Australian Government
Productivity Commission

January 2024

Making the most of the AI opportunity

Research paper 3

AI raises the stakes for data policy

Contents

The Commission acknowledges and thanks the following Commissioners and staff who have worked on the Making the most of the AI opportunity research papers: Stephen King, Rosalyn Bell, Hudan Nuch, Rebecca Stoeckel, Jeremy Kamil, Rachel Burgess, and Ritaja Das.

The Productivity Commission acknowledges the Traditional Owners of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to their Cultures, Country and Elders past and present.

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

An appropriate reference for this publication is:
Productivity Commission 2024, *Making the most of the AI opportunity: AI raises the stakes for data policy*, Research paper, no. 3, Canberra

Publication enquiries:
Phone 03 9653 2244 | Email publications@pc.gov.au

AI raises the stakes for data policy

Key points

- **Developments in artificial intelligence (AI) technologies are raising the stakes for Australia's data policy and creating new challenges.**
  - The ability for AI to collect data, make inferences, automate decisions and interact with individuals has increased the possibilities to use data to either the benefit or detriment of individuals and businesses.
  - As AI creates new capabilities, it increases the value of data access. Productivity gains will, in part, depend on whether the quality of Australian data holdings can be raised (particularly ongoing curation of datasets and interoperability) and whether access can be broadened in ways that enable data use in AI for the public benefit without undermining incentives of data holders or increasing risks to individuals.
  - Given that Australia can import AI technologies, the scope for AI-induced productivity is not solely dependent on Australian data use.
- **AI has heightened some risks in uses of personal data. Better design and enforcement of regulation to avoid harms to individuals could increase public acceptance of AI and encourage greater data-sharing.**
  - This could include greater clarity on the applicability and enforcement of existing laws in privacy, cyber and data protection; anti-discrimination, competition and consumer protection; and criminal activity.
- **Establishing clear and consistent arrangements to allow text and data mining (TDM) for the purposes of training AI models could be a major boost to AI development within Australia.**
  - The treatment of TDM, copyright and the training of AI models is currently being tested in US courts. In other major economies, exceptions from copyright are used to clarify arrangements for TDM. Australia has scope to learn from international experiences of copyright arrangements to improve accessibility of data for commercial and non-commercial uses, while protecting incentives in creative industries.
- **Development by Australian governments of a comprehensive national data strategy would help provide confidence for businesses and the community in the use of AI and other data-intensive technologies.**
  - Australia is one of the poorest performing countries in the OECD in terms of data availability, data accessibility, and government support for data re-use. A national strategy could set out agreed intentions for access and use of *all* data collected in Australia, clarifying rights of data subjects, curators, and users – thereby providing a secure basis for AI use and development.
- **Governments could play an important role as exemplars of data governance, use and sharing.**
  - The maintenance and curation of government-held data could be improved to increase the usefulness of data for AI, with sharing expanded to some private parties, subject to sufficient safeguards.

In the past 30 years, data has played an increasingly central role in both the economy and society. Stemming from large scale digitisation, connectivity, smart devices and computing infrastructure, data has become a key input into production and innovation. And while data use can generate significant value across the economy, doing so inevitably raises questions about who owns, controls and should benefit from data.

Artificial intelligence (AI) is just the latest wave of innovations to transform how data can be collected and used to generate public value. The AI models that have emerged in recent years apply advanced machine learning to increasingly sophisticated uses, including natural language processing, image recognition, recommender systems, personalised search and social media.[1] But is AI changing how we think about data from a policy perspective?

If nothing else, AI technology **raises the stakes for data policy**.

First, AI makes data more valuable. Data is a key input to the development and use of AI technologies, as the capabilities of AI models depend in part on the scale and quality of data accessible for training. In addition, AI opens up a range of possibilities for the use of data which has previously sat dormant (including unstructured text, audio and visual data). As the OECD observed, 'data that is not valuable today may become so tomorrow' (Jouanjean et al. 2020).

Second, AI is heightening data-related risks, including through increased capacity to misuse (or even weaponise) data. This has heightened concerns about privacy, cyber security and data protection (figure 1). Data that is not *risky* today may become so tomorrow.

The relationship between data and AI has also raised **new questions about data rights and incentives**.

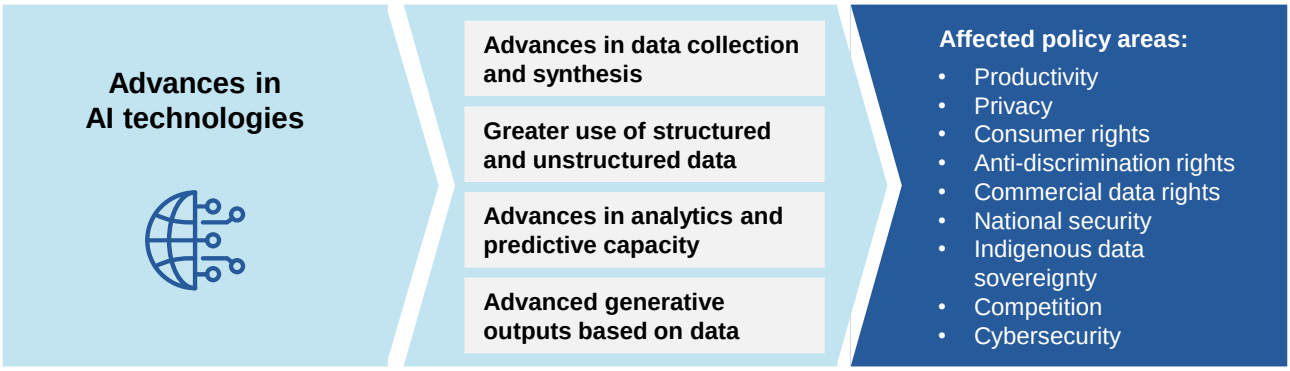
AI is changing the nature of data collection and use. Developments such as facial recognition and sensor technology have increased the potential for more intrusive collection of personal information, bringing renewed attention to issues of data rights (such as collection, control and consent). The advent of generative AI has raised questions about the intellectual property of both data inputs and creative outputs. These and related issues have been covered by a number of recent reports, focusing separately on technology, ethics, privacy, economic rights, and productivity (Attorney-General's Department 2023a; DISR 2022, 2023; IP Australia 2023; PC 2022).

So how can governments make the most of Australia's data opportunities?

The Australian Government has recently released a comprehensive strategy focusing on government use of data at the Commonwealth level (Australian Government 2023d) alongside ongoing reform agendas related to copyright, privacy, and consumer protections (Australian Government 2024). In implementing these reform agendas, some key challenges will include: aligning productivity and regulatory objectives; ensuring regulatory controls are proportionate and effective; and achieving consistency and coordination across governments. A good start would be for governments to develop a comprehensive national data strategy that sets out agreed intentions for access, maintenance and use of *all* data collected in Australia, clarifying rights of individuals and data holders and providing a secure basis for the development and use of AI and other data-intensive technologies.

[1] AI for the purpose of this paper should be distinguished from artificial *general* intelligence, which has not yet been developed and is outside the scope of this paper that considers uptake of existing AI technologies. Machine learning is a subfield of AI that describes algorithms which are capable of completing a task with minimal human instruction. Often these algorithms 'learn' to accomplish the task by being trained on example data.

