely and, where appropriate, new ethics or regulatory practice developed.

4.6 Conclusion

The use of AI technologies presents challenges to the equity, health and cohesion of Australian society. Existing inequalities could be exacerbated by the use of AI systems and, indeed, new inequalities could be generated. Additional concerns relate to discrimination, accessibility, privacy, consent and democracy. If developed without adequate safeguards, the implementation of AI could undermine Australian and New Zealand values and human rights. Conversely, responsible design and development could ensure that AI systems reflect and reinforce the Australian ethos of a 'fair go' and freedom of opportunity.

Human rights provide a framework by which to approach the safe and ethical implementation of AI technologies. Broader considerations, underpinning both the human rights framework and representative of Australian and New Zealand values, pertain to inclusion and equity. Community engagement and consultation are essential for the development of inclusive AI and public communication is required to ensure that people have the capacity to make informed decisions. Developing AI technologies with these considerations at the forefront of design and implementation would ensure that AI benefits every sector of the Australian and New Zealand population and advances human rights.

CHAPTER 5
EMERGING RESPONSES AND REGULATION

This chapter is based on input papers prepared by the generous contributions of Dr Olivia Erdélyi (AI Regulation); Nick Abrahams and Monique Azzopardi on behalf of Norton Rose Fulbright (GDPR and Regulation); Herbert Smith Freehills (Legal and Ethical Issues); Dr Olivia Erdélyi and Dr Gábor Erdélyi (Liability); Gary Lea (Liability and Algorithmic Appeal); Anne Matthew, Dr Michael Guihot and Associate Professor Nic Suzor (Appeal Algorithmic Decisions); Joy Liddicoat and Vanessa Blackwood (Privacy and Surveillance); Australian Human Rights Commission (Human Rights). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

5.1 Emerging responses to AI

The rapid development of AI in diverse fields has prompted a range of regulatory and ethical responses. This section sets out examples of developments in four areas: algorithmic transparency, development of the right to erasure, algorithmic impact assessments, and new or emerging ethical standards.

Algorithmic transparency

Algorithmic transparency means having visibility over the inputs and decision-making processes of tools relying on algorithms, programming or AI, or being able to explain the rules and calculations used by AI if these are challenged. The UK House of Commons Science and Technology Committee recommended transparency for government use of algorithms on the basis that the 'right to explanation' is a key part of accountability. The Committee recommended the default position be to publish explanations of the way algorithms work when the algorithms in question affect people's rights and liberties. The New Zealand Privacy Commissioner has recommended that new measures be included in the Privacy Bill to better safeguard the interests of people, including a new privacy principle setting the high-level expectations of fair practice and requiring algorithmic transparency in appropriate cases. Further, a review of algorithms embedded in policies that deliver public benefit has also been undertaken, suggesting how the use of algorithms can be improved for both fairness and transparency and providing a reminder of the need to take care in their use (Stats NZ, 2018). These are essential first steps in ensuring the trust and the social license that is required for governments to begin thinking about AI is established.

The right to erasure

The right to erasure is provided for to a certain extent through the General Data Protection Regulation (GDPR) and Convention 108. The New Zealand Privacy Commissioner recommended a new privacy principle on the right to erasure of personal information, recognising that:

> 'the current rights and protections available to New Zealanders are gradually weakening as technology develops. In particular, the requirement in principle 9 for information to be kept no longer than is necessary is rendered meaningless in the context of advanced algorithms and artificial intelligence. For example, the thirst of artificial intelligence systems for data will mean that agencies will want to keep all of the data that is available for increasing periods of time.'

Providing people with a right to erasure shifts the decision-making onus from agencies, who are incentivised to collect and retain information, to people who can then exert control over their own information. However, not all people have the skills, knowledge or motivation to take control of their information.

The right to erasure may affect the development of AI systems using individual information in machine learning (ML) and algorithmic development and training. It remains unclear whether the right to erasure, or the related right to data portability, will create obligations on an AI developer to delete personal information from the AI training database, or to what extent the intellectual property in the AI is linked to, or reliant on, that personal information. Further, while the data may be removed from the training data set, it may still be present in the AI model thus requiring a new data set and subsequent training phase for the model.

Algorithmic impact assessment

In Australia and New Zealand, a tool for identifying and managing privacy risks is the privacy impact assessment. Building on this concept, AI researchers have developed a practical framework for an algorithmic impact assessment (AIA), similar to impact assessment frameworks already used in data protection, privacy and human rights policy domains. They note that 'AIAs will not solve all of the problems that automated decision systems might raise, but they do provide an important mechanism to inform the public and to engage policymakers and researchers in productive conversation' (Reisman et al., 2018). In New Zealand, a digital service design standard also provides guidance for anyone who designs or provides government services, to support the provision of public services, which are easily accessible, integrated, inclusive and trusted (New Zealand Government, 2018a).

AI stocktakes

The UK House of Commons Science and Technology Committee report on algorithms in decision making contains recommendations to ensure oversight of ML-driven algorithms, including producing, publishing and maintaining a list of where algorithms with significant impacts are being used within central government. Similar work is being done in New Zealand, with a stocktake of algorithms in the public sector completed in 2018 (Stats NZ, 2018).

Legal and professional ethical frameworks

Concerns about the human rights implications of AI have led to calls for legal and professional ethical frameworks that apply to both the government and private sectors to govern the application and design of AI technologies. Statements of ethical principles, guidelines and declarations have emerged in the past decade, along with establishment of ethical advisory boards in public, private, academic and technical communities. These include the Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems and the Asilomar AI Principles, a set of 23 principles that focus on research, ethics and values, and longer-term issues such as capability caution, common good and recursive self-improvement (Future of Life Institute, 2017; IEEE Standards Association, 2018).

Other initiatives include those that are multi-lateral (Council of Europe), multi-stakeholder, by regulators (such as data protection authorities) and calls for action by individual governments. In the UK, for example, the House of Lords recommended the government introduce a statutory code of practice for the use of personal information in political campaigns, applicable to political parties and campaigns, online platforms, analytics organisations and others engaged with such processes. The committee also announced it would produce draft guidance quickly in order for the code to be fully operational before the next UK general election.

New ethical principles have emerged in the private sector. In 2018, the *New York Times* reported that thousands of Google employees were protesting the use of AI by Google to assist the Pentagon in interpreting video images that could be used to improve accuracy of drone strikes (Shane and Wakabayashi, 2018). Google responded by issuing a new set of principles to guide its design, development and deployment of AI. This included AI applications that Google would not pursue, such as weapons, surveillance technologies and technologies that cause harm (Pichai, 2018).

However, human rights advocates have criticised the principles, saying they do not go far enough, while calling for increased multi-stakeholder approaches (Eckersley, 2018). The *Toronto Declaration* is a recent example of a multi-stakeholder agreement on the human rights approach to ML systems, including AI. The Declaration focuses on the rights to equality and non-discrimination and accountability for human rights violations that arise from AI. The Declaration signatories emphasise that while the ethics discourse is gaining ground, ethics cannot replace the centrality of universal, binding and actionable human rights law and standards, which exist within a well-developed framework for remediating harms from human rights violations (Amnesty International and Access Now, 2018).

Some of the common features of these various ethical initiatives are that:

- AI should be developed for the common good to benefit humanity
- AI should operate on principles of fairness and intelligibility
- AI users should uphold the data and privacy rights of individuals and communities
- AI should be available to all (reflecting the right to benefit from scientific advances) including the education to enable benefits to accrue equally to all
- AI should never be able to operate autonomously to hurt, destroy or deceive humans.

At the same time as these new ethical norms are developing, new collaborations are forming. In September 2017, for example, the United Nations announced it would open a new office in the Netherlands to monitor the development of robotics and AI. The partnership initiative launched a working group on AI, labour and the economy, which has proposed developing:

1. a rating standard that measures an organisation's adherence to good AI ethical and compliance standards in order to promote awareness and improve practices
2. case studies to share insights on how organisations are dealing with a range of issues such as workforce displacement, the use of AI in employee vetting, ethics and transparency, and policies
3. an AI Readiness framework to help communities accelerate their ability to leverage AI technologies to minimise inequality of access to, or adoption of, AI technology.

In 2018, the Australian Human Rights Commission launched a three-year project to explore the opportunities for new technologies to protect and promote human rights and freedoms. The project is examining the challenges and opportunities for human rights of emerging technologies, and innovative ways to ensure human rights are prioritised in the design and governance of these technologies. The project is exploring issues such as bias, big data, inclusive technology and the intersection between technology, free speech and democracy. An issues paper was released in July 2018 starting a public consultation that will inform the Commission's work.

5.2 Regulation and regulatory frameworks

AI is already being used to make data-based decisions in a variety of fields – insurance vetting, loan applications; even sentencing decisions. These decisions will need to be evaluated with respect to society's desire to have important decisions be transparent, explainable and reviewable.

AI presents legal and ethical issues within two broad categories:

- responsibility for decisions made by AI systems
- issues arising from AI systems working in combination with an increasingly digital world.

Given the vast amount of work and specialised expertise needed to formulate sustainable AI policies across diverse policy domains, governments should not approach the challenge in isolation. Academic and industry stakeholders undertaking AI research and development in multidisciplinary areas possess expertise needed by governments to inform policy initiatives.

The output of AI

One category of legal and ethical issues arising from AI covers questions of responsibility and ownership that arise from what AI produces – its 'output'. In particular:

- When an AI makes a decision, it may not be transparent, explainable or reviewable in the way that decisions made by a human are. How do we respond to this?
- When AI makes a decision, who is responsible for the decision?
- Conversely, when AI creates property, who owns it?

Explainability and AI

Until recently, the 'explainability' of computer-system outputs was generally not an issue. Computers were programmed to run in accordance with a set of rules. If necessary, the basis on which decisions were made could be explained. However, decisions made by more advanced AI may not be readily explainable. Because decisions are being made by reviewing vast sets of data, and not on the basis of actual intelligent reasoning, the reason for the decision may not be explainable to humans. This means that if the decision is suboptimal or wrong– and it may be wrong if the data are flawed – then an individual affected by the decision has no way of determining this or effectively seeking review and redress.

Instances of poor data – or poor AI design – leading to wrong AI decisions have already occurred (Calo, 2017; Turchin and Denkenberger, 2018). For example, a translation engine associated the role of engineer with being male and a policing tool disproportionally targeted minority communities.

In short, AI does not always get it right. If the data processed by AI are incorrect, incomplete or biased, then the decision it makes (the output) may also be incorrect. This is true also for the algorithms that process it. Much of the data that AI is using has arisen from humans, and so inevitably bears the imprint of the inherent biases of the people who created it.

Traditionally, society has implemented processes for allowing important decisions to be reviewed. For example, almost any decision made by government, such as an application for a building permit or court decisions, including sentencing, can be reviewed by individuals affected by the decision. For reviews of this nature to be effective, the reasoning behind the decision must be explainable. This is the premise for legal rules relating to the transparency of decision making. AI can present a 'black box' problem. Increasingly, as datasets get bigger and processes more complex, it simply will not be possible to explain the reasoning behind an AI's decision, thereby compromising the ability to review decisions. In instances when important decisions are made on the basis of large datasets, consideration should be given to ensuring the accuracy of data and public confidence in the data. Data-quality regulations may facilitate data accuracy and trust. Given that AI-powered decisions are not capable of explanation or review, consideration should be given to the way in which this may be negotiated in society and by regulatory frameworks.

The responsibility of AI decisions

AI-powered systems can make decisions independently of humans. As the dissociation between the creator or operator of the system and the decisions being made by it becomes more pronounced, it will be increasingly difficult to allocate responsibility for those decisions to a particular entity. This means that when the decision has consequences that give rise to issues of responsibility – most notably, questions of legal liability – our traditional legal concepts, which require someone to be 'at fault', are no longer effective. 'No fault' schemes, such as New Zealand's no fault compensation for personal injury legislation administered by the Accident Compensation Corporation, may provide a framework for legislating the responsibility of AI decisions.

At their present stage of evolution, most AI systems would be considered to be simply 'tools', in the sense that they are controlled by humans. This aligns with traditional legal principles: if there is any liability it is attributable to the controller.[8] However, as AI use develops, and the idea of a 'controller' becomes increasingly irrelevant, this analysis will become more difficult. In the long term, when an increasing number of decisions are made by AI systems independently of humans, it is unclear who is responsible when something goes wrong and whether there should be regulation attributing responsibility for AI-based decisions.

The ownership of AI

An AI system can produce a variety of tangible and intangible outputs that can be characterised as property. Today, this mainly comprises intellectual property, such as copyright, and confidential information. However, as AI is increasingly used in combination with robotics and automation, AI will create tangible property as well.

Legally, property can only be owned by a legal entity.[9] As with liability for decisions, ownership of property arising from AI is likely to be attributed to the legal entity that 'controls' it. However, this analysis starts to break down as AI systems begin to act independently. Consideration should be given to regulatory frameworks for ownership rights and obligations in instances where property arises from AI systems independently of humans.

[8] See for example, the *Convention on the Use of Electronic Communications in International Contracts (UN)* article 12, which states that a person (whether natural or legal entity) on whose behalf a computer was programmed should ultimately be responsible for any message generated by the machine.

[9] *Copyright Act 1968* ss 32 and 35.

Human rights and AI

Another category of legal and ethical issues arising from AI relates to the concern about how increasingly powerful and pervasive AI will interact with society and individuals, and the effect it may have on our human rights. Regulatory frameworks should be underpinned by consideration for human rights, as outlined in Chapter 4.

Cybersecurity and AI

AI adds two nuances to general cybersecurity risk. There is significant potential for AI to be used maliciously to power more effective and damaging cybersecurity attacks. For example, 'spear phishing' is a type of cybersecurity attack involving an email that is specifically tailored to an individual or organisation, often using AI (Martinez, 2017). Specificity is what gives this type of attack its power, and that is achieved through AI. In addition, AI systems themselves are susceptible to cybersecurity attacks. This is true of all IT systems, but as AI becomes more integral to the making of significant decisions, this becomes a greater danger. For example, an AI-powered driverless car can be fooled by subtle alterations of road signs (Gitlin, 2017). It is also possible to develop AIs that force other AI systems into making incorrect classifications or decisions (Artificial Intelligence Committee - House of Lords, 2017). Consideration should be given to regulatory frameworks that seek to mitigate the risk of AI being used to breach systems and to protect critical AI systems from cybersecurity attacks.

Institutions

While AI can deliver benefits to society, it can also create societal risks because of its ability to disrupt existing norms. For example, AI-powered technologies have, to some degree, been involved in the displacement of workers from jobs, distorting financial markets, curating newsfeeds and creating quasi monopolies. Consideration should be given to the capacity, power and resources required by institutions (such as ASIC, the ACCC and APRA) to respond to these concerns. AI systems have the potential to perform roles that traditionally have required specific qualifications, certification or training; for example, legal advice or healthcare. Decisions should be made as to whether AI systems be allowed to perform these kinds of roles if they achieve a level of 'competence'. The monopoly risk derives from the potential for a small number of operators in a market to have the resources to adopt AI systems on an immense scale, thereby eliminating smaller players and reducing competition. It is important to consider the way in which public institutions, and the underpinning democratic principles, are protected. To achieve this, it may be necessary to imbue these institutions with additional power and responsibility.

Box 23: The 'existential threat' issue

While there is great difference of opinion on how significant the risk is, most commentators agree that to some degree uncontrolled super AI could, in the future, present a threat to our existence. One prominent response has been to attempt to 'design in' morality (Wallach and Allen, 2008). However, existential risk is considered to be well beyond the timeframe of this report.

Considerations relevant to government response

Society and government should consider how best to deal with the opportunities and risks presented by AI. It is likely that regulation will provide an effective framework to navigate the emergence of AI technologies. Equally important will be education, thought leadership and guidance, and government management. The most effective regulatory frameworks are likely to emerge as a result of an educated community and an informed discussion.

Future regulatory strategies could unite government and non-government parties and consist of a dynamically changing mix of strategies and indirect, flexible and sensitive steering processes. Society depends on large technological firms to drive technological innovation. Collaboration between government and industry could result in mutually beneficial outcomes. However, many of the legal and ethical issues discussed above will only be effective if a global approach is taken. Governments should consider strategic priorities in the field of international law and AI and partake in international institutions, initiatives and development of standards.

5.2.1 A global approach to regulating AI

AI has global impact and, as such, the regulation of AI will transcend national boundaries. International laws and norms relate to areas affected by AI, such as international trade law and human rights conventions. To address AI-related policy challenges, collaboration among different constituencies within nation states and internationally will be necessary. Policy approaches should be multidisciplinary and extend beyond traditional, distinct policy domains.

Internationally-coordinated policy action will be necessary to ensure the authority and legitimacy of the emerging body of law governing AI. Policy initiatives must be coordinated in consistent domestic and international regulatory frameworks to avoid conflicts through fragmentation and to maximise efficiency. Economically, regulatory coordination will ensure that AI is welfare enhancing, rather than aggravating existing global economic inequalities. This will ensure broad social and political support of AI regulatory frameworks (Korinek and Stiglitz, 2018).

To date, such a regulatory framework has yet to be formulated at a national, regional or international level. AI policies are developed by largely uncoordinated efforts of various academic and industry groupings as well as by government initiatives. The regulatory purviews of the agencies involved in the process are not clearly defined and issues of regulatory architecture design have not yet been addressed. As a result, AI applications are sporadically regulated. While greater regulation will be required for the application of AI within industry sectors, industry should also take proactive steps to ensure safe implementation and readiness for AI systems.

The establishment of a new intergovernmental organisation may serve as a forum for coordination to support national policymakers in the development of AI policies and frameworks. This could ensure internationally consistent AI policy approaches via the direct engagement of governments in policy debates prior to the adoption of national positions. Such an organisation should complement, and collaborate with, the diverse array of non-governmental entities involved in AI research and development, so that common approaches are informed by broad expertise.

An opportunity for Australia and New Zealand is that intergovernmental organisations are often hosted by countries considered as neutral. Australia and New Zealand are good candidates for such a role given their relatively small size and their amicable relationships with other countries.

5.2.2 Regulation and the right to privacy

The theoretical and regulatory framework for the right to information privacy is well settled in Australia and New Zealand. Privacy laws regulate in a technology neutral manner, with standards for collection, use, storage and deletion of personal information applying regardless of the nature of technology that collects and uses personal information. In general, information privacy laws in Australia and New Zealand have stood the test of time. In New Zealand, there is no general legislative framework established to directly govern or regulate AI or algorithmic tools including automated decision making (Edwards, 2018), although the Privacy Commissioner and the Government Chief Data Steward have jointly developed six key principles to support safe and effective data analytics (NZ Privacy Commissioner, 2018). Aspects of the current regulatory framework do apply to AI in New Zealand, including the information privacy principles of the *Privacy Act 1993* and other human rights obligations that apply to private and State actions involving the personal information of individuals.

Regulation should not be undertaken either too quickly or at a stage too late; rather it should keep pace with the field and emerging norms. With AI, there are challenges and opportunities for regulatory frameworks. Some of these challenges were previously presented with the emergence of other new technologies. Lessons can be learned from experiences with transparency reporting and regulating copyright with illegal file sharing online as well as the recently-proposed EU Copyright Directive.

All laws require regular review to ensure they reflect societal values and remain clear. In New Zealand, reform of the Privacy Act is underway, with a Privacy Bill introduced in early 2018. The Privacy Commissioner noted in his submission on that Bill that:

> 'the [information privacy principles] do not directly – or arguably very effectively – address the particular risks and issues created by automated decision-making processes. Nor do they require specific mitigations such as algorithmic transparency.'

5.2.3 Effects of GDPR

Industry-specific laws and regulations will be relevant to AI's deployment in Australia and New Zealand, especially in more regulated industries such as finance, healthcare and insurance. However, overseas privacy regulations, namely the new EU General Data Protection Regulation (GDPR), are likely to have one of the largest impacts and restraints on the use of AI in Australia and New Zealand.

Data has been described as the fuel for AI and to understand the relevance of privacy laws and regulations, such as the GDPR, it is necessary to firstly understand the data-centric aspects of AI. Using specific algorithms or rules, AI systems collect, sort and break down datasets to analyse them and make forecasts and decisions (UK Government Office for Science, 2016). As a technology that collects, processes and develops data, which may include personal data, privacy legislation will be relevant to AI's application and use.

The GDPR governs the collection and processing of personal data which is defined in Article 4(1) of the GDPR as:

> 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or

to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.'

In broad terms, the GDPR may apply to an entity not incorporated in the European Union (EU) where that entity:

- has an establishment in the EU (e.g. a branch office)
- processes personal data of individuals who are in the EU where such processing is related to the offering of goods or services to those individuals; or
- processes personal data of individuals who are in the EU where such processing is related to monitoring the behaviour of those individuals as far as their behaviour takes place in the EU.

As such, the GDPR has an expanded extra-territorial reach that extends to countries such as Australia and New Zealand. Importantly, entities do not need to have a physical presence in the EU to fall within the ambit of the GDPR. Moreover, Australian businesses of any size may need to comply with the GDPR, as opposed to the limited exemptions from the *Australian Privacy Act 1988* (Cth) for certain small businesses that have an annual turnover of A$3 million or less.

While the GDPR shares a number of requirements with other privacy laws, such as the Australian Privacy Act, the GDPR introduces a number of new requirements that are likely to have a significant compliance impact for entities who are captured by the new regulation. For example, the GDPR introduces increased accountability and transparency regarding the processing of personal data and enhanced data subject rights (such as the right 'to be forgotten' and the right of data portability). It also introduces a new definition of consent.

The use of AI and ML are likely to present a major challenge for compliance with privacy regimes, such as the GDPR. Such regimes are focused on transparency of processes and systems of datasets containing personal data. However, it is often difficult to obtain this transparency and to fully understand how AI systems work and the full extent of their decision-making capabilities. The potential risk of AI systems 'going rogue' and 'the robots taking over' is another concern, perhaps fuelled by science fiction rather than reality. However, these are some of the reasons why AI is an area that requires more onerous requirements and oversight under various regulatory frameworks. The regulatory implications and impacts of AI are discussed further below.

To lessen the impact and reach of the GDPR and other regulations governing personal data use, entities may consider it prudent to minimise or completely remove the processing of personal data from AI's capabilities; for example, pseudonymising or de-identifying data before it is inputted into AI systems. However, this may not always be practicable. Furthermore, de-identification (such as removing a person's name) will not be a viable solution if AI's functionalities are sophisticated enough to combine and re-identify datasets or reasonably ascertain the identity of a person based on one or a combination of datasets.

With this regulatory framework in place, it will be important for affected entities to implement appropriate technical and organisational compliance measures. The penalties are severe for non-compliance – the GDPR includes fines of up to €20 million or 4 percent of annual worldwide turnover (whichever is higher), for certain contraventions. Moreover, where an AI system causes a breach involving personal data there are legal obligations to report under both the GDPR and Australia's new notifiable data breach regime. Data breaches, whether caused by humans or machines, can adversely affect the public perception of an entity.

The GDPR is expected to affect how entities manage and process personal data, regardless of whether they are impacted by the GDPR. Compliance with regulations such as the GDPR could set the benchmark for personal data processing and compliance within Australia and New Zealand.

5.2.4 Regulatory implications for the use of AI by transnational corporations

There are several regulatory implications involved in the use of AI by corporations, whether they be transnational or not. A corporation that is not transnational could still be subject to an overseas regulation such as the GDPR.

One of the frequently cited benefits of AI and ML is its capacity to learn and make decisions without any human involvement. AI can also be used to profile people. Article 4(4) of the GDPR describes profiling as:

> 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.'

Automated decision making and profiling is subject to restriction and increased oversight under the GDPR. Subject to some exemptions, under Article 22 of the GDPR, people have the right not to be subject to a decision based solely on automated processing and profiling if it 'produces legal effects concerning him or her or similarly significantly affects him or her'. In other words, an individual cannot be subject to a decision that is made without any human involvement. While some exceptions apply, entities using automated decision making are required to implement suitable measures to safeguard the individual's rights and freedoms and legitimate interests. This includes a series of rights under Recital 71 of the GDPR in relation to profiling; including the right to an explanation of a specific decision and the right to challenge the decision. Additional restrictions apply where decisions are made based on specific categories of personal data (e.g. personal data revealing racial or ethnic origin, political opinions or religious or philosophical beliefs).

Furthermore, Articles 13(2)(f) and 14(2)(g) of the GDPR require data controllers who use personal data to make automated decisions to notify people about the existence of automated decision making, including profiling and 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'. The difficulties of complying with these obligations when using AI has been extensively reported. The complexity of AI and their associated technologies may make it difficult to understand how decisions and profiling are being undertaken by AI systems (Article 29 Data Protection Working Party, 2016).

Among other factors, Australian and New Zealand entities should identify any wholly-automated decisions that they undertake using AI and consider whether it is possible to change the process so that there is meaningful human involvement (e.g. have a sufficiently qualified and skilled person review the machine's decision) or ensure that they can satisfy one of the available exceptions under the GDPR. A data protection impact assessment may also be required. Article 35(1) of the GDPR requires that where a type of processing – and in particular where using new technologies – is likely to result in a high risk to the rights and freedoms of a person, a data protection impact assessment must be carried out. In particular, Article 35(3) of the GDPR expressly requires that a data protection impact assessment is undertaken when carrying out a 'systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person.'

Data ownership and rights

Data has become a valuable commodity. AI's data-generating capabilities present commercial opportunities for the use of data. However, many laws do not recognise data itself as a form of property that can be owned and sold. Data can only be truly owned where it constitutes intellectual property, such as copyright or a trade secret. However, there are difficulties with data constituting a work protected by copyright due to the absence of a human author.

Despite this, people still have rights associated with certain datasets; for example, personal data, confidential information or where there exists a statutory right (e.g. a right to access data under freedom of information laws). These rights are derived from a combination of contract, common law and statute. The GDPR includes enhanced rights, including data portability rights. In Australia, legislative agenda contains the new 'consumer data right', which permits certain consumers open access to specific types of data, including data held or generated by AI systems. Under this proposed legislation, the consumer would have a greater ability to access certain data concerning them. The legislation is likely to apply to banks, utilities and telecommunication companies, but may be extended beyond these sectors.

Data quality and security

AI systems do not simply process data, they also create new datasets, which may include the generation of data based on personal information. Entities using AI will need to audit and assess the accuracy and quality of those datasets. Where the datasets include any personal data, entities will need to ensure compliance with applicable privacy legislation and associated privacy principles, such as APP 10 under *the Australian Privacy Act*. Under APP 10 entities governed by the *Australian Privacy Act* must take such steps (if any) as are reasonable in the circumstances to ensure that the personal information that they collect, use or disclose is accurate, up-to-date and complete. A broadly similar principle is included in Article 5 of the GDPR.

People who collect and use data have a custodianship role, especially where that data contains confidential or personal information. Under the *Australian Privacy Act* entities must take reasonable steps to protect personal information from misuse, interference and loss, as well as unauthorised access, modification or disclosure. Reasonable steps in the context of AI might include implementing systems with information and communication technology security (Office of the Australian Information Commissioner, 2015) and regular testing of the AI system's security controls and systems.

Consent

The GDPR stipulates that consent will be required for AI systems to collect personal data on behalf of an entity. This may pose challenges. While AI systems can accommodate a 'tick a box' approach to consent (that is, they can work out whether or not someone has ticked the 'I agree' box), they may struggle to comply with the specific consent requirements under the GDPR. The GDPR requires consent to be freely given, specific, informed and an unambiguous indication of the data subject's wishes. While AI systems may be intelligent in many respects, they may lack sufficient emotional intelligence to recognise the emotions and intentions of humans. It may therefore be more difficult for AI systems to discern whether or not consent is free or represents an unambiguous indication of someone's wishes.

Intellectual property

There are a number of regulatory issues associated with the protection of AI systems and their outputs. There is a question as to whether AI computer-implemented algorithms meet the high thresholds of being novel and containing an inventive step to be eligible for patent protection. At least in Australia, courts have confirmed that mere ideas, methods of calculation, systems or plans, and certain computer-implemented business methods, are not patentable subject matter. For AI, this means that automating individual processes may not be sufficient to constitute a manner of manufacture or patentable subject matter unless the automation is an invention in itself. The concept of 'computer implemented inventions' is under currently under consideration by an expanded panel of the Federal Court of Australia (Federal Court of Australia, 2018).

Secondly, there is a question as to whether any data or works produced by AI systems constitute an original work protected by copyright. Under the Australian *Copyright Act 1968* (Cth), copyright does not protect data alone, but rather the way it is collected and put together. Compilations of data can be protected under copyright law, but only if they pass the originality test. Under Australian law, copyright does not exist in a work that is made by a machine and is effectively authorless – a human author is required.

Competition law

Concerns have been raised about the market power that technologies such as AI can provide. AI can use algorithm-pricing systems to gather and leverage vast datasets. In the right market conditions, such pricing algorithms may be used to engage in, and sustain, collusion or other anti-competitive practices that are prohibited at law (Sims, 2017). The Australian Competition and Consumer Commission has noted:

'…a profit maximising algorithm will work out the oligopolistic pricing game and, being logical and less prone to flights of fancy, stick to it… [I]f similar algorithms are deployed by competing companies, an anti-competitive equilibrium may be achieved…' (Sims, 2017)

Questions of risk and liability

AI creates a liability conundrum (see section 5.3). While some AI systems are often seen as acting autonomously and independently, they are not human. In such a scenario, who should be liable when an AI system causes an accident or other liability: should it be the programmers, manufacturers and developers of the specific AI system or someone else? The conundrum also arises from the complexity of AI systems and the interdependency between their different components, parts and layers (European Commission, 2018d). Australia and New Zealand are yet to establish meaningful precedents to address the appropriate allocation of risk and liability between the various actors involved in the development and deployment of AI systems.

5.3 Liability

Policy discussions are increasingly focused on framing responses to AI and liability in both a civil and a military context. However, there is uncertainty about the appropriate principles and methodologies to achieve regulatory change (Petit, 2017). Given the developmental stages of AI technology, it would be difficult to advance specific regulatory proposals in relation to AI liability. There is a lack of clarity about AI and its associated functions and distinct features. This means that the key parameters that could serve as benchmarks for regulation are, at best, ill-defined. Given this, premature action on AI legal liability is not advised. The development of precise and universally accepted definitions both in law, and AI, should precede concrete regulatory proposals.

In the longer term, questions arise as to when, why and to what extent, AI and smart robotic systems might be recognised as persons under the law, including assuming civil and criminal liability either with others or even alone. Presently, under Australian law, individual humans are natural persons, but some other entities are legal persons, either generally (e.g. a company registered under the *Corporations Act 2001* (Cth)) or for more limited purposes (e.g. a partnership is deemed to be a person for the purposes of Part XIC of the *Competition and Consumer Act 2010* (Cth) on telecommunications access arrangements).

5.3.1 Conceptual ambiguity in legal and AI research

Legal discussions of AI typically lack definition of AI technology. This is unsurprising, given the lack of consensus among AI researchers on a universal definition of AI technologies. Generally, there is an assumption that AI systems mimic certain aspects of human cognition, and approaches to defining AI have broadly focused on comparing AI systems' cognitive and behavioural abilities to human and rational behaviour (Russell and Norvig, 2003; Calo, 2017).

Although the absence of a universally agreed-upon definition has not hampered AI research, a consistent understanding and definition of the concept of AI and its associated functions is necessary for adequate regulation. In cases of personal injury or property damage, it is unclear whether the AI system involved or the people responsible for the AI design or distribution should be held liable.

It is necessary to consider that AI does not *know*, *think*, *foresee*, *care* or *behave* in the anthropomorphic sense; rather it applies what could be best described as *machine logic*. That is, the system identifies outputs based on a set of predefined parameters and probability thresholds through a process that is fundamentally different from human thinking. Furthermore, this type of machine reasoning always implies a certain probability of error, where the error tends to occur in – from a human perspective – unexpected ways. These errors can arise from different sources. Two examples follow.

A machine was tasked to distinguish between pictures of wolves and huskies (Ribeiro et al. in their Husky vs. Wolf experiment) (Ribeiro, Singh and Guestrin, 2016). To do so, the system was trained with 10 wolf and 10 husky pictures. All wolf pictures had snow in the background, while none of the husky pictures did. Since snow is a common element in the wolf pictures and is not present in the husky pictures, the system regarded snow as a classifier for wolves. The result is that the system predicts huskies in pictures with snow as wolves and *vice versa*.

There is potential to cheat or actively manipulate a facial-recognition system (Sharif et al., 2016). Facial recognition systems usually use neural networks to recognise

Box 24: Liability and autonomous vehicles

The complexities of AI liability can be illustrated with autonomous vehicles. Vehicle regulation in Australia is a complex amalgam of rules, standards and norms, including road rules, driver licensing, vehicle type approval and insurance (Dent, 2018).

Establishing civil liability requires that one or more persons are identified as owing a duty of care. This may be difficult in relation to AI and smart robotic systems (Gerstner, 1993). It is likely that identification of persons owing a duty of care will become significantly harder at each successive level of vehicle automation. As autonomous vehicles become legalised (National Transport Commission, 2018: 68) and legal provisions are developed, recourse to negligent actions may become less common.

Within Australian consumer law, firmware is considered as software in the defective goods context (*Ipstar Australia Pty Ltd v APS Satellite Pty Ltd* [2018] NSWCA 15). As such, AI systems supplied as vehicle firmware are likely to be treated as goods. Within this context, the actual or deemed manufacturer would be liable for safety defects relating to AI vehicle firmware. However, the more technical and restrictive definition of goods still used in many state and territory sale of goods legislation means that software is not treated as goods unless deemed merged with the goods (*Gammasonics Institute for Medical Research Pty Ltd v Comrad Medical Systems Pty Ltd* [2010] NSWSC 267). This places AI systems in a legal grey area. Until legislative changes are made it is unclear the extent to which provision of cloud or other remote AI systems might be treated as services under this heading.

The future applicability of Australian law with respect to AI may also be limited under the 'state of the art' defence – the defence that the defect could not have been discovered at the time the manufacturer supplied the goods because there was insufficient scientific or technical knowledge at that time (*Merck Sharp & Dohme (Australia) Pty Ltd* v *Peterson* [2011] FCAFC 128).

The use of autonomous vehicles also presents considerations for criminal liability. Under the *Geneva Convention on Road Traffic 1949* to which Australia and New Zealand are party, motor vehicles must have a driver and drivers must be able to control their vehicles at all times. Under the current road rules, excepting special statutory provision for vehicle trials, engagement of vehicles with conditional automation or greater would be a criminal offence in so far as the (human) driver must have proper control of the vehicle while driving (e.g. r297, *Road Rules 2014* (NSW)).

Conversely, without the development of legislation, other road rules could hypothetically cease to operate if high levels of automation were engaged. For example, if a fully autonomous vehicle stops because of a machine-unidentifiable hazard on an intersection, there could potentially be no criminal liability for obstruction of that intersection (Tranter, 2016).

In order to accommodate AI and smart robotic systems under Australian consumer law, it is necessary to clarify the application of the categories of goods and services. Further to this, it would be necessary to redefine acceptable quality with respect to consumer guarantees provided for by Australian consumer law and restrict the scope of the 'state of the art' defence.

AI personhood and accompanying rights must not be drafted or implemented in such a way as to detract from human rights and human dignity. Until AI and smart robotic systems can both uphold civic rights and responsibilities and be appropriately deterred, punished or rehabilitated for criminal law purposes, the individuals and existing legal entities that design, build, distribute and use them must be held completely responsible for them by analogy to rules on children or potentially dangerous animals (c.f. Hallevy, 2013). In short, designers, manufacturers, distributors and users should never be allowed to evade liability by simply saying 'the robot did it'.

Box 25: Case study: Algorithmic based decisions in the legal profession

Potential exists for the application of AI-based decision making within the legal justice system. AI should be capable of sophisticated legal reasoning given the structure and context of legal argument (Bench-Capon and Prakken, 2006). The application of AI to legal decision making may improve transparency, consistency and avoid the potential for ideological bias (Bench-Capon & Prakken, 2006; Cooper, 2011; Guihot, Matthew, & Suzor, 2017; Hall, 2005). However, the risks associated with automated decision making include the incapacity of algorithms to 'exercise discretion and make situational value judgments' (Schild, 1992; Leith, 1998; Broadbent et al., 2011; Lippe, Katz and Jackson, 2015; Simpson, 2016; Guihot, Matthew and Suzor, 2017). AI is not known to have strengths in exercising discretion, induction or intuition, all of which may be required to varying degrees in legal decision making (Guihot et al., 2017; Hall, 2005). AI is unlikely to have the capacity to make value judgments or to appraise and evaluate the social repercussions of the decision (Hall, 2005; Sunstein, 2001). AI may be objective, since it potentially lacks predisposition or ideological bias, but legal decision making ought to involve some normative inputs of which AI is incapable, such as evaluating the absurdity of an interpretation (Cooper, 2011). Public regulators should be alert to the spectrum of risks posed by specific applications of AI and adopt targeted strategies in their regulatory approach in order to address the risks identified.