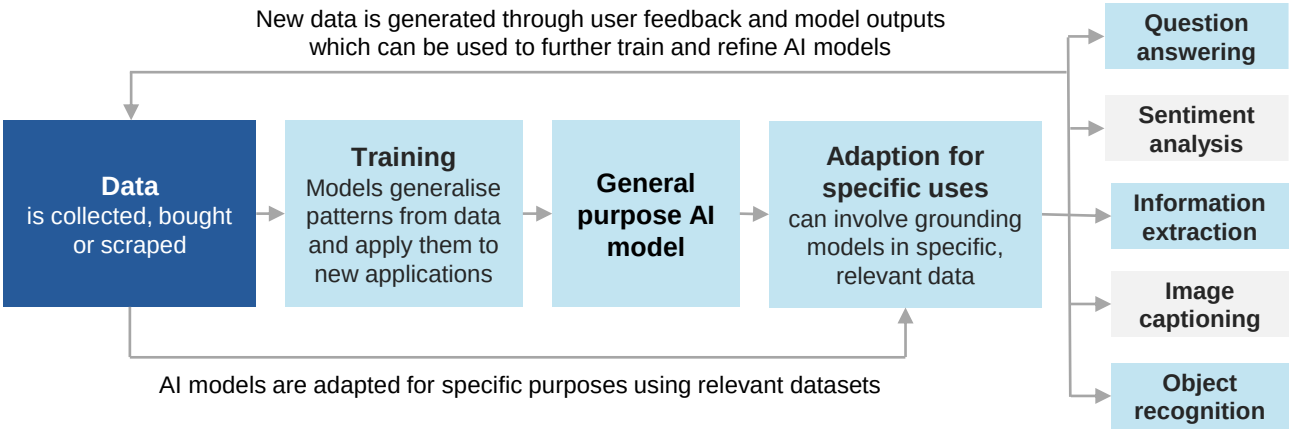
Figure 1 – A renewed focus on data will raise questions across many policy areas



1. Data is vital to AI

Data is one of the key inputs into AI technologies. Data is used to train models, adapt or customise models for narrower applications, or as part of the use of AI (i.e. models designed to analyse data). And user data, generated through interactions with AI models, can be used for further refinement and training (Zia 2023) (figure 2).

Figure 2 – Data is both an input and an output of AI which is essential for training



Source: Adapted from Merritt (2022) (non-exhaustive list of data sources and uses).

Both the volume and quality of data used to train AI models can significantly affect the usefulness of the model. Vast amounts of data have been used to train leading AI models: ChatGPT for example, is estimated to have been trained using a dataset of 100 trillion parameters (Walsh 2023); and Tesla's self-driving cars have been trained using 1.5 petabytes[2] of data, consisting of one million 10-second videos and six billion objects (Dickson 2021). Data quality is vital to improving the accuracy and reliability of AI models – many chatbots have integrated feedback loops to exploit user data to filter out poor quality outputs (Loft 2022).

In some cases, use of small, high-quality datasets may be sufficient and sometimes even preferred (given the costs associated with collecting, curating and processing data). More broadly, smaller datasets may be useful in tailoring AI models to narrower, more specific needs (such as recognising rare health conditions (Toner and Chahal 2021)).

[2] One petabyte is equal to 1,024 terabytes or over one quadrillion bytes.

Why data access arrangements are critical for AI

In the context of technological progress, arrangements to facilitate safe and ethical data access can contribute to economy-wide productivity. Greater data access can facilitate innovation by feeding into the production of new products and services. Data also enables more informed and optimised business decision making. This benefits both industry and the broader community through a lift in living standards. AI technologies create greater opportunity for data to deliver these benefits.

Improving safe and ethical data access means thinking about what rights should apply to data subjects, data holders, and data users. As the same data can be used for multiple purposes at once, the value to society of data increases with the number of users.[3] This puts data in a unique position to generate substantial productivity gains. But data holders can exclude people from access to data. It is this excludability that is the primary reason why data is often underused.

The challenge for capturing the productive potential of data lies in striking a balance between the rights of different parties (including those involved in the creation of data, those holding or curating the data, those using the data)[4] as well as potential benefits to the broader public (in the form of improved products and services and economy-wide productivity). Regulatory arrangements should be designed to optimise any tradeoffs and align the objectives of multiple parties. Poorly designed regulation can make data access more costly, and a lack of regulatory clarity can lead to uncertainty that undermines the valuable use of data.

Creating the right settings for data sharing will help motivate investments that improve the quality of data collections and allow AI technologies to spread throughout our economy. Some uses may result in heightened risks for **personal data** that must be addressed (section 2). And as the nature of data collection and use evolves, **commercial terms of data-sharing**, particularly about intellectual property (section 3) will become increasingly important. In resolving these issues, governments can be **valuable role models** for data collection, management, and sharing.

Does AI need Australian data?

The role of Australian data in AI development and use should not be overstated – without (greater) access to Australian data, AI models will continue to be developed overseas, and to deliver productivity benefits to Australia.

But to the extent that demand for AI technologies increases demands for accessible data, AI could be a catalyst for Australia to get more value from its data holdings. For many applications, the use of Australian data would increase the relevance and grounding of AI applications to our needs. It may also be the case that Australian data is more useful than we currently understand it to be in the development of productive and useful AI models.[5] More broadly, access to Australian data could improve the **quality of data** available to train AI models. For some applications, such as medical diagnosis or autonomous vehicles, the quality of data used to train AI models can be more important to the accuracy of the model than the quantity of data (Gomede 2023).

[3] The private value of a specific data set to a business may fall if rival businesses access the same data, even though the social value increases, leading private data holders to exclude access to the data they hold.

[4] For example, a patient and a diagnostic laboratory may jointly create data through medical testing requested by the patient's GP and paid for by the government through Medicare; the data may be simultaneously held by the patient's GP, the laboratory and potentially other parties such as a hospital or third-party app provider; and users may include other medical professionals, government or researchers.

[5] One argument for making data more available is that opportunities to use it are largely unknown until the data sources themselves are better known, and until data users have been able to undertake discovery of data (PC 2017, p. 2).

Businesses may be able to capture much of the productive potential of AI technologies using their own data. Data is already an input for many businesses although Australia ranks relatively poorly in the OECD on business use of data analytics (PC 2022, p. 20) – the addition of AI technologies would expand the potential for productive use of data, generating more and higher-quality insights.

2. Sharing personal data in the AI era

AI diffusion will escalate some risks to individuals and society. These risks stem from both AI-specific and complementary technological developments including technical capabilities in collecting data (including by inference, re-identifying previously de-identified data) and using data (including making invasive predictions about individuals and discrimination). Two risks typically raised relate to concerns about **privacy,** and more generally, the **use or misuse of personal information** (to the detriment of an individual or society).

Failing to manage the risks posed by greater AI uptake would erode public trust and confidence (with long-run consequences for data availability and quality if individuals elect to withhold or distort their data), result in material harms to society and individuals, stall further technological progress, and reduce the productivity benefits from AI.

Privacy and consent

The right to privacy in Australia is established by the International Covenant on Civil and Political Rights and is regulated at the national level through the *Privacy Act 1988* (Cth), as well as through a range of legislation within most states and territories. Privacy relates to the protection of personal information[6] with regard to its collection, use, storage and disclosure (Attorney-General's Department nd).