Mechanising decision making through algorithms raises questions about what could be lost: to what extent 'an algorithm can have a heart', or 'deal with the unexpected, quirky or unique individual that may require appeals to a sense of justice?' (Simpson, 2016). To ensure a balanced decision making process, the development of algorithms in legal decision making should focus on the optimal combination of AI and humans (Lippe, Katz and Jackson, 2015; Guihot, Matthew and Suzor, 2017). AI should not be expected to make reliable, definitive legal decisions that entail the exercise of discretionary judgments; resolution of 'conflicting arguments', or 'ambiguous and contradictory evidence' (Schild, 1992; Zeleznikow, 2000); or the interpretation of facts or data (Oskamp and Tragter, 1997; Zeleznikow, 2000). Rather, the use of algorithms in legal decision making could be limited to applications that better inform human decisions (Schild, 1992; Oskamp and Tragter, 1997; Zeleznikow, 2000). For example, Article 22 of the EU's General Data Protection Regulation creates a new right for people 'not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'(European Commission, 2016; Guihot, Matthew and Suzor, 2017). The implication, at least in Europe, is that humans must somehow be involved in decision making, although how effective this is likely to be, remains to be seen.

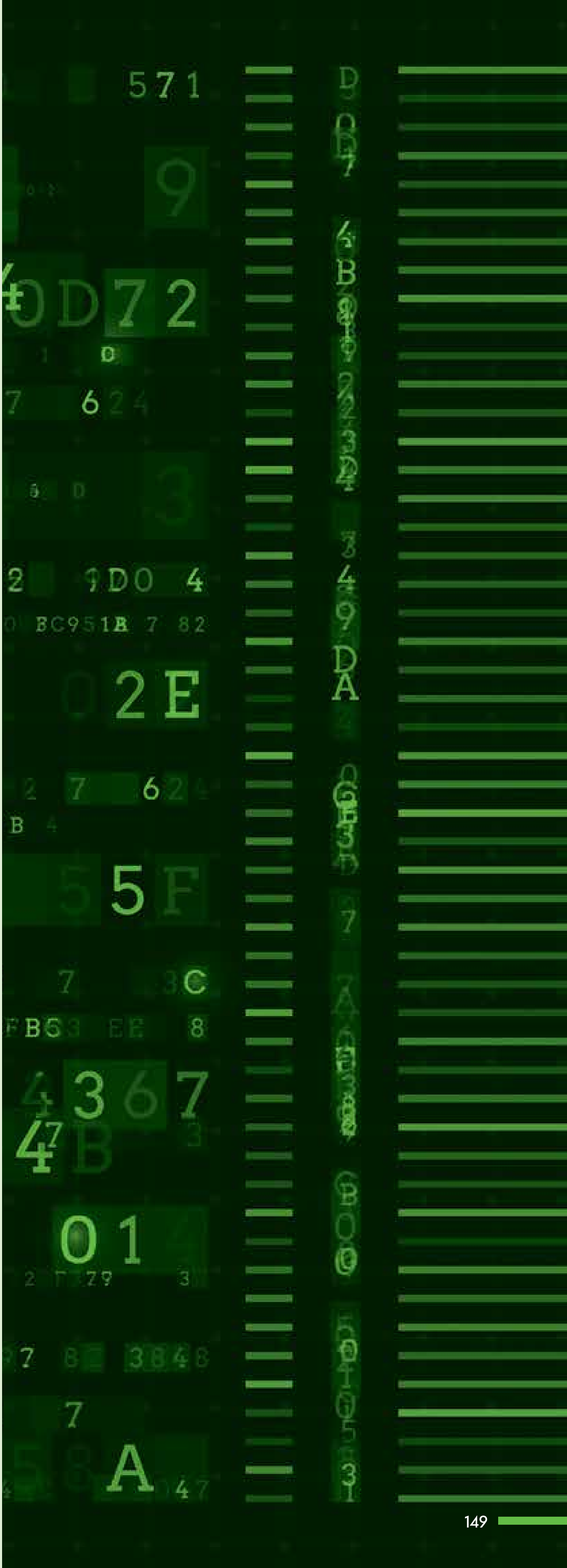
While consistency in legal decision making does sound like an admirable goal, it may be problematic should it lead to standardisation (Hall, 2005). Standardisation in automated legal decision-making processes has been seen to have a regulatory effect on people involved in the decision-making process,

[10] Consider for example, Amanda Schaffer, 'Robots That Teach Each Other' (2016) 119 *MIT Technology Review* 48, where Schaffer explains data sharing goals to improve robot ability; Will Knight, 'Shared Robot Knowledge' (2016) 119 *MIT Technology Review* 25, 26.

including those who are required to implement the decision (Oskamp and Tragter, 1997).

The design of the appeal process should begin with the careful and considered design of the decision-making process itself. Developers and regulators require a deeper understanding of the social and ethical contextual framework and the users' needs of the decision-making system (Oskamp and Tragter, 1997). Concerns for the development of systems with deep contextual understanding will become more pressing where AI systems share information with each other to improve their own process.[10] It is best to avoid establishing a new centralised authority to deal with challenges to automated decision-making processes. Given the potential for the rapid uptake of AI with broad applications, a new, centralised authority would soon find itself in a situation where it was required to be a ministry for a wide variety of departments.

Regulators adopting algorithmic decision-making processes should have a clear path for internal challenge or human review. If Australian and New Zealand regulators take a similar approach to the EU by requiring human involvement in decision making, this would both significantly reduce the risks discussed above and preclude the need for radically new mechanisms to facilitate challenges to algorithmic decisions of AI. The decision would be made by human relying on a range of inputs, only one of which would be the algorithmic system.



patterns in big datasets; specifically, to identify differences between millions of faces. In an experiment, a pair of glasses with a colourful frame was used to interfere with the system's pattern recognition. It did not just block the *view* to crucial parts of the faces but, due to the colourful frame, gave the system the impression of some misleading patterns. In this way, the facial-recognition system made mistakes despite claiming a high confidence.

Another frequently discussed but poorly defined concept used in the context of AI liability is the *black box* attribute of certain ML-based AI systems. There are two types of AI: 'black box', of which there is little knowledge about the inner workings, and transparent systems, which are reasonably explainable. Central to any form of legal liability, is the foreseeability requirement; this pertains to whether or not the action involves some sort of mental element. To hold an individual responsible for harm requires that the individual can anticipate that harm as it is not possible to intend for, or be negligent about, that which cannot be foreseen. Black box systems yield results that may be unforeseeable, whereas transparent systems can be deemed to foreseeably lead to an outcome that may be undesired. Contemplating ways in which AI systems could potentially be held liable without conceptual clarity on this attribute is problematic, as it crucially affects the foreseeability requirement. Intuitively, one would assume that while foreseeability cannot be given in the case of black box systems, it should not be a major problem if we are working with a transparent system, where we can comprehend the system's every move.

Further complexities arise because the notion of transparency is itself subject to considerable conceptual ambiguity. Three distinct model properties are used to facilitate *ex ante* transparency; namely simulatability, decomposability, and algorithmic transparency (Lipton, 2018). In *simulatability* we assume that a person can reflect the whole ML model at once. In *decomposability*, each part of the ML model (input, parameter, and calculation) admits an intuitive explanation. And finally, in *algorithmic transparency* we require a full understanding of the learning algorithm itself, i.e., we expect to fully understand and reconstruct each and every step it makes. This analysis suggests that each of these notions of transparency may well require different levels of expertise in order to establish foreseeability. Additionally, these cases must be distinguished from *ex post* transparency and interpretability, that is, when we are able to understand how the system has achieved a given output, for example, or to seek explanation for an unforeseen, and from an *ex ante*perspective perhaps, even unforeseeable, outcome. This does not mean that we can fully back-trace every step the ML model did. A final aspect regarding transparency is that there is always a trade-off between AI performance and transparency. Transparent models usually have much simpler structures than black-box models.

Discussion of these conceptual ambiguities provides an indication of the problems in devising policy initiatives on AI liability. Lawyers may have to accept that foreseeability – the primary benchmark for imposing liability – needs to be replaced in the context of AI, or face a different set of unexpected challenges. The law will need to be adapted to the changing realities of our AI-driven world; our guiding principles should be the core societal values we intend to preserve. The design of AI-related policies, whether for liability or in any other area, will require a broader perspective, accounting for multidisciplinary imperatives incollaboration with multiple stakeholders.

5.3.2 Ability to appeal algorithmic decisions

Decisions generated by AI technologies are dependent on the use of algorithms. The use of AI decision making has a broad range of applications in public and private sectors. For example, algorithmic-based decisions may be used to determine health treatments, the outcome of loan applications or the granting of bail applications. The case study below examines the potential use of AI in the legal justice system. The Commonwealth practice guide on Automated Assistance in Administrative Decision-Making highlights the importance of accuracy, accountability and transparency in algorithmic decision processes. Important considerations also include algorithmic fairness (Zou and Schiebinger, 2018). Individuals subject to algorithmic-based decisions may wish to review or appeal the decision. Developing clear frameworks would facilitate this appeal process.

To allow people to appeal algorithmic decisions, it is necessary that they are informed when algorithmic decision making has occurred. This information should be accompanied by a basic explanation of the way in which the algorithm works and what factors were considered. Presently, due to the process of deep learning, it can be difficult to identify which information was used in the algorithmic decision-making process. With current legal decision making, transparency is of paramount concern and a significant feature of review and appeal processes. Concerns for transparency of the algorithmic evaluation or lack thereof will become increasingly critical if decision making with legal ramifications is automated by algorithms. However, advances in technological processes indicate (Castelvecchi, 2016) that this will not long remain a barrier to algorithmic transparency.

People should be provided with a clear and simple pathway for appeal. For example, in accordance with section 495A of the Migration Act 495A, automated computer-based decisions may be appealed via the same process of challenging decisions made by the Minister. As the use of AI decision making is likely to become increasingly common, a standardised appeal process within public and private sectors would be useful.

5.4 Independent body

The regulatory issues and implications related to the use of AI by transnational corporations and other entities are complex and varied. As disruptive technologies such as AI become more prevalent, we are likely to see increased regulation. Governance and regulatory mechanisms could be assisted by an independent body that could be established to identify key areas for regulation and response. 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.

An independent body could constitute a collaborative space where STEM and HASS disciplines could determine how the demands of personal, community and national interests may change rapidly as a result of AI adoption. In addition, the independent body could assess the way in which governance could be structured to avoid being left behind technological and social changes. Interdisciplinary work undertaken by the body could draw on the social sciences to assess how political systems can adjust, anticipate and manage inevitable future change. A clear national direction, which integrates planning, regulation and innovation, could help ensure that AI is developed in a manner that specifically addresses national needs.

In New Zealand, the Artificial Intelligence Forum of New Zealand (AIFNZ) is an industry-led body that includes representatives from academia and government. It largely focuses on enabling the implementation and development of AI in New Zealand. The AIFNZ seeks to raise awareness and capabilities of AI and contributes to the social and political debate on AI's broader implications for society. It is a member of the Partnership on AI – an international industry consortium established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences.

An independent body in Australia could contribute to shaping both domestic and international AI policies. There may be a need for an independent body to provide institutional leadership on the development and deployment of AI in Australia – promoting what the Australian Human Rights Commission has described as 'responsible innovation' (Australian Human Rights Commission, 2018b). Such a body could play an oversight role in the design, development and use of AI and associated technologies that would help protect human rights in Australia and at the same time foster technological innovation. Such an organisation would be a forum for collaboration and be independently led, drawing together stakeholders from government, industry, the public, and academia, uniting both HASS and STEM disciplines. Its roles and functions could include rule making, monitoring, enforcement and dispute resolution. The organisation could establish a new governance model that covers the various stakeholders' interests and relationships, encompassing a framework that harnesses the private sector's insight and influence, while also protecting human rights (Elmi and Davis, 2018). Additionally, the body could provide direction and support for governance mechanisms, conduct research for the development of technology and policy, and facilitate research partnerships and initiatives. Given this, consideration should be given to:

- the establishment of a government-supported AI institute, tasked with further researching legal, ethical and other issues arising from AI
- out of that initiative, the government facilitating the development of an overarching set of values and principles to guide the response to AI issues
- on the basis of those values and principles, overseeing the creation of guidelines and frameworks for the development of regulations that can be provided to relevant departments, sectors and industries
- where appropriate, encouraging industry-specific regulations tailored to the specific issues that AI applications are creating.

There are a number of existing bodies in Australia that could be expanded to incorporate these functions. For instance, Data61 already has a significant role in data innovation, builds collaborative partnerships and networks between government, industry and academia, conducts research to inform decision making, and develops new products and platforms. Standards Australia could also play a role. In addition to developing national technical standards, Standards Australia currently acts as Australia's representative at international standards fora, such as the International Standards Organisation (ISO) and International Electrotechnical Commission (IEC). Forming international agreement on the definition and standards of AI technologies could occur through the ISO and IEC. Indeed, ISO/IEC WD 22989 Artificial Intelligence – Concepts and Terminology is in development and Standards Australia has already initiated national AI projects.

An independent body could be well placed to examine the way in which governments and industry can adjust, anticipate and manage change resulting from AI to the benefit of society. Establishing this independent body would enable Australia to provide global leadership in AI governance models and potentially initiate global governance measures.

5.4.1 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.

The independent body could be 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.

CHAPTER 6
DATA

This chapter is based on input papers prepared by the generous contributions of Associate Professor Lyria Bennet Moses and Amanda Lo (Data Collection, Consent and Use); Dr Ian Opperman (Re-identification of Anonymised Data); Dr Vanessa Teague and Dr Chris Culnane (Data Storage and Security); Associate Professor Mark Burdon (Information Privacy); Data61 (Data Integrity, Standards and Ethics); Professor Fiona Wood (Data Linkage). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

6.1 Introduction

Data are essential to the development and operation of AI technologies. AI and machine learning (ML) require large datasets to learn from and generate outputs, and skilled practitioners need data to develop the AI itself. Advances in core fields of data-driven AI, including ML, 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 humans. The field of data science and informatics is continuing to grow, with increasing demand for skilled data science experts, engineers and cybersecurity expertise at an all-time high. As the costs associated with the collection, storage and analysis of data reduces, a rapid change is occurring in the exploration and uptake of digital technologies and data has become an increasingly valuable commodity. Australia and New Zealand's public and private sectors are increasingly premised on the collection, control, and use of data – often personal and sensitive – between people and organisations or between people and governments. 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. Platform companies such as Google and Facebook rely on these data to generate revenue in various ways, while governments can analyse data to better understand citizens' concerns and needs. Much of these data are not *provided* by people per se, but rather *generated* through various internet-enabled technologies and services, such as smart technologies and services in homes, workplaces, cities and governments that produce continuous streams of data.

AI’s data-generating capabilities present commercial opportunities for the use and leverage of that data. Big data has attracted global attention through datasets offering new insights on patterns and trends that were previously intractable. If used appropriately, the new technologies using big data could generate new potentials, however at the same time, big data can have significant methodological and ethical limitations, social and political implications and epistemological challenges (Crawford, Miltner and Gray, 2014). For example, algorithmic decision-making tools raise concerns of bias and discrimination, while AI systems capable of deriving personal information from multiple datasets point to technical and legal challenges regarding tracing the ‘provenance’ of data. Developments in our ability to rapidly collect, analyse and safely share data between people and organisations – without compromising individual privacy – will support the development of such AI-enabled, targeted services, by enabling organisations to understand our particular needs and characteristics, from observing our data.

Policies that affect data collection and sharing inevitably affect the development of AI. In Australia, policy discussions have been focused on bolstering data innovation – in which data, including personal information, is treated as a tradeable asset to stimulate growth in digital economies (Productivity Commission, 2016: 47) at the potential expense of data protection and privacy. Key issues facing data protection and privacy include governance and regulation of aggregated data, which involves protecting aggregated datasets from ‘de-anonymisation’ by AI systems; data sovereignty, which refers to the storage and security of national datasets; and data integrity and portability, which relate to an individual’s right to obtain, reuse or delete personal data. Data protection and data innovation need not be considered mutually exclusive goals; enhanced consumer protections on the collection and control of data may have the flow-on effect of stimulating competitive digital economies and innovation.

6.2 Collection, consent and use

Data underpins AI, and the quality, complexity, availability and origins of data will influence the accuracy and validity of the AI-based systems it powers. A 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 to have been impacted by bias. This may be the bias of the contributor, whose data are 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.

Trust in the integrity of data is essential for a dataset to be consistent, reliable and effectively contribute to an AI or ML-based technology. This involves ensuring 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.

One of the main challenges for AI is centred on concerns about unintended consequences of sharing data including appropriate use and interpretation and unauthorised disclosure or use of data. Aggregation and anonymisation of individual data is a common approach used to reduce the risk of personal disclosure within a dataset.

6.2.1 Identification and access to personal data

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

Information is considered personal if it is about an individual who is identifiable or reasonably identifiable. Personal information encompasses a broad range of information and might include name, email address or unique identifiers such as photos or videos. A further element of personal information is sensitive personal information, which often encompasses information or opinion about an individual's health, race or ethnicity, political opinion, religious beliefs, sexual orientation or criminal record. In this case, algorithmic frameworks can detect identifiable people from a range of data because the detection is based on the ability to categorise information with little analytical recourse as to how the information was generated.

However, the situation is more complex when considering reasonably identifiable information. Reasonably identifiable information refers to identification arising from data aggregation processes. In this case, data that do not readily identify an individual can be aggregated to enable re-identification. By doing this, an AI system can determine whether the aggregated output is 'about' an individual. For example, mobile phone metadata can be used to identify individual life-style patterns (Isaacman et al., 2011) and can therefore result in the re-identification of an individual. In these situations, understanding the social context of data generation is crucial, as is understanding the capabilities, resources and abilities of the data aggregating organisation.

Identification of individuals is a risk, not specific to AI, but arising from the proliferation of detailed personal data used by AI systems. Simple algorithms can be used for re-identification of individuals, such as data linking (Culnane, Rubinstein and Teague, 2017). Indeed, the more data that are available about an individual, the easier it will be to re-identify their record and data. While this is something that humans can do already, AI is highly effective at finding latent patterns in data, allowing it to re-identify data quickly and on a large scale. The pace of development in AI and the increasing detail of data collected about individuals outstrips the progress of de-identification. This results in datasets becoming easier to re-identify over time and the risk increases due to a combination of algorithmic progress in AI and the increasing availability of auxiliary data.

AI's ability to identify personal information is a complex technical and legal issue. In Australia, the Consumer Data Right is beginning to roll out across industries to ensure that consumers have the right to safely access their personal data and authorise the transference of their data to third parties. The Consumer Data Right will apply to specific data sets and is aimed at empowering the consumer with the use of their own data while also improving the flow of information in the economy, encouraging competition and creating opportunities. The Right focuses on the consumers choice and ability to share their data rather than a business's right to share consumer data. An example of how this Right provides benefit to the consumer would include a consumer freedom to use a comparison website for home loans. In the future, it is possible that an AI or ML framework could be used to assist with tracing the 'provenance' of the re-identification process described above. However, the degree of legal interpretation skills required are still such that the ultimate identification of personal information will still remain a human analytical task, particularly given the legal uncertainty regarding interpretative processes of categorisation.

6.2.2 Data aggregation

Data aggregation may involve linking datasets or mining information from continuous streams of data generated by internet-enabled technologies. Data aggregation can present both opportunities and risks for people, organisations and governments. For example, the accumulation and aggregation of large amounts of data will provide a more accurate insight into the complexities of social life, which can enhance policy and service insights (Executive Office of the President and National Science and Technology Council Committee on Technology, 2016). Enhanced insights into activities and increased access to data and analytical outputs, could also enable better choice-making mechanisms for people (Productivity Commission, 2016: 84). For example, smartphones can now monitor driving behaviours, including distance driven, driving speed, location, how abruptly the car brakes and phone use during driving (Canaan, Lucker and Spector, 2016). By providing drivers with these data or by supplying customers with automated reminders and real-time coaching to track safe driving behaviours, individual driving habits could be improved. This has obvious benefits for the insured individual, the insurer and society at large (Clarke and Libarikian, 2014).

The combination of enhanced forms of data collection and analysis are also giving rise to improved knowledge for resource allocation (Productivity Commission, 2016: 89). For example, smart grids operate in conjunction with smart meters. Smart meters provide a number of benefits for both consumer and supplier alike because they generate near to real-time data on energy consumption. For the supplier, the collective use of smart grids provides a more detailed understanding of electricity demand at every stage in the grid. The activities of the individual, the building and the environment are connected, and it becomes possible to see the effects of individual action in the home and its related impact across the grid.

However, the increasing prevalence of data accumulation, particularly in the public sector, is giving rise to concerns regarding key public policy issues (British Academy and The Royal Society, 2017: 42). Examples include the mandatory opt-out process of the My Health Record implementation; the use of census data for government-wide data analytics and automated welfare debt collection processes. Concerns have been raised regarding data accumulation strategies in the private sector, particularly in relation to data-driven customer services. Collective monitoring of these services (Yeung, 2016; Calo, 2017: 423) may lead to new forms of surveillance (Zuboff, 2015; Yeung, 2016: 10; Cohen, 2017). For example, sensors and cameras embedded in vehicles can detect driver states such as emotion, frustration and fatigue (el Kaliouby, 2017; Goadsuff, 2018). These sensor technologies can detect risky, impulsive or inattentive patterns of decision making (Canaan, Lucker and Spector, 2016). However, it is possible for organisations to derive intimate knowledge from these data, perhaps inadvertently (Calo, 2017: 421).

Data governance structures could help ensure the appropriate use and handling of data, including determining legally acceptable bounds of data aggregation involving personal information.

6.2.3 Data governance in an age of big data

6.2.3.1 Data anonymisation

Data anonymisation allows information in a database to be manipulated in a manner that makes it difficult to identify data subjects (Ohm, 2010: 1701). This is often achieved by ensuring personal identifiers are removed from the datasets (Australian Computer Society, 2017). These techniques are often used by data controllers to anonymise data before release to protect an individual's sensitive information. However, this faith in anonymisation has been criticised (Ohm, 2010: 1704), because it is usually possible to reverse engineer or de-anonymise data that has been de-identified (Narayanan and Shmatikov, 2008; Ohm, 2010: 1708; Srivatsa and Hicks, 2012).

Protecting data from de-anonymisation requires reliable protection from data breaches, which remain an ongoing problem for both commercial and governmental data holders. The advent of the Internet of Things (which refers to the proliferation of internet-enabled technologies in everyday use) and ubiquitous computing will lead to burgeoning databases and new vulnerabilities. In the near term we can anticipate that new kinds of data will be collected for the purposes of ML and automated decision making, generating new stockpiles of data to be targeted for theft. Policymakers will need to respond to these changes.

6.2.3.2 Data protection and privacy

Data innovation should not progress at the expense of data protection and privacy. Information privacy law could play a role in defining and determining the acceptable bounds of data aggregation, especially where personal information is involved. As data collection becomes increasingly widespread in public and private sectors, Australia and New Zealand's information privacy laws will need to be reconsidered.

Existing privacy laws regulate personal data, which is generally defined as information that makes an individual identifiable. However, it is not easy to determine whether certain information is personal data because people can be re-identified when de-identified data is cross-matched with other datasets (Australian Computer Society, 2017). The principal legislation governing privacy and data protection in Australia is the federal *Privacy Act 1988* (Cth), which regulates the handling of personal information by the private sector and federal government agencies.[11] It contains 13 Australian Privacy Principles based on the 1980 OECD Guidelines and the EU Directive.

Australian Privacy Principles collectively govern collection, use, disclosure, storage, security, access and correction of personal information. Personal information is defined in the *Privacy Act 1988* (Cth) (s6) as 'information or an opinion about an identified individual or an individual who is reasonably identifiable: (a) whether the information or opinion is true or not; and (b) whether the information or opinion is recorded in a material form or not.' The *Privacy Act* and other state and territory privacy legislation views information in binary terms, meaning the data must either be personal or non-personal. The extent to which information can identify an individual will differ between datasets, and legislation in Australia differs in how it deals with this challenge. Context is relevant in classifying information as personal. However, in some instances the wording in the legislation can be viewed as suggesting that the context in which people are identifiable from a dataset is an intrinsic property of that dataset.[12] While the contextual definition helps to ensure appropriate data governance, challenges arise when the same dataset may fall into the definition only at particular times or in certain circumstances.

There is a challenge to ensure that the benefits of aggregated data are harnessed without undermining an individual's right to privacy. The current privacy framework in Australia and New Zealand emphasises consent, or individual control, over personal data. Under the system of 'notice and consent' (Tene and Polonetsky, 2013: 260), the data subject (the user) is given notice, often in the form of a privacy policy, of the intended use of data at the time of data collection. However, the notice and consent model is problematic for a number of reasons (Solove, 2013; Barocas and Nissenbaum, 2014). For example, it is well documented that consumers often do not read detailed privacy policies (Nissenbaum, 2010: 105; Cate and Mayer-Schönberger, 2013: 67), while oversimplified policies can fail to explain privacy choices meaningfully (Nissenbaum, 2011: 36). Even with sufficient information, consumers are likely to trade off long-term privacy for short-term benefits (Acquisti and Grossklags, 2005).

[11] Public sectors of various states and territories are governed by separate legislations: *Information Privacy Act 2014* (ACT), *Privacy and Personal Information Protection Act 1998* (NSW), *Information Act* (NT), *Information Privacy Act 2009* (Qld), *Personal Information and Protection Act 2004* (Tas), and *Privacy and Data Protection Act 2014* (Vic). South Australia issued administrative rules requiring compliance with a set of Information Privacy Principles, while in Western Australia, some privacy principles are included in the *Freedom of Information Act 1992* (WA). See Office of the Australian Information Commissioner, 'Other privacy jurisdictions' at https://www.oaic.gov.au/privacy-law/other-privacy-jurisdictions

[12] Unlike legislation in the ACT, NT or (after 2012) the Commonwealth, the definition of personal information in Queensland, Victoria and NSW states a person must be identifiable 'from the information'. It is possible that these words mean information does not become personal information merely because there is potential for linking with other information. When a similar wording used to exist in the Commonwealth Act, former OAIC guidance suggested such strict interpretation was inappropriate. The former guidance is no longer accessible.

Moreover, the value of an individual's personal information is often unknown by both the organisation and the individual at the time of collection when consent is usually requested. This may make it difficult for the data controller (an organisation that determines the purpose for which personal data are processed) to specify upfront the types of purposes that the data may be used for. Additionally, there could be new data controllers (or third-party organisations) who use the data after collection depending on how the data are combined and processed. Future purposes and use by new data controllers are often unexpected and would require amended consent, which is likely a costly and complex exercise.

The current system places a heavy burden on individual users to self-manage their privacy in the context of numerous entities collecting their data (Cate and Mayer-Schönberger, 2013: 68; Solove, 2013: 1888). While data controllers are in a position to analyse risks, people generally lack the information or expertise. This imbalance could lead data controllers to exploit privacy risks to their advantage. Penalties for data misuse or opportunities for redress for breaches of privacy could discourage sharing or publication of identifiable data without consent.

However, weighing the costs of privacy protection and the benefits of big data innovation is not straightforward. The benefits and harms of privacy choices are distributive in a society (Strahilevitz, 2013). One privacy choice could benefit some to the detriment of others. If consent places the responsibility on individuals, then individuals are likely to make privacy choices in isolation from broader social factors. This may inadvertently create the 'tyranny of the minority' (Barocas and Nissenbaum, 2014: 61), where a small number of people who volunteer information make it possible for knowledge to be inferred about the majority who have withheld consent.

Improved data protection policies are required which inform the individuals of how their data are being used, the conclusions being drawn from it, and a right to access, correct, and delete their data. Due to the covert nature of re-identification, is it necessary that companies are able to establish and demonstrate the provenance of the data they use. It should be beholden on them, and therefore indirectly on the supplier of the data, to demonstrate that data were collected with consent and is permitted to be used for the purpose intended.

The Australian government is forming a National Data Advisory Council that will help the National Data Commissioner create laws that will help govern the release of data, along with protections for privacy, as part of a Data Sharing and Release Act (Australian Government, 2018e).

6.2.3.3 Data licences

Privacy policies often contain highly individualised and specific terms on how data are collected, processed and used. In the copyright domain, there are six standardised Creative Commons (CC) licences, which reflect different combinations of lawful uses and conditions (Creative Commons Australia, 2013). A similar licensing framework could potentially be applied to personal data, in which a limited number of licence types specify the different terms of data usage, for example:

- Use limited to entity to which data are provided *and purposes closely aligned with purpose of collection*. Data deleted when no longer required for that purpose
- Use limited to entity to which data are provided, *but purposes can be related to primary purpose*. Data deleted when no longer required for that purpose

- Use limited to entity to which data are provided *but uses can be unrelated to primary purpose of collection*. Data deleted in accordance with ordinary company policy
- Data can be shared with related entities and use can be unrelated to primary purpose of collection. Data will not necessarily be deleted after any particular fixed period
- Data can be broadly shared and used provided it is de-identified under the data controller's data risk governance framework (or some general standard).

These could each be associated with more detailed 'standard conditions' privacy policies. The advantage of standardisation is that computers could be programmed to communicate directly with each other, without human intervention, automatically negotiating terms of data management between a data subject and a data controller based on relatively simple settings.

6.2.3.4 Reforming the data consent model

A review identified several possible reforms to the current legislative framework for data consent (Cate and Mayer-Schönberger, 2013). First, the burden of privacy management could be shifted away from data subjects to data controllers. This involves reduced emphasis on individual consent and increased priority on disclosure to a regulator or a central repository. In addition, data controllers could demonstrate accountability through 'responsible data stewardship', resulting in higher privacy standards for compliance.

The Australian Computer Society (ACS) has suggested the focus should not be on the data itself, but the impact from the use of data. This would be a move away from examining who owns the data, to the 'rights, roles, responsibilities, and limitations for those who access data in the various processes from collection, use, sharing and storage' (Australian Computer Society, 2017).

An idea gaining popularity are Data Trusts. A data trust takes the concept of a legal trust and applies it to data. The trustors could be individuals or organisations that hold the data and grant some rights they have to control the data to a set of trustees who then make decisions on who has access to the data and what the data can be used for. This legal structure provides independent stewardship of data for the benefit of society or organisations (Artificial Lawyer, 2018).

6.2.3.5 Privacy certification and rating

Australia's Chief Scientist, Dr Alan Finkel, has proposed the creation of a recognised mark for consumer technologies called the 'Turing Certificate', which would indicate whether the technology adheres to certain ethical standards (Finkel, 2018c). It has been suggested that the Turing Certificate would be a voluntary system suitable only for low-risk consumer technology, such as smartphone applications and digital home assistants (Finkel, 2018b). The mark would be informed by standards developed by experts in consultation with consumer and industry groups, and when applied, the products, procedures and processes of the applicant company would be reviewed by an independent auditor. Privacy standards could also be embedded into part of the ethical certification process.

Another method that may help consumers make better purchasing decisions would be a privacy-friendly label. Currently, Energy Star ratings assess the energy efficiency of electrical appliances. A similar system could be used to demonstrate whether a technology application is privacy-friendly. It is possible that the visualisation of privacy risks could make privacy choices more accessible to the average consumer and potentially increase the transparency and disclosure of privacy risks by data controllers. Inspired by

food nutrition labels, Cranor (2012) proposed the design of standardised and simplified privacy nutrition labels to replace privacy policies for consumers. Labelling may even encourage competition between data controllers to provide more privacy-friendly solutions, including computer-to-computer negotiations over data management terms.

6.2.3.6 Legislative reform for aggregated data

The Australian Data Sharing and Release Bill, which is being drafted based on recommendations from the Productivity Commission, aims to create a new data governance framework that enables researchers to harness the value of government data (Department of the Prime Minister and Cabinet, 2018). A report from Allens Hub provided the following recommendations for possible reforms related to the aggregation of data (The Allens Hub for Technology, Law & Innovation, 2018):

- Rationalisation of the current patchwork of laws about how government shares information internally and externally, and clarification of the Bill's relationship with existing data protection laws
- Clarifying definitions and concepts relating to data sharing and release
- Acknowledging quality, context, community perspectives and amenability of data to reuse
- Ensuring decisions are based on principles of fairness and justice given the risks of uneven data availability
- Developing a data ethics framework and accountability mechanism and increasing education and training to promote responsible data use.

6.3 Data integrity, standards and interoperability

The provenance, quality, and integrity of data will become more important in an environment where the collection, disclosure, and analysis of data become continually blurred. Data, and personal information in particular, are not simply *provided* by people but rather *generated* automatically by internet-enabled sensors, devices, and appliances (Andrejevic and Burdon, 2015). Data can also become dated, and thus so can the AI models that are based on them. In this environment, issues of data integrity will become more visible and thus more accountable across a much wider network of participants.

Establishing accepted and common standards that ensure data integrity can accelerate the potential for data sharing and linkage. This may in turn offer opportunities for the development of new technologies and predictive systems for individual, societal and system-wide needs.

The use of common metadata registries, such as those conforming with international standard ISO11179, will facilitate the accurate capture and management of descriptive and structural health metadata (including assumptions and methodologies used in data capture) and 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 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'. Adherence to these principles is largely supported by legislative frameworks.

6.3.1 Data standards

In 2017, the World Economic Forum 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 (World Economic Forum, 2017).

International standards will affect various industries. Organisations are therefore observing practice overseas, and within, leading AI developers to identify and set best-practice standards. The GDPR will apply to

Box 26: Power of data linkage

Data linkage is a powerful tool that can significantly impact on individual wellbeing. Professor Fiona Stanley and Professor Carol Bower used data-linkage epidemiology studies to guide the basic science investigating the role of folate in birth defects.

They investigated the way in which folate in a mother's diet could reduce the incidence of neural tube defects. Birth defects of the brain or spinal cord happen early during pregnancy, often before women know they are pregnant. Such defects range from anencephaly, the improper development of the brain and skull, to spina bifida, the disordered formation of the spinal column. This research contributed to global studies and precipitated the Australian federal and state governments' 2007 introduction of the compulsory enrichment of bread-making flour with folate (Bower, 2014).

Enabled by data linkage, this research has also been used to generate health benefits within areas of heart disease and cancer.

all organisations that have an establishment in, trade with, or collect information on, the EU and it will be increasingly important for industries working internationally to understand the requirements.

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

6.3.2 Data portability

Data portability refers to the ability of an individual to obtain, reuse or transfer personal data from and between different organisations and services. The right to data portability is enshrined in Article 20 of the EU GDPR. Under this Article, an individual can request their personal data from a data collector in a structured, commonly used, and machine-readable format for their own use. Consumers are protected from having their data stored in closed platform silos that are incompatible with other platforms, which has the effect of locking the consumer into a service provider (Article 29 Data Protection Working Party, 2017).

Conceivably, the right to data portability will encourage the adoption of common data storage and data-processing standards across different services, organisations, and IT environments (Article 29 Data Protection Working Party, 2017). Portability standardisation is intended to empower people by providing them with more control over their data and also to foster competition between data collectors by making it easier for consumers to switch between different service providers (Article 29 Data Protection Working Party, 2017). Therefore, while the right to data portability is viewed as an important update to traditional information privacy rights, it also has a significant innovation-oriented focus that seeks to enhance consumer protections and stimulate competitive digital economies.

The types of personal data covered by the GDPR right to data portability include:

- Personal information actively and knowingly provided (e.g. name and address)
- Observed data arising from the use of a service or a device (e.g. search histories, traffic data, location data and raw data from wearable devices).