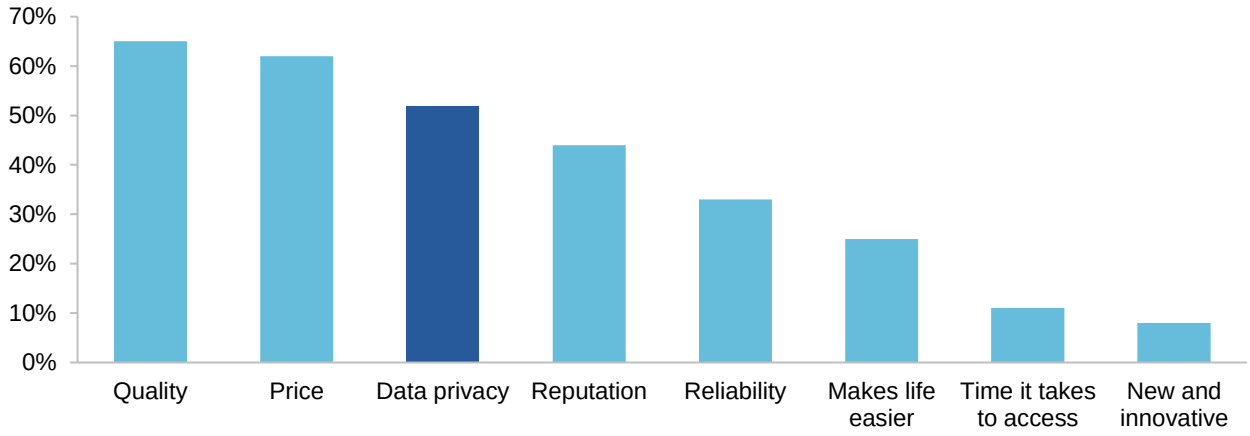
While the right to privacy is sometimes considered inalienable, there are many circumstances where individuals choose to forgo their privacy by sharing personal data in return for goods, services, or for public benefit. For example, Facebook collects significant data from its users, and still about 11.6 million Australians used, and provided some personal information to Facebook in 2022 (Hughes 2023). This willingness is also stated for data use in the general public interest. A recent survey revealed that over 75% of Australians would be willing to share personal health information to advance medical research (Research Australia 2021, p. 11).

As AI develops, there are likely to be direct trade-offs made between privacy and product or service quality. Such trade-offs are not straightforward: survey evidence suggests that when choosing products and services, Australians place a high priority on privacy – albeit ranking it behind quality and price (figure 3).

[6] 'Personal information' is defined under the Act as 'information or an opinion about an identified individual, or an individual who is reasonably identifiable' (Privacy Act 1988 Part II Division 1).

Figure 3 – Australians rank data privacy highly, but behind quality and price

Share of respondents ranking top 3 for importance when choosing a product or service



Source: OAIC (2023, p. 28).

Trade-offs also occur between the indivdual's privacy and the broader public benefit – for example, if individuals withheld their location data from online map services to protect their personal privacy, the macro-level impact would be inaccurate traffic information and poor mapping services (that is, a lower quality product). Similar trade-offs could occur with regard to personal information and the quality of medical and other research.[7]

Consent frameworks can play a valuable role in allowing individuals to have some degree of control over their privacy, and in establishing social licence for data use.[8] However, the threshold for consent is low and there are questions about its effectiveness – particularly where consumers have few alternatives to accessing services.

AI is challenging consent frameworks by incentivising riskier data collection practices and increasing the use of higher risk data. Web scraping is one way that AI developers have been able to access the vast amounts of both structured data and unstructured data needed to train generative models however, scraping sensitive data without consent has landed at least one developer in legal trouble (OAIC 2021).

Advances in AI-enabled facial recognition technology (FRT) also incentivise the collection of biometric data which poses a higher risk to privacy and human rights.[9] Collection of sensitive information is allowed where it is reasonably necessary for one or more of an entity's functions or activities. And consent can be as simple as remaining in a space where FRT is being used and has been signposted. For example, those who attend

[7] The use of detailed health data could improve health treatments (by avoiding misdiagnosis or wrongful medication) but could lead to discrimination or a loss of privacy in general. Allowing individuals to exclude their data from use in population-level medical modelling could result in biased models (and subsequent harm).

[8] Consent is currently only required under the *Privacy Act 1988* (Cth) for a limited range of collections, uses and disclosures of personal information, including sensitive information, and can be express or implied (for example, consent can include remaining in a location which uses video surveillance).

[9] The advent of FRT will likely make activities or functions requiring the use of sensitive biometric information more frequent and expand the scope of such activites. For example, FRT can analyse consumer sentiment and inform business decisions about what products to stock and where to display them. Wide use of FRT could see our biometric information being collected across public facing industries including the retail and hospitality sectors.

a sporting event at one of the many Australian sports grounds using FRT may unknowingly be trading their biometric information for entry to the event.

For consent to be meaningful and informed, individuals must be given legitimate alternatives that still achieve the intended objectives (such as fraud control or maintaining public safety). For example, providing spaces within stadiums free from the use of FRT would allow attendees to make meaningful choices about their willingness to have their biometric information collected or scanned. This will not always be practicable, and as such, other areas of regulation and enforcement will play an important role.

There is also a question as to what it means to provide consent at a given point in time, given that data access may have ongoing – and evolving – implications. Risks and potential benefits for data subjects may change over time, and full information may not be available at the time that consent decisions are made. As such, policymakers may need to consider both the provision and retraction of consent – both of which will have practical implications for the training of AI models (discussed below).

Consent frameworks alone are not sufficient to deliver broader public benefits of data use. Over time, regulation will need to reflect social norms about privacy and consent in the digital world. Some degree of digital 'visibility' may be unavoidable, just as in the physical world. Governments will need to consider to what extent privacy concerns about technologies such as FRT (and AI more broadly) also stem from the potential *misuse* of sensitive information (discussed below).

The potential misuse of personal information

The expanding capabilities of AI technologies and increasing AI uptake could increase the range and scale of potential harms (box 1). The risks of misuse are a threat to genuine consent – it is questionable whether individuals would knowingly accept the broad benefits of data-sharing (such as enhanced services) if it resulted in individual vulnerability to serious risks (such as cybercrime or discrimination). At a policy level, it would be difficult to argue that personal data should be more freely shared if doing so significantly increased the risk of scams, identity theft, or other crime.

Risks that arise from the use of personal and sensititve information[10] are not only regulated through privacy law, but also consumer law and anti-discrimination laws (among others). This provides a lot of avenues for government action to make AI safer, bolster data protections, and improve enforcement of existing regulations to build trust in data sharing and use. It also provides significant scope for inconsistencies in requirements and mounting cumulative burdens on those seeking to use data for AI.

It is important to consider how regulation can help build the trustworthiness of data sharing arrangements, rather than simply reducing the extent of data use in the economy. That is, where the public is confident that relevant regulation offers adequate protection for their data, there may be a greater willingness to share data.[11] The risk of misuse of personal information is not distributed evenly across society, and regulations should be designed and administered to address both the needs of the most vulnerable and the broader public benefit.

[10] 'Sensitive information' is defined in the *Privacy Act 1988* (Cth) as information or an opinion about an individual's racial or ethnic origin, political opinions, membership of a political association, religious beliefs or affilitations, philosophical beliefs, membership of a trade union, sexual orientation or practices or criminal record that is also personal information, health information, genetic information, biometric templates or biometric information that is to be used for the purpose of automated biometric verification or biometric identification.

[11] Iceland has used strong data protections to create a data haven and attract investment in information-based technologies like AI (Gaedtke 2014).

Box 1 – Data risks and AI

There are a range of risks that exist around data, regardless of AI. However, some of these risks may be increased through the development of AI without well-designed rules around data.