However, the right to data portability does not extend to all circumstances and has limitations:

- The right does not apply when the data processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the data controller, or when a data controller is exercising its public duties or complying with a legal obligation (Article 29 Data Protection Working Party, 2017)
- The right only applies to digital data provided to a data controller by an individual. It therefore does not cover personal information acquired by the controller from other sources
- More importantly, the right does not apply to portability regarding profiling or analytics work undertaken by organisations collecting data. It therefore does not include 'inferred' or 'derived' data where an algorithmic assessment has been made about an individual based on behavioural monitoring (Information Commissioner's

Office (UK), 2018). Accordingly, while the right seeks to increase the control people have over the use of their personal data, including creating new options of consumer-oriented trade, the purpose of the right, and indeed the GDPR in general, is to regulate personal data rather than competition in the EU data ecosystem (Article 29 Data Protection Working Party, 2017; Lynskey, 2017). Corporations can thus still safeguard their competitive advantage by being able to retain algorithmically-driven insights.

The Australian Government is developing a new right to data portability, much like Article 20 of the GDPR. The development of an Australian data portability right is important because of unfolding Australian policy developments via the Productivity Commission's recent report on 'Data Availability and Use' (Productivity Commission, 2016) and the Open Banking Review. Both appear to herald a new response to Australian information-privacy regulation that places a much greater emphasis on consumer protection as a desired societal outcome of information privacy law.

The Australian Productivity Commission's Comprehensive Right is focused on expanding consumer control and use of data to stimulate digital economy innovations that are separate from information privacy regulatory models. Both the EU and Australian policy positions will give rise to a much greater focus on the exchange of information to customers, which will have the flow-on effect of establishing legal standards of data compatibility and interoperability.

6.4 Data storage and security

Large and often sensitive datasets required by AI will necessitate appropriate data storage and input methods. Data handling considerations are not unique to AI. The issue of data storage and security would exist regardless of whether AI is applied to the data. However, it becomes even more difficult to solve when an AI system is dependent on access to all the data.

6.4.1 Onshore and offshore data storage

Data can be stored onshore (locally) or offshore (overseas). Data sovereignty is the concept that information is subject to the laws of the nation within which it is stored. Data-localisation legislation requires network providers to store original or copies of collected data about internet users, on servers located within the jurisdiction. These laws have been justified to ensure the privacy and security of citizens' data, provide better information security against foreign intelligence agencies, and support domestic law enforcement activities (Selby, 2017).

Data-localisation measures vary in scope (Chander & Lê, 2015). Countries such as China, Russia, and Indonesia, have enacted broad data-localisation laws requiring most personal information and data to be stored within their respective borders. Most countries, however, have narrow data-localisation laws, imposing the requirement only on certain types of personal information and specific industry sectors.[13] For example, Australian laws are narrow and require electronic health records to be stored locally (Australian Government, 2018e – My Health Records Act 2012 (Cth): s77). The transfer, processing or handling of such data outside of Australia is permitted only if such records do not include 'personal information in relation to a consumer' or 'identifying information of an individual or entity' (Australian Government, 2018e – My Health Records Act 2012 (Cth): s77).

If the data are stored offshore and not end-to-end encrypted, there is a possibility that the data are *readily available* to the government of the country in which the data are stored. This is true even when the cloud storage provider offers encryption at rest (inactive data or data that is not moving). The legal jurisdiction covering the data matters when there are no globally agreed privacy standards. If the data becomes available in an unencrypted form on an offshore server, it presents a problem for effective privacy oversight and may hinder appropriate redress for people whose information is included in the dataset. If the data are stored offshore but has end-to-end encryption (with keys held in Australia) then it is assumed that the *encrypted data are* available to the government of the country in which it is stored. If the encryption is sound, this may be considered an acceptable risk. However, it is important to note that most systems for end-to-end encrypted file storage expose some metadata, such as who accessed a file and when. Even for end-to-end encrypted data, some countries are considering laws that would force software companies to provide their government with a secondary mechanism of access to that encrypted data. It will be important to consider not purchasing encryption software

[13] For example, in Europe, different governments require different types of data to be stored locally. These range from financial records, gambling winnings and user transactions, and government records as discussed by Selby. Other countries impose restrictions to data collected from specific sectors, such as financial, health and medical information, online publishing, and telecommunications data. See Cohen, Hall, & Wood, 2017.

from countries with such laws. If data are stored onshore in Australia, there is still no guarantee that it will be secure. Data breaches occur often, with attackers from within Australia and overseas. End-to-end encrypted cloud storage is a good tool to protect the data, along with standard mechanisms for secure access and deletion.

6.4.2 Secure data storage

Providing technological solutions to ensuring reasonably secure storage of data, while allowing appropriate access for analysis, is an active area of research. There are several main directions, including: traditional access control; differential privacy and secure multiparty computation. These areas are not mutually exclusive and can be applied together. For example, secure research environment with formally restricted access control could use differential privacy to perturb answers before showing them to an analyst, and use secure computation for analysis on datasets stored elsewhere.

Secure multiparty computation

Secure (multiparty) computation uses cryptography to allow two (or more) computers to evaluate a function on each of their private inputs, without revealing what those inputs are. For example, a set of pharmacists could compute the total number of sales of a particular medication, without revealing their individual sale totals. This does not guarantee that the resultant answer protects privacy: if the computation is an election outcome, and the vote is unanimous, then this reveals exactly how everyone voted. Secure computation has numerous practical applications and has been used by Google, which partnered with a third party to compute the total number of users who had seen an advertisement and subsequently bought the item in a store (Ion et al., 2017). Crucially, they were able to do this without revealing who the customers were, or even how many had seen the advertisement or visited the store.

Secure computation platforms are freely available online (Damgård et al., 2012; Ejgenberg et al., 2012). Some use (partially) homomorphic encryption, which means that some computations (such as addition) can be performed while the data remains encrypted. However, their computational speed is limited – some simple computations run quite fast, but more complex ML algorithms rapidly become infeasible.

Differential privacy

Differential privacy addresses the complementary problem: it limits the amount of information that can be leaked by the answer to a query about any particular individual. In its simplest form it consists of randomly perturbing the algorithm's output to introduce uncertainty about its true value, hence hiding individual details (Dwork, Roth and others, 2014). In very large datasets, local differential privacy can still yield accurate results: each individual input is randomly perturbed first, then the algorithm is applied to the differentially-private data. Both Apple and Google have run example projects using these techniques (Abadi et al., 2016), in addition to academic research.

Differential privacy represents a bound-on information leakage, not a guarantee of perfect privacy. If the same data are reused across multiple differentially-private mechanisms, information about people can be more accurately inferred.

Combining techniques from cryptography and multiparty computation with differential privacy is an area of research. Many federated data analysis platforms borrow some techniques from each, though not all are designed on rigorous and provable security guarantees.

Data storage is not solely a technology issue. Unless an entity is held accountable for data breaches and failures to protect and secure data, then there is little motivation for them to do so. Many entities appear unconcerned about data security, which has led to numerous and increasingly serious data breaches despite advances in security, cryptography and information security management. Appropriate regulations to hold entities accountable may change this.

6.4.3 Data sovereignty and multinational companies

Technical aspects

Data-localisation laws could create technical difficulties for multinational companies seeking to generate business insights from data collected across multiple jurisdictions. Many companies store data in 'the cloud', making it difficult for companies to see where the data are stored and processed (Synytsky, 2017). However, to comply with the laws, companies need to know precisely what type of data are stored, and in what location.

Legal compliance

Countries with broad data-localisation laws create privacy standards for data collected within their jurisdiction. This means multinational companies could have the additional burden of complying with privacy standards unique to each country on top of international and regional privacy legislative frameworks. For example, China's Cybersecurity Law (CSL) introduces restrictions on cross-border data transfers that differ from international privacy regimes such as the European Union's GDPR and the voluntary Asia-Pacific Economic Cooperation Cross-Border Privacy Rules (Sacks, 2017).

Under CSL, network operators and operators of critical information infrastructures are required to store personal information and other important data that are collected and generated in China within the jurisdiction. Such data can be stored or provided overseas for business reasons only if it is truly necessary and the operators conduct a self-security assessment or pass an official security assessment when a threshold test is met (Chin et al., 2018). The security assessment is based on a two-pronged test (Chin et al., 2018): firstly, whether the transfer is lawful, legitimate, and necessary; and secondly, the risk of transfer is evaluated by looking at the nature of the data and the likelihood and impact of security breaches involving such data.

While Europe's GDPR and CSL appear to have similar cross-border transfer tests, there are material differences (Zhang, 2018). CSL does not provide for derogations that are found in the GDPR. Neither does the CSL contain mechanisms in the GDPR such as Binding Corporate Rules[14] and standard data protection clauses for companies to gain approval.[15] Lastly, data-localisation laws are likely to increase compliance costs since companies engaged in data collection from different countries will have to build local data centres in each jurisdiction.

This is not to say that data-localisation laws may not be rational for individual countries seeking to protect citizen data and ensure local access (e.g. by intelligence and law-enforcement agencies). However, an international framework with consistent data protections and clear rules for transnational access would resolve some of these issues.

[14] Binding Corporate Rules allow multinational companies to transfer personal data out of the European Union within the same corporate group to countries that do not have an adequate level of data protection.

[15] Standard contractual clauses are used to transfer data outside the EU and are deemed to provide sufficient data protection by the European Commission.

6.4.4 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 underused due to challenges in data linkage – in particular, the potential for breaches of privacy.

There is an opportunity for government and industry to share and leverage datasets across organisations, for building 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 ML 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.

There is an opportunity to establish an ecosystem of federated ML technologies across government and industry, based on the use of open formats and application-programming interfaces, which 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 platform and the Australian Government's NationalMap federated spatial visualisation platform.

Box 27: Data federation in practice

CSIRO's Data61 is 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, known as MAGDA (making Australian government data available), supports better ways for locating and accessing data from across the country, combining these with personal data for more targeted analytics.

Further, IT companies (such as IBM) are also investing in federation technologies to provide a unified interface to diverse data (Lin and Haas, 2002).

6.5 Data management and disposal

Australia and New Zealand's information privacy principles guarantee certain protections for individuals. Traditional information privacy law provides protections that seek to imbue fairness in the exchange of personal information. People have a limited range of process rights that provide a degree of control over how personal information is collected, handled and used by data collectors. Individuals can access and amend collected personal information, request to see personal information held about them and ask that 'out of date' information about them be deleted or amended. Similarly, data collectors are obliged to inform users about when and why collections are undertaken, to collect personal information only for relevant and specified purposes, to store personal information securely and to ensure that subsequent uses are in accord with the purpose of collection.

The question is whether these protections will still have the same substantive application in structures of *automated* collection and analysis. Along with the traditional types of information privacy protections highlighted above, the EU GDPR introduces several enhanced information privacy protections for people relating specifically to automated profiling, which would include an AI decision-making context (Article 29 Data Protection Working Party, 2016). These include:

- Articles 13 and 14 provide enhanced transparency measures that require data controllers to inform people about the existence and scope of automated decision making
- Articles 17 and 18 provide the ability to rectify or erase personal information used as part of an algorithmic output and the output itself
- Articles 21 and 22 provide rights to object to data processing, particularly in providing a right not to be subject to a decision based solely on automated processing, including profiling.

It is unclear exactly how Article 22 will apply (Artificial Intelligence Committee - House of Lords, 2017; Veale and Edwards, 2017; Kaminski, 2018), though it has been argued that it establishes a general prohibition of decision making based solely on automated processing, unless certain exemption situations arise (Article 29 Data Protection Working Party, 2017).

While some of the GDPR protections are similar to the protections of Australian information privacy law – namely the Australian Privacy Principles (APPs 12 and 13 regarding access and correction) – the regulatory focus in the EU on automated processing is novel. One of the perennial criticisms of the Australian *Privacy Act* is that it is under-litigated and therefore does not have significant judicial consideration of how the key protections and components of the Act should be interpreted (Burdon and McKillop, 2013). As such, it is unclear whether the Australian framework would provide the same degree of protections to personal information in an AI-processing and decision-making context.

As society moves towards a digital future where new forms of individual data are collected, stored and used, detailed historical accounts of individual activities and behaviours will increase. The above discussion also raises questions around 'digital death' – that is, who has access to accounts and ownership of digital assets after death. There may also be issues of automated and inferred decision making of deceased persons based on extended, historical data holdings. This could lead to arguments about the creation or identification of new forms of legal identity predicated on decision-making inferences regarding the historical accumulation of deceased individual life-long data repositories.

6.5.1 Managing disclosure risk

Organisations around the world – including the Australian Bureau of Statistics – use the 'Five Safes' framework (Desai, Ritchie and Welpton, 2016) to help make decisions about the appropriate use of data considered confidential or sensitive. The framework has five dimensions: safe people, safe projects, safe setting, safe data and safe outputs.

- **Safe People** – refers to the knowledge, skills, and incentives of the users to store and use the data appropriately (is the person accessing the data appropriately authorised or trusted to use it in an appropriate manner).
- **Safe Projects** – refers to the legal, moral, and ethical considerations surrounding use of the data (is the data being used for an appropriate purpose).
- **Safe Setting** – refers to practical controls on the way data are accessed (does the access facility limit unauthorised use. At one extreme, researchers may be restricted to using the data in a supervised physical location, while on the other, there are no restrictions on data downloaded from the internet).
- **Safe Data** – refers primarily to the potential for identification in the data (has appropriate and sufficient protection been applied to the data). It could also refer to the sensitivity of the data itself or to the quality of the data and the conditions under which it was collected.
- **Safe Outputs** – refers to the residual risk in publishing results derived from sensitive data (are the statistical results non-disclosive).

The proliferation of artificially-intelligent algorithms suggests the need to further modify the Five Safes framework. In the world of AI, the Safe People element may be replaced with algorithms. The environment an algorithm operates in may be very different from that of a human researcher, and the restrictions and scrutiny placed on an algorithm may be far more intrusive than those that can be applied to a human. Consequently, some of the implicit assumptions in the Five Safes framework need to be re-examined. A proposal from the Australian Computer Society (outlined in Figure 10) suggests the following:

- **Safe Algorithms** – for an artificially intelligent algorithm, the behaviours and associated access conditions can be enforced under many circumstances, but will need supervision if adapting over time. Any biases that develop also need to be monitored.
- **Safe Project**s – this category still refers to the legal, moral, and ethical considerations surrounding the use of data. 'Grey' areas might exist when exploitation of data may be acceptable if an overall public good is realised or with consent from the person who has provided the project outcome (knowledge) or who benefits from the AI-driven service. The Safeness of the project that an algorithm undertakes should be known before the algorithm is applied to the data.

- **Safe Setting** – when the researcher is an algorithm, the operating environment can be locked, disconnecting the algorithm from other sources of input. This, however, does not allow evaluation of any biases in the algorithm itself.
- **Safe Data** – when the observer is an algorithm, the context which the algorithm brings to the data can be strictly limited through limiting access to other datasets.
- **Safe Outputs** – there is a distinct difference to be further examined – whether the output from an algorithm is single and discrete or feeds an operational loop (such as a steering algorithm or cruise control algorithm).

The potential for continuous 'learning' by algorithms presents challenges. It has been noted numerous times that AI is prone to amplify sexist and racist biases from the real world (Reese, 2016; Cossins, 2018) and evolve to positions well beyond those intended by developers. A Safe Algorithm needs to be constantly monitored for their Safe Level – which may change over time or be recalibrated. As AI and technology evolves, it could be appropriate to recalibrate the elements of a safe framework to help make decisions about effective use of data that is confidential or sensitive.

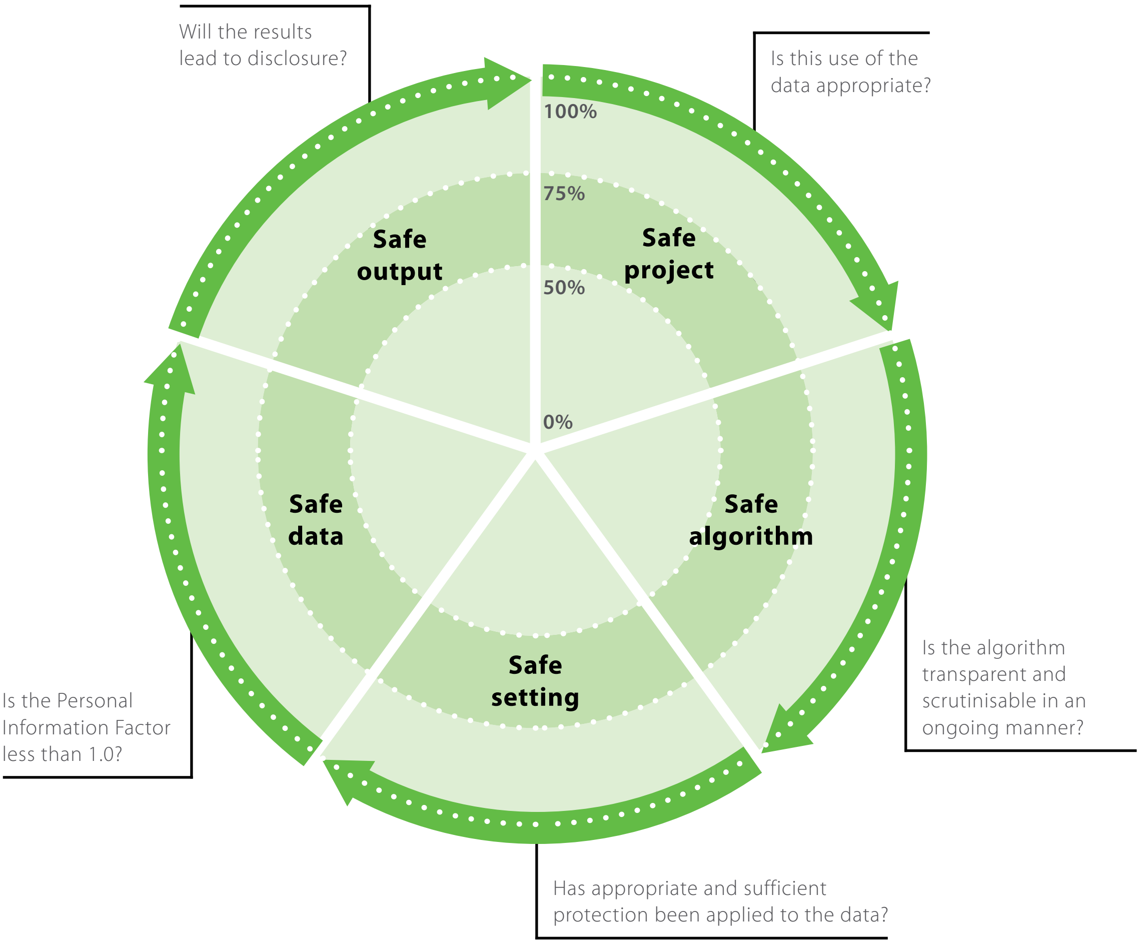

Will the results lead to disclosure?
Is this use of the data appropriate?
100%
75%
50%
0%
Safe output
Safe project
Safe data
Safe algorithm
Safe setting
Is the algorithm transparent and scrutinisable in an ongoing manner?
Is the Personal Information Factor less than 1.0?
Has appropriate and sufficient protection been applied to the data?

Figure 10. Five Safes framework for algorithms

Adapted from: Australian Computer Society, 2018b.

6.6 Building on our strengths

Australia and New Zealand have a significant opportunity to be among the leading technology developers and adopters of data-driven AI systems and technologies. The countries have well established, and globally recognised, strengths in some of the key, data-driven capability areas core to AI, including data sharing or federation, trustworthy systems, ML, image analytics, natural language processing and automation. In addition to this, Australia and New Zealand are culturally diverse and serves as a desirable population in which to gather robust datasets, a core requirement for unbiased and effective AI.

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

Realising this opportunity will require investment in a focused and coordinated effort, linking national AI capabilities and domain knowledge to the particular challenges and use cases for AI identified by industry and government. This will only be achieved by actively seeding and nurturing a deep partnership between government, 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, with control of the raw data continuing to reside with its custodian organisation.

In New Zealand, information sharing and information matching are two separate frameworks in the Privacy Act. The information-sharing framework provides for the authorisation and oversight of Approved Information Sharing Agreements. The information-matching framework provides a detailed set of rules dealing with the supervision and operation of authorised information-matching programs, with the Privacy Commissioner having a regulatory role to monitor the use of data matching by government departments. In 2011, the New Zealand Government approved new principles for managing the data and information it holds and approved a Declaration on Open and Transparent Government. Further, there is a new Privacy Bill before the New Zealand Parliament, and a new role of Government Chief Data Steward has been created.

Supported through the National Innovation and Science Agenda, initiatives such as Platforms for Open Data 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, while preserving the data's confidentiality and integrity.

To take full advantage of the opportunities presented by data-driven AI, governments, businesses and the community will need to increase their levels of awareness, adoption and acceptance of AI's use. This 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, so that opportunities are maximised, without trust being compromised.

CHAPTER 7
INCLUSIVE DATA

This chapter is based on input papers prepared by the generous contributions of Professor James Maclaurin and Dr John Zerilli (Discrimination and Bias); Professor Maggie Walter and Professor Tahu Kukutai (Indigenous Data Sovereignty); Associate Professor Reeva Lederman (Trust); Professor Mark Andrejevic (Trust and Accessibility); Dr Oisín Deery and Katherine Bailey (Ethics, Bias and Statistical Models). The original input papers and views of the experts listed can be found on the ACOLA website (www.acola.org).

7.1 Introduction

A challenge to understanding risks for particular population groups is the quality of data that is available about those groups. Human rights treaty bodies have repeatedly highlighted the need for governments to better collect and use data on gender, ethnicity, race, age and physical or mental disability (Commissioner for Human Rights, 2018). AI systems need large datasets that may be expensive to build or purchase, or which may exclude open data sources, resulting in data that are of variable quality or drawn from a narrow set of sources. Data on which AI is trained may exclude people about whom data are not collected or not collected well, thereby embedding bias (Buolamwini and Gebru, 2018).

Even where good data are available, the design or deployment of AI-learning systems may result in discrimination in other ways (Commissioner for Human Rights, 2018). For example, developers may build a model with inadvertent or indirect discriminatory features, without human oversight or without the ability for a human to intervene at key decision-making points, with unpredictable or opaque systems or with unchecked intentional direct discrimination (World Economic Forum, 2018b). Recent research by Buolamwini and Gebru (2018) demonstrates that some existing commercial AI applications have embedded race and gender biases. For example, testing Microsoft, IBM and Chinese company, MegVii, for accuracy of gender in facial recognition revealed accuracy rates for Caucasian men of more than 95 percent but only 20-35 percent for darker skinned women (Buolamwini and Gebru, 2018).

7.2 Risks of data-driven bias in AI

7.2.1 Discrimination based on data aggregation

Over the past decade, there has been an unprecedented acceleration in the sophistication and uptake of various algorithmic decision-making tools, which draw on aggregated data. Examples include automated music and TV show recommendations, product and political advertising and opinion polling, medical diagnostics, university admissions, job placement and financial services. However, the use of aggregated data in these contexts carries the risk of amplifying discrimination and bias, and problems of fairness arise (see for example, Hajian, Bonchi and Castillo, 2016; O'Neil, 2016; Corbett-Davies et al., 2017b). This may be a bias in the algorithm or a bias in the input data that is reflected in what the algorithm learns and subsequently applies.

There is a widespread belief that algorithmic decision-making tools are more objective because they are less biased than human decision makers. Such assertions imply that legal protection against unfair discrimination might not be relevant to 'objective' algorithmic decision making. Human prejudice and algorithmic bias differ in character, but both are capable of generating unfair and discriminatory decisions. Tackling this problem will be particularly challenging owing to the contested nature of fairness and discrimination.

To assess the risks of bias in automated decision making, one must begin by looking at bias in human decision making. Research into human decision making has generated important results over the past thirty years (Pomerol and Adam, 2008). It is now understood that human prejudice is the result of various failures of reasoning (Arpaly, 2003). For example, we often reason probabilistically from very small samples and we regularly fail to update our beliefs in light of new information (Fricker, 2007; Gendler, 2011). At other times we abandon probabilistic reasoning altogether, relying instead on 'generic' reasoning (Begby, 2013), judging that groups have particular characteristics irrespective of information about the frequency of those traits (Leslie, 2017). These generic judgements are harmful as they are largely insensitive to evidence (Greenwald and Banaji, 1995; Saul, 2013). For example, long-held beliefs about the criminality of certain culturally and linguistically diverse groups are not usually overcome by merely supplying evidence of the inaccuracy of such beliefs (Bezrukova et al., 2016). Moreover, emotions exert a powerful influence on human decision making (Damasio, 1994) and negative emotions, like fear, make us particularly prone to prejudice.

There are federal laws in Australia that prohibit various discriminatory grounds of reasoning. These include the Racial Discrimination Act, the Sex Discrimination Act (protecting also gender, marital status and sexual orientation), the Age Discrimination Act and the Disability Discrimination Act (Khaitan, 2015). However, prejudice and resulting discrimination also affect the operation and institutions of the law itself. Research suggests that the tendency to be unaware of one's own predilections is present even in those with regular experience of having to handle incriminating material in a sensitive and professional manner (McEwen, Eldridge and Caruso, 2018).

The problem of discrimination is widespread and complex, and to date we have had legal

Box 28: Case study: Bias in natural language-processing AI systems

A new technique has been developed for representing the words of a language which is proving useful in many NLP tasks, such as sentiment analysis and machine translation. The representations, known as word embeddings, involve mathematical representations of words that are trained from millions of examples of actual word usage. For example, a good set of representations would capture the relationship 'king is to man as queen is to woman' by ensuring that a particular mathematical relationship holds between the respective vectors (specifically, king – man + woman = queen).

Such representations are at the core of Google's translation system, although they are representations of entire sentences, not just words. According to researchers at the Google Brain Team, this new system 'reduces translation errors by more than 55-85 percent on major language pairs measured on sampled sentences from Wikipedia and news websites' (Wu, Y., et al., 2016) and can even perform translations between language pairs for which no training data exists.

However, researchers at Boston University and Microsoft Research (Bolukbasi et al., 2016) noticed that Google's Word2Vec dataset was producing seemingly sexist outputs. For example, just as the relationships 'man is to woman as king is to queen,' and 'sister is to woman as brother is to man,' were captured by word embeddings, so too were the relationships 'man is to computer programmer protections that are generally accepted to be effective, even though it is difficult to assess their actual efficacy on the accuracy and fairness of public decision making. The use of such tools rests on the assumption that behaviours and experiences are universal and measurable. But even standardised tools – or 'structured professional judgments' as they are known – present a bias in how individuals are perceived, how behaviours are formulated and how decisions are informed (Tamatea, 2016). It is in this context that algorithmic-decision tools have been vigorously promoted (Palk, Freeman and Davey, 2008; Craig and Beech, 2009; Baird and Stocks, 2013; Hardt, Price and Srebro, 2016; Lawing et al., 2017).

7.2.2.1 Algorithmic bias

Algorithmic decision-making tools may fail to reduce bias in decision making (Angwin, 2016; Crawford and Calo, 2016; Lum and Isaac, 2016; O'Neil, 2016; Shapiro, 2017). Algorithms designed to be accurate and fair routinely assess individual creditworthiness, desirability as employees, reliability as tenants, and value as customers. However, their probabilistic accuracy may in fact militate *against* fairness in most cases (Corbett-Davies, Pierson, Feller and Goel, 2017; Corbett-Davies, Pierson, Feller, Goel, et al., 2017). Bias in algorithmic decision makers can be either intrinsic or extrinsic (similar to humans), but differs in character from the corresponding human failings. It is useful to distinguish intrinsic and extrinsic bias in decision-making systems.

Intrinsic bias is built-in in the development of the AI system or results from inputs causing permanent change in the system's structure and rules of operation. For example, a

as woman is to homemaker,' and 'father is to doctor as mother is to nurse.'

In order to produce accurate outputs, NLP systems relying on word embeddings need to learn the biases in the bodies of text on which they are trained (Caliskan, Bryson and Narayanan, 2017). Thus, if these models are to successfully learn the relationships that exist between words in actual uses of language, they must learn relationships that are biased. Bias in the texts on which a model is trained are naturally going to be captured in the geometry of the word-embeddings vector space. There is a risk that the application of this technology may exacerbate or amplify biases within the data (Bolukbasi et al., 2016).

One way to address the underlying cause may be to address systemic bias in society, rather than in the NLP systems themselves. It has been suggested that 'one perspective on bias in word embeddings is that it merely reflects bias in society, and therefore one should attempt to de-bias society rather than word embeddings' (Bolukbasi et al., 2016). However, that result is not something that can be achieved by means of a statistical model, if it can be achieved at all.

Therefore, caution is advised with the output from statistical models. The developers and the users of any statistical model must not regard the model's output as more objective than the human-produced data on which it is trained. Additionally, developers and users must take this into account, especially in cases where bias in the data are impossible or difficult to eliminate.

human resources system designed by a male team to implement a set of rules that fail to accommodate the needs of female employees is intrinsically biased in its design. Ingrained unconscious prejudice in human reasoners that is effectively impervious to counter-evidence is also intrinsic. Intrinsic bias can occur:

- as a result of prejudiced developers or of ill-conceived software development
- from the inherent constraints imposed by the technology itself (Friedman and Nissenbaum, 1996)
- if the data are represented in a manner that might have unexpected effects on the output of an algorithm. For example, an algorithm that polls companies represented in an alphabetical list leads to increased business for those earlier in the alphabet (Mittelstadt et al., 2016)
- from the result of programming errors, such as when poor design in a random number generator causes particular numbers to be favoured (Mittelstadt et al., 2016)
- as a result of fundamental historical bias, as when an algorithm is tied to rules that reflect current science, law or social attitudes.

Extrinsic bias derives from a system's inputs in a way that does not effect a permanent change in the system's internal structure and rules of operation. The output of such systems might be inaccurate or unfair but the system remains 'rational' in that new evidence is capable of correcting the fault. The recent explosion in the use of AI is largely driven by the development of algorithms that are not rule-based in the style of expert systems, but instead are capable of learning. Such 'deep learning' networks can avoid intrinsic bias insofar as they can learn from their mistakes; but the cost of being able to learn is vulnerability to extrinsic bias. This has become a pressing issue in the development of ethical AI (Friedman and Nissenbaum, 1996; Johnson, 2006). Extrinsic bias results from the fact that such apparently objective tools derive their power from historical data and hence actually aggregate decisions made by the very people whose potentially biased decision making we are seeking to supplant (Citron and Pasquale, 2014). Extrinsic bias can occur:

- errors and biases latent in 'dirty' data tend to be reproduced in the outputs of machine learning (ML) tools (Diakopoulos, 2015; Barocas and Selbst, 2016b). This is a significant problem, and one that is compounded by copyright and intellectual property laws that limit access to better quality training data (Levendowski, 2017)
- from the use of unrepresentative datasets. For example, face recognition systems trained predominantly on Caucasian faces might reject the passport application photos of culturally and linguistically diverse people (Griffiths, 2016). Speech recognition systems, too, are known to make more mistakes decoding female voices than male ones (Tatman, 2016). Such situations arise from a failure to include members of diverse social groups in training data. The obvious solution is to diversify the training sets (Crawford and Calo, 2016; Klingele, 2016), although there are political and legal barriers preventing this (Levendowski, 2017)
- when the diversification of training data presents a difficult technical problem. Demographic parity is achieved when a dataset is equally representative of two groups (e.g. men and women). However, where fairness is sought regarding many different identity characteristics, it is impossible to achieve demographic parity for all of them
- if the data available is strongly skewed in favour of a particular demographic

group, discarding data in order to achieve demographic parity is likely to decrease the accuracy of the system (Corbett-Davies, Pierson, Feller, Goel, et al., 2017).

Not all 'dirty' data suffers from being unrepresentative. COMPAS scores, based on questionnaires completed by prisoners, are predictive of risk of reoffending, but a recent study in the US shows a strong correlation between COMPAS score and race (Larson et al., 2016). African Americans routinely have higher scores and so find it harder to get parole. The effects of historical injustice are writ large in such statistics. African Americans are likely to have lower incomes, to live in crime-ridden neighbourhoods, and to have diminished educational opportunities. This vicious circle is exacerbated by previous discriminatory patterns of policing (Crawford and Calo, 2016; Larson et al., 2016; Lum and Isaac, 2016). This bias does not originate from unrepresentative data, which could be corrected by including more culturally and linguistically diverse groups in the training set. It stems from intrinsic human bias, with machines simply inheriting the bias. So, an algorithm that accurately predicts recidivism also unfairly penalises an already disadvantaged group. Moreover, because of these persistent correlations between race and disadvantage, modern AI, harnessing big data and ML, persistently detects race even when it receives no data specifically about this protected category (Veale and Edwards, 2017).

Research in data science shows that we can develop algorithms that are, in some sense, fairer. The challenge, however, is that different notions of fairness are in conflict, meaning that it appears to be impossible to be fairer in every sense of that term (Hardt, Price and Srebro, 2016; Corbett-Davies, Pierson, Feller, Goel, et al., 2017; Kleinberg, Mullainathan and Raghavan, 2017).

7.3 Indigenous data sovereignty

AI is data driven and therefore relies on ongoing access to data. However, such data requires input from individuals or devices owned by individuals. Questions therefore arise such as who owns the data? How should it be used? Who should have access to the data and under what circumstances? And who makes the decisions about the ownership, use, control and access to data and its value? These questions have been of increasing concern and interest for Indigenous peoples around the globe.

The data used in AI is a socio-cultural artefact that is the product of human subjectivities (Walter and Andersen, 2013). The construction of algorithmic rules involve choices about which assumptions are incorporated and which are not. How those choices fall is fundamentally linked to the epistemic and ontological realities of algorithm designers and data generators. Therefore, AI rules often resemble their creators in terms of their prioritisation of knowledge holders and sources, and their perspective of how the social and cultural world operates. In the vast majority of cases those creators are not Aboriginal and Torres Strait Islander or Māori (Kukutai and Walter, 2015).

Indigenous data sovereignty (IDS) is a response to the intensification of data collected about Indigenous people and issues of importance to them, whether by commercial, government, research entities, NGOs or international agencies. IDS is concerned with the rights of Indigenous peoples to own, control, access and possess data that derives from them, and which pertain to their members, knowledge systems, customs or territories (Kukutai and Taylor, 2016; Snipp, 2016). IDS is supported by Indigenous peoples' inherent rights of self-determination and governance over their peoples, country (including lands, waters and sky) and resources as described in the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). Implicit in IDS is the desire for data to be used in ways that support and enhance the collective wellbeing of Indigenous peoples. In practice, that means Indigenous peoples need to be the decision makers regarding how data about them are used or deployed, including within social-program algorithms.

Indigenous peoples are included in a diverse range of data aggregations, from self-identified political and social groupings (e.g. tribes, ethnic and racial groups), to clusters of interest defined by data analysts and controllers. The definition of Indigenous identity varies across datasets, administrative regimes and cultures. Indigenous communities may have social processes of deciding who is included, but these systems do not necessarily scale to big data or data-matching technologies. Moreover, AI systems create models and inferences from sources that Indigenous communities themselves might not have the ability to see or use and which may be incomplete. For example, economic data might fail to show informal economies in Indigenous communities, where particular cultural arrangements influence how resources are accrued and distributed. Definitions of household and family may also differ from those assumed in data processing. Indigenous families might therefore share resources in ways that may be invisible in electronic transaction records, leading to incorrect assumptions about their vulnerability.

Indigenous identifiers need not be explicitly included in algorithms for Indigenous peoples to experience the disproportionate impacts of AI-informed decision making. For example, a study on child maltreatment in New Zealand using predictive risk modelling excluded ethnicity as it added little explanatory power to the models once socioeconomic risk factors were accounted for (Vaithianathan et al., 2013). However, Māori children were much more likely to be exposed to the risk factors associated with maltreatment, reflecting inequities in access to the determinants of wellbeing. More broadly, techniques of collecting data from device use, wearable technology or sensors embedded in the built environment may recapitulate what McQuillan (2017: 101) calls 'the capture of a territory'. For McQuillan (2017: 101), these data-capturing processes mirror 'historical colonialism' in that their 'effect […] is to shift the locus of control and decision making' from Indigenous populations to the colonisers. Both Australian and Aotearoa New Zealand governments are using algorithms and tools such as predictive risk modelling in a wide variety of frontline services. Despite an increasing call for transparency and accountability in machine-driven decision-making (Lepri et al., 2017), the logic underlying algorithms is rarely accessible to the communities that they affect (Eubanks, 2018a).

While AI systems can produce numerous positive outcomes for society, the marginalised social, cultural and political location of Māori and Aboriginal and Torres Strait Islander peoples suggest that these outcomes will not be shared equally. We are unlikely to see, for example, the immediate benefits of precision diagnostics and AI-assisted surgery in the strained public systems where most of our Indigenous populations receive health care. The considerable risks embedded in the ubiquity of AI are also unevenly distributed, and there are significant challenges for Māori and Aboriginal and Torres Strait Islander peoples relating to bias, stigma and accountabilities. For these reasons, Indigenous people need to be included in the discussions of data sovereignty and the management of data that may be used for decision-making purposes.

Box 29: IDS movements in Australia and New Zealand

IDS movements are active in Australia and Aotearoa New Zealand and are grappling with the complexities of Indigenous-data usage in AI. In Australia, the Maiam nayri Wingara Indigenous Data Sovereignty Collective, in partnership with the Australian Institute of Indigenous Governance, issued a communique from a 2018 national meeting of Aboriginal and Torres Strait Islander leaders. This communique stated the demand for Indigenous decision and control of the data ecosystem including creation, development, stewardship, analysis, dissemination and infrastructure. In Aotearoa New Zealand the Te Mana Raraunga Māori Data Sovereignty Network Charter asserts Māori rights and interests in relation to data and requires the quality and integrity of Māori data and its collection. Māori have often been the subject of intrusive data surveillance and misuse but have well-tested 'tikanga' (ethics, processes, principles) on the protection and sharing of knowledge for collective benefit. Groups like Te Mana Raraunga are exploring ways that tikanga can be used to rethink scientific approaches to data governance, use and validation. For a country that aspires to be a 'world leader in the trusted, inclusive and protected use of shared data' (New Zealand Data Futures Forum, 2018), issues relating to ethics, trust and confidence are both timely and critical in New Zealand. For advocates of Māori data sovereignty, the goal is not only to protect Māori people and communities from future harm and stigma, but also to safeguard Māori knowledge and intellectual property rights, and to ensure that public data investments create benefits and value in a fair and equitable manner that Māori can fully share in.

7.4 Trust

AI will depend on the confidence that society places in the technology. The issue of trust in AI systems raises a number of definitional problems, including trust that the algorithms will produce the desired output, trust in the values underlying the system, trust in the way data in the system are protected and secured, and trust that the system has been developed for the good of all stakeholders. Such questions of trust take users far beyond the simple matter of whether they believe the technology works.

Trust is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party (Mayer, Davis and Schoorman, 1995). When trust is discussed with respect to technology, there are similar expectations that people can give themselves over to the technology and it will perform reliably in a predetermined way.

The problem of trust in technology and automation is not new (see Lee and See, 2004). However, the complexity of AI means that it is more difficult for users to gain a deep understanding of the technology and consequently, can lead to additional issues of trust. The potential benefits of AI for health, wellbeing and other areas of society, mean that issues of trust need to be explored and dealt with further to ensure they do not create any unfounded barriers to use.

AI systems offer great potential benefits in a diverse range of application areas from transportation, finance, security, legal practice, to medicine and the military. Most of the systems under consideration in these fields involve what is termed 'weak AI' in that it assists in the performance of specific tasks that involve probabilistic reasoning, visual or contextual perception and can deal with complexity in ways that far outpace the human mind. AI systems are not yet able to deal with ethical judgements, the effective management of social situations or mimic all facets of human intelligence. Nonetheless, they still provide significant opportunities to increase our ability to make effective use of data.

AI systems that try to anticipate human needs are mostly found in household systems that use data to determine or anticipate the need. For example, AiCure reminds patients to take medication and confirms their compliance (Hengstler, Enkel and Duelli, 2016). Other examples include household robots that can fetch, deliver and clean. In a clinical setting, AI health systems include applications that can, for example, potentially replace the work of radiologists by performing diagnoses (Hsieh, 2017), or applications that simulate some of the features of a human psychologist (D'Alfonso et al., 2017).

Car manufacturers are well on the way to developing autonomous and semi-autonomous vehicles. BMW already has a semi-autonomous vehicle on the market, Daimler has a fully autonomous truck planned for 2025 and Nurnberg in Germany has operated a fully autonomous train since 2008. In the military, there are scenarios where lone mission commanders direct unmanned military vessels controlled by AI, and in the US, the Defense Advanced Research Project Agency is working on ways to use AI to extract military information from visual media captured in the field and turn available photos and videos into useable sources of intelligence.

The areas described above – health, transportation and military services, alongside other areas discussed in Chapter 2 of the report – are central to society's safety and wellbeing. People are protective of these areas of their lives and are reluctant to cede control to automatous devices. Thus, trust is an important issue in the acceptance and adoption of such systems.

While trust has traditionally been a concept used to describe human-to-human interactions, studies have shown that it is valid to use the concept of trust to describe the way in which the relationship between humans and computers, or automation, is mediated (see for example Zuboff, 1989). Trust in what were previously human-led processes (where trust was previously not guaranteed) needs to somehow be extended to a new environment where the same processes are now automated. Trust is also difficult to achieve where complex algorithms are being implemented and a full understanding of the technology is often hard to attain (Lee and See, 2004). Lack of trust in automating technologies, including AI, can lead to misuse or disuse, which can compromise safety or profitability (Lee and See, 2004).

Trust in AI depends on several factors. Firstly, the technology needs to have proven reliability: 'a technology based on the delegation of control will not be trusted if it is flawed' (Hengstler, Enkel and Duelli, 2016). In AI applications, useability, reliability and consistent operation all engender trust (Siau and Wang, 2018). Users of automation consider four factors to be important in trust:

- Performance (what the technology does), including specifically operational safety and maintenance of data security (Lee and See, 2004; Hengstler, Enkel and Duelli, 2016)
- Process, including useability and whether or not it can be trialled (Lee and See, 2004)
- Purpose, or why the technology was developed and whether it benefits the consumer (Hengstler, Enkel and Duelli, 2016) and is visible (such as the automated train, Rogers, 2003)
- Designs that humanise technologies are more trustworthy. Robotic designs need to make users feel that they have a significant level of control (Hengstler, Enkel and Duelli, 2016).

People experience greater feelings of trust if the innovating firm is known to them (Hengstler, Enkel and Duelli, 2016). Consequently, positive brand identification is important, but firms also need to build relationships with consumers through information provision and their involvement in project development. This issue highlights the difference in two forms of trust in AI –trust in the technology and trust in the technology provider (Siau and Wang, 2018). Both are important to whether or not users are willing to interact with AI.

A further important factor in trust is the notion of explainability, where the actions of the AI are easily understood by humans. AI is being used by systems to arrive at important decisions in the lives of people, such as admission to education or provision of finance. Increasingly, consumers are calling for the right to an explanation in of decisions made by AI, but legal frameworks are yet to respond adequately.

Previous work suggests that:

- people seek explanations of AI when cases are contrastive (they wonder why one thing happened and not another)
- people use their cognitive biases to selective explanations for how AI performs
- people are not always swayed by the most likely explanation for how AI has behaved, unless they understand the cause of the most likely explanation
- explanations for AI are social and are influenced by a person's beliefs (Miller, 2017).

Trust in AI will be seen to be dependent on how much developers respond to these problems of explainability.

7.5 Access to personal data

A nationwide survey conducted in Australia in 2014 revealed that when it comes to large-scale data collection, there is strong support (over 90 percent) for greater control over personal information and for more information about how it is being used (Andrejevic, 2014). Greater transparency in this context does not mean simply letting people know that their information is being harvested. It means providing them with a clear idea about *how* it is being used – a key point with respect to the development of data-driven AI systems. An individual's personal data profile, in isolation, does not provide information about how it interacts with the data of millions of others (Turow, Hennessy and Bleakley, 2008). Because of the emergent character of AI decision-making processes, it is not possible to specify in advance the affect that particular forms of data may have on life-impacting decisions.

The 2014 survey also indicated strong support (over 90 percent) for the ability to request that one's personal data be deleted from a particular database. As discussed above, the individual's right to access, reuse or delete personal data is enshrined in the EU GDPR. According to the aforementioned survey, people should be able to have some control over their information, even when it is collected in a transactional context. In practice, this right depends upon forms of knowledge that are difficult to obtain in the case of third-party data collection. It also depends upon a largely outdated conception of personal information (Andrejevic and Burdon, 2015). As discussed earlier, it is possible to 're-identify users' in meaningful ways (ways that can be tracked back to name, address and other specific personal information) by aggregated data from multiple unrelated datasets. The 'right to be forgotten' may retain some meaning in the case of a search engine like Google, but how does one request to have the record of one's clickstream or browsing history removed and how does one determine which companies have a copy of it?

When it comes to government and law-enforcement access, if information about a particular individual is requested, existing restrictions on the collection and use of personal data can be used as a foundation for determining access. However, increasingly, targeted monitoring is replaced by group or classification monitoring: the request to access all information about those who fit a particular behavioural profile. In many cases, this profile may not even contain what is conventionally considered to be personally identifiable information. However, as already established, it is possible to re-identify personal information from non-personally identifiable information.

This poses serious issues for regulation of access because standard protections for personal information rely on the model of targeted information collection. In these cases, it might be more appropriate to monitor use than access – that is, to determine which decisions can be made based on data mining and which ones are ruled out. Or, a regulatory decision could be made regarding which types of information are available for automated forms of decision making and which are ruled out. For example, a decision might be made to rule out the use of genetic information in hiring decisions. Some of these decisions might fall within existing regulatory regimes, to the extent that some classes of information would amount to decision making based on categories that are protected from discrimination (e.g. certain genetic markers might have high correlation with ethnic background and their use in decision-making processes could constitute discrimination).

By definition, automated systems generate 'emergent' outcomes – that is, they discern patterns and correlations that cannot be deduced in advance (which is the entire point of enlisting such systems). So, for example, a job-screening system might determine that the web browser used to submit a job application correlates more strongly with subsequent job performance than the content of the application. The finding is useful because it is unanticipated, but it would not be possible to inform applicants in advance before the finding is generated. Once the finding is generated, informing applicants after the fact is useless. Once again, the structural issue here suggests that regulation of use may be more meaningful than the attempt to provide informed consent (which would state something like, 'all data collected from this application will be used in conjunction with existing datasets by automated systems to predict future job performance' and would not lead to meaningful informed consent).

The logic of automated decision making lends itself to the use of data for unanticipated purposes. There are large potential benefits to having the data accessible for this use. For example, it might be determined that certain lifestyle patterns can be used to anticipate and intervene pre-emptively in the treatment of some illnesses. Finding these new connections would require speculative data mining. Once again, it will likely become necessary to regulate use (by data class or decision class, or both – that is, to say that some forms of data cannot be used speculatively or that some decisions cannot rely solely on AI-generated recommendations).

Although it is possible to require the deletion of data, the declining cost of storage and the potential future value of linking existing datasets to reveal new information and patterns provide incentive to data collectors to retain information.

It will be increasingly difficult to regulate data collection because of the proliferation of internet-enabled devices and contexts in which data are generated, gathered and stored. As new forms of information collection emerge, it will be difficult for regulatory regimes to catch up: should information about an individual's mood, anxiety levels or emotional expressions be protected? What about their biological responses captured by personal fitness devices like Fitbit? The key challenge for regulators will be to develop guidelines that can be applied to the development of new forms of monitoring. It might be decided, for example, that biometric information should be unavailable to advertisers. This is unlikely to happen, but it indicates the type of decision that a society might make in order to set guidelines for controlling the implementation of new forms of automated decision making.

7.6 Initiatives by the Australian Government

Research still needs to be undertaken to establish ways to design suitable objectives into machine-learning approaches which will consider – often conflicting – ethical imperatives, such as reducing racial, gender or ideological bias, valuing privacy and ensuring reliability.

In response to the Productivity Commission's comprehensive inquiry into Data Availability and Use, the Australian Government Department of Prime Minister and Cabinet released a report outlining approaches to address ethical issues associated with the use of data across government, community and industry (Australian Government, 2018k). The Australian Government has committed A$65 million over the forward estimates to reform the Australian data system and introduce a range of measures to implement the Productivity Commission's recommendations. 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, and to generate and promote innovation while adhering to best practice ethical use. The government has committed to the following:

1. A consumer data right as a new competition and consumer measure to allow people 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 that will streamline data sharing and release, subject to strict data privacy and confidentiality provisions.

There are also existing frameworks for government departments that use automated decisions. These frameworks will 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 (Elvery, 2017). In addition, a 2004 report to the Attorney General outlines 27 principles for government departments that use automated decision-making processes (Administrative Review Council, 2004). These include drawing a clear distinction between decisions that require discretion (that should not be automated) and situations that require a large volume of decisions, where the facts are already established (that are suitable for automation). An inter-agency report 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 (Department of Finance and Administration, 2007).

CHAPTER 8
CONCLUSION

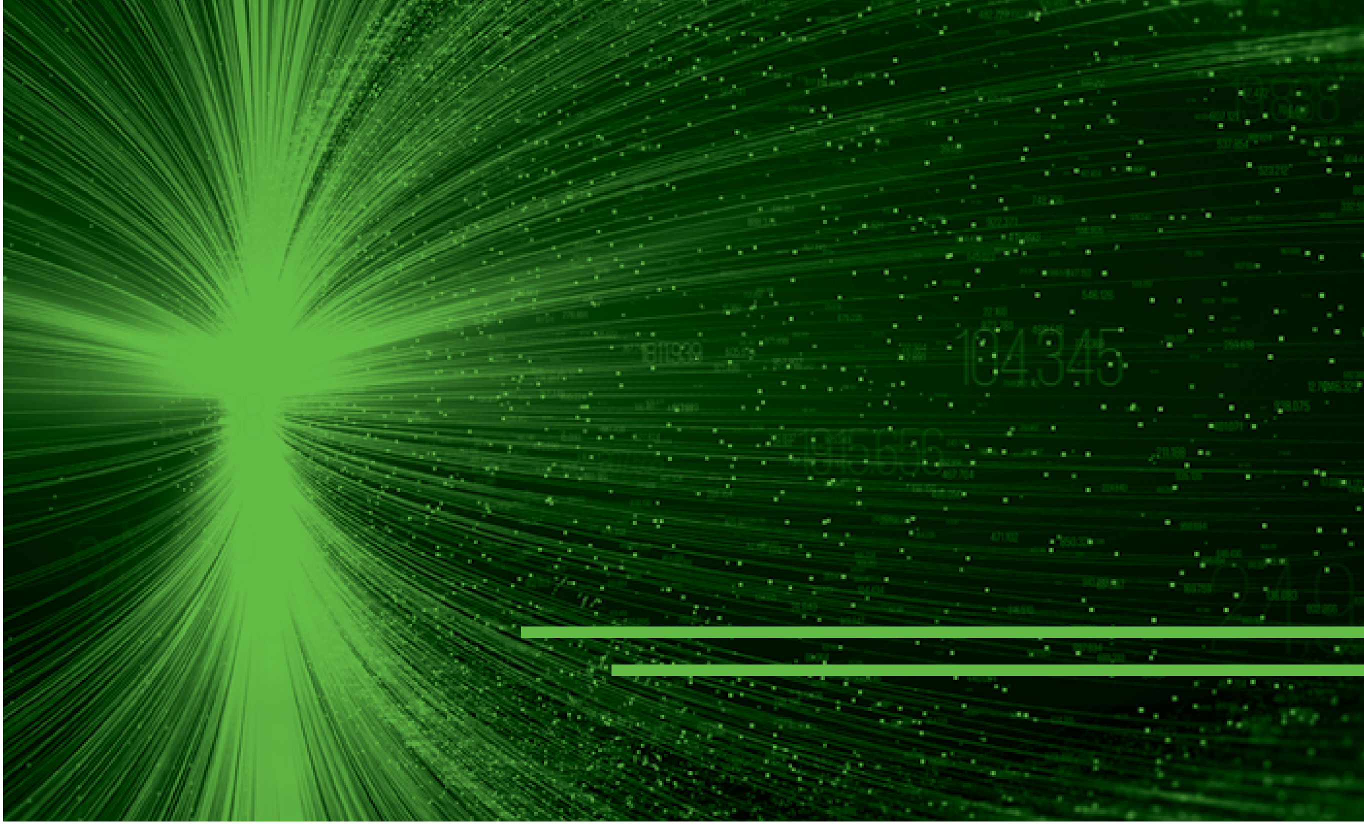
The application and implementation of AI are already underway and set to evolve in exciting ways. This report has mapped many of these directions, identifying the opportunities and the challenges that will accompany implementation.

Australia and New Zealand are well equipped to advance AI. Both countries are establishing new AI-based products and services across sectors including finance, health, arts, transport and manufacturing. People are interacting with AI-powered chatbots, music is being composed using AI, the mining sector is using AI to automate operations and there is increasing use of precision agriculture devices to collect and analyse data on crops and livestock. The breadth of applications provides new social and economic opportunities. However, other countries are outspending both Australia and New Zealand on AI technology development. To adopt and apply AI technologies effectively, we will need to understand the technologies at a level that will only come from engagement in development.

Australia and New Zealand's leading expertise in AI development and application, alongside expertise in relevant ethical, legal and social considerations, will be key to the safe implementation and advancement of AI. Development and implementation of AI will benefit from the insights of many disciplines. Knowledge will need to be exchanged across institutions, public and private sectors and geographical locales. Successfully navigating the ethical, legal and social considerations of AI will be a defining task for the field. Facilitating this interdisciplinary connectivity through national, independently-led, AI bodies would provide a platform for innovation and the required collaboration of HASS and STEM specialists.

Successful implementation of AI will need to be developed within sustainable, ethical and socially responsible boundaries that prevent development of undesired technologies. Many opportunities and challenges will be played out in global fora, so Australia and New Zealand will need to ensure they participate in the development of safe AI frameworks. National frameworks based on integrity, principles of fairness and wellbeing will also be important.

Understanding what kind of society Australia or New Zealand want to be will be critical to the development of national frameworks and will require engagement with society to help shape this vision. A process like the 'Australia



2020 summit', but in this instance focused on a single topic, could help define the desired AI-enabled society. What can be said with confidence is that AI development should be centred on addressing inequity, improving prosperity and on continued betterment.

Vision for Australia and New Zealand

AI research and implementation are transnational and the data and expertise it relies on cross borders. However, it is unavoidable that Australia and New Zealand will be affected by forces we cannot control. Given appropriate investment in research and development, by calling upon homegrown expertise and by attracting world-quality talent, we can play an important role in guiding the international development of AI.

This development should emphasise our values. In this report, we have placed equity and wellbeing at the heart of the development of our vision for AI. AI can magnify existing bias and lock the already disadvantaged out of further opportunities. However, if its development and implementation are guided by a concern for equity, it can play an important role in minimising these problems. AI promises enormous benefits to Australia and New Zealand, but these benefits can only be realised if they are shared.

Our position as relatively small countries with diverse populations provides advantages that can be exploited. We make an ideal test bed for new developments and an ethical AI strategy should enable us to attract significant overseas investment. Our reputation as a forward-looking, open and liberal society also allows us to play an important role in the development of international frameworks for regulating AI. We have the opportunity to ensure that the development of AI does not come at the expense of human rights, either at home or internationally. An AI strategy that places equity at its forefront will strengthen our international reputation in this arena and ensure that we are not left behind by some of the most important developments of the 21st century.

APPENDIX 1
AUSTRALIAN AI CAPABILITIES AND INITIATIVES

Publicly released national plan

None yet released.

Key documents

- The effective and ethical development of artificial intelligence: An opportunity to improve our wellbeing (2018), Australian Council of Learned Academies (ACOLA)
- Artificial Intelligence: Australian Ethics Framework (Discussion Paper), Data 61
- Artificial Intelligence: Technology Roadmap, Data61
- Australia 2030: Prosperity Through Innovation (2018), Innovation and Science Australia
- A Robotics Roadmap for Australia (2018), Australian Centre for Computer Vision
- Human Rights & Technology Project, Australian Human Rights Commission (2018-2020)

Research and development

Percentage of GDP spent on R&D (2015): 1.88 percent (OECD, 2018b)

R&D spend (2015): US$25.4 billion (The World Bank, 2018)

Share of global IP5 AI patent families (2000-2005): 0.72 percent (OECD, 2017b)

Share of global IP5 AI patent families (2010-2015): 0.44 percent (OECD, 2017b)

The 2015-16 Excellence in Research for Australia evaluations show that many Australian universities are undertaking world-class AI and image processing research activities (ARC, 2015). While only the Australian National University was ranked well above world standard (ERA Score 5), 12 Australian institutions are ranked above world standard (ERA Score 4) and a further 12 are considered to be at world standard (ERA Score 3). Only three Australian universities performed below world standard (ERA Score of 2) (ARC, 2015). Data61, the CSIRO data science consultancy, is another hot spot of AI R&D in Australia. Data61 has the highest concentration of data scientists in Australia and an emphasis on industry engagement and the application of data science including AI and ML to real world problems (Innovation and Science Australia, 2017).

Australia does not rank in the top ten countries worldwide by volume of AI publications, but is ranked 7th globally for field-weighted citation impact for papers published between 2011 and 2015, which indicates that it is performing highly (Times Higher Education, 2017).

Australia filed 0.4 percent of AI-related IP5 patents between 2010 and 2015, down slightly from 0.7 percent during 2000-05. This is well behind the share of leading countries such as Japan (27.9 percent) and South Korea (17.5 percent) and is comparable with countries such as Switzerland (0.4 percent) and Italy (0.3 percent).

Gross expenditure on research and development in Australia was estimated to be around 1.88 percent of GDP in 2015; lower than the OECD average of 2.36 percent.

Policy, laws, government

The Economist Automation Readiness Index Ranking 2018: 10th (70.4)

Oxford Insights Government AI Readiness Index: 8th (7.48)

With Singapore, Australia is considered to be at the forefront of AI development and experimentation in the Asia-Pacific (APAC) region (FTI Consulting, 2018). However, policy and legal reforms to support and regulate the use of AI in Australia remain fragmented. The 2018-19 budget contained a A$29.9 million investment in AI, including the creation of a technology roadmap, a standards framework and a national AI ethics framework (Australian Government, 2018d). In an innovation roadmap published in May 2018, Innovation and Science Australia recommended that the Australian Government's Digital Economy Strategy 'prioritise the development of advanced capability in AI and ML in the medium to long-term to ensure growth of the cyber–physical economy' (Innovation and Science Australia, 2017). The Digital Economy Strategy, *Australia's Tech Future*, was released in December 2018 and focuses on four key areas: developing Australia's digital skills and leaving no one behind; how government can better deliver digital services; building infrastructure and providing secure access to high-quality data; and maintaining our cyber security and reviewing our regulatory system (Australian Government, 2018a).

In 2018, the Australian Council of Learned Academies (ACOLA) was awarded a project grant by the Australian Government through the Australian Research Council's Linkage Learned Academies Special Projects (LASP) program Supporting Responses to Commonwealth Science Council Priorities (project number CS170100008). The project, *The effective and ethical development of artificial intelligence*, examines the technological, social, cultural, legal, ethical, and economic implications of the deploying artificial intelligence in Australia and New Zealand. Through the final report, ACOLA will provide an evidence base to support government decision making and help to ensure that safe and responsible implementation can provide maximum benefit for the economic and societal wellbeing of Australia and New Zealand.

The Automation Readiness Index assesses countries' preparedness for the augmentation and substitution of human activity presented by autonomous technologies that is expected to occur in the next 20-30 years. It ranks Australia 10th of the 25 countries assessed, with a score of 70.4/100. Australia was ranked 7th for its innovation environment, 11th for education policy and equal 10th for labour market policies. The study notes the importance of state governments in achieving readiness in countries with decentralised political structures, such as Australia. It commends the New South Wales Government's leadership including their proactive efforts in studying and experimenting with the application of AI technologies in education (The Economist Intelligence Unit, 2018a).

UK-based consultancy, Oxford Insights, ranked Australia 8th of 35 countries (between Japan and New Zealand) in their Government AI Readiness Index. This index provides a broad indicator of how prepared the national government is for implementing AI in its public service delivery and is based on a composite score derived from nine metrics related to public service reform, the economy and skills and digital infrastructure. The UK and the US top the list, although some key jurisdictions including China and Singapore were not assessed (Stirling, Miller and Martinho-Truswell, 2017). In 2018, the Australian Government awarded a A$1 billion contract to IBM to develop AI, blockchain and cloud initiatives for government agencies (IBM, 2018).

The Australian Human Rights Commission is undertaking a major research project examining the impacts of technology on human rights, with a particular focus on AI technology (Australian Human Rights Commission, 2018a).

Australian states and territories are considering the impacts of AI. The Victorian Parliament launched the all-party parliamentary group on AI to explore the opportunities and challenges that AI will present to the state. The New South Wales Department of Education has commissioned researchers to investigate how to best prepare young Australians for the future impacts of AI on society, and the ethics of AI (Walsh, 2017; Buchanan et al., 2018; Parliament of Victoria, 2018).

The Federal government is forming a National Data Advisory Council that will help the National Data Commissioner create laws that will help govern the release of data, along with protections for privacy, as part of a Data Sharing and Release Act (Australian Government, 2018e). This will have significant ramifications for data science and AI research in Australia.

Societal response

Most Australians know little about AI and related technologies. A survey by Ipsos revealed that attitudes towards AI are mostly positive or neutral (Riolo and Bourgeat, 2018). However, there were concerns about the risks of driverless vehicles, the use of robots in the armed forces and the use of AI in financial markets. The potential for robots and AI to replace jobs was also viewed negatively by the majority of respondents (Riolo and Bourgeat, 2018). Australians are also likely to believe that customer service is becoming too automated and impersonal (Chatterton, 2018).

Attitudes towards autonomous vehicles (AVs) also provide a useful proxy for people's trust in AI systems. A study of attitudes towards driverless vehicles reported that 37 percent of survey respondents were positive about AVs, 23 percent negative with the remaining 40 percent neutral (Pettigrew, 2018). A 2017 survey found that 51 percent of Australian men and 41 percent of women would travel in an AV (Roy Morgan, 2017). A 2016 survey of Victorian road users found that 74 percent of participants were concerned about the technology in AV failing and more than half of the respondents said they would not be comfortable in a car that could completely drive itself (Page-Smith and Northrop, 2017). A global poll of 28 countries conducted by Ipsos found that Australians are less optimistic about the perceived benefits of AVs than people from other countries. They are also more likely to trust governments to regulate AVs over the companies that design and manufacture them (Wade, 2018).

Australia's Chief Scientist, Alan Finkel, has stated a voluntary ethical AI certification could support trust in AI for low-risk applications (Finkel, 2018a).

Industry uptake

Asgard, Roland Berger estimate of AI start-ups: 27 start-ups (16th) (Roland Berger and Asgard, 2018)

A 2017 report, *Amplifying Human Potential: Towards Purposeful Artificial Intelligence*, which surveyed 1,600 senior business decision makers in organisations with more 1,000 employees or more than $500 million in annual revenue, across seven countries (China, India, Germany, US, UK, France and Australia), revealed that Australian organisations were the least likely of those surveyed to have plans to deploy AI-related technologies (21 percent of respondents) (Infosys, 2017). A similar but smaller survey conducted in 2017, found that Australia is skewed towards later adoption than the rest of the world. However, respondents predicted increased investment and use of AI processes and offerings over the next five years (daisee, 2017).

Australia also lags behind on automation, with only 9.1 percent of publicly-listed firms engaging in this field. This is significantly lower than the level of engagement in leading countries such as Switzerland (25.1 percent), the US (20.3 percent) and the UK (12.3 percent) (AlphaBeta, 2017). However, Australia is recognised as a world leader in the deployment of automation in the mining sector (Australian Centre for Robotic Vision, 2018).

The Australian Centre for Robotic Vision identified around 1,100 Australian companies engaged in the robotics sector across diverse sectors including manufacturing, services, healthcare, resources, infrastructure, agriculture, the environment, space and defence. Data from 442 of these companies indicated that they employ almost 50,000 Australians and generate more than A$12 billion revenue annually (Australian Centre for Robotic Vision, 2018).

A global survey of AI start-ups found 27 based in Australia, placing it 16th globally. The US dominates, with almost 1400 AI start-ups listed, followed by China (383) and Israel (362) (Roland Berger and Asgard, 2018).

Workforce skills and training

Canadian AI consultancy, Element AI, determined that there are 22,000 PhD-educated AI-experts globally, of whom 657 were in Australia. The leading countries were the US (9,010 experts) and the UK (1,861 experts), although the company notes that experts from Asia are likely to be underrepresented as it uses data from LinkedIn, which has a higher penetration in the US and other English-speaking countries. The study also found that of the 5,400 researchers who had presented at recent international AI conferences, 76 were based in Australia (Element AI, 2018). Part of the Federal Government's A$29.9 million investment will support research projects and PhD scholarships in AI and machine learning (Australian Government, 2018d).

Digital infrastructure

Global Open Data Index: 2/94 (79 percent) (Open Knowledge International, 2016)

The Inclusive Internet Index 2018 Ranking: 25/86 (The Economist Intelligence Unit, 2018b)

Australia is ranked equal second to Great Britain by the Global Open Data Index. The index measures the openness of government data by assessing whether key datasets are openly licensed, machine readable, easily downloadable, up-to-date, publicly available and free of charge. Australia scored 79 percent overall, with a majority of its datasets fully open (Open Knowledge International, 2016).

Australia is ranked 25th of 86 countries in the Inclusive Internet Index 2018. Australia was ranked 12th for the availability metric due to good infrastructure, but only 28th for both affordability (the cost of access relative to income and the level of competition in the Internet marketplace) and readiness (the capacity to access the internet, including skills, cultural acceptance and supporting policy) (The Economist Intelligence Unit, 2018b).

Australia ranked 50th globally, with an average connection speed of 11.1 Mb/s in Akamai's *Q1 2017 State of the Internet Connectivity Report*. This was eighth fastest amongst countries in the Asia-Pacific region, slower than New Zealand (seventh in the region with an average speed of 14.7Mb/s) and less than half the average speed of global leader South Korea (28.6 Mb/s) (Akamai, 2017). For mobile connections, Australia performs significantly better with the highest average mobile connection speed in the Asia Pacific region at 15.7 Mb/s, just beating Japan at 15.6 Mb/s (Akamai, 2017). OpenSignal's *State of LTE* February 2018 report, which focuses on the amount of time users have access to a particular network rather than geographical coverage, ranks Australia 13th for availability of a 4G network (OpenSignal, 2018).

Region	Unique IPv4 addresses	Average connection speed (Mbps)	Average peak connection speed (Mbps)	% above 4 Mbps	% above 10 Mbps	% above 15 Mbps
Australia	10,538,918	11.1	55.7	81%	35%	19%

Table 1: Australia's state of internet connectivity

From Akamai, 2017.

APPENDIX 2
NEW ZEALAND AI CAPABILITIES AND INITIATIVES

Publicly released national plan

None yet released, though the national crown innovation entity has released a key white paper (Callaghan Innovation, 2018).

Key documents

Artificial Intelligence: Shaping a Future New Zealand (2018), AI Forum New Zealand

Thinking Ahead: Innovation Through Artificial Intelligence (2018), Callaghan Innovation

Research and development

Percentage of GDP spent on R&D (2015): 1.28 percent (OECD, 2018b)

R&D spend (2015): US $2.6 billion (The World Bank, 2018)

Gross expenditure on research and development in New Zealand was 1.28 percent of GDP in 2015; lower than the OECD average of 2.36 percent.

The AI Forum has identified five New Zealand universities working on AI research.

The University of Technology, Auckland, is developing language and speech technologies, as well as mind theory. The university has developed 'neuromorphic' data processing technologies modelled on brain processes, and is researching robotics vision, unmanned aerial vehicles and bee monitoring (The AI Forum of New Zealand, 2018).

The University of Otago has established an interdisciplinary research centre to examine the benefits and problems associated with AI, and related ethical issues. The Centre for Artificial Intelligence and Public Policy will focus on urgent AI issues. A relationship with the government is likely to be formalised (Gibb, 2018; The AI Forum of New Zealand, 2018). The University of Otago is also researching computer vision and human models of memory and language (The AI Forum of New Zealand, 2018).

The University of Auckland has developed 'life-like artificial systems'. The research includes the development of the virtual digital baby, BabyX, and has resulted in the creation of start-up Soul Machines, which creates avatars that act as interfaces for digital platforms. The university is also researching game AI, applied AI case reasoning, multi-agent systems and data mining (The AI Forum of New Zealand, 2018).

Victoria University of Wellington undertakes research into machine learning, neural networks, data mining and cognitive science, as well as projects on evolutionary computation (The AI Forum of New Zealand, 2018).

A number of applications are being researched, with an emphasis on agriculture or biosecurity. A University of Canterbury researcher is developing an AI that can identify from photos invasive insects, plants and fungi on imported goods (LiveNews, 2018). The university is also working on machine learning and algorithm engineering, as well as brain-computer interfaces to examine microsleeps (The AI Forum of New Zealand, 2018).

Policy, laws, government

The Economist Automation Readiness Index Ranking 2018: N/A

Oxford Insights Government AI Readiness Index: 9/35 (7.38)

New Zealand is ranked 9th of 35 countries in the Government AI Readiness Index 2018, which provides a broad indicator of the national government's capacity to implement AI in its public service delivery. Its score of 7.38 places it just behind Australia. The UK and the US top the list with scores of 8.40 and 8.21 respectively (Stirling, Miller and Martinho-Truswell, 2017).

The New Zealand Government intends to develop an ethical framework and action plan to manage the opportunities and challenges presented by AI. Despite this, Oxford Insights ranks the New Zealand Government 9th of 35 OECD governments for its capacity to absorb and exploit the potential of AI technologies (Stirling, Miller and Martinho-Truswell, 2017). The Government supported the AI Forum of New Zealand – an independent organisation with representatives from academia, industry and government – to analyse the potential impact and opportunity of AI on New Zealand's society and economy. This report, released in May 2018, examines the AI landscape globally and in New Zealand; discusses the potential economic benefits, labour market impacts, and social implications of AI in New Zealand; and provides recommendations to assist policymakers to advance the AI ecosystem (The AI Forum of New Zealand, 2018). The report recommends actions to:

- forge a coordinated AI strategy for New Zealand
- create awareness and understanding of AI
- support the adoption of AI by industry and government
- improve access to trusted, high-quality data sources
- grow the AI talent pool
- address the potential legal, ethical, and social effects of AI.

Government representatives have signalled an intention to rapidly develop the AI plan (New Zealand Government, 2018b).

The University of Otago is undertaking a three-year multi-disciplinary project investigating the implications of AI technologies on New Zealand law and public policy (University of Otago, 2018). The New Zealand Law Foundation has established an Information Law and Policy Project [ILAPP], with NZ$2 million of funding available since 2016 to develop law and policy around IT, data, information, artificial intelligence and cyber-security.

In May 2018, the New Zealand Human Rights Commission released a paper for public discussion on privacy and data issues. It outlined approaches to formulating policy frameworks for algorithms and privacy, citing international bodies. It also emphasised the need to consider privacy safeguards for metadata (New Zealand Human Rights Commission, 2018).

New Zealand leads a group of seven digital nations which are investigating enhancing digital government based on open markets and open source principles. Estonia, Israel, New Zealand, South Korea and the UK were the original five members, and Canada and Uruguay joined in 2018 (Digital Government New Zealand, 2018).

Societal response

The New Zealand AI Forum notes that 'AI raises many new ethical concerns relating to bias, transparency and accountability. AI will have long term implications for core legal principles like legal responsibility, agency and causation' (The AI Forum of New Zealand, 2018).

A Samsung poll on technology adoption in New Zealand found that around a third of respondents would be open to using AI assistants in smart homes. Over 50 percent of respondents believed AI could help them save time each week. Around two-thirds were worried about being hacked or having their voice stolen. The most popular automated task for a smart home was setting alarms and locks when people leave (Paredes, 2018).

Industry uptake

Asgard, Roland Berger estimate of AI start-ups: 6 start-ups (equal 27th) (Roland Berger and Asgard, 2018)

A global survey of AI start-ups found 6 based in New Zealand, ranking it equal 27th globally. The US dominates, with almost 1,400 AI start-ups, followed by China (383) and Israel (362) (Roland Berger and Asgard, 2018).

A New Zealand AI Forum survey found that 20 percent of organisations had adopted AI systems. However, respondents were overwhelming large enterprises that have invested significantly in IT. These organisations were most commonly implementing AI systems to:

- improve business processes, including financial analytics and reporting
- augment current applications
- automate processes, including transactions and customer service interfaces (e.g. chatbots)

enhance cybersecurity (The AI Forum of New Zealand, 2018)

New Zealand's innovation agency, Callaghan Innovation, predicts that AI will affect key industry sectors including:

- an extreme impact on agriculture, enabling smart and more efficient application of water and sprays, optimised animal health monitoring, and improved crop yield prediction
- a medium impact on the digital sector, including applications across the finance, accounting, legal and e-commerce sectors
- a high impact on the energy sector, enabling system and cost optimisation, and smart grids
- an extreme impact on the health sector, including use in augmented diagnoses and personalised healthcare.