Consumer profiling and manipulation

Digital consumer manipulation has been described as a firm's ability to 'use data and behavioural research to exploit the biases, emotions and vulnerabilities of consumers' (Manwaring 2018, p. 145). While consumer manipulation is nothing new, the potential for and effectiveness of such manipulation grows with the scale of data collection, the power of AI to make inferences, and the types of interactions that are possible (e.g. chatbots). Large language models can provide 'false but authoritative-sounding statements' to consumers when making purchaising decisions (DP-REG 2023, p. 4). The widescale diffusion of AI could make these issues more pervasive.

Bias and discrimination

Technology has heightened the potential for biases to enter commercial decisions. Several studies have found that the racial make-up of the AI training dataset has a bearing on how race is treated by models (Angileri et al. 2019; Gebru 2020; Sham et al. 2022). In Australia, concerns have been raised that Aboriginal and Torres Strait Islander people being overrepresented in some administrative data would result in overrepresentation in data analysis through new technologies (Walter et al. 2020).

Scope for inaccurate predictions for a given individual may also be heightened by making generalised inferences from large datasets of personal information. From a business perspective, such inaccuracies may be irrelevant at an aggregate level. Some evidence suggests that discriminatory pricing based on personal information is likely to be profitable at scale – one estimate showed that Netflix could (hypothetically) improve its profits by 0.14% by setting individualised prices based on personal characteristics (Shiller 2014).

Discrimination may create consumer 'winners and losers'. For example, improved data on individual driving behaviour could lead to better drivers paying lower premiums for car insurance, while worse drivers pay more. Such discrimination may be socially undesirable and banned, as is the case for private health insurance in Australia. It may also undermine the viability of insurance markets by limiting the ability to pool risk.

Poor data and digital literacy

Some misuse or mismanagement of data is accidental or negligent. While firms have both a commercial and reputational incentive to safely handle and use data, not all firms will be prepared to do so. Some may be unprepared for increasing cyber security risks, or unaware of the liabilities they hold for personal data in their possession. They may be ill-informed on the potential for bias to result from flawed algorithms, poor data or poor inferences.

Given the rapid development of AI, the keenness to adopt sophisticated technologies by business could outpace levels of both AI literacy and data competence. The increasing use of AI to make inferences from personal data to customise products and services will mean that any data mismanagement would have increasingly tangible impacts on people's lives.

Safer AI could open up more valuable uses of data

A challenge for the regulation of data use is that more accessible data and greater capabilities in using data both contribute to productivity, and yet:

- wide use of data, particularly with inadequate regulation, could create risks to data holders and individuals
- risks to data holders and individuals could create incentives to reduce data-sharing, with negative implications for productivity.

It is important to consider that a range of existing laws already put boundaries on how personal data can be used (Solomon and Davis 2023). Amended or new regulation will only be necessary where existing laws are inadequate to meet the risks associatd with AI (Paper 2, *The challenges of regulating AI*). Any new limits should be considered in terms of their effectiveness in preventing harm, including the likelihood and seriousness of that harm, and the extent of forgone benefits.

Some safeguards will be integrated into AI development

As a small economy with relatively few developers of AI, Australia should not rush to enact regulation that requires idiosyncratic Australian standards and principles in the design of AI models. Such as step would impose a cost on AI developers and users that may prevent the spread of AI technologies and complementary innovations to Australia (Paper 2, *The challenges of regulating AI*).

All major economies are developing regulatory approaches to AI, many of which will affect the quality and safety of AI models imported to Australia. Europe for example, has adopted the 'privacy-by-design' approach in their General Data Protection Regulation (GDPR), as has the United States in exploring ways to step-up their privacy requirements in the context of AI (Lima and Zakrzewski 2023). Overseas regulation which integrates safeguards into AI models will have spillover effects for Australia. For Australian firms and individuals to make informed choices in their use of AI, AI applications will need to be transparent about which foundation models they are based on and which international standards have been met. Commercial incentives are likely to encourage this kind of transparency, as developers have incentives to advertise both the power and trustworthiness of their applications.

Explainable AI and effective enforcement to encourage more data use for AI

Concerns about personal data use in AI are heightened by the extent to which AI technologies present a 'black box' of decision-making. While a degree of opacity is likely to continue, some regulatory approaches are now promoting the development of 'explainable AI', which can 'provide explanations for their decisions or predictions to human users' (Saranya and Subhashini 2023).

In Europe, the GDPR includes a right to 'meaningful information about the logic involved' when an individual is significantly affected by automated decision-making.[12] The Australian Government has agreed in-principle that individuals have a right to an explanation of what an organisation or government agency has done with their personal information (Australian Government 2023b, p. 30).[13] How detailed this right would be or whether it would explain how particular personal data is treated within an algorithm, is unclear. It remains to be seen how feasible implementation of these rights will be as technology develops.

Increased clarity around AI use in decision-making will shape the role that regulators play in enforcement. For example, current enforcement of anti-discrimination laws relies heavily on complaints from victims –

[12] General Data Protection and Regulation 2016, articles 13-15.

[13] The Privacy Act Review Report recommended a right to explanation (proposal 18.1), although it would differ to that of the GDPR (Attorney-General's Department 2023a, p. 11).

identifying AI-enabled discrimination may only be feasible if businesses are required to explain their use of algorithms. Regulators such as the Australian Human Rights Commission could potentially be equipped to play a more proactive role in identifying illegal discrimination, either by interrogating how the AI models function, or by testing the outcomes of AI use.

There could be value in bolstering the capabilities (and potentially, powers) of some regulators, to allow more proactive assessment of AI models in addition to existing complaints-based processes. In addition, ongoing investments in the technical capabilities will be needed for a range of regulators. Proactive collaboration between regulatory bodies (as well as with industry stakeholders) will be a useful first step in building capabilities and understanding. The ongoing work of the Digital Platform Regulators – including the eSafety Commissioner in preventing serious misuse of personal data, and the ACCC in preventing online scams – provides a valuable example (DP-REG 2023). The Office of the Australian Information Commissioner (OAIC; the national regulator for privacy and freedom of information), will be crucial throughout the early phase of AI diffusion in Australia – not just for safeguarding protections in the use of personal and sensitive data – but for instilling confidence in frameworks that enable greater data use.

Data protections to address residual risks

While AI could be made safer and more explainable, efforts to clarify or amend existing regulation, or to boost enforcement capabilities may take time. There is a risk that governments, under pressure to act quickly, introduce unduly restrictive data policies in areas where other measures may be more appropriate.

As AI technologies continue to develop, gaps in regulatory safeguards around data collection and use may become apparent. In some ways, AI poses a fundamental challenge to how privacy and data protection laws can work in practice. For example, the way that AI uses data is more akin to reading data than storing it – as such, the right to erasure (which was proposed by the Privacy Act Review Report and agreed in-principle by government (Australian Government 2023b, p. 31)) could be subject to practical limitations.[14]

More difficult questions relate to how the use of biometric data should be regulated – particularly in the context of FRT, limited consent, the capacity for mass surveillance, and the inaccurancy of some current technologies.

3. Commercial terms for data-sharing

The advance of AI technologies over time has established new means to collect, share, and use data in productive ways. The productive and commercial value of data has increased – particularly forms of data that were previously underutilised. This has focused renewed attention on who benefits from data – and how rights and benefits might be afforded to data creators, subjects, holders, users, and the broader public.

Copyright, text and data mining, and AI training data

Uncertain or restrictive copyright arrangements could limit (or make more costly) data access for the training of AI models. At the same time, concerns have been raised as to whether the training of AI models on

[14] The practical feasibility of erasing personal data from AI models will evolve along with both model design and data techniques. Methods for 'approximate deletion' of personal data have been developed for some specifications of AI models (Izzo et al. 2021). The more complex the model is, the more challenging it becomes to delete data (Myers 2021).

copyrighted materials[15] is a violation of existing copyright law and broader conceptions of fairness, and whether it creates poor incentives for creative industries.

Major economies are still determining to what extent copyrighted materials can and should be used for the purposes of training AI. While the Bletchley Declaration established agreement between major economies on addressing many of the data-related risks that AI poses, it made no mention of intellectual property.

In the US, where several general purpose AI models have been developed and trained, text and data mining is subject to 'fair use' restrictions that permit limited use of copyright protected material if the use is deemed 'transformative'. Case law has yet to establish how 'fair use' would apply to the training of AI models, although a number of law suits have been instigated to challenge the 'fair use' of copyright materials in training AI models. For example, a suit filed by the New York Times against Microsoft and Open AI alleges that such training has gone beyond 'fair use', in part because the outputs of the models in question are not always transformative – and when prompted, the models have been shown to reproduce verbatim significant excerpts of text that would usually be subject to a paywall (2023).

In several other jurisdictions, text and data mining (TDM) exceptions have been implemented to allow the use of copyrighted materials in training AI. However, approaches vary significantly. While some exceptions apply only to non-commercial or other limited uses, others include commercial uses and/or attempt to facilitate the use of data to train AI models. Japan has made a deliberate effort to ensure AI development is unfettered by copyright (European Alliance for Research Excellence 2018). The EU has provided an exception for scientific use, and a broader exception that is subject to opt-out by copyright holders (Margoni and Kretschmer 2022). And in the United Kingdom, TDM exceptions remain limited to non-commercial use, as plans to expand exceptions to commercial use in 2023 were abandoned (Montagnon and Cho 2023).

In Australia, the *Copyright Act 1968* (Cth) does not contain an infringement exception for TDM. Without such an exception, it is unclear the extent to which copyright materials can be used to train AI models. It may be argued that infringements have occurred where the AI training process has involved the copying, digitising or reformatting of copyright material without permission (ALRC 2013b). There may be scope for some text mining to be covered by 'fair dealing exceptions', although these would be limited to 'reasonable portions' of the copyrighted work. The issue was considered by the Australian Law Reform Commission report on *Copyright and the Digital Economy* a decade ago.

> … The reach of the fair dealing exceptions may not extend to text mining if the whole dataset needs to be copied and converted into a suitable format. Such copying would be more than a 'reasonable portion' of the work concerned. Nor is it clear whether copying for text mining would fall under the exception relating to temporary reproduction of works as part of a technical process, under s 43B of the Copyright Act, but it seems unlikely. (ALRC 2013a, para. 8.47)

Others have argued that Australian courts would likely follow US courts in excepting TDM from copyright, at least for non-commercial purposes, suggesting that:

> In light of the limited Australian guidance on fair dealing and TDM, predicting how Australian courts will treat TDM under the fair dealing exception is speculative. However, if TDM is utilised for non-profit research purposes, it will likely be considered a fair dealing under the Act. (Ford 2020, p. 46)

[15] Copyright protects an original form or way an idea or information is expressed (such as through writing, visual images, music or moving images), not the idea, data or information itself. However, copyright could apply to a database under Australian law if, for example, it is a literary work; expressed in material form; is original; and is connected with Australia (Fitzgerald and Dwyer 2012).

In the absence of a clear legal framework, publishers will likely apply different requirements with regard to text mining. For example, Reddit announced plans to begin charging users to access the site for text mining for commercial purposes (Isaac 2023) and Elsevier (an academic publishing company) has integrated TDM rights into their subscription agreement (SPARC nd). Typically, AI developers using copyrighted material as training data would need to obtain legal access to the data and to verify that the rights have not been reserved to make reproductions for TDM purposes and keep the copies made only as long as necessary for TDM purposes (DLA Piper 2023). In practice, this is likely to add limitations, regulatory burden, and some uncertainty about the use of Australian data for AI training (in addition to other commercial and non-commercial uses). Given the scale of data needed to train AI models, the task of identifying ownership, rights and remuneration for all data used in training would be substantial – potentially requiring new forms of royalty collection organisations, or new approaches to be taken by existing collection societies.

While industry stakeholders have suggested that Australia revisit the issue (Google 2023), recent consultations by the Australian Government showed that data users (e.g. technology companies and researchers) tended to be more in favour of reform than copyright holders (Attorney-General's Department 2023b). It is worth noting that aside from AI-related concerns, there are broader concerns about how well existing copyright laws are enforced within the digital economy more broadly (for instance, where artworks have been subject to unauthorised reproduction in online marketplaces).[16]

Trade-offs between copyright protection and AI development

Restrictive copyright arrangements can present a significant barrier to AI development. In addition, unclear or varied conditions for data access can increase the search costs associated with data, and the transaction costs of undertaking research (and other data-intensive activities). Some have argued the cost of acquiring data, as well as other limitations to access, could 'favour the development of biased AI systems', noting that:

> … it may be economically attractive for EU-based developers to train their algorithms on older, less accurate, biased data, or import AI models already trained abroad on unverifiable data. (Margoni and Kretschmer 2022, p. 700)

While more accessible data would help the development of AI models, concerns have been raised about ethics and fairness (Taylor 2023), violation of moral rights (Matulionyte 2023),[17] and the inclusion in training datasets of artworks that have intangible personal and cultural value. For instance, generative AI could use Indigenous art as training data and subsequently produce culturally unsafe outputs (Carlson and Richards 2023).

Others have argued that such arrangements pose threats to creative industries (DACS 2023). If the potential payoffs to creative work are reduced, so too are the incentives to undertake that work. In this sense, some TDM exceptions are more straightforward than others – non-commercial use (such as research) or commercial uses in non-competing services would appear to pose little threat to the copyright holder.

[16] Beyond AI-related copyright infringement concerns, creative industries face other challenges in the digital economy. For example, some Australian artists have alleged that copyright for their designs and artworks were infringed by sellers on online marketplaces (Carroll and Healey 2023; Keating 2023). Such infringements do not necessarily require any use of AI, so are in some sense distinct from the issue of using copyright materials to train AI models. Notwithstanding, if generative AI models are designed to effectively reproduce copyright works, this could create copyright infringement at scale.

[17] Moral rights are defined by the *Copyright Act 1968* (Cth) as including the right of attribution, the right against false attribution, and the right to integrity. The latter includes the right to ensure that a work is not subjected to derogatory treatment that is harmful to the author's honour or reputation.

The implications of TDM exceptions are more complicated for generative AI, given its ability to produce outputs that may compete with creators whose works were used in training.[18] This has been addressed through several different approaches overseas: while Japanese regulation relies on the concept of 'non-enjoyment' use, the EU approach allows for an opt-out mechanism for copyright holders.

In many cases, the *training* of AI models is not an inherent threat to existing revenue streams for artists and authors, in contrast to the advent of music streaming that supplanted album sales. However, it is difficult to generalise across all generative AI, given that models will differ in the extent to which outputs reproduce or derive particular works or artists. In most jurisdictions internationally, questions remain as to how to balance these issues. Even where TDM exceptions have been established, it is not always clear how courts will apply the law with regard to commercial generative AI (International Copyright Issues and Artificial Intelligence 2023).

Australia should reconsider TDM arrangements

It is beyond the scope of this paper to investigate the various possibilities to clarify the legal framework for the use of copyrighted material. It is clear however, that Australia has historically lagged behind major economies in allowing for TDM exceptions from copyright for research purposes, and has fallen further behind as jurisdictions update their regulations to meet the opportunities of AI.[19] This is not to say that the most liberal regimes are necessarily applicable to Australia. Nor does it suggest that exceptions to copyright are the only available mechanism to improve arrangements for data use. Rather, there is scope to learn from international experiences to establish TDM arrangements that better facilitate commercial and non-commercial uses of data.