In particular, Callaghan Innovation considers it important for New Zealand businesses to explore machine learning and deep learning AI technologies (Callaghan Innovation, 2018). The agency offers innovation support services to business including access to AI specialists.

Workforce skills and training

Canadian AI consultancy, Element AI, determined that there are 22,000 PhD-educated AI-experts globally, of whom only 85 were in New Zealand. The leading countries were the US (9,010 experts) and the UK (1,861 experts) (Element AI, 2018).

The New Zealand AI Forum notes that 'there is an acute worldwide shortage of machine learning experts with competition for talent' (The AI Forum of New Zealand, 2018).

In 2017, New Zealand had 2,166 postgraduate students in computer science or IT at the honours level and 1,405 at the masters or PhD level. These numbers are expected to be boosted via ICT graduate schools hosted by several high profile universities (The AI Forum of New Zealand, 2018).

Digital infrastructure

Global Open Data Index: 8/94 (68 percent)

The Inclusive Internet Index 2018 Ranking: Not ranked

New Zealand is ranked equal 8th for the openness of government data sources in the Global Open Data Index. Its overall score of 68 percent indicates a generally positive attitude towards open data, but only 13 percent of the assessed data sets were completely open, which suggests that there are some shortcomings in data practices. This score is comparable to countries such as Canada (69 percent).

New Zealand ranked 27th globally with an average connection speed of 14.7 Mb/s in Akamai's *Q1 2017 State of the Internet Connectivity Report*. This was the seventh fastest speeds recorded amongst countries in the Asia-Pacific region, and faster than Australia's average of 11.1 Mb/s, though significantly slower than the average speeds of global leader South Korea (28.6 Mb/s) (Akamai, 2017).

New Zealand had average mobile connection speed 13.0 Mb/s in the first quarter of 2017, ranking it third behind Australian and Japan in the Asia Pacific region (Akamai, 2017). However, it ranks poorly in OpenSignal's *State of LTE* February 2018 report, with only 69.07 percent availability of its 4G network.[186] Geographically, New Zealand's 4G networks now provide access to about 90 percent of its population (Akamai, 2017).

Region	Unique IPv4 addresses	Average connection speed (Mbps)	Average peak connection speed (Mbps)	% above 4 Mbps	% above 10 Mbps	% above 15 Mbps
New Zealand	2,047,756	14.7	70.8	91%	52%	32%

Table 2: New Zealand's state of internet connectivity

From Akamai, 2017.

GLOSSARY

agency	a term used in social and political science to denote an individual's capacity for choice within a given context
aggregated data	refers to the process of gathering data from multiple sources and condensing that data into report-based or summarized form. This may involve linking 'static' data sets, or mining information from continuous streams of data generated by Internet-enabled technologies
algorithm	a set of mathematical processes used by machines to perform calculation, processing and decision making
algorithmic bias	an occurrence where an algorithm reflects and reproduces human bias. Human bias can be replicated in the algorithm as a result of coding decisions or the use of biased data
algorithmic decision-making	decision-making assisted by AI techniques such as ML
algorithmic transparency	algorithmic transparency means having visibility over the inputs and decision-making processes of tools relying on algorithms, programming or AI, or being able to explain the rules and calculations used by AI if these are challenged
anonymisation of data	the process of encrypting or removing personally sensitive or identifiable information from data sets, so as to ensure protection of privacy
anti-trust policies	laws that seek to regulate the behaviour of corporations that result in anti-competitive behaviour
artificial general intelligence	also known as 'generalised AI', this refers to the potential future capacity of AI to conduct and perform intelligent action, or thinking, to the same extent and ability as humans
artificial neural networks	a key component of ML that seeks to replicate the process of human learning using mathematical models comprised of a network of nodes representative of artificial neurons.
assistive technology	devices or systems that enable people with a disability to perform tasks that would otherwise not be possible
augmented reality	the use of technology to 'augment' or alter an individual's visual, auditory or olfactory experience of the real world environment
automation	the process by which a procedure is performed by a machine or technology without the need for human intervention
autonomy	the capacity to engage in self-governance. In an AI context, autonomy may refer to the capacity of AI to independently (or, in some cases, semi-independently) make decisions
big data	very large data sets which are unable to be stored, processed or used via traditional methods. Frequently determined in relation to data volume, variety, velocity and veracity.
black box	a term used to describe technologies whose underlying functions, processes, and outputs are obscured from the user's view, or made deliberately opaque

blockchain	a distributed, publicly accessible database of information that is spread over multiple computers, and that updates itself automatically
chatbots	an AI-powered computer program that can communicate and conduct a conversation either via voice or text, often by drawing on NLP techniques
cloud storage	the storage of data on servers that can be accessed remotely via the Internet
computer vision	an AI technique wherein AI systems have the capacity to 'see', identify, analyse and process images in a similar fashion to humans
cryptography	refers to a broad set of techniques for encrypting sensitive or personal information
data controller	an entity within an organisation that controls the procedures and purposes of data collection and usage
data governance	the people, processes, and technologies that ensure effective data management within an organisation
data integrity	the people, processes, and technologies that ensure the accuracy and consistency of data within an organisation
data linkage	the process of aggregating different data sets in order to derive common information about people, places, and events
data mining	the process of extracting anomalies, patterns and correlations from large data sets
data portability	the ability of an individual to obtain, reuse, or transfer personal data from and between different organisations and services
data provenance	a lineage of the records, entities, and systems that produce data
data sovereignty	the concept that information is subject to the laws of the nation within which it is stored
data subjects	an end user whose personal data is subject to collection and analysis
data surveillance	the process of collecting and analysing data without the owner's direct consent
deep learning	an ML system with multiple layers of neural networks
dialogue systems	a system designed to converse with a human user
digital inclusion	the project of ensuring all people can be included in, and ultimately benefit from, advances in digital technologies
digital infrastructure	the technical infrastructures required to support the implementation and integration of digital technologies throughout a society
digital technologies	technologies whose underlying processes are informed by digital binary – that is, 1s and 0s
digital tools	digital services and software interfaces that enable people to author and edit content
dirty data	a data set that contains errors or is inaccurate, incomplete, inconsistent and unstructured
equity	furthering the concept of equality, equity recognises the different needs and circumstances of each individuals and provides individuals with the resources needed in order to realise a fair outcome

explainability	ensuring that the actions, outputs, and decision-making processes of an AI system are transparent and easily understood by humans
facial recognition	AI systems that can compare, identify and verify an individual from an image or video
fake news	false news stories that can involve deliberate disinformation or propaganda, frequently spread via social media and designed to appear as genuine news reports
federated learning	a technique – developed by Google – of extracting data for the purposes of AI development, without compromising privacy
FinTech	intelligent financial service technologies
general purpose technologies	technologies that have widespread applications and uses, such as electricity and the Internet
inclusive design	a series of design principles which seek to accommodate and involve those experiencing difference, disability or disadvantage
Indigenous data sovereignty	the right of Indigenous People to govern the collection, generation, ownership and use of their data
intelligent virtual agents	computer-controlled assistants that can interact with humans
internet of things (IoT)	refers to the proliferation of Internet-enabled devices and technologies. These devices and technologies can produce, analyse and share large quantities of data through sensors and user interactions
interoperability	the capacity of systems to connect, share and exchange data, and utilise exchanged information
long term	a timeframe of greater than 20 years
machine learning (ML)	the ability of computers to execute tasks through processes of 'learning' that derive inspiration from (but are not reducible to) human intelligence and decision-making. ML involves the capacity of machines to process and adapt rapidly and independently to large quantities of data, without being explicitly programmed to do so.
medium term	a timeframe of 10 to 15 years
messy data	see 'dirty data'
metadata	data that provide information about other data; for example, a digital image may include metadata that provides information about the resolution of the image, when it was created, who the author is and so on
meta-intelligence	the ability to develop an understanding of what knowledge is in different contexts
micro credentialing	mini-qualifications obtained online through tertiary and job training institutions
narrow AI	also known as 'weak AI', narrow AI refers to AI systems that are good at a highly specific task or range of tasks
natural language processing (NLP)	encompasses all AI technologies related to the analysis, interpretation and generation (of text-based) natural language

personalised medicine	also known as 'precision medicine', personalised medicine is an umbrella term that encompasses medical and scientific techniques for targeted and tailored medical treatment of individuals
platforms	digital infrastructures and intermediaries that enable various entities to create, interact and transact in diverse ways, and whose revenue models are often premised on the extraction and usage of data
predictive analysis	a technique that uses data to forecast outcomes
predictive risk modelling	an automated algorithmic process used to predict outcomes. A risk score is determined and applied to the probably of an adverse event occurring
profiling	in information science, profiling refers to the construction of a user's profile via techniques of data analysis and mining
short term	a timeframe of 5 years
smart devices	internet-enabled devices (see 'Internet of Things')
smart grids	an electricity supply network that uses Internet-enabled technologies to communicate between customers, distributors, retailers and emergency response units
softbots	an abbreviation for 'software robot', a program that is imbued with the capacity to act on behalf of another user, organisation or program
spear phishing	an email attack intended to steal data from a specific individual or organisation
superhuman AI	also known as artificial emergent intelligence
supervised learning	where an AI learns a function from data labelled by humans, or is taught a function directly by a human
systemic bias	a form of bias that is deeply embedded in the underlying structure of a society or institution
unmanned aerial vehicles	aircraft that are autonomous, or remote controlled, but do not have a human pilot on board (e.g. drones).
unsupervised learning	where an AI learns a function independent of human intervention or guidance, by improving its actions against a well-defined objective

ABBREVIATIONS

ABC	Australian Broadcasting Corporation
ABS	Australian Bureau of Statistics
ACCC	Australian Competition and Consumer Commission
AI	artificial intelligence
AIFNZ	Artificial Intelligence Forum of New Zealand
APIs	application programming interfaces
APPs	Australian privacy principles
APRA	Australian Prudential Regulation Authority
ASIC	Australian Securities and Investments Commission
AVs	autonomous vehicles
CEDA	Committee for Economic Development of Australia
COMPAS	correctional offender management profiling for alternative sanctions
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CSL	China's Cybersecurity Law
DARPA	Defense Advanced Research Project Agency
EU	European Union
FDA	Food and Drug Administration (USA)
FinTech	financial service technologies
GDP	gross domestic product
GDPR	general data protection regulation
GFC	global financial crisis
HASS	humanities, arts and social sciences
ICESCR	International Covenant on Economic, Social and Cultural Rights
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IP	intellectual property
ML	machine learning
MOOCs	massive open online courses
NHS	National Health Service (UK)
NLP	natural language processing
OECD	Organisation for Economic Co-operation and Development
SMEs	small-to-medium sized enterprises
STEM	science, technology, engineering and mathematics
UDHR	Universal Declaration of Human Rights
UK	United Kingdom
UNCRPD	United Nations Convention on the Rights of Persons with Disabilities
US	United States
WEF	World Economic Forum
WTO	World Trade Organisation

REFERENCES

Abadi, M. *et al.* (2016) 'Deep Learning with Differential Privacy', in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*. New York, New York, USA: ACM Press, pp. 308–318. doi: 10.1145/2976749.2978318.

Abbott, R. and Bogenschneider, B. (2017) 'Should Robots Pay Taxes? Tax Policy in the Age of Automation', *Harvard Law Policy Review*, 12(1), pp. 145–75.

ABC (2018) 'The AI Race'. Australia: ABC. Available at: http://www.abc.net.au/tv/programs/ai-race/.

Acquisti, A. and Grossklags, J. (2005) 'Privacy and Rationality in Individual Decision Making', *IEEE Security and Privacy Magazine*, 3(1).

Administrative Review Council (2004) *Automated Assistance in Administrative Decision Making*. Available at: https://www.arc.ag.gov.au/Documents/AAADMreportPDF.pdf.

Agrawal, A. (2018) *The economics of artificial intelligence*, *McKinsey Quarterly*. Available at: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-economics-of-artificial-intelligence.

Agrawal, A., Gans, J. S. and Goldfarb, A. (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence*. Available at: https://hbr.org/product/prediction-machines-the-simple-economics-of-artificial-intelligence/10195-HBK-ENG.

Akamai (2017) *akamai's [state of the internet] Q1 2017 Report*. Available at: https://www.akamai.com/fr/fr/multimedia/documents/state-of-the-internet/q1-2017-state-of-the-internet-connectivity-report.pdf.

Alam, N. and Kendall, G. (2017) *Are robots taking over the world's finance jobs?* Available at: http://theconversation.com/are-robots-taking-over-the-worlds-finance-jobs-77561.

Alan, A. *et al.* (2014) 'A Field Study of Human-Agent Interaction for Electricity Tariff Switching', p. 8. Available at: https://eprints.soton.ac.uk/360820/1/main.pdf.

Alelo (2018) *Alelo*. Available at: https://www.alelo.com/.

Alfano, M., Carter, J. A. and Cheong, M. (2018) 'Technological seduction and self-radicalization', *Journal of the American Philosophical Association*, pp. 1–30. Available at: http://eprints.gla.ac.uk/165209/.

Alge, B. J. and Hansen, S. D. (2014) 'Workplace monitoring and surveillance research since 1984: A review and agenda', in Coovert, M. D. and Thompson, L. F. (eds) *The Psychology of Workplace Technology*. Routledge.

Allen, G. and Chan, T. (2017) *Artificial Intelligence and National Security*. Available at: https://www.belfercenter.org/sites/default/files/files/publication/AI NatSec - final.pdf.

AlphaBeta (2017) *The Automation Advantage*. Available at: https://www.alphabeta.com/wp-content/uploads/2017/08/The-Automation-Advantage.pdf.

Ames, M. (2018) 'Deconstructing the algorithmic sublime', *Big Data & Society*, 5(1). doi: 10.1177/2053951718779194.

Amnesty International and Access Now (2018) *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems*. Available at: https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf.

Ananny, M. and Crawford, K. (2018) 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability', *New Media & Society*, 20(3), pp. 973–989.

Anderson, M. (2015a) 'Technology Device Ownership: 2015', *Pew Research Center*, 29 October. Available at: www.pewinternet.org/2015/10/29/technology-device-ownership-2015.

Anderson, M. (2015b) 'The Demographics of Device Ownership', *Pew Research Center*, 29 October. Available at: http://www.pewinternet.org/2015/10/29/the-demographics-of-device-ownership/.

Andrejevic, M. (2014) 'Big data, big questions: the big data divide', *International Journal of Communication*, 8, pp. 1678–1689.

Andrejevic, M. and Burdon, M. (2015) 'Defining the sensor society', *Television & New Media*, 16(1), pp. 19–36.

Angwin, J. *et al.* (2016) 'Machine Bias', *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Angwin, J. (2016) 'Making Algorithms Accountable', *ProPublica*, 1 August. Available at: https://www.propublica.org/article/making-algorithms-accountable.

Angwin, J., Scheiber, N. and Tobin, A. (2017) 'Dozens of companies are using Facebook to exclude older workers from job ads', *ProPublica*, 20 December. Available at: https://www.propublica.org/article/facebook-ads-age-discrimination-targeting.

Aoun, J. E. (2017) *Robot-proof: Higher education in the age of artificial intelligence*. Cambridge, MA: MIT Press.

ARC (2015) *State of Australian university research report 2015–2016. Volume 1 ERA National Report*. Available at: https://www.arc.gov.au/sites/g/files/net4646/f/minisite/static/4551/ERA2015/downloads/ARC03966_ERA_ACCESS_151130.pdf.

Arpaly, N. (2003) *Unprincipled virtue: An inquiry into moral agency*. Oxford: Oxford University Press.

Article 29 Data Protection Working Party (2016) 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679'. Available at: https://www.pdpjournals.com/docs/887862.pdf.

Article 29 Data Protection Working Party (2017) 'Guidelines on the right to data portability'. Available at: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233.

Artificial Intelligence Committee - House of Lords (2017) *AI in the UK: ready, willing and able?* Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm.

Artificial Lawyer (2018) 'Official Launch of Groundbreaking Data Trusts For AI Training', *Artificial Lawyer*, 21 November. Available at: https://www.artificiallawyer.com/2018/11/21/official-launch-of-groundbreaking-data-trusts-for-ai-training/.

Austin, J. (2006) 'How much risk can we take? The misuse of risk assessment in corrections', *Federal Probation*, 70(2), pp. 58–63. Available at: http://www.uscourts.gov/sites/default/files/70_2_9_0.pdf.

Australian Bureau of Statistics (2017) *Household Income and Wealth, Australia, 2015-16*. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by Subject/6523.0~2015-16~Main Features~Household Income and Wealth Distribution~6.

Australian Capital Territory Legislative Assembly (2004) 'ACT Human Rights Act 2004'. Available at: https://www.legislation.act.gov.au/a/2004-5.

Australian Centre for Robotic Vision (2018) *Robotic Roadmap*. Available at: https://www.roboticvision.org/wp-content/uploads/Robotics-Roadmap_FULL-DOCUMENT.pdf.

Australian Competition and Consumer Commission (2018) *Digital platforms inquiry*. Available at: https://www.accc.gov.au/focus-areas/inquiries/digital-platforms-inquiry.

Australian Computer Society (2017) *Data Sharing Frameworks: Technical White Paper*. Available at: https://www.acs.org.au/content/dam/acs/acs-publications/ACS_Data-Sharing-Frameworks_FINAL_FA_SINGLE_LR.pdf.

Australian Computer Society (2018a) *Australia's IoT Opportunity: Driving Future Growt*h. Available at: https://www.acs.org.au/content/dam/acs/acs-publications/ACS-PwC-IoT-report-web.pdf.

Australian Computer Society (2018b) *Privacy in Data Sharing: A Guide for Business and Government*. Available at: https://www.acs.org.au/content/dam/acs/acs-publications/Privacy in Data Sharing - final version.pdf.

Australian Government (2011) *4446.0 - Disability (Labour Force), Australia, 2009*, *Australian Bureau of Statistics*. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4446.0main+features92009.

Australian Government (2014a) *8146.0 - Household Use of Information Technology, Australia, 2012-13*, *Australian Bureau of Statistics*. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/8146.0Chapter32012-13.

Australian Government (2014b) *Australian Industry Report 2014*. Available at: https://www.industry.gov.au/data-and-publications/australian-industry-report-2014.

Australian Government (2016) *8146.0 - Household Use of Information Technology, Australia, 2016-17*, *Australian Bureau of Statistics*. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0.

Australian Government (2017) *Child protection and Aboriginal and Torres Strait Islander children*, *Australian Institute of Family Studies*. Available at: https://aifs.gov.au/cfca/publications/child-protection-and-aboriginal-and-torres-strait-islander-children.

Australian Government (2018a) *Australia's Tech Future: Delivering a strong, safe and inclusive digital economy*. Available at: https://www.industry.gov.au/sites/default/files/2018-12/australias-tech-future.pdf.

Australian Government (2018b) *Australian Competition and Consumer Commission*. Available at: https://www.accc.gov.au/.

Australian Government (2018c) *Budget 2018-19*. Available at: https://www.budget.gov.au/2018-19/content/bp2/download/bp2_combined.pdf.

Australian Government (2018d) *Budget 2018-2019. Budget overview*. Available at: https://www.budget.gov.au/2018-19/content/overview.html.

Australian Government (2018e) 'Expressions of interest now open for the National Data Advisory Council', *Department of the Prime Minister and Cabinet*, 4 July. Available at: https://www.pmc.gov.au/news-centre/public-data/expressions-interest-now-open-national-data-advisory-council.

Australian Government (2018f) *My Health Records Act 2012 (Cth)*. Available at: https://www.legislation.gov.au/Details/C2018C00509.

Australian Government (2018g) 'Office of Future Transport Technologies Revealed', *Minister for Infrastructure, Transport and Regional Development*, 4 October. Available at: http://minister.infrastructure.gov.au/mccormack/releases/2018/october/mm178_2018.aspx.

Australian Government (2018h) *Office of the Australian Information Commissioner*. Available at: https://www.oaic.gov.au/.

Australian Government (2018i) *Office of the eSafety Commissioner*. Available at: https://www.esafety.gov.au/.

Australian Government (2018j) 'Strengthening the national data system', *Department of the Prime Minister and Cabinet*, 1 May. Available at: https://www.pmc.gov.au/news-centre/public-data/strengthening-national-data-system (Accessed: 18 June 2018).

Australian Government (2018k) *The Australian Government's response to the Productivity Commission Data Availability and Use Inquiry*. Available at: https://dataavailability.pmc.gov.au/sites/default/files/govt-response-pc-dau-inquiry.pdf.

Australian Human Rights Commission (1986) 'Australian Human Rights Commission Act 1986'.

Australian Human Rights Commission (2018a) *Human Rights and Technology*. Available at: https://www.humanrights.gov.au/our-work/rights-and-freedoms/projects/human-rights-and-technology.

Australian Human Rights Commission (2018b) *Human Rights and Technology Issues Paper*. Available at: https://tech.humanrights.gov.au/consultation.

Autor, D. (2015) 'Why Are There Still So Many Jobs? The History and Future of Workplace Automation', *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 29(3), pp. 3–30.

Baird, J. and Stocks, R. (2013) 'Risk assessment and management: Forensic methods, human results', *Advances in Psychiatric Treatment*, 19, pp. 358–365.

Baker, J. (2014) 'Inventor creates mind-controlled wheelchair which steers away from obstacles using an artificial intelligence system', *The Daily Telegraph*, 13 September. Available at: https://www.dailytelegraph.com.au/news/nsw/inventor-creates-mindcontrolled-wheelchair-which-steers-away-from-obstacles-using-an-artificial-intelligence-system/news-story/d6b23b35afdac8b1860a0a1903e2fe9b.

Bakhshi, H. *et al.* (2017) *The Future of Skills: Employment in 2030*. London. Available at: https://futureskills.pearson.com/research/assets/pdfs/technical-report.pdf.

Baldassarre, G. *et al.* (2014) 'Intrinsic motivations and open-ended development in animals, humans, and robots: an overview', *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.00985.

Barocas, S. and Nissenbaum, H. (2014) 'Big Data's End Run around Anonymity and Consent', in Stodden, V., Lane, J., and Nissenbaum, H. (eds) *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. New York: Cambridge University Press.

Barocas, S. and Selbst, A. D. (2016a) 'Big Data's Disparate Impact', *California Law Review*, 104, pp. 671–732. doi: 10.2139/ssrn.2477899.

Barocas, S. and Selbst, A. D. (2016b) 'Big Data's Disparate Impact', *SSRN Electronic Journal*. doi: 10.2139/ssrn.2477899.

Bastos, M. T. and Mercea, D. (2017) 'The Brexit Botnet and User-Generated Hyperpartisan News', *Social Science Computer Review*. doi: 10.1177/0894439317734157.

Bedi, G. *et al.* (2015) 'Automated analysis of free speech predicts psychosis onset in high-risk youths', *npj Schizophrenia*, 1(1), p. 15030. doi: 10.1038/npjschz.2015.30.

Beer, D. (2009) 'Power through the algorithm? Participatory web cultures and the technological unconscious', *New Media & Society*, 11(6), pp. 985–1002.

Begby, E. (2013) 'The epistemology of prejudice', *Thought*, 2(2), pp. 90–99.

Bench-Capon, T. and Prakken, H. (2006) 'Argumentation', in Lodder, A. R. and Oskamp, A. (eds) *Information Technology and Lawyers: Advanced Technology in the Legal Domain, from Challenges to Daily Routine*. Springer.

Bennet, J. *et al.* (2018) *Current State of Automated Legal Advice Tools*.

Bernstein, A. (2018) 'To increase Canadian innovation, take a lesson from our AI successes', *The Globe and Mail*, 14 May. Available at: https://www.theglobeandmail.com/business/commentary/article-to-increase-canadian-innovation-take-a-lesson-from-our-ai-successes/.

Best, S. (2018) '"World's first intelligent BIN" sorts your recycling for you - and it could be available in the UK soon', *Mirror*, 29 May. Available at: https://www.mirror.co.uk/tech/worlds-first-intelligent-bin-sorts-12615626.

Bezrukova, K. *et al.* (2016) 'A meta-analytical integration of over 40 years of research on diversity training evaluation', *Psychological Bulletin*, 142(11), pp. 1227–1274.

Bhidé, A. (2010) *A Call for Judgment: Sensible Finance for a Dynamic Economy*. Available at: https://global.oup.com/academic/product/a-call-for-judgment-9780199756070?cc=au&lang=en&.

Bickmore, T., Pfeifer, L. and Schulman, D. (2011) 'Relational Agents Improve Engagement and Learning in Science Museum Visitors', in, pp. 55–67. doi: 10.1007/978-3-642-23974-8_7.

Biggs, T. (2018) 'The social robot that could help save indigenous languages', *Sydney Morning Herald*, 4 June. Available at: https://www.smh.com.au/technology/the-socialrobot- that-could-help-save-indigenous-languages-20180601-p4ziyj.html.

Bilbrough, N. R. (2014) 'The FDA, Congress, and Mobile Health Apps: Lessons from DSHEA and the Regulation of Dietary Supplements', p. 45. Available at: https://core.ac.uk/download/pdf/56360421.pdf.

Birnbaum, M. and Fung, B. (2017) 'E.U. fines Google a record $2.7 billion in antitrust case over search results', *Washington Post*. Available at: https://www.washingtonpost.com/world/eu-announces-record-27-billion-antitrust-fine-on-google-over-search-results/2017/06/27/1f7c475e-5b20-11e7-8e2f-ef443171f6bd_story.html?utm_term=.6296044ffbc5.

Black, C. F. (2011) *The Land is the Source of the Law: A Dialogic Encounter with Indigenous Jurisprudence*. Routledge.

Black, C. F. (2018) 'Thinking about Artificial Intelligence through an Indigenous Jurisprudential Lens', in *Seminar presented at the Melbourne School of Government, 24 July, Melbourne University.*

Blank, A. *et al.* (2015) *Ethical issues for Māori in predictive risk modelling to identify new-born children who are at high risk of future maltreatment*. Available at: https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/research/predicitve-modelling/.

Bloomberg (2018) 'Biggest AI Startup Boosts Fundraising to $1.2 Billion', *Bloomberg News*, 31 May. Available at: https://www.bloomberg.com/news/articles/2018-05-31/world-s-biggest-ai-startup-raises-1-2-billion-in-mere-months.

Bloomberg New Energy Finance (2018) *Anglo Using 'Digital Twins', Robotics to Boost Mining: Q&A*, *Bloomberg NEF*. Available at: https://about.bnef.com/blog/anglo-using-digital-twins-robotics-boost-mining-qa/.

Bobo, L. D. and Thompson, V. (2006) 'Unfair by Design: The War on Drugs, Race, and the Legitimacy of the Criminal Justice System', *Social Research: An International Quarterly*, 73(2), pp. 445–472. Available at: https://www.jstor.org/stable/40971832.

Bolukbasi, T. *et al.* (2016) 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings'. Available at: http://arxiv.org/abs/1607.06520.

Booth, R. *et al.* (2017) 'Russia used hundreds of fake accounts to tweet about Brexit, data shows', *The Guardian*, 15 November. Available at: https://www.theguardian.com/world/2017/nov/14/how-400-russia-run-fake-accounts-posted-bogus-brexit-tweets.

Borg, K. and Smith, L. (2018) 'Digital inclusion and online behaviour: five typologies of Australian internet users', *Behaviour & Information Technology*. Taylor & Francis, 37(4), pp. 367–380. doi: 10.1080/0144929X.2018.1436593.

Borland, J. and Coelli, M. (2017) 'Are Robots Taking Our Jobs?', *Australian Economic Review*. Wiley/Blackwell, 50(4), pp. 377–397. doi: 10.1111/1467-8462.12245.

Boston, J. and Gill, D. (2018) *Social Investment: A New Zealand Policy Experiment*. doi: 10.7810/9781988533582.

Boulos, M. N. K. *et al.* (2014) 'Mobile phone and health apps: state of the art, concerns, regulatory control and certification.', *Online Journal of Public Health Informatics*, 5(3).

Bower, C. (2014) *The Folate Story*, *Telethon Kids Institute*. Available at: https://www.telethonkids.org.au/news--events/news-and-events-nav/2014/january/the-folate-story/ (Accessed: 26 September 2018).

Bowles, N. (2016) 'I Think My Blackness is Interfering: Does Facial Recognition Show Racial Bias?', *The Guardian*, 9 April. Available at: https://www.theguardian.com/technology/2016/apr/08/facial-recognition-technology-racial-bias-police.

Boyd, R. and Holton, R. J. (2017) 'Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation?', *Journal of Sociology*. SAGE Publications Ltd, p. 1440783317726591. doi: 10.1177/1440783317726591.

Brady, M. (1984) 'Artificial intelligence and robotics', p. 40. Available at: https://www.sciencedirect.com/science/article/pii/000437028590013X.

Bratton, B. H. (2015) *The Stack: On Software and Sovereignty*. Cambridge MA: The MIT Press.

Brea, E. *et al.* (2013) *Lightweight assistive manufacturing solutions: improving Australia's manufacturing competitiveness*.

Breakstone, M. (2017) 'Automatic Speech Recognition: Artificial Intelligence, Big Data, and the race for Human Parity', *Machine Learnings*, 26 June. Available at: https://machinelearnings.co/automatic-speech-recognition-artificial-intelligence-big-data-and-the-race-for-human-parity-a68a0350440f.

Breland, A. (2018) 'Week ahead: Tech giants to testify on extremist content'. Available at: http://thehill.com/business-a-lobbying/368775-week-ahead-in-tech-youtube-twitter-and-facebook-to-face-senate-commerce.

Bresnahan, T. and Trajtenberg, M. (1995) 'General Purpose Technologies "Engines of Growth"?', *Journal of Econometrics*, 65(1), pp. 83–108.

British Academy and The Royal Society (2017) *Data management and use: Governance in the 21st century*. Available at: https://royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf.

Broadbent, E. *et al.* (2011) 'Mental Schemas of Robots as More Human-Like Are Associated with Higher Blood Pressure and Negative Emotions in a Human-Robot Interaction', *International Journal of Social Robotics*, 3(3), pp. 291–297. doi: 10.1007/s12369-011-0096-9.

Broadbent, E. *et al.* (2016) 'Benefits and problems of health-care robots in aged care settings: A comparison trial', *Australasian Journal on Ageing*, 35(1), pp. 23–29. doi: 10.1111/ajag.12190.

Brooke, E. and Wissinger, E. (2017) 'Mythologies of Creative Work in the Social Media Age: Fun, Free, and "Just Being Me"', *International Journal of Communication*, (11), pp. 4652–4671.

Brooks, R. (2017) *The Seven Deadly Sins of AI Predictions*, *MIT Technolgoy Review*.

Brown, A. and Guttmann, R. (2017) *Ageing and Labour Supply in Advanced Economies*. Available at: https://www.rba.gov.au/publications/bulletin/2017/dec/5.html.

Bruckner, M., LaFleur, M. and Pitterl, I. (2017) *The Impact of the Technological Revolution on Labour Markets and Income Distribution*. Available at: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/2017_Aug_Frontier-Issues-1.pdf.

Brynjolfsson, E. and McAfee, A. (2016) 'The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies', p. 336. Available at: https://www.amazon.com/Second-Machine-Age-Prosperity-Technologies/dp/0393350649.

Buchanan, J. *et al.* (2018) *Preparing for the best and worst of times*. Available at: https://education.nsw.gov.au/our-priorities/innovate-for-the-future/education-for-a-changing-world/research-findings/future-frontiers-analytical-report-preparing-for-the-best-and-worst-of-times/Future-Frontiers_University-of-Sydney-exec-summary.pdf.

Buck, T. (2018) 'Germany to spend €3bn on boosting AI capabilities', *Financial Times*, 17 December. Available at: https://www.ft.com/content/fe1f9194-e8e3-11e8-a34c-663b3f553b35.

Bughin, J. *et al.* (2017) *Artificial Intelligence: The Next Digital Frontier?* Available at: https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced Electronics/Our Insights/How artificial intelligence can deliver real value to companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

Bughin, J. *et al.* (2018) *Notes from the AI frontier: Modeling the impact of AI on the world economy*. Available at: https://www.mckinsey.com/~/media/McKinsey/Featured Insights/Artificial Intelligence/Notes from the frontier Modeling the impact of AI on the world economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx.

Buolamwini, J. and Gebru, T. (2018) *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Available at: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Burdon, M. and McKillop, A. (2013) 'The Google Street View Wi-Fi Scandal and Its Repercussions for Privacy Regulation', *Monash University Law Review*, 39, pp. 702–738.

Burrell, J. (2016) 'How the machine "thinks": Understanding opacity in machine learning algorithms', *Big Data & Society*, 3(1), pp. 1–12.

Byambasuren, O. *et al.* (2018) 'Prescribable mHealth apps identified from an overview of systematic reviews', *npj Digital Medicine*, 1(1), p. 12. doi: 10.1038/s41746-018-0021-9.

Caliskan, A., Bryson, J. J. and Narayanan, A. (2017) 'Semantics derived automatically from language corpora contain human-like biases', *Science*, 356(6334), pp. 183–186. doi: 10.1126/science.aal4230.

Callaghan Innovation (2018) *Thinking Ahead: Innovation through Artificial Intelligence*. Available at: https://www.callaghaninnovation.govt.nz/sites/all/files/ai-whitepaper.pdf.

Calo, R. (2017) 'Artificial Intelligence Policy: A Primer and Roadmap', *UC Davis Law Review*, 51, pp. 404–429.

Calvo, R. and Peters, D. (2014) *Positive Computing: Technology for Wellbeing and Human Potential*. Available at: https://rafael-calvo.com/projects/publications/.

Canaan, M., Lucker, J. and Spector, B. (2016) *Opting in: Using IoT connectivity to drive differentiation: The Internet of Things in insurance*. Available at: https://www2.deloitte.com/content/dam/insights/us/articles/innovation-in-insurance-iot/DUP2824_IoT_Insurance_vFINAL_6.6.16.pdf.

Cannataci, J. (2018) 'Special Rapporteur on the Right to Privacy Joseph Cannataci, End of Mission Statement of the Special Rapporteur on the Right to Privacy at the Conclusion Of his Mission to the United Kingdom of Great Britain and Northern Ireland', *United Nations Human Rights: Office of the High Commissioner*, 29 June. Available at: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23296&LangID=E (Accessed: 3 August 2018).

Capgemini (2018) *Artificial Intelligence Decoded*. Available at: https://www.capgemini.com/ar-es/resources/digital-transformation-review-artificial-intelligence-decoded/.

Capgemini Consulting (2017) *Unleashing the potenital of Artificial Intelligence in the Public Sector*. Available at: https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf.

Cardoso, C. *et al.* (2007) *Inclusive design toolkit*. Edited by J. Clarkson et al. Engineering Design Centre Department of Engineering, University of Cambridge. Available at: https://www-edc.eng.cam.ac.uk/downloads/idtoolkit.pdf.

Carlstrom, V. (2018) 'The World's First Electric and Autonomous Container Ship is Being Build in Norway - to Replace 100 Diesel Trucks a Day', *Business Insider Nordic*. Available at: https://nordic.businessinsider.com/the-worlds-first-electric-and-autonomous-container-ship-is-being-built-in-norway--to-replace-100-diesel-trucks-a-day--/.

Cascio, W. F. and Montealegre, R. (2016) 'How Technology Is Changing Work and Organizations', *Annual Review of Organizational Psychology and Organizational Behavior*, 3(1), pp. 349–375. doi: 10.1146/annurev-orgpsych-041015-062352.

Castelvecchi, D. (2016) 'Can we open the black box of AI?', *Nature News*, 538(7623). Available at: https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731.

Cate, F. H. and Mayer-Schönberger, V. (2013) 'Notice and consent in a world of Big Data', *International Data Privacy Law*, 3(2).

CEDA (2015) *Australia's future workforce?* Available at: http://www.ceda.com.au/Research-and-policy/All-CEDA-research/Research-catalogue/Australia-s-future-workforce.

Center for Inclusive Design and Environmental Access (2011) *Goals of Universal Design*, *University at Buffalo, School of Architecture and Planning*.

Centre for Excellence in Universal Design (2018) *Homepage*, *National Disability Authority, Centre for Excellence in Universal Design*. Available at: http://universaldesign.ie/.

Chander, A. and P. Lê, U. (2015) 'Data Nationalism', *Emory Law Journal*, 64(3), pp. 677–739.

Chatila, R. *et al.* (2017) 'The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems [Standards]', *IEEE Robotics & Automation Magazine*, 24(1), pp. 110–110. doi: 10.1109/MRA.2017.2670225.

Chatterton, M. (2018) 'Impersonal service? Ensuring AI enhances rather than diminishes the customer experience', *Ipsos MORI*, 14 June. Available at: https://www.ipsos.com/ipsos-mori/en-uk/impersonal-service-ensuring-ai-enhances-rather-diminishes-customer-experience.

Chen, S. (2018) 'China's schools are quietly using AI to mark students' essays ... but do the robots make the grade?', *South China Morning Post*, 27 May. Available at: https://www.scmp.com/news/china/society/article/2147833/chinas-schools-are-quietly-using-ai-mark-students-essays-do.

Cheok, A. D., Levy, D. and Karunanayaka, K. (2016) 'Lovotics: Love and Sex with Robots', in Karpouzis, K. and Yannakakis, G. (eds) *Emotion in Games*. Springer.

Chin, M. *et al.* (2018) *China's Cybersecurity Law*, *ReedSmith*. Available at: https://www.reedsmith.com/en/perspectives/2018/01/chinas-cybersecurity-law.

Chou, J., Murillo, O. and Ibars, R. (2017) 'How to Recognize Exclusion in AI', *Microsoft - The Inclusive Design Team*, 27 September. Available at: https://medium.com/microsoft-design/how-to-recognize-exclusion-in-ai-ec2d6d89f850.

Christian, J. (2018) 'Why Is Google Translate Spitting Out Sinister Religious Prophecies?', *Motherboard*, 21 July. Available at: https://motherboard.vice.com/en_us/article/j5npeg/why-is-google-translate-spitting-out-sinister-religious-prophecies.

Christie's (2018) *Is artificial intelligence set to become art's next medium?*, *Christie's*. Available at: https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx.

Chui, M. *et al.* (2018) *Notes from the AI frontier: insights from hundreds of use cases*.

Chui, M., Manyika, J. and Miremadi, M. (2016) *Where machines could replace humans—and where they can't (yet)*, *McKinsey Quarterly*. Available at: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet.

CIO New Zealand (2018) 'AI Helps Harmoney improve ability to assess credit risk', *CIO New Zealand*, 19 January. Available at: https://www.cio.co.nz/article/632369/ai-helps-harmoney-improve-ability-assess-credit-risk/.

Citron, D. and Pasquale, F. (2014) 'The Scored Society: Due Process for Automated Predictions', *Washington Law Review*, 9(1), pp. 1–33.

City of New York (2018) *Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City*, *City of New York*. Available at: https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by.

Claburn, T. (2017a) 'Suit filed against state fraud detection vendor', *Detroit Free Press*. Available at: https://www.freep.com/story/news/local/michigan/2017/03/02/suit-filed-against-state-fraud-detection-vendor/98646934/

Claburn, T. (2017b) 'Fraud detection system with 93% failure rate gets IT companies sued', *The Register*. Available at: https://www.theregister.co.uk/2017/03/08/fraud_detection_system_with_93_failure_rate_gets_it_companies_sued/

Clarke, R. and Libarikian, A. (2014) *Unleashing the value of advanced analytics in insurance*, *McKinsey&Company*. Available at: https://www.mckinsey.com/industries/financial-services/our-insights/unleashing-the-value-of-advanced-analytics-in-insurance.

Clegg, C. W. (2000) 'Sociotechnical principles for system design', *Applied Ergonomics*, 31(5), pp. 463–477. doi: 10.1016/S0003-6870(00)00009-0.

Cohen, B., Hall, B. and Wood, C. (2017) 'Data Localization Laws and their Impact on Privacy, Data Security and the Global Economy', *Antitrust*, 32(1), pp. 107–114.

Cohen, J. E. (2017) 'The Biopolitical Public Domain: the Legal Construction of the Surveillance Economy', *Philosophy & Technology*, 31(2), pp. 213–233.

Cohn, M. (2017) 'Drones could soon get crucial medical supplies to patients in need', *Baltimore Sun*. Available at: http://www.baltimoresun.com/health/maryland-health/bs-hs-drones-for-blood-20161223-story.html.

Collins, M. (2018) *Te Hiku Media project teaching machines to speak te reo Maori*, *New Zealand Herald*.

Colvin, G. (2015) *Humans Are Underrated: What High Achievers Know That Brilliant Machines Never Will*. New York: Penguin.

Commissioner for Human Rights (2018) *Safeguarding human rights in the era of artificial intelligence*. Available at: https://www.coe.int/en/web/commissioner/-/safeguarding-human-rights-in-the-era-of-artificial-intelligence?inheritRedirect=true (Accessed: 11 October 2018).

Commonwealth of Australia (2018) *Senate Select Committee on the Future of Public Interest Journalism*. Available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Future_of_Public_Interest_Journalism/PublicInterestJournalism/Report.

Conn, A. (2017) *Can we properly prepare for the risks of superintelligent AI?*, *Future of Life Institute*. Available at: https://futureoflife.org/2017/03/23/ai-risks-principle/.

Connor, N. (2017) 'Legal Robots Deployed in China to Help Decide Thousands of Cases', *The Telegraph*, 4 August. Available at: https://www.telegraph.co.uk/news/2017/08/04/legal-robotos-deployed-china-help-decide-thousandscases.

Cooper, B. (2011) 'Judges in Jeopardy: Could IBM's Watson Beat Courts at Their Own Game?', *Yale Law Journal Forum*, 87, pp. 97–99.

Corbett-Davies, S., Pierson, E., Feller, A. and Goel, S. (2017) 'A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear', *Washington Post*, 17 October.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., *et al.* (2017) 'Algorithmic Decision Making and the Cost of Fairness', in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*. New York, New York, USA: ACM Press, pp. 797–806. doi: 10.1145/3097983.3098095.

Corcoran, C. M. *et al.* (2018) 'Prediction of psychosis across protocols and risk cohorts using automated language analysis', *World Psychiatry*. Hoboken: John Wiley and Sons Inc., 17(1), pp. 67–75. doi: 10.1002/wps.20491.

Cossins, D. (2018) *Discriminating algorithms: 5 times AI showed prejudice*, *New Scientist*. Available at: https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/.

Craig, L. A. and Beech, A. (2009) 'Best practice in conducting actuarial risk assessments with adult sexual offenders', *Journal of Sexual Aggression*, 15(2), pp. 193–211.

Cranor, L. F. (2012) 'Necessary but not Sufficient: Standardized Mechanisms for Privacy Notice and Choice', *Journal on Telecommunications and High Technology Law*, 10(2), pp. 273–308.

Crawford, K. and Calo, R. (2016) 'There is a blind spot in AI research', *Nature*, 538(7625). Available at: https://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805.

Crawford, K., Miltner, K. and Gray, M. L. (2014) 'Critiquing big data: Politics, Ethics, Epistemology', *International Journal of Communication*, pp. 1663–1672.

Creative Commons Australia (2013) *About the Licences*. Available at: https://creativecommons.org.au/learn/licences/.

Crockford, T. (2018) 'Brisbane AI to Help with Cancer Treatment in an Australian First', *Brisbane Times*. Available at: https://www.brisbanetimes.com.au/national/queensland/brisbane-ai-to-help-in-cancer-treatment-in-an-australian-first-20180728-p4zu76.html.

CSIRO (2015a) *Creating a 'Digital Homestead'*, *CSIRO*. Available at: https://www.csiro.au/en/Research/AF/Areas/Sustainable-farming/Precision-agriculture/Digital-Homestead.

CSIRO (2015b) *Our Top 10 Inventions*. Available at: https://www.csiro.au/en/About/History-achievements/Top-10-inventions.

Culnane, C., Rubinstein, B. and Teague, V. (2017) 'HEALTH DATA IN AN OPEN WORLD', p. 23. Available at: https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf.

D'Alfonso, S. *et al.* (2017) 'Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health', *Frontiers in Psychology*. Available at: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00796/full.

daisee (2017) *Outlook on the Australian AI market landscape in Australia*. Available at: https://www.daisee.com/daisee-2017-australian-ai-report/.

Damasio, A. R. (1994) *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam's Sons.

Damgård, I. *et al.* (2012) 'Multiparty computation from somewhat homomorphic encryption', in *Advances in Cryptology--CRYPTO 2012*. Springer, pp. 643–662.

Danks, D. and London, A. J. (2017) 'Algorithmic Bias in Autonomous Systems', in *In Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne. Available at: https://www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias-Distrib.pdf.

Dassen, T. and Hajer, M. A. (2014) *Smart About Cities - Visualising the Challenge for 21st Century Urbanism*. Available at: https://www.amazon.co.uk/Smart-About-Cities-Visualising-Challenge/dp/9462081484.

Dastin, J. (2018) 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women', *Reuters*. Available at: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Data for Black Lives (2018) *About Data for Black Lives*. Available at: http://d4bl.org/about.html (Accessed: 30 November 2018).

Davenport, T. H. and Short, J. E. (1990) 'The New Industrial Engineering: Information Technology and Business Process Redesign', p. 17. Available at: https://sloanreview.mit.edu/article/the-new-industrial-engineering-information-technology-and-business-process-redesign/.

Dawes, R. (1994) *House of Cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Deakin University (2018) *How Artificial Intelligence is Revolutionising Conservation*, *this*. Available at: http://this.deakin.edu.au/innovation/how-artificial-intelligence-is-revolutionising-conservation (Accessed: 29 November 2018).

Deloitte (2017) *AI-augmented human services*. Available at: https://www2.deloitte.com/content/dam/insights/us/articles/4152_AI-human-services/4152_AI-human-services.pdf.

Deloitte Access Economics (2018) *ACS Australia's Digital Pulse: Driving Australia's international ICT competitiveness and digital growth*. Available at: https://www.acs.org.au/content/dam/acs/acs-publications/aadp2018.pdf.

Deloitte University Press (2016) *The Future of Moblity: What's Next?* Available at: https://www2.deloitte.com/content/dam/insights/us/articles/3367_Future-of-mobility-whats-next/DUP_Future-of-mobility-whats-next.pdf.

Dennis, K. and Urry, J. (2013) *After the Car*. 1. ed. Oxford: Polity Press.

Dent, C. (2018) 'Taking the Human Out of the Regulation of Road Behaviour', *Sydney Law Review*, 41(1), p. 39.

Department of Finance and Administration (2007) *Automated Assistance in Administrative Decision-Making: Better Practice Guide*. Available at: https://www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf.

Department of the Prime Minister and Cabinet (2018) *New Australian Government Data Sharing and Release Legislation: Issues paper for consultation*. Available at: https://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation.

Desai, T., Ritchie, R. and Welpton, F. (2016) 'Five Safes: designing data access for research', *Working Paper*. Available at: http://eprints.uwe.ac.uk/28124/.

Desmet, P. M. A. and Pohlmeyer, A. E. (2013) 'Positive Design: An Introduction to Design for Subjective Well-Being', p. 15. Available at: http://www.ijdesign.org/index.php/IJDesign/article/viewFile/1666/587.

DeVoss, C. C. (2017) 'Artificial intelligence can expedite scientific communication and eradicate bias from the publishing process', *LSE Impact Blog*. Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2017/05/11/artificial-intelligence-can-expedite-scientific-communication-and-eradicate-bias-from-the-publishing-process/.

Diakopoulos, N. (2015) 'Algorithmic Accountability'. Available at: http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/algorithmic_accountability_final.pdf.

Diakopoulos, N. (2016) 'Accountability in algorithmic decision making', in *Communications of the ACM*, pp. 56–62.

Dietvorst, B. J., Simmons, J. P. and Massey, C. (2016) 'Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them', *Management Science*, 64(3), pp. 1155–1170. doi: 10.1287/mnsc.2016.2643.

Diez, W. (2017) *Wohin steuert die deutsche Automobilindustrie?* Available at: https://www.amazon.com.au/Wohin-steuert-die-deutsche-Automobilindustrie-ebook/dp/B01N8S1JQN.

Digital Government New Zealand (2018) *D7 group of digital nations*. Available at: https://www.digital.govt.nz/digital-government/international-partnerships/d7-group-of-digital-nations/.

Digital Promise (2018) *Educator Micro-credentials*, *Digital Promise: Accelerating Innovation in Education*. Available at: https://digitalpromise.org/initiative/educator-micro-credentials/ (Accessed: 21 August 2018).

Digital Skills Forum (2018) *About the Forum*, *Digital Skills Forum New Zealand*. Available at: https://digitalskillsforum.nz/about/ (Accessed: 29 November 2018).

Digital Transformation Agency (2018) *Digital Transformation Strategy*. Available at: https://www.dta.gov.au/digital-transformation-strategy.

Dillet, R. (2018) 'France wants to become an artificial intelligence hub', *Tech Crunch*. Available at: https://techcrunch.com/2018/03/29/france-wants-to-become-an-artificial-intelligence-hub/.

Domingo, M. C. (2012) 'An overview of the Internet of Things for people with disabilities', *Journal of Network and Computer Applications*, 35(2), pp. 584–596. doi: https://doi.org/10.1016/j.jnca.2011.10.015.

Downey, A. B. (2016) *Think Complexity, volume 2*. Needham, Massachusetts: Green Tea Press.

Dressel, J. and Farid, H. (2018) 'The Accuracy, Fairness, and Limits of Predicting Recidivism', *Science Advances*. Available at: http://advances.sciencemag.org/content/4/1/eaao5580.

Dudenhöffer, F. (2008) *Wer kriegt die Kurve?* Available at: https://content-select.com/de/portal/media/view/57208804-1128-4028-84ff-5d03b0dd2d03.

Dunlop, T. (2016) *Why the Future Is Workless*. Sydney: NewSouth.

Dutta, R. *et al.* (2013) 'Deep cognitive imaging systems enable estimation of continental-scale fire incidence from climate data', *Scientific Reports*, 3. Available at: https://www.nature.com/articles/srep03188.

Dutton, T. (2018) 'An Overview of National AI Strategies', *Medium*, 28 June. Available at: https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

Dwork, C., Roth, A. and others (2014) 'The algorithmic foundations of differential privacy', *Foundations and Trends®in Theoretical Computer Science*. Now Publishers, Inc., 9(3--4), pp. 211–407.

Eckersley, P. (2018) 'How Good Are Google's New AI Ethics Principles?', *Electronic Frontier Foundation*, 7 June. Available at: https://www.eff.org/deeplinks/2018/06/how-good-are-googles-new-ai-ethics-principles.

Edwards, J. (2018) 'Submission to the Justice and Law Select Committee on the Privacy Bill'. Office of the Privacy Commissioner, New Zealand.

Egan, P. (2017) 'Michigan Integrated Data Automated System Experiences 93 Percent Error Rate During Nearly Two Years of Operation', *Detroit Free Press*. Available at: https://www.govtech.com/data/Michigan-Integrated-Data-Automated-System-Experiences-93-Percent-Error-Rate-During-Nearly-Two-Years-of-Operation.html

Ejgenberg, Y. *et al.* (2012) 'SCAPI: The Secure Computation Application Programming Interface.', *IACR Cryptology EPrint Archive \url{https://eprint.iacr.org/2012/629.pdf}*, 2012, p. 629.

Ekbia, H. R., Nardi, B. and Sabanovic, S. (2015) 'On the Margins of the Machine: Heteromation and Robotics', in, p. 12. Available at: https://www.ideals.illinois.edu/handle/2142/73678.

Element AI (2018) *Global AI Talent Report 2018*. Available at: http://www.jfgagne.ai/talent.

Elliott, A. and Urry, J. (2010) *Mobile Lives: SELF, EXCESS AND NATURE*. Available at: https://www.chapters.indigo.ca/en-ca/books/mobile-lives-self-excess-and/9780415480222-item.html.

Elmi, N. and Davis, N. (2018) 'How governance is changing in the 4IR', *World Economic Forum*, 13 January. Available at: https://www.weforum.org/agenda/2018/01/agile-governance-changing-4ir-public-private-emerging-technologies.

Elvery, S. (2017) *How algorithms make important government decisions — and how that affects you*, *ABC*. Available at: http://www.abc.net.au/news/2017-07-21/algorithms-can-make-decisions-on-behalf-of-federal-ministers/8704858.

Eriksson-Zetterquist, U Lindberg, K. and Styhre, A. (2009) 'When the good times are over: Professionals encountering new technology', *Human Relations*, 62(8), pp. 1145–1170.

Esteva, A. *et al.* (2017) *Dermatologist-level classification of skin cancer with deep neural networks*. Available at: https://www.nature.com/articles/nature21056.

Eubanks, V. (2018a) *Automating Inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Eubanks, V. (2018b) 'The Digital Poorhouse', *Harper's Magazine*. Available at: https://harpers.org/archive/2018/01/the-digital-poorhouse/.

European Commission (2016) *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da*, *EU Law*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679.

European Commission (2017) *Summary report of the public consultation on Building a European Data Economy*, *European Commission, Digital Single Market - Counsulation Results*. Available at: https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-building-european-data-economy (Accessed: 16 August 2018).

European Commission (2018a) *2018 reform of EU data protection rules*. Available at: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en.

European Commission (2018b) *Artificial Intelligence: Commission Outlines a European Approach to Boost Investment and Set Ethical Guidelines*, *Press Release*.

European Commission (2018c) 'Artificial Intelligence for Europe', *Digital Single Market News*, 25 April. Available at: https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.

European Commission (2018d) *European Commission Staff Working Document: Liability for emerging digital technologies*, *Digital Single Market News*. Available at: https://ec.europa.eu/digital-single-market/en/news/european-commission-staff-working-document-liability-emerging-digital-technologies (Accessed: 18 July 2018).

European Group on Ethics in Science and New Technologies (2018) *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. Available at: http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

Executive Office of the President and National Science and Technology Council Committee on Technology (2016) *Preparing for the Future of Artifical Intelligence*. Available at: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

Eyers, J. (2018) *Chatbots just the beginning for AI in banking*.

Federal Court of Australia (2018) *Applications For File*, *Commonwealth Courts Portal*. Available at: https://www.comcourts.gov.au/file/Federal/P/NSD734/2018/actions (Accessed: 30 November 2018).

Ferrein, A. and Meyer, T. (2012) 'A Brief Overview of Artificial Intelligence in South Africa', *AI Magazine*, 33(1). Available at: https://doi.org/10.1609/aimag.v33i1.2357.

Filtered (2018) *Filtered*. Available at: https://www.filtered.ai/.

Finkel, A. (2018a) 'Artificial Intelligence – a matter of trust', in *Keynote address at Committee for Economic Development of Australia event 'Artificial Intelligence: potential, impact and regulation" 18 May'*. Sydney, NSW: Office of the Chief Scientist. Available at: https://www.chiefscientist.gov.au/2018/05/speech-artificial-intelligence-a-matter-of-trust/.

Finkel, A. (2018b) 'Setting the scene: What role for human rights in a new age of technology?', in *Human Rights & Technology*.

Finkel, A. (2018c) 'What will it take for us to trust AI?', *Agenda*, 12 May. Available at: https://www.weforum.org/agenda/2018/05/alan-finkel-turing-certificate-ai-trust-robot/.

Fisher, D. (2017) 'Uber Fights Seattle's Push To Make It Bargain With The Teamsters', *Forbes*, 16 March. Available at: https://www.forbes.com/sites/danielfisher/2017/03/16/uber-asks-can-seattle-really-make-us-bargain-with-the-teamsters/#52b1e4bc22f6.

Ford, L. (2012) 'Between Indigenous and Settler Governance', in Ford, L. and Rowse, T. (eds) *Locating Indigenous self-determination in the margins of settler sovereignty: an introduction*. Routledge, p. 11.

Ford, M. (2016) *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.

Fraedrich, E. and Lenz, B. (2016) 'Societal and Individual Acceptance of Autonomous Driving', p. 20. Available at: https://elib.dlr.de/110349/1/FRAEDRICH_LENZ_2016_aF_Acceptance.pdf.

Frey, C. B. and Osborne, M. A. (2013) *The Future of Employment: How Susceptible are Jobs to Computerisation?* accessed 12th January 2016. Available at: www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.

Frey, C. B. and Osborne, M. A. (2017) 'The future of employment: How susceptible are jobs to computerisation?', *Technological Forecasting and Social Change*, 114, pp. 254–280. doi: https://doi.org/10.1016/j.techfore.2016.08.019.

Fricker, M. (2007) *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press.

Friedman, B. and Nissenbaum, H. (1996) 'Bias in computer systems', *ACM Transactions on Information Systems*, 14(3), pp. 330–347.

Fritz, L. (2016) 'How Much Digitalization Can a Human Tolerate?', in *Conference Proceedings Trends in Business Communication*. Wiesbaden: Springer Gabler, pp. 107–113.

FTI Consulting (2018) *Artificial Intelligence: The Race Is On The Global Policy Response To A*. Available at: https://euagenda.eu/upload/publications/untitled-128126-ea.pdf.

Future of Life Institute (2017) *Asilomar AI Principles*, *Future of Life Institute*. Available at: https://futureoflife.org/ai-principles/.

G2 Crowd (2018) *Artificial Intelligence Software*. Available at: https://www.g2crowd.com/categories/artificial-intelligence (Accessed: 10 October 2018).

Gartry, L. (2018) 'Robots Ready to Start Killing Crown-of-Thorns Starfish on Great Barrier Reef', *ABC News*. Available at: https://www.abc.net.au/news/2018-08-31/crown-of-thorns-starfish-killing-robot-great-barrier-reef-qld/10183072.

Gendler, T. (2011) 'On the epistemic costs of implicit bias', *Philosophical Studie*, 156(1), pp. 33–63.

Gerstner, M. E. (1993) 'Liability Issues with Artificial Intelligence Software', *Santa Clara Law Review*, 33(1), pp. 239–269.

Gibb, J. (2018) 'Chance to lead ethical use of AI', *Otago Daily Times*, 8 May. Available at: https://www.odt.co.nz/news/dunedin/campus/university-of-otago/chance-lead-ethical-use-ai.

Giddings, S. (2014) *Gameworlds: Virtual Media and Children's Everyday Play*. Bloomsbury Academic.

Gillingham, P. (2015) 'Predictive risk modelling to prevent child maltreatment and other adverse outcomes for service users: Inside the "black box" of machine learning', *The British Journal of Social Work*, 46(4), pp. 1044–1058.

Gillingham, P. and Graham, T. (2016) 'Designing electronic information systems for the future: Social workers and the challenge of New Public Management', *Critical Social Policy*. SAGE PublicationsSage UK: London, England, 36(2), pp. 187–204. doi: 10.1177/0261018315620867.

Gitlin, J. M. (2017) 'Hacking street signs with stickers could confuse self-driving cars', *arstechnica*, 9 February. Available at: https://arstechnica.com/cars/2017/09/hacking-street-signs-with-stickers-could-confuse-self-driving-cars/.

Goadsuff, L. (2018) *Emotion AI Will Personalize Interactions*, *Gartner*. Available at: https://www.gartner.com/smarterwithgartner/emotion-ai-will-personalize-interactions/.

Google (2009) 'Personalized Search for everyone', *Google Official Blog*, 4 December. Available at: https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html.

Greenfield, A. (2017) *Radical Technologies: The Design of Everyday Life*. Available at: https://www.amazon.com/Radical-Technologies-Design-Everyday-Life/dp/178478043X.

Greenwald, A. and Banaji, M. (1995) 'Implicit social cognition: Attitudes, self-esteem, and stereotypes', *Psychological Review*, 102(1), pp. 4–27.

Griffiths, J. (2016) *New Zealand passport robot thinks this Asian man's eyes are closed*, *CNN*. Available at: https://edition.cnn.com/2016/12/07/asia/new-zealand-passport-robot-asian-trnd/index.html.

Gross, J. A. (2017) 'Unmanned subs, sniper drones, gun that won't miss: Israel unveils future weapons', *Times of Israel*, 5 September. Available at: https://www.timesofisrael.com/unmanned-subs-and-sniper-drones-israel-unveils-its-weapons-of-thefuture.

Grote, G. and Kochan, T. (2018) *Here's how we can make innovation more inclusive*. Available at: https://www.weforum.org/agenda/2018/06/here-s-how-we-can-make-innovation-more-inclusive/.

Guess, A., Nyhan, B. and Reifler, J. (2018) *Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 presidential campaign*. Available at: http://www.dartmouth.edu/~nyhan/fake-news-2016.pdf.

Guihot, M., Matthew, A. F. and Suzor, N. P. (2017) 'Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence', *Vanderbilt Journal of Entertainment & Technology Law*, 20(2), pp. 385–456.

Guy, S. and Shove, E. (2000) *The sociology of energy, buildings and the environment: Constructing knowledge, designing practice*. London and New York: Routledge.

Hajian, S., Bonchi, F. and Castillo, C. (2016) 'Algorithmic bias: From discrimination discovery to fairness-aware data mining', in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 2125–2126.

Haldane, A. (2015) 'Labour's Share'. Speech by Andrew Haldane given at the Trades Union Congress, London. Available at: https://www.bankofengland.co.uk/speech/2015/labours-share.

Hall, M. J. J. (2005) 'Supporting Discretionary Decision-Making with Information Technology: A Case Study in the Criminal Sentencing Jurisdiction', p. 36. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=777745.

Hall, W. and Pesenti, J. (2017) *Growing the Artificial Intelligence Industry in the UK*. Available at: https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.

Hallevy, G. (2013) *When Robots Kill*. Northeastern University Press.

Halloran, T. (2017) 'How Atlassian went from 10% female technical graduates to 57% in two years', 12 December. Available at: https://textio.ai/atlassian-textio-81792ad3bfbf.

Harari, Y. N. (2017) *Homo Deus: A Brief History of Tomorrow*. New York: HarperCollins.

Hardt, M., Price, E. and Srebro, N. (2016) 'Equality of Opportunity in Supervised Learning'. Available at: http://arxiv.org/abs/1610.02413.

Hassenzahl, M. (2010) 'Experience Design: Technology for All the Right Reasons', *Synthesis Lectures on Human-Centered Informatics*, 3(1), pp. 1–95. doi: 10.2200/S00261ED1V01Y201003HCI008.

Hastie, T., Tibshirani, R. and Friedman, J. (2008) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hausmann, R. and Hidalgo, C. A. (2010) 'Country Diversification, Product Ubiquity, and Economic Divergence', p. 45. Available at: file:///C:/Users/Bryan/Downloads/RWP10-045_Hausmann_Hidalgo.pdf.

Hengstler, M., Enkel, A. and Duelli, S. (2016) 'Applied artificial intelligence and trust — The case of autonomous vehicles and medical assistance devices', *Technological Forecasting & Social Change*, 105, pp. 105–120. Available at: http://daneshyari.com/article/preview/896383.pdf.

Hennig, N. (2016) 'Natural User Interfaces and Accessibility', in *Mobile Learning Trends: Accessibility, Ecosystems, Content Creation*. ALA TechSource. Available at: https://journals.ala.org/index.php/ltr/article/view/5969/7598.

de Hert, P. and Christianen, K. (2013) *Progress Report on the Application of the Principles of Convention 108 to the Collection and Processing of Biometric Data*. Available at: https://www.coe.int/en/web/data-protection/reports-studies-and-opinions (Accessed: 3 August 2018).

HM Government (2017) *Industrial Strategy: building a Britain fit for the future*. Available at: https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future.

Hoadley, D. and Lucas, N. (2018) *Artificial Intelligence and National Security*. Available at: http://www.crs.govr45178.

Hodson, R. (2016) 'Precision medicine', *Nature*, 537(S49). Available at: https://www.nature.com/articles/537S49a.

Hoermann, S. *et al.* (2017) 'Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review', *Journal of Medical Internet Research*, 19(8), p. e267. doi: 10.2196/jmir.7023.

Hollier, S. (2007) *The Disability Divide: a study into the impact of computing and internet-related technologies on people who are blind or vision-impaired*. Curtin University. Available at: http://hdl.handle.net/20.500.11937/214.

Hollier, S. and Abou-Zahra, S. (2018) 'Internet of Things (IoT) As Assistive Technology: Potential Applications in Tertiary Education', in *Proceedings of the Internet of Accessible Things*. New York, NY, USA: ACM (W4A '18), p. 3:1--3:4. doi: 10.1145/3192714.3192828.

Hsieh, P. (2017) 'AI In Medicine: Rise Of The Machines', *Forbes*, 30 April. Available at: https://www.forbes.com/sites/paulhsieh/2017/04/30/ai-in-medicine-rise-of-the-machines/#21f3d901abb0.

Hsu, C.-Y. *et al.* (2017) 'Zero-Effort In-Home Sleep and Insomnia Monitoring using Radio Signals', in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.

Hsu, E. L. (2014) 'The sociology of sleep and the measure of social acceleration', *Time & Society*, 23(2), pp. 212–234.

Hu, L. and Chen, Y. (2017) 'Fairness at Equilibrium in the Labor Market', *Fairness, Accountability, and Transparency in Machine Learning*. Available at: http://arxiv.org/abs/1707.01590.

Huet, N. (2013) 'France's Carmat implants its first artificial heart in human', *Reuters*. Available at: https://uk.reuters.com/article/us-carmat-implant/frances-carmat-implants-its-first-artificial-heart-in-human-idUKBRE9BJ11L20131220?feedType=RSS&feedName=healthNews&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%253A+reuters%252FUKHealthNews+%2528News+%25.

Human Rights Watch (2012) *Losing Humanity: The Case against Killer Robots*. Available at: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

IBM (2015) *Q&A with an accessibility research pioneer Chieko Asakawa: 'AI is going to allow blind people to see the world'.*, *IBM Featured Interviews*. Available at: https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/chieko-asakawa.html.

IBM (2018) 'Australian Federal Government signs a $1B five-year agreement with IBM', *IBM News Releases*, 5 July. Available at: https://www-03.ibm.com/press/au/en/pressrelease/54124.wss.

IBM Research (2018) *Science for Social Good. Applying AI, cloud and deep science toward new societal challenges.*, *Responsibility at IBM*. Available at: https://www.research.ibm.com/science-for-social-good/#projects (Accessed: 30 November 2018).

IEEE Standards Association (2018) *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. Available at: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html.

Inclusion Australia (2018) 'My Health Record – More time, better information', 27 July. Available at: https://www.inclusionaustralia.org.au/my-health-record-more-time-better-information/.

Information Commissioner's Office (UK) (2018) *Guide to the General Data Protection Regulation (GDPR)*.

Infosys (2017) 'Amplifying Human Potential: Towards Purposeful Artificial Intelligence', p. 20. Available at: https://www.infosys.com/aimaturity/Documents/amplifying-human-potential-CEO-report.pdf.

Innes, J. M. and Bennett, R. G. (2010) 'Training the professional psychologist: Is there a need for a reappraisal of the nature of education within a scientific paradigm?', in *Fourth International Conference on Psychology Education*. Sydney.

Innes, J. M. and Morrison, B. W. (2017) 'Projecting the future impact of advanced technologies on the profession: Will a robot take my job?', *InPsych*, 39(2), pp. 34–35.

Innovation and Science Australia (2017) *Australia 2030: Prosperity through innovation*. Australian Government, Canberra. Available at: https://industry.gov.au/Innovation-and-Science-Australia/Documents/Australia-2030-Prosperity-through-Innovation-Full-Report.pdf (Accessed: 11 April 2018).

Institute of International Finance (2018) *Machine Learning in Credit Risk*. Available at: https://www.iif.com/publication/regulatory-report/machine-learning-credit-risk.

'International Covenant on Civil and Political Rights' (1966), p. 999 UNTS 171.

'International Covenant on Economic, Social and Cultural Rights' (1966), p. 993 UNTS 3.

International Data Corporation (2016) *Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending Guide*. Available at: https://www.idc.com/getdoc.jsp?containerId=IDC_P33198.

Ion, M. *et al.* (2017) 'Private Intersection-Sum Protocol with Applications to Attributing Aggregate Ad Conversions'.

Isaacman, S. *et al.* (2011) 'Identifying Important Places in People's Lives from Cellular Network Data', in *Pervasive Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 133–151.

Japanese Government (2017) *Artificial Intelligence Technology Strategy (Report of Strategic Council for AI Technology)*.

Jenson, O. B. (2007) 'City of Layers: Bangkok's Sky Train and How It Works in Socially Segregating Mobility Patterns', *Swiss Journal of Sociology*, 33(3), pp. 387–405.

Jervis-Bardy, D. (2018) 'Drone Delivers Service to Fly to Canberra's Northern Suburbs', *The Canberra Times*. Available at: https://www.canberratimes.com.au/national/act/drone-delivery-service-to-fly-to-canberra-s-northern-suburbs-20181108-p50erl.html.

Johnson, J. A. (2006) *Technology and pragmatism: From value neutrality to value criticality, SSRN Scholarly Paper, Rochester, NY: Social Science Research Network*. Available at: http://papers.ssrn.com/abstract=2154654.

Jones, O. (2017) 'We Should All be Working a Four-Day Week. Here's Why', *The Guardian*. Available at: https://www.theguardian.com/commentisfree/2017/nov/16/working-four-day-week-hours-labour.

Jovanovic, B. and Rousseau, P. (2005) *General Purpose Technologies*. Available at: https://doi.org/10.3386/w11093.

Kahneman, D. and Klein, G. (2009) 'Conditions for intuitive expertise.: A failure to disagree', *American Psychologist*, 64, pp. 515–526.

Kalantre, S. S. *et al.* (2018) *Machine Learning techniques for state recognition and auto-tuning in quantum dots*, *arXiv:1712.04914*. Available at: https://arxiv.org/abs/1712.04914.

el Kaliouby, R. (2017) *We Need Computers with Empathy*, *MIT Technolgoy Review*. Available at: https://www.technologyreview.com/s/609071/we-need-computers-with-empathy/.

Kamel Boulos, M. N. *et al.* (2014) 'Mobile medical and health apps: state of the art, concerns, regulatory control and certification', *Online Journal of Public Health Informatics*, 5(3). doi: 10.5210/ojphi.v5i3.4814.

Kaminski, M. (2018) 'The Right to Explanation, Explained', *U of Colorado Law Legal Studies Research Paper*, 18(24).

Katz, L. and Margo, R. (2014) 'Technical Change and the Relative Demand for Skilled Labour', in Bouston, L. P., Frydman, C., and Margo, R. (eds) *Human Capital in History*. Chicago: University of Chicago Press, pp. 15–57.

Keddell, E. and Davie, G. (2018) 'Inequalities and child protection system contact in Aotearoa New Zealand: developing a conceptual framework and research agenda', *Social Sciences*, 7(6), p. 89.

Kehl, D., Guo, P. and Kessler, S. (2017) *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, *HLS Student Papers*. Available at: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041.

Kelly, R. (2018) 'AI welcome to NZ homes, but privacy remains primary concern: survey'. NewsBeezer. Available at: https://newsbeezer.com/newzealand/ai-welcome-to-nz-homes-but-privacy-remains-primary-concern-survey/.

Kennedy, B. and Innes, M. (2005) 'The teaching of psychology in the contemporary university: Beyond the accreditation guidelines', *Australian Psychologist*, 40, pp. 159–169.

Kenney, M. and Zysman, J. (2016) 'The Rise of the Platform Economy', *Issues in Science and Technology*, XXXII(3). Available at: http://issues.org/32-3/the-rise-of-the-platform-economy/.

Kewley-Port, D. (1999) 'Application of current speech recognition technology to nonstandard domains', *The Journal of the Acoustical Society of America*. Acoustical Society of America, 106(4), p. 2130. doi: 10.1121/1.428012.

Khaitan, T. (2015) *A theory of discrimination law*. Oxford: Oxford University Press.

Kleinberg, J., Mullainathan, S. and Raghavan, M. (2017) 'Inherent trade-offs in the fair determination of risk scores', in *8th Conference on Innovations in Theoretical Computer Science*. Available at: https://arxiv.org/pdf/1609.05807.pdf.

Klingele, C. (2016) 'The promises and perils of evidence-based corrections', *Notre Dame Law Review*, 91(2), pp. 537–584.

Knight, W. (2017) 'Biased Algorithms Are Everywhere, and No One Seems to Care', *MIT Technology Review*, 12 July. Available at: https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/.

Knight, W. (2018) *Canada and France plan an international panel to assess AI's dangers*, *MIT Technolgoy Review*.