At the same time, international TDM arrangements will have direct spillover effects for Australian creative industries. AI models imported from overseas will continue to be used in Australia, many of which will have been trained on copyrighted materials through TDM exceptions. If generative AI is a threat to creative industries (as some have argued), then this will likely be the case regardless of whether Australian data are used in training.

Establishing clear and consistent arrangements that facilitate text and data mining for the purposes of training AI models would be a major boost to AI development within Australia. There could be significant benefits from greater use of data for non-commercial and/or commercial purposes. At the same time, any consideration of new exceptions to copyright should consider how they would affect the incentives to further creativity and innovation in the economy.

Challenges to copyright emerging from AI are slated to be pursued in 2024 through the establishment of a copyright and artificial intelligence reference group (Dreyfus 2023). It will be important for the Australian Government to consider how copyright should be treated with regard to training data (inputs) into AI models, as distinct from how copyright law can be better enforced with regard to any produced works (outputs).

New potential avenues for commercial data sharing

An increase in sharing of private-sector data would be a significant move forward for increasing data analytics in Australia, as a first step toward greater use of AI. Much private sector data is co-created by

[18] Concern about technologies competing with human creative outputs is not new. In 1982 the UK Musicians Union passed a resolution to ban synthesisers over concerns they would replace orchestral players (Jones 2023).
[19] None of the 84 randomised clinical trials of AI in healthcare published internationally between 2018-2023, were conducted in Australia (AAAIH 2023, p. 18).

consumers and businesses (such as through shopping transactions or website traffic) and it is typically assumed (without legal basis) that the data collector has an exclusive right to that data.

Private sector data is sometimes shared voluntarily by data holders where they benefit from doing so or are required to do so. For example, banks share data with superannuation funds, insurance companies and fraud reporting agencies. But sharing is done in a very limited way, with a wealth of private-sector data that could be shared safely for the benefit of individuals or the community more broadly.

Regardless of the available community-wide benefits of private-sector data sharing, many businesses avoid sharing their data, so as to not benefit competing firms, to limit liability when there is uncertainty about their data obligations, or to avoid reputational damage where the social license for data sharing does not yet exist.

Roles for government in private-sector data sharing

Governments have a role to play in lifting data sharing to socially optimal levels, but the task is not simple. It requires answering fundamental questions about who has rights over and benefits from data, and reassessing the costs and benefits of data sharing in the context of AI's evolving capabilities. Data sharing efforts will also need to establish social acceptance on a reoccurring basis to ensure sustainability.

The Commission has previously discussed some of these issues in the *Data Availability and Use* report (2017). The diffusion of AI only increases the importance of pursuing data access and creates greater impetus for governments to consult with industry and consumers and ensure any new private-sector data sharing mechanisms are practical and are socially acceptable.

The Consumer Data Right (CDR) establishes data sharing in the banking and energy sectors by extending some rights to individuals (primarily the rights to access and move data pertaining to them between accredited third parties) and limiting the scope of data holders to exclude others from accessing data. As AI diffuses across the economy, the benefits to the community of data sharing in other sectors will become clearer. CDR is not necessarily the tool to achieve this: freeing up data individual by individual is a slow process for increasing data access; and much of the useful data that could be used in AI applications is not consumer data (such as operational data, research and development data and supply chain data). But the underlying principle that data access can be improved by limiting the excludability of data and creating data rights for others remains a useful framework for considering how to further progress private-sector data sharing. Moving beyond the concept of excludable ownership over data would unleash the productive potential of private-sector data.

4. Governments as role models in data curation and safe sharing

Governments hold significant amounts of data in Australia. Commonwealth agencies alone have seen an average growth of 328% a year in data collection (How 2023). Even more data is generated through publicly funded activities and organisations. While some of this data is shared (for example, through ABS publications) much of it is heavily restricted.

Public-sector data, as a public asset, should be used to return the greatest benefit possible to the Australian public. Safe sharing and use of this data would facilitate greater returns from data holdings – enabling innovation in government services, supporting research efforts and bolstering productivity.

Much of the legislation to enable safe sharing of public-sector data is in place. But more attention on data set curation would improve the quality and accessibility of data for AI (including those data sets that are

collected under public sector funding but held by third parties such as universities and non-government service providers), and extension of the range of potential accredited data users beyond governments and academics would open up public sector data for AI applications.

Better data-sharing between governments

While the *Data Availability and Transparency Act 2022* (Cth) (DAT Act) provides a legal basis for sharing public-sector data between public entities, this is not always achieved. In many instances data-sharing is stifled by lack of protocols or incompatible systems. In some cases, the inability of governments to share data effectively can have direct impacts on service provision. For instance, lack of mutual recognition of teachers across states and territories limits capacity to track practitioners through the education system.

Some examples of better government data-sharing have emerged. For instance:

- PeopleWA links administrative data including health, education, justice and other data to provide richer and more accessible datasets to approved users, including government agencies, researchers and non-profit organisations for the benefit of Western Australians (Government of Western Australia 2023).
- In the US, the State of Oklahoma also has developed a centralised data hub working with Google Cloud which allows government agencies to opt-in to sharing their data and viewing data from other agencies in a cloud-based, scalable and secure environment (Google Cloud nd). The data hub has unified 23 petabytes of data and allows different agencies to draw on more comprehensive information to better perform their functions.[20]
- One US county has allowed their birth records, child protective services, homeless services and justice system records to be used in a machine learning algorithm to predict risk levels of abuse and neglect for newborn babies and assign families different levels of preventative support (Pauwels 2021). The program had a 92% predictive value accuracy during its pilot, demonstrating substantial capacity for improved family interventions (p. 11).

The Office of the National Data Commissioner is currently working to catalogue government data holdings which will create emore transparency about data assets in the public sector, although the extent to which agencies contribute remains to be seen (DTA 2023).

Modernising record-keeping within government

While public sector use of AI is beyond the scope of this paper, it should be noted that governments risk forgoing the benefits of AI for the wider public if the data they collect is not ready for AI use.

The Australian Public Service has been noted as underperforming in data management (How 2023). A 2022 survey of agencies' data practices highlighted poor performance in 'describing information assets', and 'use, reuse and interoperability' – two vital data practices for AI readiness (ANAO 2023, p. 73). Agencies performing the lowest were those with cultural or heritage functions and smaller operational functions (ANAO 2023, p. 29) suggesting these areas are most likely to miss out on AI-induced benefits if governments do not uplift public-sector data management practices.

The Commission (2017, 2022) has raised data interoperability alongside data sharing – in particular in the healthcare sector with the opportunity provided by My Health Record. The *Data Availability and Use* report recommended data management standards to support increased data availability and dataset quality

[20] For example, in the event of a natural disaster, data from the Oklahoma State Department of Education could be used to identify children who would normally receive funding for school lunches and targeted funding could be provided to those families to help keep their children nourished while kept away from school.

(PC 2017). In 2021, the National Archives of Australia introduced a set of standards for Australian Government agencies to better manage information assets. The impact of this policy on the interoperability and accessibility of public sector data will take time owing to the many legacy systems still in use and the volume of data being governed.

Sharing data beyond government entities

The extension of the DAT Act to allow data sharing with trusted private entities, individuals and unincorporated bodies would aid the diffusion of AI technologies. This could include gradual sharing, starting with accredited private organisations using data for policy and research purposes to achieve social objectives before allowing commercial use of the data. The DAT Act also provides some guardrails around the sharing of public sector data through an accreditation system and regulation by the National Data Commissioner (Office of the National Data Commissioner nd).

The Commission's *Productivity Inquiry* recommended ways to share data from government-funded services (recommendation 4.4), including 'identifying relevant data that could be safely shared and linked to benefit individuals receiving services' and 'setting technical standards for data sharing to promote interoperability' (2023). It was recommended that healthcare data be targeted in the first instance, including provisions for: opting out of the system; software compatibility and standards; and de-identification of data for use in system planning. For AI, healthcare applications have the advantage of potentially offering comparatively large datasets that could be safely used with AI for very high public benefit.

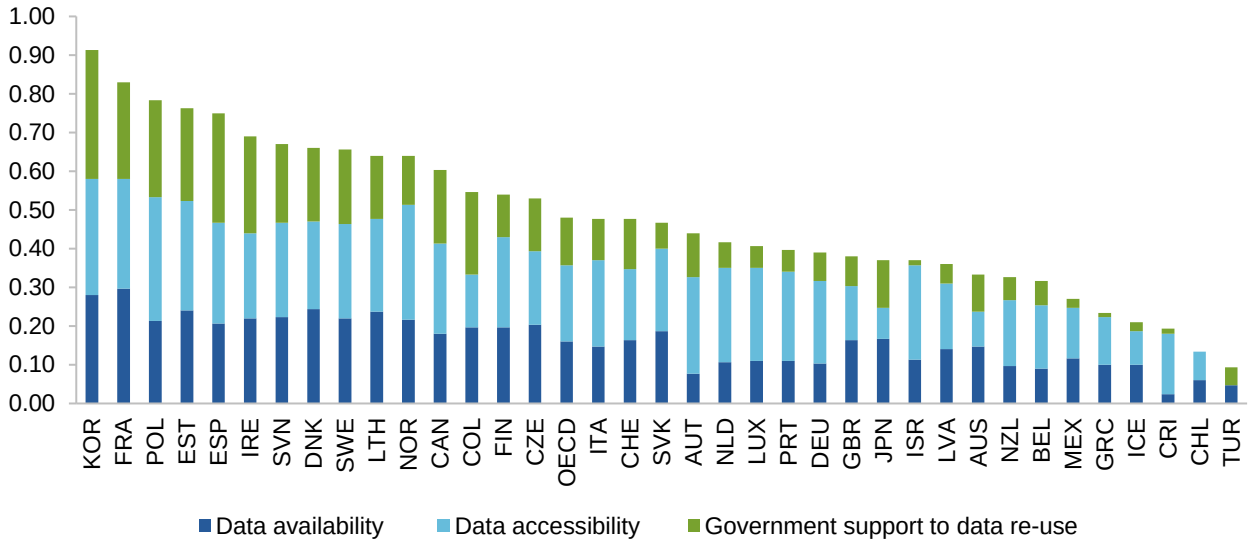
To support seamless service delivery, safe sharing of data held by government-funded service providers outside of healthcare – such as school education, childcare, aged care, criminal justice, community services and infrastructure contracts – should also be investigated and facilitated by the Australian Government. Experimental data sharing with low-risk data sets is especially important in the context of AI where technologies, digital literacy and applications are rapidly evolving.

There is likely to be a significant amount of de-identified, non-personal and non-confidential data that could be shared with minimal risk. This type of data can be sufficient (and sometimes desirable due to the limited liability created by it) for improved delivery of many products and services, including through AI applications (PC 2017, p. 12).

Towards a national data strategy

Australia has been a front runner internationally in development of frameworks that enable safe data access – think CDR, the DAT Act and Privacy legislation at the Commonwealth level, the data integration services provided by nine accredited data service providers nationally, and intra government sharing of data under the NSW Data Sharing Act. However, data use remains constrained: there are significant gaps in these frameworks (such as in relation to data held by commercial and taxpayer subsidised entities), and some aspects appear untested in light of new technologies (for example, given the way AI can use data without storing a copy). The OECD (2023) recently ranked Australia among the poorest performing countries in terms of data availability, data accessibility, and government support for data re-use (figure 4).

Figure 4 – Australia ranks poorly in data availability, access and government support to data re-use



Source: OECD (2023).

The recent implementation of the Data and Digital Government Strategy (2023c) will be a valuable framework to coordinate the use of data across different agencies of the Australian Government.[21] The Strategy aims to 'manage data as a valuable national asset' and includes a roadmap to 2030 of 41 initiatives (such as Digital ID, a Data Ethics Framework, and a Data Governance Framework) (Australian Government 2023a, 2023c). There would be value in not only developing similar strategies in state and territory governments, but also in aligning approaches and agreeing principles between Australian governments. As such, there would be value in extending the Data and Digital Government Strategy to establish a national strategy for data across all Australian governments.[22]

Such a strategy would hold even greater potential to enhance productivity were it to encompass the access and use of data across the economy (not limited to government). That is, a national strategy could set out agreed intentions for access, maintenance and use of *all* data collected in Australia, and to clarify rights of data subjects, curators and users.[23] The EU, for example, has attempted to address many such issues – across both government and non-government data – through the Data Governance Act and the Data Act.[24]

[21] The *Data and Digital Government Strategy* (2023c) is the Australian Government's current vision statement on data use, and includes a comprehensive roadmap of policy initiatives. The Strategy is distinct from its (now superseded) precursors: the *Australian Data Strategy* (2022) and the *Digital Government Strategy* (2021).

[22] Noting that a range of data- and AI-related issues are being considered on an ongoing basis by the *Data and Digital Ministers Meeting.*

[23] While the now superseded *Australian Data Strategy* (2022) was designed as the Australian Government's 'whole-of-economy vision for data', the Commission is not suggesting that data policy should revert to that strategy. Rather, a whole-of-economy vision for data would need to not only be genuinely economy-wide in scope, but also address issues such as data rights of different parties.

[24] In the EU, where it is estimated that 80% of 'industrial data' is never used, the European Parliament (2023) is aiming to increase data sharing through the Data Governance Act (adopted April 2022) and the Data Act (adopted November 2023). These acts aim to create trust in data sharing, addressing barriers to the use and reuse of 'industrial data'. The Data Act clarifies who can use data and under which conditions, including making it more difficult for companies limit

A good start in addressing these gaps and issues would be for all Australian governments to build on the Data and Digital Government Strategy and to develop a comprehensive *national* data strategy. The strategy would draw on the social objectives encapsulated in a range of existing legislation (such as laws on privacy, cyber and data protection; anti-discrimination, competition and consumer protection). Once developed, all future regulations and guidelines around data use and data analytics could refer to the agreed principals of the national data strategy. In this way, the data strategy could provide a secure and consistent basis for the development and use of AI and other data-intensive technologies.

References

AAAIH (Australian Alliance for Artificial Intelligence in Healthcare) 2023, A National Policy Roadmap for Artificial Intelligence in Healthcare, AAAIH.

ALRC (Australian Law Reform Commission) 2013a, Copyright and the Digital Economy, DP79.

—— 2013b, Text and data mining, Government, Australian Law Reform Commission, https://www.alrc.gov.au/publication/copyright-and-the-digital-economy-dp-79/8-non-consumptive-use/text-and-data-mining/ (accessed 4 December 2023).

ANAO (Australian National Audit Office) 2023, Management of Information Assets, Performance audit, 28 June, 44, Auditor-General Report.