Korinek, A. and Stiglitz, J. E. (2018) 'Artificial Intelligence and Its Implications for Income Distribution and Unemployment', in Agrawal, A. K., Gans, J., and Goldfarb, A. (eds) *The Economics of Artificial Intelligence: An Agenda*. National Bureau of Economic Research, University of Chicago Press. Available at: http://www.nber.org/chapters/c14018.

Kratochwill, T., Doll, E. and Dickson, W. (1991) '5. Use of Computer Technologyin Behavioral Assessments', *The Computer and the Decision-Making Process*. Available at: http://digitalcommons.unl.edu/buroscomputerdecision/7 (Accessed: 11 October 2018).

Kukutai, T. and Taylor, J. (2016) 'Data sovereignty for Indigenous peoples: current practice and future needs', in Kukutai, T. and Taylor, J. (eds) *Indigenous Data Sovereignty: Toward an agenda*. Canberra: ANU Press, pp. 1–24.

Kukutai, T. and Walter, M. (2015) 'Recognition and indigenizing official statistics: Reflections from Aotearoa New Zealand and Australia', *Statistical Journal of the IAOS*, 31(2), pp. 317–326. Available at: https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji896.

Kuma, R. (2018) *Prediciting Earthquakes Using Artificial Intelligence and Neural Networks*, *Acadgild*. Available at: https://acadgild.com/blog/predicting-earthquakes-artificial-intelligence.

Kunze, L. *et al.* (2018) 'Artificial Intelligence for Long-Term Robot Autonomy: A Survey', *IEEE Robotics and Automation Letters*, 3(4), pp. 4023–4030. doi: 10.1109/LRA.2018.2860628.

Labrecque, R. M. *et al.* (2014) 'The Importance of Reassessment: How Changes in the LSI-R Risk Score Can Improve the Prediction of Recidivism', *Journal of Offender Rehabilitation*, 53(2), pp. 116–128. doi: 10.1080/10509674.2013.868389.

Larson, J. *et al.* (2016) 'How we analyzed the COMPAS recidivism algorithm', *ProPublica*, 23 May. Available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Lasi, H. *et al.* (2014) 'Industrie 4.0', *WIRTSCHAFTSINFORMATIK*, 56(4), pp. 261–264. doi: 10.1007/s11576-014-0424-4.

Lawing, K. *et al.* (2017) 'Use of Structured Professional Judgment by Probation Officers to Assess Risk for Recidivism in Adolescent Offenders', *Psychological Assessment*, 29(6), pp. 652–663.

Lee, J. D. and See, K. A. (2004) 'Trust in Automation: Designing for Appropriate Reliance', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), pp. 50–80. doi: 10.1518/hfes.46.1.50_30392.

Leetaru, K. (2018) 'Is Twitter Really Censoring Free Speech?', *Forbes*, January. Available at: https://www.forbes.com/sites/kalevleetaru/2018/01/12/is-twitter-really-censoring-freespeech/#3b3416ea65f5.

Leith, P. (1998) 'The Judge and the Computer: How Best "Decision Support"?", *Artificial Intelligence and Law*, 6(2–4), pp. 289–309. Available at: https://link.springer.com/article/10.1023%2FA%3A1008226325874#citeas.

Lepri, B. *et al.* (2017) 'Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges', *Philosophy & Technology*. Available at: https://doi.org/10.1007/s13347-017-0279-x.

Leslie, S. (2017) 'The original sin of cognition: Fear, prejudice, and generalization', *Journal of Philosophy*, 114, pp. 393–421.

Leswing, K. (2017) 'Apple just revealed how its iPhone-recycling robot "Liam" works', *Business Insider Australia*, 21 April. Available at: https://www.businessinsider.com.au/apple-liam-iphone-recycling-robot-photos-video-2017-4?r=US&IR=T.

Levendowski, A. (2017) *How copyright law can fix artificial intelligence's implicit bias problem*, *Washington Law Review*. Available at: https://ssrn.com/abstract=3024938.

Levy, N. (2017a) 'Due deference to denialism: explaining ordinary people's rejection of established scientific findings', *Synthese*. doi: 10.1007/s11229-017-1477-x.

Levy, N. (2017b) 'Nudges in a post-truth world', *Journal of Medical Ethics*, 43(8), p. 495 LP-500. Available at: http://jme.bmj.com/content/43/8/495.abstract.

Lewis-Kraus, G. (2016) 'The Great A.I. Awakening', *New York Times*, 14 December. Available at: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Lewis, P. (2018) '"Fiction is outperforming reality": how YouTube's algorithm distorts truth', *The Guardian*, 2 February. Available at: https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth.

Li, X. *et al.* (2015) 'Towards Reading Hidden Emotions: A comparative Study of Spontaneous Micro-expression Spotting and Recognition Methods'. doi: 10.1109/TAFFC.2017.2667642.

Lilienfeld, S. O. *et al.* (2014) 'Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness', *Perspectives on Psychological Science*, 9(4), pp. 355–387.

Lin, E. and Haas, L. (2002) 'IBM federated database technology', *IBM*, 13 November. Available at: https://www.ibm.com/developerworks/data/library/techarticle/0203haas/0203haas.html.

Lippe, P., Katz, D. M. and Jackson, D. (2015) 'Legal By Design: A New Paradigm for Handling Complexity in Banking Regulation and Elsewhere in Law', *Oregon Law Review*, 93(4), pp. 833–854. Available at: file:///C:/Users/Bryan/Downloads/SSRN-id2539315.pdf.

Lipton, Z. C. (2018) 'The Mythos of Model Interpretability', *Queue*, 16(3).

LiveNews (2018) 'New Zealand Research – Name that pest – Using AI to rapidly ID pest species and biosecurity risks', 18 June. Available at: https://livenews.co.nz/2018/06/18/new-zealand-research-name-that-pest-using-ai-to-rapidly-id-pest-species-and-biosecurity-risks/.

Lohr, S. (2018) 'Facial Recognition is Accurate, if You're a White Guy', *The New York Times*, 9 February. Available at: https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

Lonstein, W. (2018) *Drone technology: The good, the bad and the horrible*, *Forbes*. Available at: https://www.forbes.com/sites/forbestechcouncil/2018/01/22/drone-technology-the-good-the-bad-and-the-horrible/2/#2ab8516d8fa4 (viewed 18 June 2018).

Lucas, G. M. *et al.* (2014) 'It's only a computer: Virtual humans increase willingness to disclose', *Computers in Human Behavior*, 37, pp. 94–100. doi: 10.1016/j.chb.2014.04.043.

Luckin, R. (2018) *Machine Learning and Human Intelligence The future of education for the 21st century*. Available at: https://www.ucl-ioe-press.com/books/education-and-technology/machine-learning-and-human-intelligence/.

Lum, K. (2017) 'Rise Of The Racist Robots – How AI Is Learning All Our Worst Impulses', *The Guardian*. Stephen Buryani for The Guardian, 18 August. Available at: https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses.

Lum, K. and Isaac, W. (2016) *To predict and serve?*, *Royal Statistical Society: In Detail*. doi: 10.1111/j.1740-9713.2016.00960.x.

Luxton, D. D. (2015) *Artificial Intelligence in Behavioral and Mental Health Care*. Available at: https://www.elsevier.com/books/artificial-intelligence-in-behavioral-and-mental-health-care/luxton/978-0-12-420248-1.

Lynskey, O. (2017) 'Aligning data protection rights with competition law remedies? The GDPR right to data portability', *European Law Review*, 42, pp. 793–814.

Mackenzie, D. and Wajcman, J. (1999) 'Introductory essay: Social shaping of technology', in Mackenzie, D. and Wajcman, J. (eds) *Social Shaping of technology*. Milton Keynes: Open University Press, pp. 3–27.

Maguire, B. (2018) 'Bruce Maguire personal communication'.

Malewar, A. (2018) 'Self-driving AI wheelchair to aiding people with disabilities', *Tech Explorist*, 13 April. Available at: https://www.techexplorist.com/self-driving-ai-wheelchair-aiding-people-disabilities/13525/.

Mandate for the International Panel on Artifical Intelligence (2018) *Prime Minister of Canada*.

Manning, J. (2018) *How AI is disrupting the banking industry*. Available at: https://internationalbanker.com/banking/how-ai-is-disrupting-thebanking-industry/.

Marin-Guzman, D. and Bailey, M. (2017) *Automation could add $2.2 trillion to Australian economy by 2030*, *Australian Financial Review*. Available at: https://www.afr.com/news/economy/automation-could-add-22-trillion-to-australian-economy-by-2030-20170808-gxrks6.

Marr, B. (2018) 'How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read'. Available at: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#566d765860ba.

Martinez, J. (2017) 'Spear-Phishing Attacks: What You Need to Know', *PCMagazine*, 12 June. Available at: https://au.pcmag.com/feature/48348/spear-phishing-attacks-what-you-need-to-know.

Mason, P. (2016) 'The Racist Hijacking of Microsoft's ChatBot Shows the Internet Teems with Hate', *The Guardian*, 29 March. Available at: https://www.theguardian.com/world/2016/mar/29/microsoft-tay-tweets-antisemitic-racism.

Mattern, F. and Floerkemeier, C. (2010) 'From the Internet of Computers to the Internet of Things BT - From Active Data Management to Event-Based Systems and More: Papers in Honor of Alejandro Buchmann on the Occasion of His 60th Birthday', in Sachs, K., Petrov, I., and Guerrero, P. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 242–259. doi: 10.1007/978-3-642-17226-7_15.

Maurer, M. *et al.* (eds) (2016) *Autonomous Driving*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-48847-8.

May, M. (2018) 'Breaking Down Accessibility, Universality, and Inclusion in Design', 4 February.

Mayer, R. C., Davis, J. H. and Schoorman, F. D. (1995) 'An Integrative Model of Organizational Trust', *The Academy of Management Review*, 20(3), p. 709. doi: 10.2307/258792.

Mays, J., Marston, G. and Tomlinson, J. (2016) *Basic Income in Australia and New Zealand*. New York: Springer.

McClure, P., Sinclair, S. and Aird, W. (2015) *A New System for Better Employment and Social Outcomes: Report of the Reference Group on Welfare Reform to the Minister for Social Services*. Available at: https://www.dss.gov.au/sites/default/files/documents/02_2015/dss001_14_final_report_access_2.pdf.

McEwen, R., Eldridge, J. and Caruso, D. (2018) 'Differential or deferential to media? The effect of prejudicial publicity on judge or jury', *International Journal of Evidence and Proof*, 22(2), pp. 124–143.

McKelvey, F. and Gupta, A. (2018) 'Here's how Canada can be a global leader in ethical AI', *The Conversation*. Available at: https://theconversation.com/heres-how-canada-can-be-a-global-leader-in-ethical-ai-90991.

McKinsey&Company (2016) *Automotive revolution - perspective towards 2030*. doi: 10.1365/s40112-016-1117-8.

McKinsey&Company (2017) *Digital Australia: Seizing opportunities from the Fourth Industrial Revolution*.

McKinsey Global Institute (2017) *How artificial intelligence can deliver real value to companies*. Available at: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies.

McQuillan, D. (2017) 'The Anthropocene, resilience and post-colonial computation', *Resilience*, 5(2), pp. 92–109.

MedicMind (2018) *MedicMind*. Available at: https://www.medicmind.tech/ (Accessed: 29 November 2018).

Meehl, P. E. (1954) *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Mehta, A. (2018) 'Google Pledges Not to Develop AI Weapons', *The Telegraph*. Available at: https://www.telegraph.co.uk/technology/2018/06/08/google-pledges-not-develop-ai-weapons/.

Melbourne Water Corporation (2018) *Driving Efficiency with Artificial Intelligence and Machine Learning*, *Melbourne Water*. Available at: https://www.melbournewater.com.au/what-we-are-doing/news/driving-efficiency-artificial-intelligence-and-machine-learning.

Mesiter, J. (2018) *AI Plus Human Intelligence Is The Future Of Work*, *Forbes*. Available at: https://www.forbes.com/sites/jeannemeister/2018/01/11/ai-plus-human-intelligence-is-the-future-of-work/#530fd68a2bba.

Mesko, B. (2017) 'The role of artificial intelligence in precision medicine', *Expert Review of Precision Medicine and Drug Development*. Taylor & Francis, 2(5), pp. 239–241. doi: 10.1080/23808993.2017.1380516.

Metcalf, J., Keller, E. and Boyd, D. (2016) *Perspectives on big data, ethics, and society*. Available at: https://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/.

Metz, C. and Satariano, A. (2018) 'Silicon Valley's Giants Take Their Talent Hunt to Cambridge', *The New York Times*, 3 July. Available at: https://www.nytimes.com/2018/07/03/technology/cambridge-artificial-intelligence.html.

Meyer, G. and Shaheen, S. (eds) (2017) *Disrupting Mobility*. Cham: Springer International Publishing (Lecture Notes in Mobility). doi: 10.1007/978-3-319-51602-8.

Michie, S. *et al.* (2017) 'Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop', *Journal of medical Internet research*, 19(6).

Miconi, T. (2017) 'The Impossibility of Fairness: a Generalized Impossibility Result for Decisions'. Available at: https://arxiv.org/abs/1707.01195.

Microsoft (2018a) *FATE: Fairness, Accountability, Transparency, and Ethics in AI*. Available at: https://www.microsoft.com/en-us/research/group/fate/ (Accessed: 30 November 2018).

Microsoft (2018b) 'Microsoft partners with SRL Diagnostics to expand AI Network for Healthcare to pathology', *Microsoft News Center India*, 11 September. Available at: https://news.microsoft.com/en-in/microsoft-partners-with-srl-diagnostics-to-expand-ai-network-for-healthcare-to-pathology/.

Mihailidis, A., Barbenel, J. C. and Fernie, G. (2004) 'The efficacy of an intelligent cognitive orthosis to facilitate handwashing by persons with moderate to severe dementia', *Neuropsychological Rehabilitation*. Routledge, 14(1–2), pp. 135–171. doi: 10.1080/09602010343000156.

Mihailidis, A., Fernie, G. R. and Barbenel, J. C. (2001) 'The use of artificial intelligence in the design of an intelligent cognitive orthosis for people with dementia.', *Assistive Technology*. United States, 13(1), pp. 23–39. doi: 10.1080/10400435.2001.10132031.

Milanovic, B. (2016) *Global Inequality: A New Approach for the Age of Globalization*. Cambridge, MA: Harvard University Press.

Miller, T. (2017) *Explanation in Artificial Intelligence: Insights from the Social Sciences*, *CoRR*. Available at: http://arxiv.org/abs/1706.07269.

Mittelstadt, B. D. *et al.* (2016) 'The ethics of algorithms: Mapping the debate', *Big Data and Society*, 16, pp. 1–21.

Mokyr, J., Vickers, C. and Ziebarth, N. L. (2015) 'The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?', *Journal of Economic Perspectives*, 29(3), pp. 31–50. doi: 10.1257/jep.29.3.31.

Molster, C. *et al.* (2012) 'An Australian Approach to the Policy Translation of Deliberated Citizen Perspectives on Biobanking', *Public Health Genomics*, 15(2), pp. 82–91. Available at: https://www.karger.com/DOI/10.1159/000334104.

Monash Unviersity (2018) *All Abord the Future of Sustainable Transport*, *Monash Unviersity*. Available at: https://www.monash.edu/engineering/about-us/news-events/latest-news/articles/2018/all-aboard-the-future-of-sustainable-transport.

Mordor Intelligence (2017) *United States Artificial Intelligence in Medicine Market - Growth, Trends, and Forecasts (2017 - 2022)*, *Mordor Intelligence Healthcare Industry Report*. Available at: https://mordorintelligence.com/industry-reports/united-states-artificial-intelligence-in-medicine-market (Accessed: 1 September 2018).

Morrison, B. W., Innes, J. M. and Morrison, N. M. V (2017) 'Current advances in robotic decisionmaking: Is there such a thing as an intuitive robot?', in *Australian Psychological Society Industrial and Organisational Psychology Conference*. Sydney.

Mueller, R. S. (2018) 'United States Of America V. Internet Research Agency'. (Case 1:18-cr-00032-DLF). District Of Columbia.

Mukherjee, S. (2017) *Stanford's Artificial Intelligence Is Nearly as Good as Your Dermatologist*. Available at: http://fortune.com/2017/01/26/stanford-ai-skin-cancer/.

Munn, N. (2016) 'How mass surveillance harms societies and individuals - and what you can do about it', *Canadian Journalists for Free Expression*, 8 November. Available at: https://www.cjfe.org/how_mass_surveillance_harms_societies_and_individuals_and_what_you_can_do_about_it.

Narayanan, A. and Shmatikov, V. (2008) 'Robust de-anonymization of large sparse datasets', in *IEEE Symposium on Security and Privacy*, pp. 111–125.

Natale, S. and Ballatore, A. (2017) 'Imagining the thinking machine', *Convergence: The International Journal of Research into New Media Technologies*, p. 135485651771516. doi: 10.1177/1354856517715164.

National Transport Commission (2018) *Changing Driving Laws to Support Automated Vehicles. Policy Paper*. Available at: http://www.ntc.gov.au/Media/Reports/(B77C6E3A-D085-F8B1-520D-E4F3DCDFFF6F).pdf.

Naveed, K., Watanabe, C. and Neittaanmäki, P. (2017) 'Co-evolution between streaming and live music leads a way to the sustainable growth of music industry – Lessons from the US experiences', *Technology in Society*, 50, pp. 1–19.

Ndiomewese, I. (2018) 'This is Nigeria's first ever Artificial Intelligence hub', *Techpoint.Africa*, 11 June. Available at: https://techpoint.africa/2018/06/11/nigerias-first-ever-artificial-intelligence-hub/.

New Zealand Data Futures Forum (2018) *Harnessing the economic and social power of data*. Available at: http://datafutures.co.nz/assets/Uploads/Data-Futures-Forum-Key-recommendations.pdf.

New Zealand Government (2018a) *Digital service design standard*, *DIGITAL.GOVT.NZ*. Available at: https://www.digital.govt.nz/standards-and-guidance/digital-service-design-standard/ (Accessed: 3 December 2018).

New Zealand Government (2018b) 'Government will move quickly on AI action plan', 2 May. Available at: https://www.beehive.govt.nz/release/government-will-move-quickly-ai-action-plan.

New Zealand Human Rights Commission (2018) *Privacy, Data and Technology: Human Rights Challenges in the Digital Age*. Available at: https://www.hrc.co.nz/files/5715/2575/3415/Privacy_Data_Technology_-_Human_Rights_Challenges_in_the_Digital_Age_FINAL.pdf.

Newton, C. (2017) 'How YouTube Perfected the Feed', *The Verge*, 30 August. Available at: https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-video-recommendation-personalized-feed.

Nezami, E. and Butcher, J. N. (2000) 'Chapter 16 - Objective Personality Assessment', in Goldstein, G. and Hersen, M. B. T.-H. of P. A. (Third E. (eds). Amsterdam: Pergamon, pp. 413–435. doi: https://doi.org/10.1016/B978-008043645-6/50094-X.

Nguyen, C. T. (2018) 'Cognitive islands and runaway echo chambers: problems for epistemic dependence on experts', *Synthese*, pp. 1–19. doi: 10.1007/s11229-018-1692-0.

Nicholson, B. (2018) *P.W. Singer: Adapt Fast, or Fail*, *Real Clear Defence*. Available at: https://www.realcleardefense.com/articles/2018/07/07/pw_singer_adapt_fast_or_fail_113585.html.

Nickerson, R. S. (1998) 'Confirmation bias: A ubiquitous phenomenon in many guises.', *Review of General Psychology*. US: Educational Publishing Foundation, 2(2), pp. 175–220. doi: 10.1037/1089-2680.2.2.175.

Nissenbaum, H. (2010) *Privacy in Context: technology, policy, and the integrity of social life*. Stanford Law Books.

Nissenbaum, H. (2011) 'A Contextual Approach to Privacy Online', *Daedalus*, 140(4), pp. 32–48.

NITI Aayog (2018) *National Strategy for Artificial Intelligence*. Available at: http://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf.

Nixon, D. (2017) *Is the economy suffering from the crisis of attention?*, *Bank Underground*. Available at: https://bankunderground.co.uk/2017/11/24/is-the-economy-suffering-from-the-crisis-ofattention/.

Norton Rose Fulbright (2018) *eDiscovery*, *Dispute resolution and litigation: eDiscovery*. Available at: http://www.nortonrosefulbright.com/au/our-services/dispute-resolution-and-litigation/ediscovery/.

NSTC (2016) *The National Artificial Intelligence Research and Development Strategic Plan*. Available at: https://www.nitrd.gov/news/national_ai_rd_strategic_plan.aspx.

Nyhan, B. and Reifler, J. (2013) *Which corrections work? Research results and practice recommendations*, *New America Foundation, Media Policy Initiative*. Washington, D.C. Available at: https://www.dartmouth.edu/~nyhan/nyhan-reifler-report-naf-corrections.pdf.

NZ Privacy Commissioner (2018) *Principles for safe and effective use of data and analytics*, *Office of the Privacy Commissioner*. Available at: https://www.privacy.org.nz/news-and-publications/guidance-resources/principles-for-the-safe-and-effective-use-of-data-and-analytics-guidance/.

Office of the Auditor General (2016) *Performance Audit: Michigan Integrated Data Automated System (MiDAS)*, Available at: https://audgen.michigan.gov/wp-content/uploads/2016/06/rs641059315.pdf

O'Mallon, F. (2017) 'NSW police visit homes of people on secret watchlist without cause', *The Sydney Morning Herald*, 11 November. Available at: https://www.smh.com.au/national/nsw/nsw-police-visit-homes-of-people-on-secret-watchlist-without-cause-20171107-gzgcwg.html.

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. Available at: https://www.amazon.com/Weapons-Math-Destruction-Increases-Inequality/dp/0553418815.

Oak, E. (2016) 'A Minority Report for Social Work? The Predictive Risk Model (PRM) and the Tuituia Assessment Framework in addressing the needs of New Zealand's Vulnerable Children', *British Journal of Social Work*. Oxford University Press, 46(5), pp. 1208–1223. doi: 10.1093/bjsw/bcv028.

OECD (2009) *Sickness, disability and work: Keeping on track in the economic downturn – Background paper, High-Level Forum, Stockholm, 14-15 May*. Available at: https://www.oecd.org/els/emp/42699911.pdf.

OECD (2017a) *Education at a Glance 2017*. Available at: https://www.oecd-ilibrary.org/education-at-a-glance-2017_5jfrn2shpfxt.pdf?itemId=%2Fcontent%2Fpublication%2Feag-2017-en&mimeType=pdf.

OECD (2017b) *OECD Science, Technology and Industry Scoreboard 2017*, *Science, Technology and Industry Scoreboard 2017*. Available at: http://www.oecd.org/sti/oecd-science-technology-and-industry-scoreboard-20725345.htm.

OECD (2018a) *Income Inequality*, *Organisation for Economic Co-operation and Development*. Available at: https://data.oecd.org/inequality/income-inequality.htm.

OECD (2018b) *Main Science and Technology Indicators*, *OECD, Directorate for Science, Technology and Innovation*. Available at: http://www.oecd.org/sti/msti.htm.

OECD (2018c) *Poverty Rate*, *Organisation for Economic Co-operation and Development*. Available at: https://data.oecd.org/inequality/poverty-rate.htm.

Office of the Australian Information Commissioner (2015) *Chapter 11: APP 11 — Security of personal information*, *APP Guidelines*. Available at: https://www.oaic.gov.au/agencies-and-organisations/app-guidelines/chapter-11-app-11-security-of-personal-information (Accessed: 18 July 2018).

Office of the Australian Information Commissioner (2016a) *Australian Community Attitudes to Privacy Survey*. Available at: https://www.oaic.gov.au/resources/engage-with-us/community-attitudes/acaps-2017/acaps-2017-report.pdf.

Office of the Australian Information Commissioner (2016b) *Privacy shortcomings of Internet of Things businesses revealed*. Sydney, Australia.

Office of the Chief Scientist (2016) *Australia's STEM Workforce*. Available at: https://www.chiefscientist.gov.au/2016/03/report-australias-stem-workforce/.

Office of the Children's Commissioner (2015) *State of care 2015*.

Office of the Director of National Intelligence (2017) *Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution*. Available at: https://www.intelligence.senate.gov/sites/default/files/documents/ICA_2017_01.pdf.

Ohm, P. (2010) 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', *UCLA Law Review*, 57, pp. 1707–1778.

Open Knowledge International (2016) *The Global Open Data Index*. Available at: https://index.okfn.org/.

OpenAI (2018) *OpenAI - home*. Available at: https://openai.com/.

OpenSignal (2018) *The State of LTE (February 2018)*. Available at: https://opensignal.com/reports/2018/02/state-of-lte.

Oskamp, A. and Tragter, M. W. (1997) 'Automated Legal Decision Systems in Practice: The Mirror of Reality', *Artificial Intelligence and Law*, 5(4), pp. 291–322. doi: 10.1023/A:1008298517919.

Osman, M. (2010) 'Controlling uncertainty: A review of human behavior in complex dynamic environments.', *Psychological Bulletin*, 136(1), pp. 65–86. doi: 10.1037/a0017815.

Page-Smith, J. and Northrop, A. (2017) 'A Community Survey of Attitudes Towards Autonomous Vehicles', in *Proceedings of the 2017 Australasian Road Safety Conference 10th– 12th October, Perth, Australia*, p. 2. Available at: http://acrs.org.au/files/papers/arsc/2017/Page-Smith_00165_EA.pdf.

Palk, G. R., Freeman, J. E. and Davey, J. D. (2008) 'Australian forensic psychologists' perspectives on the utility of actuarial versus clinical assessment for predicting recidivism among sex offenders', in *Proceedings 18th Conference of the European Association of Psychology and Law*. Maastricht, The Netherlands.

Palmer, J. (2017) *Quantum technology is beginning to come into its own*, *The Economist*. Available at: https://www.economist.com/news/essays/21717782-quantum-technology-beginning-come-its-own.

Palmier-Claus, J. E. *et al.* (2012) 'The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application', *BMC Psychiatry*, 12(1), p. 172. doi: 10.1186/1471-244X-12-172.

Paredes, D. (2018) 'AI welcome in NZ homes, but privacy remains prime concern: survey', *CIO New Zealand*, 9 July. Available at: https://www.cio.co.nz/article/643508/ai-welcome-nz-homes-privacy-remains-prime-concern-survey/.

Pariser, E. (2011) *The Filter Bubble: What the Internet is hiding from you*. Penguin Group.

Parker, S. K. (2014) 'Beyond Motivation: Job and Work Design for Development, Health, Ambidexterity, and More', *Annual Review of Psychology*, 65(1), pp. 661–691. doi: 10.1146/annurev-psych-010213-115208.

Parker, S. K., Van den Broeck, A. and Holman, D. (2017) 'Work Design Influences: A Synthesis of Multilevel Factors that Affect the Design of Jobs', *Academy of Management Annals*, 11(1), pp. 267–308. doi: 10.5465/annals.2014.0054.

Parliament of Victoria (2018) 'Artificial Intelligence group launched', *Parliament News*, 7 March. Available at: https://www.parliament.vic.gov.au/about/news/4029-artificial-intelligence-group-launched.

Partnership on AI (2018) *About Us*. Available at: https://www.partnershiponai.org/about/.

Pasquale, F. (2015) *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Peng, T. (2018) 'McKinsey Report: AI Promises Added Value of Up to US$5.8 Trillion', *Synced: AI Technology & Industry Review*, 23 April. Available at: https://medium.com/syncedreview/mckinsey-report-ai-promises-added-value-of-up-to-us-5-8-trillion-80cc0043ebf6.

Petit, N. (2017) 'Law and Regulation of Artificial Intelligence and Robots - Conceptual Framework and Normative Implications'. Available at: http://dx.doi.org/10.2139/ssrn.2931339.

Pettigrew, S. (2018) 'Driverless cars really do have health and safety benefits, if only people knew', *The Conversation*, 5 July. Available at: https://theconversation.com/driverless-cars-really-do-have-health-and-safety-benefits-if-only-people-knew-99370.

Pichai, S. (2018) 'AI at Google: our principles', *Google Comapny News*, 7 June. Available at: https://www.blog.google/technology/ai/ai-principles/.

Plantera, F. and Di Stasi, L. (2017) 'A conversation with Marten Kaevats on e-governance and Artificial Intelligence', *e-estonia*, October. Available at: https://e-estonia.com/a-conversation-with-marten-kaevats-on-e-governance-and-artificial-intelligence/.

Pomerol, J.-C. and Adam, F. (2008) 'Understanding human decision making: A fundamental step towards effective intelligent decision support', in Phillips-Wren, G., Ichalkaranje, N., and Jain, L. C. (eds) *Intelligent decision making: An AI-based approach*, pp. 41–76.

Porter, E. (2017) 'Why Big Cities Thrive, and Smaller Ones Are Being Left Behind', *The New York Times*. Available at: https://www.nytimes.com/2017/10/10/business/economy/big-cities.html.

Powell, D. (2018) *How Mikaela Jade built augmented reality startup Indigital from deep in Kakadu National Park*, *Smart Company*.

Pretz, K. (2018) 'New IEEE Courses on Ethics and AI and Autonomous Systems', *The Institute*, 20 April. Available at: http://theinstitute.ieee.org/resources/products-and-services/new-ieee-courses-on-ethics-and-ai-and-autonomous-systems.

Prince, K. and Butters, R. P. (2014) *Brief Report: An Implementation Evaluation of the LSI-R as a Recidivism Risk Assessment Tool in Utah*. Available at: https://socialwork.utah.edu/wp-content/uploads/sites/4/2016/11/LSI-R-Summary-Report-Final-v2.pdf.

'Privacy Act 1988 (Cth)' (no date).

Productivity Commission (2016) *Data Availability and Use*. Available at: https://www.pc.gov.au/inquiries/completed/data-access#report.

Purdy, M. and Dougherty, P. (2017) 'Why Artificial Intelligence is the Future of Growth', p. 27. Available at: https://www.accenture.com/t20170927T080049Z__w__/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.PDFla=en.

PwC (2017) *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.

Quiggin, J. (2017) *Financing a UBI/GMI*. Available at: https://johnquiggin.com/2017/11/23/financing-a-ubigmi/.

Rabesandratana, T. (2018a) *Emmanuel Macron wants France to become a leader in AI and avoid 'dystopia'*, *Science*. Available at: http://www.sciencemag.org/news/2018/03/emmanuel-macron-wants-france-become-leader-ai-and-avoid-dystopia.

Rabesandratana, T. (2018b) 'With €1.5 billion for artificial intelligence research, Europe pins hopes on ethics', *Science*, April. Available at: http://www.sciencemag.org/news/2018/04/15-billion-artificial-intelligence-research-europe-pins-hopes-ethics.

Railway Gazette (2018) 'SNCF Targets Autonomous Trains in Five Years', *Railway Gazette*. Available at: https://www.railwaygazette.com/news/traction-rolling-stock/single-view/view/sncf-targets-autonomous-trains-in-five-years.html.

Redden, E. S., Elliott, L. R. and Barnes, M. J. (2014) 'Robots: The new teammates', in Coovert, M. D. and Thompson, L. F. (eds) *The Psychology of Workplace Technology*. New York: Routledge, pp. 185–208.

Reese, B. (2018) *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity*. New York: Atria Books.

Reese, H. (2016) *Why Microsoft's 'Tay' AI bot went wrong*, *Tech Republic*. Available at: https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/.

Reeves, B. and Nass, C. (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.

Reisman, D. *et al.* (2018) *Algorithmic impact assessments: A practical framework for public agency accountability*, *AI Now Institute*. Available at: https://ainowinstitute.org/aiareport2018.pdf.

Rexford, J. and Kirkland, R. (2018) *The role of education in AI (and vice versa)*, *McKinsey&Company*. Available at: https://www.mckinsey.com/featured-insights/artificial-intelligence/the-role-of-education-in-ai-and-vice-versa (Accessed: 21 July 2018).

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016) '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, New York, USA, pp. 1135–1144. Available at: http://doi.acm.org/10.1145/2939672.2939778.

Rifkin, J. (2015) *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*. Available at: https://www.amazon.com/gp/product/1137280115/ref=dbs_a_def_rwt_hsch_vapi_taft_p1_i0.

Riolo, K. and Bourgeat, P. (2018) *Brave New World: Are consumers ready for AI*. Available at: https://www.ipsos.com/sites/default/files/ct/publication/documents/2017-09/IAA-AI_Report_v3.pdf.

RioTinto (2018) *Rio Tinto Achieves First Delivery of Iron Ore with World's Largest Robot*, *Rio Tinto*. Available at: https://www.riotinto.com/media/media-releases-237_25824.aspx.

Robertson, G. (2018) *Future of Work Tripartite Forum Champions Skills Shift Programme*, *Beehive.govt.nz*. Available at: https://www.beehive.govt.nz/release/future-work-tripartite-forum-champions-skills-shift-programme.

Robinson, H., Broadbent, E. and MacDonald, B. (2016) 'Group sessions with Paro in a nursing home: Structure, observations and interviews', *Australasian Journal on Ageing*, 35(2), pp. 106–112. doi: 10.1111/ajag.12199.

Rogers, E. (2003) *Diffusion of innovations*. New York: Free Press.

Roland Berger and Asgard (2018) *Artificial Intelligence – A strategy for European startups. Recommendations for Policymakers*. Available at: file:///C:/Users/laure/Downloads/roland_berger_ai_strategy_for_european_startups.pdf.

Rolls-Royce plc (2016) *Autonomous ships: The next step*. Available at: https://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/rr-ship-intel-aawa-8pg.pdf.

Ross, C. and Swetlitz, I. (2017) *IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close*, *Stat News*. Available at: https://www.statnews.com/2017/09/05/watson-ibm-cancer/.

Roy Morgan (2017) 'Australians want driverless cars - NOW', 7 April. Available at: http://www.roymorgan.com/findings/7209-sotn-auto-driverless-cars-april-2017-201704061939.

Royal Bank of Canada (2018) *Humans Wanted: How Canadian youth can thrive in the age of disruption*. Available at: https://www.rbc.com/dms/enterprise/futurelaunch/_assets-custom/pdf/RBC-Future-Skills-Report-FINAL-Singles.pdf.

Rubery, J. (2018) *Automation has the potential to improve gender equality at work, The Conversation*, *The Conversation*. Available at: https://theconversation.com/automation-has-the-potentialto-improve-gender-equality-at-work-96807.

Rural Industries Research and Development Corporation (2016) *Artificial Intelligence*. Available at: https://www.agrifutures.com.au/wp-content/uploads/publications/16-038.pdf.

Russell, S. J. and Norvig, P. (2003) *Artificial Intelligence, A Modern Approach*. Second Edi. University of Michigan Press.

Ryan, R. M. and Deci, E. L. (2000) 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.', *American Psychologist*, 55(1), pp. 68–78. doi: 10.1037/0003-066X.55.1.68.

Ryan, R. M. and Deci, E. L. (2017) *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. Available at: https://www.amazon.com/Self-Determination-Theory-Psychological-Motivation-Development/dp/1462528767.

Sabbagh, D. (2018) 'Facebook to expand inquiry into Russian influence of Brexit', *The Guardian*, 18 January. Available at: https://www.theguardian.com/technology/2018/jan/17/facebook-inquiry-russia-influence-brexit.

Sacks, S. (2017) *Beyond the Worst-Case Assumptions on China's Cybersecurity Law*, *Centre for Strategic & International Studies*. Available at: https://www.csis.org/blogs/technology-policy-blog/beyond-worst-case-assumptions-chinas-cybersecurity-law.

Safe Work Australia (2015) *Handbook - Principles of Good Work Design*. Available at: https://www.safeworkaustralia.gov.au/doc/handbook-principles-good-work-design.

Santos, I. (2016) *Labor market polarization in developing countries: challenges ahead*, *The World Bank*. Available at: http://blogs.worldbank.org/developmenttalk/labor-market-polarization-developing-countries-challenges-ahead.

Saul, J. (2013) 'Implicit bias, stereotype threat, and women in philosophy', in Hutchison, K. and Jenkins, F. (eds) *Women in philosophy: What needs to change?* Oxford: Oxford University Press.

Scharre, P. and Horowitz, M. (2018) *Artificial Intelligence: What Every Policymaker Needs to Know*. Available at: https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know.

Schild, U. J. (1992) *Expert systems and case law*. Ellis Horwood Limited.

Schneider, T., Hong, G. H. and Le, A. Van (2018) *Land of the rising robots*. Available at: https://www.imf.org/external/pubs/ft/fandd/2018/06/japan-labor-force-artificial-intelligence-and-robots/schneider.pdf.

Schumpeter, J. (1975) *Capitalism, Socialism, and Democracy*. New York: Harper.

Schwab, K. (2017) *The Fourth Industrial Revolution*. Available at: https://www.penguin.co.uk/books/304971/the-fourth-industrial-revolution/.

Scudellari, M. (2017) *Lidar-Equipped Autonomous Wheelchairs Roll Out in Singapore and Japan*, *IEEE Spectrum*. Available at: https://spectrum.ieee.org/transportation/self-driving/lidar-equipped-autonomous-wheelchairs-roll-out-in-singapore-and-japan (Accessed: 20 September 2018).

Selby, J. (2017) 'Data localization laws: trade barriers or legitimate responses to cybersecurity risks, or both?', *International Journal of Law and Information Technology*, 25(3), pp. 213–232.

Senate Community Affairs References Committee (2017) *Delivery of outcomes under the National Disability Strategy 2010-2020 to build inclusive and accessible communities*. Available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/ Community_Affairs/AccessibleCommunities/Report.

Sentas, V. and Pandolfini, C. (2017) *Policing Young People in NSW: A Study of the Suspect Targeting Management Plan. A Report of the Youth Justice Coalition NSW*.

Sepuloni, C. (2018) 'New Privacy, Human Rights and Ethics framework essential step in safe data use', *New Zealand Government Media Release*, 10 May. Available at: https://www.beehive.govt.nz/release/new-privacy-human-rights-and-ethics-framework-essential-step-safe-data-use.

Shah, S. (2018) 'China Uses Facial Recognition to monitor Ethnic Minorities', *Bloomberg News*, 18 January. Available at: https://www.bloomberg.com/news/articles/2018-01-17/china-said-to-test-facial-recognition-fence-in-muslim-heavy-area.

Shane, S. and Wakabayashi, D. (2018) '"The Business of War": Google Employees Protest Work for the Pentagon', *The New York Times*, 4 April. Available at: https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html.

Shapiro, A. (2017) 'Reform predictive policing', *Nature*, 541, pp. 458–460.

Sharif, M. *et al.* (2016) 'Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition', in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. New York, New York, USA, pp. 1528–1540. Available at: http://doi.acm.org/10.1145/2976749.2978392.

Shen, X. (2018) 'Facial recognition camera catches top businesswoman "jaywalking" because her face was on a bus', *Abacus News*. Available at: https://www.abacusnews.com/digital-life/facial-recognition-camera-catches-top-businesswoman-jaywalking-because-her-face-was-bus/article/2174508.

Shoham, Y. *et al.* (2017) *AI Index*, *Stanford University*. Available at: https://aiindex.org/2017-report.pdf.

Siau, K. and Wang, W. (2018) 'Building Trust in Artificial Intelligence, Machine Learning, and Robotics', *Cutter Business Technology Journal*, 31(2), pp. 47–53.

Simpson, B. (2016) 'Algorithms or advocacy: does the legal profession have a future in a digital world?', *Information & Communications Technology Law*, 25(1), pp. 50–61. doi: 10.1080/13600834.2015.1134144.

Sims, R. (2017) 'The ACCC's approach to colluding robots', in *Can robots collude?* Available at: https://www.accc.gov.au/speech/the-accc's-approach-to-colluding-robots.

Singer, N. (2018) 'Tech's Ethical "Dark Side": Harvard, Stanford and Others Want to Address It', *The New York Times*. Available at: https://www.nytimes.com/2018/02/12/business/computer-science-ethics-courses.html.

Singer, P. W. (2009) *Wired for War*. Available at: http://wiredwar.pwsinger.com/.

Smith, B. (2018) *Facial recognition technology: The need for public regulation and corporate responsibility*. Available at: https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/.

Smith, D. (2018) 'Putin's chef, a troll farm and Russia's plot to hijack US democracy', *The Guardian*, 18 February. Available at: https://www.theguardian.com/us-news/2018/feb/17/putins-chef-a-troll-farm-and-russias-plot-to-hijack-us-democracy.

Smith, G. J. D. (2014) 'Opening the Black Box The Work of Watching', p. 202. Available at: https://www.taylorfrancis.com/books/9781134085750.

Snipp, M. (2016) 'What does data sovereignty imply: what does it look like?', in Kukutai, T. and Taylor, J. (eds) *Indigenous Data Sovereignty: Towards an agenda*, pp. 39–56.

Solomon, S. (2017) 'Nvidia sees Israel as a key to leadership in AI technologies', *The Times of Israel*, 2 August. Available at: https://www.timesofisrael.com/nvidia-sees-israel-as-key-to-leadership-in-ai-technologies/.

Solove, D. J. (2013) 'Privacy Self-Management and the Consent Dilemma', *Harvard Law Review*, 126.

Spencer, D. (2018) 'Fear and Hope in an Age of Mass Automation: Debating the Future of Work', *New Technology, Work and Employment*, 33(1), pp. 1–12.

Srivatsa, M. and Hicks, M. (2012) 'Deanonymizing mobility traces: Using social network as a side-channel', in *Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 628–637.

Srnicek, N. and Williams, A. (2016) *Inventing the Future: Postcapitalism and a World without Work*. London: Verso.

State Council (2017) *A Next Generation Artificial Intelligence Development Plan (Chinese)*.

Stats NZ (2018) *Algorithm Assessment Report*. Available at: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf.

Steele, C. (2018) 'The Real Reason Voice Assistants Are Female (and Why it Matters)', *PC Mag Australia*. Available at: http://au.pcmag.com/opinion/51146/the-real-reason-voice-assistants-are-female-and-why-it-matters.

Stern, J. (2017) 'Alexa, Siri, Cortana: The Problem with All-Female Digital Assistants', *The Wall Street Journal*, 21 February. Available at: https://www.wsj.com/articles/alexa-siri-cortana-the-problem-with-all-female-digital-assistants-1487709068.

Stilgoe, J. (2018) 'Machine learning, social learning and the governance of self-driving cars', *Social Studies of Science*, 48(1), pp. 25–56. doi: 10.1177/0306312717741687.

Stirling, R., Miller, H. and Martinho-Truswell, E. (2017) *Government AI Readiness Index*, *Oxford Insights*. Available at: https://www.oxfordinsights.com/government-ai-readiness-index/ (Accessed: 1 September 2018).

Stobbs, N., Hunter, D. and Bagaric, M. (2017) 'Can Sentencing Be Enhanced by the Use of Artificial Intelligence?', *Criminal Law Journal*, 41.

Stoyanov, S. R. *et al.* (2015) 'Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps', *JMIR mHealth and uHealth*, 3(1), p. e27. doi: 10.2196/mhealth.3422.

Stoyles, M. (2017) *MiCare launches aged care robot to benefit workers*, *Australian Ageing Agenda*.

Strahilevitz, L. J. (2013) 'Toward a Positive Theory of Privacy Law', *Harvard Law Review*, 126, pp. 2010–2042.

Stubbs, A. (2017) 'Automation, Artificial Intelligence, and the God/Useless Divide', *Perspectives on Global Development and Technology*, 16(6), pp. 700–716.

Sunstein, C. R. (2001) 'Of Artifical Intelligence and Legal Reasoning', *University of Chicago Law School Roundtable*, 8. Available at: http://dx.doi.org/10.2139/ssrn.289789.

Sunstein, C. R. (2017) *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Supreme Court of Wisconsin (2016) 'STATE of Wisconsin, Plaintif-Respondent, v. Eric L. LOOMIS, Defendant-Appellant'. Available at: http://www.courts.ca.gov/documents/BTB24-2L-3.pdf.

Susskind, R. and Susskind, D. (2015) *The Future of the Professions How Technology Will Transform the Work of Human Experts*.

Synergies Economic Consulting (2018) *The robotics and automation advantage for Queensland*. Available at: https://cms.qut.edu.au/__data/assets/pdf_file/0006/783888/Synergies-summary.pdf.

Synytsky, R. (2017) *GDPR and Data Localization: The Significant (and Often Unforeseen) Impact on the Cloud*, *SC Media: The Cybersecurity Source*. Available at: https://www.scmagazine.com/home/blogs/executive-insight/gdpr-and-data-localization-the-significant-and-often-unforeseen-impact-on-the-cloud/.

Tamatea, A. J. (2016) 'Culture is our business: Issues and challenges for forensic and correctional psychologists', in *ANZFSS 23rd International Symposium on the Forensic Sciences: Together InForming Justice*. Auckland.

Tatman, R. (2016) *Google's speech recognition has a gender bias*, *Making Noise and Hearing Things*. Available at: https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/.

Tene, O. and Polonetsky, J. (2013) 'Big Data for All: Privacy and User Control in the Age of Analytics', *Northwestern Journal of Technology and Intellectual Property*, 11(5), pp. 237–273.

Tett, G. (2017) 'An anthropologist in the boardroom', 21 April. Available at: https://www.ft.com/content/38e276a2-2487-11e7-a34a-538b4cb30025.

Thaler, R. H. and Sunstein, C. R. (2009) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin.

The AI Forum of New Zealand (2018) *Artificial intelligence: Shaping a future for New Zealand*. Available at: https://aiforum.org.nz/wp-content/uploads/2018/07/AI-Report-2018_web-version.pdf.

The Allens Hub for Technology Law & Innovation (2018) *Response to Issues Paper on Data Sharing and Release*. Available at: available on request.

The Economist Intelligence Unit (2018a) *The Automation Readiness Index*, *The Automation Readiness Index*. Available at: http://www.automationreadiness.eiu.com/.

The Economist Intelligence Unit (2018b) *Inclusive Internet Index*. Available at: https://theinclusiveinternet.eiu.com/.

The Guardian (2018) 'Data Science Nigeria opens 1st Artificial Intelligence Hub in Unilag', *The Guardian Nigeria*, 6 June. Available at: https://guardian.ng/technology/data-science-nigeria-opens-1st-artificial-intelligence-hub-in-unilag/.

The Law Foundation New Zealand (2018) *Information Law & Policy Project*, *The Law Foundation New Zealand*. Available at: https://www.lawfoundation.org.nz/?page_id=2381.

The World Bank (2018) *World Bank Open Data*. Available at: https://data.worldbank.org/.

Thomas, J. *et al.* (2017) *Measuring Australia's digital divide: the Australian digital inclusion index 2017*. Melbourne. doi: 10.4225/50/596473db69505.

Thomsen, S. (2018) 'South Australia Just Put a Driverless Bus on Public Roads', *Business Insider*. Available at: https://www.businessinsider.com.au/south-australia-just-put-a-driverless-bus-on-public-roads-2018-6.

Threadgold, S. (2018) 'Creativity, Precarity and Illusio: DIY Cultures and "Choosing Poverty"', *Cultural Sociology*, 12(2), pp. 156–173.

Times Higher Education (2017) 'Which Countries and Universities are Leading AI Research?', 22 May. Available at: https://www.timeshighereducation.com/data-bites/which-countries-and-universities-are-leading-ai-research.

Tomaszewski, W., Perales, F. and Xiang, N. (2017) 'Career guidance, school experiences and the university participation of young people from low socio-economic backgrounds', *International Journal of Educational Research*, 5, pp. 11–23.

Tranter, K. (2016) 'The Challenges of Autonomous Motor Vehicles to Queensland Road and Criminal Laws', *Queensland University of Technology Law Review*, 16(2), pp. 59–81. doi: 10.5204/qutlr.v16i2.655.

Trilling, B. and Fadel, C. (2012) *21st Century Skills: Learning for Life in Our Times*. Available at: https://www.amazon.com/21st-Century-Skills-Learning-Times/dp/1118157060.

Turchin, A. and Denkenberger, D. (2018) 'Classification of global catastrophic risks connected with artificial intelligence', *AI & SOCIETY*. doi: 10.1007/s00146-018-0845-5.

Turkle, S. (2012) *Alone Together: Why We Expect More from Technology and Less from Each Other*. Available at: https://www.amazon.com/Alone-Together-Expect-Technology-Other/dp/0465031463.

Turner, A. (2018) *Capitalism in the age of robots: work, income and wealth in the 21st-century*. Available at: https://www.ineteconomics.org/research/research-papers/capitalism-in-the-age-of-robots-work-income-and-wealth-in-the-21st-century.

Turow, J., Hennessy, M. and Bleakley, A. (2008) 'Consumers' understanding of privacy rules in the marketplace', *Journal of consumer affairs*, 3, pp. 411–424.

UK Government (2018) *Policy Paper: AI Sector Deal*. Available at: https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal.

UK Government Office for Science (2016) *Artificial Intelligence – Opportunities and Implications for the Future of Decision Making*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf.

UMR Research (2018) *The Privacy Concerns and Sharing Data survey (2018), Commissioned by the Office of the Privacy Commissioner and conducted by UMR Research*. Available at: https://www.privacy.org.nz/news-and-publications/surveys/privacy-survey-2018/ (Accessed: 12 August 2018).

United Nations (2006) *Convention on the Rights of Persons with Disabilities (CRPD): Article 9 - Accessibility, United Nations*. Available at: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-9-accessibility.html.

United Nations Women (2017) *Making innovation and technology work for women*. Available at: http://www.unwomen.org/en/digital-library/publications/2017/7/making-innovation-and-technology-work-for-women.

Université de Montréal (2017) *The Montréal Declaration: Responsible AI, Montréal Declaration: Responsible AI*. Available at: https://www.montrealdeclaration-responsibleai.com/the-declaration (Accessed: 18 September 2018).

University of Otago (2018) *Artificial Intelligence and Law in New Zealand*. Available at: http://www.cs.otago.ac.nz/research/ai/AI-Law/.

US Federal Reserve (2011) *Supervision and Regulation Letters: SR 11-7: Guidance on Model Risk Management*, Board of Governors of the Federal Reserve System. Available at https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm

Vaithianathan, R. *et al.* (2013) 'Children in the public benefit system at risk of maltreatment: Identification via predictive modelling', *Am J Prev Med*, 45(3), pp. 354–359.

Veale, M. and Edwards, L. (2017) 'Slave to the Algorithm? Why a "Right to Explanation" is Probably Not the Remedy You Are Looking For', *Duke Law and Technology Review*, 18.

Victorian Government (2006) 'Victorian Charter of Human Rights and Responsibilities Act'. Available at: http://www5.austlii.edu.au/au/legis/vic/consol_act/cohrara2006433/.

Vijayakumar, S. (2017) *Algorithmic Decision-Making: to what extent should computers make decisions for society?, Harvard Politics*. Available at: http://harvardpolitics.com/covers/algorithmic-decision-making-to-what-extent-should-computers-make-decisions-for-society/.

Vincent, J. (2018) 'MIT is investing $1 billion in an AI college', *The Verge*. Available at: https://www.theverge.com/2018/10/15/17978056/mit-college-of-computing-ai-interdisciplinary-research.

Vinyals, O. *et al.* (2017) 'Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), pp. 652–663. doi: 10.1109/TPAMI.2016.2587640.

Vrieze, S. I. and Grove, W. M. (2010) 'Multidimensional assessment of criminal recidivism: Problems, pitfalls, and proposed solutions.', *Psychological Assessment*, 22(2), pp. 382–395. doi: 10.1037/a0019228.

Wade, M. (2018) 'Australia more cautious about driverless cars than many nations: poll', *Sydney Morning Herald*, 6 April. Available at: https://www.smh.com.au/national/australia-more-cautious-about-driverless-cars-than-many-nations-poll-20180405-p4z80g.html.

Wajcman, J. (2002) 'Addressing technological change: The challenge to social theory', *Current Sociology*, 50(3), pp. 347–363.

Wajcman, J. (2008) 'Life in the fast lane? Towards a sociology of technology and time', *The British journal of sociology*, 59(1), pp. 59–77.

Wallach, W. and Allen, C. (2008) *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press. doi: 10.1093/acprof:oso/9780195374049.001.0001.

Walsh, T. (2017) *The AI Revolution*. Available at: https://education.nsw.gov.au/media/exar/The_AI_Revolution_TobyWalsh.pdf.

Walter, M. and Andersen, C. (2013) *Indigenous Statistics: A Quantitative Research Methodology*. New York: Routledge.

West, D. and Lockley, A. (2016) 'Implementing Digital Badges in Australia: The Importance of Institutional Context BT - Foundation of Digital Badges and Micro-Credentials: Demonstrating and Recognizing Knowledge and Competencies', in Ifenthaler, D., Bellin-Mularski, N., and Mah, D.-K. (eds) *Foundation of Digital Badges and Micro-Credentials*. Cham: Springer International Publishing, pp. 467–482. doi: 10.1007/978-3-319-15425-1_26.

Wilkins, R. (2017) *The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 15*. Available at: https://melbourneinstitute.unimelb.edu.au/__data/assets/pdf_file/0010/2437426/HILDA-SR-med-res.pdf.

Winfield, A. F. (2016) *Written evidence submitted to the UK Parliamentary Select Committee on Science and Technology Inquiry on Robotics and Artificial Intelligence*. Available at: http://eprints.uwe.ac.uk/29428.

World Economic Forum; The Boston Consulting Group (2015) *New Vision for Education Unlocking the Potential of Technology*. Available at: http://www3.weforum.org/docs/WEFUSA_NewVisionforEducation_Report2015.pdf.

World Economic Forum (2016) *The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*. Available at: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf (Accessed: 18 September 2018).

World Economic Forum (2017) *The Global Risks Report 2017*. Available at: https://www.weforum.org/reports/the-global-risks-report-2017.

World Economic Forum (2018a) *Harnessing Artificial Intelligence for the Earth*. Available at: http://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf.

World Economic Forum (2018b) *How to Prevent Discriminatory Outcomes in Machine Learning*. Available at: http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf.

World Inequality Database (2018) *World Inequality Database - Australia*. Available at: https://wid.world/country/australia/.

World Wide Fund for Nature Australia (2017) *Can Technology Save the Planet? A Discussion Paper by WWF-Australia*. Available at: https://www.wwf.org.au/knowledge-centre/resource-library/resources/can-technology-save-the-planet#gs.EqaR4Gfg.

Wu, X. and Zhang, X. (2016) 'Automated Inference on Criminality using Face Images'.

Wu, Y. *et al.* (2016) 'Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation'. Available at: http://arxiv.org/abs/1609.08144.

Xu, V. X. and Xiao, B. (2018) 'Chinese authorities use facial recognition, public shaming to crack down on jaywalking, criminals', *ABC News*, 20 March. Available at: http://www.abc.net.au/news/2018-03-20/china-deploys-ai-cameras-to-tackle-jaywalkers-in-shenzhen/9567430.

Yeung, K. (2016) '"Hypernudge": Big Data as a mode of regulation by design', *Information, Communication & Society*, pp. 1–19.

Zeleznikow, J. (2000) 'Building Decision Support Systems in Discretionary Legal Domains', *International Review of Law, Computers & Technology*, 14(3), pp. 341–356. doi: 10.1080/713673368.

Zhang, X. (2018) *Cross-Border Data Transfers: CSL vs. GDPR*, *ReedSmith*. Available at: https://www.reedsmith.com/en/perspectives/2018/01/cross-border-data-transfer-csl-vs-gdpr.

Zhao, S. (2006) 'Humanoid social robots as a medium of communication', *New Media & Society*, 8(3), pp. 401–419.

Zou, J. and Schiebinger, L. (2018) 'AI can be sexist and racist — it's time to make it fair', *Nature*, 18 July. Available at: https://www.nature.com/articles/d41586-018-05707-8.

Zuboff, S. (1989) *In The Age Of The Smart Machine: The Future Of Work And Power*. New York: Basic Books.

Zuboff, S. (2015) 'Big other: surveillance capitalism and the prospects of an information civilization', *Journal of Information Technology*, 30, pp. 75–89.

WORKING GROUP

Professor Toby Walsh FAA (Co-chair)

Professor Walsh is a leading researcher in AI. He was named by *The Australian* as a rock star of Australia's digital revolution. He is Scientia Professor of Artificial Intelligence at UNSW, leads the Algorithmic Decision Theory group at Data61, Australia's Centre of Excellence for ICT Research, and is Guest Professor at TU Berlin. He has been elected a fellow of the Australian Academy of Science, and has won the prestigious Humboldt research award as well as the NSW Premier's Prize for Excellence in Engineering and ICT. He has previously held research positions in England, Scotland, France, Germany, Italy, Ireland and Sweden.

He regularly appears in the media talking about the impact of AI and robotics. He is passionate that limits are placed on AI to ensure the public good. In the last two years, he has appeared in TV and the radio on the ABC, BBC, Channel 7, Channel 9, Channel 10, CCTV, CNN, DW, NPR, RT, SBS, and VOA, as well as on numerous radio stations. He also writes frequently for print and online media. His work has appeared in *New Scientist*, *American Scientist*, *Le Scienze*, *Cosmos*, *The Conversation* and *The Best Writing in Mathematics*. His twitter account has been voted one of the top ten to follow to keep abreast of developments in AI. He often gives talks at public and trade events including CeBIT, the World Knowledge Forum, TEDx, and Writers Festivals in Melbourne, Sydney and elsewhere. He has played a leading role at the UN and elsewhere on the campaign to ban lethal autonomous weapons (aka 'killer robots').

Professor Neil Levy FAHA (Co-Chair)

Neil Levy is Professor of Philosophy at Macquarie University, as well as a Senior Research Fellow at the Uehiro Centre for Practical Ethics, University of Oxford. Before coming to Macquarie, he was Head of Neuroethics at the Florey Institute of Neuroscience and Mental Health.

He works, or has worked, in many different areas of philosophy, ranging from continental philosophy through to applied ethics and philosophy of mind. His work has a special focus on the implications of the sciences of mind for ethics and for human agency. He has published more than 200 articles and book chapters, as well as 7 books with major presses. As well as publishing in philosophy, his work has appeared in high-profile medical and cognitive-science journals. His most recent book is *Consciousness and Moral Responsibility* (Oxford University Press, 2014). In 2009, he was awarded the Australia Museum Eureka Award for Research in Ethics.

Professor Genevieve Bell FTSE

Professor Bell is the Director of the 3A Institute, Florence Violet McKenzie Chair and a Distinguished Professor at the Australian National University (ANU) as well as a Vice President and Senior Fellow at Intel Corporation. Professor Bell is a cultural anthropologist, technologist and futurist, best known for her work at the intersection of cultural practice and technology development.

Professor Bell joined the ANU's College of Engineering and Computer Science in February 2017, after having spent 18 years in Silicon Valley helping guide Intel's product development by developing the company's social science and design research capabilities.

Professor Bell now heads the newly established Autonomy, Agency and Assurance (3A) Institute, launched in September 2017 by the ANU in collaboration with CSIRO's Data61, in building a new applied science relating to the management of AI, data and technology and their impact on humanity.

Professor Bell is the inaugural appointee to the Florence Violet McKenzie Chair at the ANU, named in honour of Australia's first female electrical engineer, which promotes the inclusive use of technology in society. Professor Bell also presented the highly acclaimed ABC Boyer Lectures for 2017, in which she investigated what it means to be human, and Australian, in a digital world.

Professor Bell completed her PhD in cultural anthropology at Stanford University in 1998.

Professor Anthony Elliott FASSA

Professor Elliott is Dean of External Engagement at the University of South Australia, where he is Executive Director of the Hawke EU Centre and Research Professor of Sociology. Professor Elliott is also Global Professor of Sociology (Visiting) in the Graduate School of Human Relations, Keio University, Japan and Visiting Professor of Sociology at University College Dublin, Ireland.

Anthony Elliott was born in Australia and holds a BA Honours degree from the University of Melbourne and a PhD from Cambridge University, where he was supervised by Lord Anthony Giddens, architect of Third Way progressive politics. Professor Elliott was formerly Director of the Hawke Research Institute at UniSA (2012-2016), and Associate Deputy Vice-Chancellor (Research) and Head of the Department of Sociology at Flinders University (2006-2012).

Professor Elliott contributes to media worldwide: among others, he has recently been interviewed by the BBC World Service, *The Sunday Times*, ABC Radio National, *The Australian*, BBC Radio 4, GMTV Sunday, as well as European and North American radio and television networks.

Professor James Maclaurin

Professor Maclaurin is a member of the Department of Philosophy and Associate Dean for Research in the Humanities at the University of Otago. He received his Doctorate in Philosophy of Science from the Australian National University. A longstanding advocate for Humanities education, he was instrumental in the development of the University of Otago's Bachelor of Arts and Science degree. His research focuses on conceptual and ethical issues posed by scientific innovation as well as the process of distilling academic research into public policy in disciplines such as public health, economics, ecology, computer and information science.

He is co-director of the Centre for Artificial Intelligence and Public Policy and co-signatory to the University of Otago's memorandum of understanding on research into the social, ethical and legal effects of AI, with the New Zealand Government Department of Internal Affairs. He is a principal investigator on the Artificial Intelligence and Law in New Zealand Project which is funded under the New Zealand Law Foundation's Information Law and Policy Project. He is also a member of the Bioethics Panel for Predator Free New Zealand 2050.

Professor Iven Mareels FTSE

Since February 2018, Professor Mareels is the Lab Director, IBM Research Australia. He is an honorary Professor at the University of Melbourne. Prior to this he was the Dean of Engineering at the University of Melbourne (2007-2018).

He received the PhD in Systems Engineering from the Australian National University in 1987, and the Master of Engineering (Electromechanical) from Gent University in 1982.

At IBM Research Australia he is focused on developing the next generation of artificial intelligence remaining true to the motto "Famous for science and vital to IBM". The AI application domains he pursues are health and medical systems, financial services, and the Internet-of-Things. The main implementation modality is to build on and to exploit IBM's cloud infrastructure, and edge computing assets.

Iven is a Commander in the Order of the Crown of Belgium, a Fellow of The Australian Academy of Technology and Engineering; The Institute of Electrical and Electronics Engineers (USA), the International Federation of Automatic Control and Engineers Australia. He is a Foreign Member of the Royal Flemish Academy of Belgium for Science and the Arts.

Professor Fiona Wood AM FAHMS

Professor Wood has been a burns surgeon and researcher for the past 20 years and is Director of the Burns Service of Western Australia. She is a Consultant Plastic Surgeon at Fiona Stanley Hospital (previously at Royal Perth Hospital) and Princess Margaret Hospital for Children, co-founder of the first skin cell laboratory in WA, Winthrop Professor in the School of Surgery at The University of Western Australia, and co-founder of the Fiona Wood Foundation (formerly The McComb Foundation).

Professor Wood's greatest contribution and enduring legacy is her work pioneering the innovative 'spray-on skin' technique (Recell), which greatly reduces permanent scarring in burns victims. Professor Wood patented her method in 1993 and today the technique is used worldwide. In October 2002, Fiona was propelled into the media spotlight when the largest proportion of survivors from the 2002 Bali bombings arrived at Royal Perth Hospital. She led a team working to save 28 patients suffering from between 2 and 92 percent body burns, deadly infections and delayed shock.

Fiona was named a Member of the Order of Australia (AM) in 2003. In 2005, she won the Western Australia Citizen of the Year award for her contribution to Medicine in the field of burns research. That same year her contribution to burns care was recognised through Australia's highest accolade when she was named Australian of the Year for 2005.

PEER REVIEW PANEL

This report has been reviewed by an independent panel of experts. Members of this review panel were not asked to endorse the report's conclusions and findings. The Review Panel members acted in a personal, not organisational, capacity and were asked to declare any conflicts of interest. ACOLA gratefully acknowledges their contribution.

Professor Nikola Kasabov FRSNZ

Nikola K Kasabov is the Director of the Knowledge Engineering & Discovery Research Centre and Personal Chair of Knowledge Engineering in the School of Engineering, Computing and Mathematical Science at AUT. He is a Fellow of the Royal Society of New Zealand, Fellow of the Institute of Electrical and Electronic Engineers, and a distinguished visiting Fellow of the Royal Academy of Engineering, UK. He has published 600 works and has most recently invented the first neuromorphic spatio-temporal data machine called NeuCube. His main interests are in the areas of: computational intelligence, neuro-computing, bioinformatics, neuroinformatics, speech and image processing, novel methods for data mining and knowledge discovery.

Emeritus Professor Russel Lansbury AO FASSA

Russell Lansbury is Emeritus Professor of Work and Employment Relations at Sydney University Business School where he was Associate Dean, Research. He holds a PhD from the London School of Economics and has been a Senior Fulbright Scholar at Harvard and MIT. His early research was on the impact of computerisation in the airline industry. He has been a research associate in the International Motor Vehicle Project at MIT His most recent research is on the impact of autonomous mining on the workforce, skills and work organisation in Australia and Sweden. He recently served on the advisory board of a major EU research project on the 'intelligent mine of the future' and its social and technological implications. His publications include jointly authored and edited books which include 'After Lean Production: Changing Employment Practices in the Global Auto Industry' Cornell Uni Press and 'Working Futures' Federation Press.

Professor Huw Price FBA FAHA

Huw Price is Bertrand Russell Professor of Philosophy and a Fellow of Trinity College at the University of Cambridge. Before moving to Cambridge in 2011 he was ARC Federation Fellow and Challis Professor of Philosophy at the University of Sydney. In Cambridge he is Academic Director of the Leverhulme Centre for the Future of Intelligence, and co-founder of the Centre for the Study of Existential Risk. He is a Fellow of the British Academy and the Australian Academy of the Humanities, and on the Board of the new Ada Lovelace Institute, London.

His publications include Facts and the Function of Truth (Blackwell, 1988), Time's Arrow and Archimedes' Point (OUP, 1996), Naturalism Without Mirrors (OUP, 2011), Expressivism, Pragmatism and Representationalism (CUP, 2013), and a range of articles in journals such as Nature, Science, Philosophical Review, The Journal of Philosophy, Mind, and The British Journal for the Philosophy of Science. He is also co-editor of three collections published by Oxford University Press: Causation, Physics, and the Constitution of Reality (2007, co-edited with Richard Corry); Making a Difference (2017, co-edited with Helen Beebee and Chris Hitchcock); and The Practical Turn (2017, co-edited with Cheryl Misak).

ACKNOWLEDGEMENTS

ACOLA and the Expert Working Group offer their sincere gratitude to the many experts from Australia, New Zealand and further afield who have extensively contributed to this report by way of input papers. The contributions and expertise from these experts has helped shape and develop the final report. Further information of these contributions can be found in 'evidence gathering'.

We also gratefully acknowledge the expertise and contributions from our project stakeholders. In particular, we would like to acknowledge Dr Alan Finkel, Sarah Brown, Dr Adam Wright and Dr Kate Boston from the Office of the Chief Scientist; Edward Santow and the Australian Human Rights Commission; Adrian Turner and Data61. We also thank our peer reviewers for the time and effort they have provided in reviewing the report.

We would particularly like to thank the Australian Research Council (project number CS170100008), the Department of Industry, Innovation and Science, and the Department of Prime Minister and Cabinet for their financial support and contributions to the project.

Our thanks to the EWG who put a great deal of time, effort, and insight into coordinating the report's conceptualisation and production, and also to the ACOLA team, in particular Dr Lauren Palmer, Dr Angus Henderson, Dr Alexandra James, Dr Benjamin Nicoll, Michelle Steeper and Dr Marc Rands (Royal Society of Te Apārangi), who made significant contributions to supporting the EWG and managing the project.

EVIDENCE GATHERING

Workshops and meetings were held across Australia during this project. Many people have contributed their time and expertise to the project through written submissions, meetings with members of the Expert Working Group and participating in the workshops.

The views expressed in this report do not necessarily reflect the opinions of the people and organisations listed in the following sections.

Workshops

The ACOLA Artificial Intelligence Project held two workshops:

- Initial scoping workshop: held 22 September 2017 to discuss the scope of the Horizon Scanning project;
- Second scoping workshop: held in Sydney on 28 June 2018, with project advisors and the Expert Working Group; and
- Synthesis workshop: held in Melbourne on 24 August 2018, with project advisors and the Expert Working Group to synthesise the submissions received (below).

Stakeholders consulted at workshops

We thank the following stakeholders for their time and participation in the ACOLA Artificial Intelligence Project workshops:

Adam Wright

Adrian Turner

Alan Finkel

Ben Macklin

Dave Dawson

Edward Santow

Jessica Hartmann

Neil Williams

Rachael Frost

Sarah Brown

Susanne Busch

Written submissions

As part of the evidence-gathering to support the development of the report, a call for input was sent to experts in the field. The development of the report has been made possible through their generous contributions. ACOLA and the Expert Working Group would like to sincerely thank the following people.

Agriculture (Australia)
John Billingsley

Agriculture (Australia)
Salah Sukkarieh

Agriculture (New Zealand)
Mengjie Zhang

AI and Trade
Ziyang Fan and Susan Aaronson

Appeal Algorithmic Decisions
Anne Matthew, Michael Guihot and Nic Suzor

Arts and Culture
Thomas Birtchnell

Data Collection, Consent and Use
Lyria Bennet Moses and Amanda Lo

Data Integrity, Standards and Ethics
Data61

Data Storage and Security
Vanessa Teague and Chris Culnane

Defence, Security and Emergency Response
Adam Henschke

Defence, Security and Emergency Response
Seumas Miller

Defence, Security and Emergency Response
Reuben Steff and Joe Burton

Disability
Sean Murphy and Scott Hollier

Discrimination and Bias
James Maclaurin and John Zerilli

Economic and Social Inequality
Greg Marston and Juan Zhang

Economic and Social Inequality
Nik Dawson

Education and Training
Rose Luckin

Employment and the Workforce
Alexander Lynch of behalf of Google Australia

Employment and the Workforce
Ross Boyd

Employment and the workforce
Robert Holton

Energy
Sylvie Thiebaux

Environment
John Quiggin

Environment
Iven Mareels

Ethics, Bias and Statistical Models
Oisín Deery and Katherine Bailey

Fake News
Neil Levy

Finance
Mark Lawrence

FinTech
Koren O'Brien

FinTech
Mark Pickering and Dimitrios Salampasis

FinTech
Westpac Technology,

GDPR and Regulation
Nick Abrahams and Monique Azzopardi
on behalf of Norton Rose Fulbright

Geopolitics
Nicholas Davis and Jean-Marc Rickli

Government
3A Institute led by Robert Hanson

Global Governance
Andrea Renda

Health and Aged Care
Federico Girosi

Health and Aged Care
Bruce MacDonald, Elizabeth Broadbent and
Ho Seok Ahn

Human AI Relationship
Hussein Abbass

Human Autonomy in AI Systems
Rafael Calvo, Dorian Peters and Richard Ryan

Human Rights (Australia)
Australian Human Rights Commission

Human Rights (New Zealand)
Joy Liddicoat

Inclusive Design
Manisha Amin and Georgia Reid

Indigenous Data Sovereignty
Maggie Walter and Tahu Kukutai

Indigenous Peoples
Ellie Rennie

Information Privacy
Mark Burdon

Legal and Ethical Issues
Herbert Smith Freehills

Legal Services
Julian Webb, Jeannie Patterson,
Annabel Tresise and Tim Miller

Liability and Algorithmic Decisions
Gary Lea

Liability
Olivia Erdélyi and Gábor Erdélyi

Machine Learning
Robert Williamson

Machine Learning
Anton van den Hengel

Mining
Chris Goodes, Adrian Pearce and Peter Scales

Natural Language Processing
Tim Baldwin and Karin Verspoor

Privacy and Surveillance
Joy Liddicoat and Vanessa Blackwood

Psychological and Counselling Services
Mike Innes

Public Communications
Mark Alfano

Quantum Machine Learning
Lloyd Hollenberg

Regulation
Olivia Erdélyi

Re-identification of Anonymised Data
Ian Opperman

Robotics
Alberto Elfes, Elliot Duff, David Howard,
Fred Pauling, Navinda Kottege, Paulo Borges,
Nicolas Hudson

SMEs and Start-ups
Tiberio Caetano Andrew Stead

Training the Next Generation of AI Researchers
Mark Reynolds

Transformations of Identity
Anthony Elliot

Transformations of Identity
Eric Hsu and Louis Everuss

Transport and Mobility
David Bissell

Transport and Mobility
Malene Freudendal-Petersen and
Robert Martin

Transport and Mobility
Michael Cameron

Transport and Mobility
Sven Kesselring, Eriketti Servou, Dennis Zuev

Trust
Reeva Lederman

Trust and Accessibility
Mark Andrejevic

Universal Design
Jane Bringolf

Work Design
Sharon Parker

HORIZON SCANNING

ACOLA™

AUSTRALIAN COUNCIL OF LEARNED ACADEMIES