Angileri, J Brown, M Dipalma, J Dancy, C and Ma, Z 2019, Ethical Considerations of Facial Classification: Reducing Racial Bias in AI.

Attorney-General's Department 2023a, Privacy Act Review 2022.

—— 2023b, Third Roundtable on Copyright.

—— nd, Privacy, Government, Attorney-General's Department, https://www.ag.gov.au/rights-and-protections/privacy, https://www.ag.gov.au/rights-and-protections/privacy (accessed 4 December 2023).

Australian Government 2021, Digital Government Strategy: Accelerating the Digital Future, Canberra.

—— 2022, Australian Data Strategy: The Australian Government's whole-of-economy vision for data, Canberra.

—— 2023a, Data and digital foundations, Data and Digital Government Strategy, https://www.dataanddigital.gov.au/strategy/missions/data-and-digital-foundations (accessed 11 January 2024).

—— 2023b, Government Response | Privacy Act Review Report.

—— 2023c, Roadmap, Data and Digital Government Strategy, https://www.dataanddigital.gov.au/sites/default/files/2023-12/Data%20and%20Digital%20Roadmap%20v1.0.pdf (accessed 11 January 2024).

—— 2023d, Simple, secure and connected public services, Data and Digital Government Strategy, https://www.dataanddigital.gov.au/ (accessed 4 January 2024).

—— 2024, Safe and responsible AI in Australia consultation: Australian Government's interim response, Canberra.

Carroll, G and Healey, N 2023, 'Australian artist Tank accuses online marketplace Temu of stealing design, selling for less than $7', ABC News.

DACS 2023, DACS warns that new text and data mining exception will undermine licensing opportunities for visual artists and weaken copyright, https://www.dacs.org.uk/latest-news/dacs-warns-that-new-text-and-data-mining-exception?category=For+Artists&title=N (accessed 11 November 2023).

Dickson, B 2021, Tesla AI chief explains why self-driving cars don't need lidar, TechTalks, https://bdtechtalks.com/2021/06/28/tesla-computer-vision-autonomous-driving/ (accessed 23 October 2023).

DISR (Department of Industry, Science and Resources) 2022, Australia's AI Ethics Principles, Government, https://www.industry.gov.au/node/91877, https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles (accessed 25 July 2023).

—— 2023, Safe and responsible AI in Australia, June.

DLA Piper 2023, Can generative AI rely on the text and data mining (TDM) exception for its training?, https://www.lexology.com/library/detail.aspx?g=a04a436f-3225-4f77-9bce-ee0f2a608ad0 (accessed 9 November 2023).

DP-REG 2023, DP-REG joint submission to DISR AI discussion paper, Submission to the Department of Industry, Science and Resources.

Dreyfus, M 2023, Copyright and AI reference group to be established, Media Release.

DTA (Digital Transformation Agency) 2023, Australian Government Data Catalogue, Government, Australian Government Architecture, https://architecture.digital.gov.au/australian-government-data-catalogue (accessed 6 December 2023).

European Alliance for Research Excellence 2018, Japan amends its copyright legislation to meet future demands in AI and Big Data, https://eare.eu/japan-amends-tdm-exception-copyright/ (accessed 9 November 2023).

Eurpoean Parliament 2023, Boosting data sharing in the EU: what are the benefits?, News, https://www.europarl.europa.eu/news/en/headlines/society/20220331STO26411/boosting-data-sharing-in-the-eu-what-are-the-benefits (accessed 4 January 2024).

data sharing through claims of 'trade secrets'. It also addresses when individual consumers have rights to access the data they have generated (similar to the Consumer Data Right in Australia).

Gomede, E 2023, Data Quality vs. Data Quantity: The Crucial Balance for Artificial Intelligence, Medium, 14 September, https://medium.com/@evertongomede/data-quality-vs-data-quantity-the-crucial-balance-for-artificial-intelligence-faed8b0eaea4 (accessed 5 December 2023).

Google 2023, Response to the Copyright Enforcement Review issues paper, 7 March, Submissions to the Copyright Enforcement Review, Google.

Google Cloud nd, State of Oklahoma Case Study, Google Cloud, https://www.oaic.gov.au/__data/assets/pdf_file/0029/94295/OAIC_Annual-Report-2022-23.pdf (accessed 17 November 2023).

Government of Western Australia 2023, PeopleWA, Government, Government of Western Australia, https://www.wa.gov.au/organisation/department-of-the-premier-and-cabinet/office-of-digital-government/peoplewa (accessed 17 November 2023).

Hughes, C 2023, Australia: number of Facebook users 2022, Statista, https://www.statista.com/statistics/304862/number-of-facebook-users-in-australia/ (accessed 23 October 2023).

International Copyright Issues and Artificial Intelligence 2023, Transcript from Online Webinar on July 26, 2023.

IP Australia 2023, Generative AI and the IP rights system provocation series, Australian Government, Canberra.

Myers, A 2021, A New Approach to the Data-Deletion Conundrum, Stanford University Human-Centred Artificial Intelligence.

OAIC (Office of the Australian Information Commissioner) 2021, Clearview AI breached Australians' privacy, Government, Office of the Australian Information Commissioner, https://www.oaic.gov.au/newsroom/clearview-ai-breached-australians-privacy (accessed 27 July 2023).

—— 2023, Australian Community Attitudes to Privacy Survey 2023, 8 August, Office of the Australian Information Commissioner, https://www.oaic.gov.au/engage-with-us/research-and-training-resources/research/australian-community-attitudes-to-privacy-survey/australian-community-attitudes-to-privacy-survey-2023 (accessed 23 October 2023).

OECD 2023, 2023 OECD Open, Useful and Re-usable data (OURdata) Index: Results and Key Findings, OECD Public Governance Policy Papers.

Office of the National Data Commissioner nd, Data Availability and Transparency Act 2022, Government, Office of the National Data Commissioner, https://www.datacommissioner.gov.au/law/dat-act (accessed 10 November 2023).

Pauwels, E 2021, Allegheny Country Department of Human Services: case study, Case study, October, UNICEF, New York.

PC (Productivity Commission) 2017, Data Availability and Use, Report no. 82, Canberra.

—— 2022, 5-year Productivity Inquiry: Australia's data and digital dividend, Interim Report 2, Canberra, August.

—— 2023, 5-year Productivity Inquiry: Australia's data and digital dividend, Inquiry Report no. 100, Vol. 4, Canberra.

Research Australia 2021, 2021 Public Opinion Poll on Health & Medical Research & Innovation, September, Research Australia.

Solomon, L and Davis, 2023, The State of AI Governance in Australia, May, Human Technology Institute, The University of Technology Sydney, https://www.uts.edu.au/human-technology-institute/news/report-launch-state-ai-governance-australia (accessed 24 July 2023).

SPARC nd, Developments in Publishers' Text and Data Mining (TDM) Policy, SPARC, https://sparcopen.org/our-work/developments-in-tdm-policy/ (accessed 14 December 2023).

Zia, T 2023, Enhancing Language Models: How Your Feedback Transforms LMs like ChatGPT, Techopedia, 3 July, https://www.techopedia.com/enhancing-language-models-how-your-feedback-transforms-lms-like-chatgpt (accessed 20 December 2023).

2023, The New York Times Company v Microsoft Corporation, OpenAI Inc, OpenAI LP, OpenAI GP, LLC, OpenAI LLC, OpenAI Opco LLC, OpenAI Global LLC, OAI Corporation LLC, and OpenAI Holdings LLC., 1:23-cv-11195, https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf (accessed 2 January 2024).