Submission No 41

INQUIRY INTO ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES

Organisation: CSIRO

Date Received: 27 October 2023

Australia's National
Science Agency

CSIRO submission to Inquiry into artificial intelligence (AI) in New South Wales

CSIRO Submission 23/834

October 2023

Enquiries should be addressed to:

1. Response to New South Wales Government Select Committee's Inquiry

A. The current and future extent, nature, and impact of AI in New South Wales.

Artificial intelligence (AI) presents transformative and profound implications for economies, societies, and individual lives. However, the test of relevance and impact for AI models lie in their implementation within real-world contexts. AI is already transforming industries globally, but overall adoption in Australia is still emerging (CSIRO, 2023). The path from AI models to value in business is a lengthy and risky path. Larger corporations are leading in AI-based innovations due to cost, risk, data and infrastructure availability. Meanwhile, small business is finding it more difficult to afford innovations.

Areas where adoption is occurring include:

- **Generative AI:** software development, art, literacy
- **Responsible AI:** impact investing, business membership associations, assurance
- **Autonomous driving and operations**: automated mining, ports, logistics, and transportation
- **Computer vision:** healthcare, security, and agriculture and food
- **Chatbots:** customer services, ChatGPT usage at work.

Below we have identified general trends and insights in AI development and implications for New South Wales (NSW).

General trends and insights in AI development:

- **AI is getting better fast:** AI models are getting increasingly more capable at a rapid pace (Zhao et al, 2023). The accelerated progress in AI models is backed by the fast-growing AI research community, and the two continue to reinforce each other. We expect to see ongoing growth in both AI capabilities and AI research communities in the next few years.
- **AI becomes both bigger and smaller:** New AI models being developed grow stronger with size, while other models continue to be optimised to be smaller and more efficient so they can be deployed on smaller devices with a smaller energy footprint (Murshed et al., 2021). For example, OpenAI's GPT-4, released in 2023, is speculated to have 1.76 trillion parameters. Conversely, Edge AI represents an approach focusing on the reduction of model parameters and complexity, as well as the conservation of computational resources, such as Extended Reality (XR)/Virtual Reality (VR) systems, Internet of Things (IoT) devices, sensors, and embedded systems.

- **Use of AI is democratised, yet creation is more centralised:** Open-sourced AI models and Cloud APIs/services for AI, such as Github, HuggingFace, ChatGPT and Google Cloud Platform are today widely and publicly accessible**.** However, training a Large Language Model reportedly costs millions of US dollars combing hardware (GPU cluster), data collection, and labour costs (dev team) (Leswing, 2023). As a result, opportunities related to creating large AI models effectively become capital heavy endeavours, which pose great financial challenges for individuals, SMEs, or startups to take on.
- **Industry now leads AI innovations:** Academia ignited great advancement in AI with fundamental innovations (e.g., LeNet-5) driving AI development in the past decade, but industry is now leading scaling up, training, and deploying impactful AI systems (Tesla AutoPilot, Open AI ChatGPT) (Eastwood, 2023).
- **Regulations have a fundamental impact on the development of AI:** AI is intrinsically dependent on large-scale training datasets, and in many cases, direct links between its outputs and its training data are difficult to establish (Trend, 2023). This poses challenges in fitting AI into existing or new data and privacy regulation frameworks (see Section G). A report (EPRS, 2020) from the EU Parliament which explores the relationship between the GDPR regulatory framework to AI acknowledged that "a number of AI-related data protections issues are not explicitly answered in the GDPR, which may lead to uncertainties and costs, and may needlessly hamper the development of AI applications".
- **AI for robotics:** The advancement of Deep Learning and AI has revolutionised robotic manipulation and autonomous learning. For example, through deep reinforcement learning, robots have been able to learn from raw sensory input, practicing and refining their manipulation skills autonomously in the real world. This method has led to significant improvements in the efficiency and adaptability of robotic systems, allowing them to learn a wide array of tasks without extensive human intervention or reprogramming.
- **AI for science:** CSIRO recently published a report (Hajkowicz et al., 2022) and journal paper (Hajkowicz et al., 2023) about adoption patterns for AI in science and capability development pathways. The research shows that AI has been playing a transformative role in advancing the natural sciences, physical sciences, life sciences, social sciences and the arts and humanities. AI is enhancing human comprehension of phenomena across all disciplines and wide-ranging spatial and temporal scales. The rising AI adoption by researchers, and the many success stories, has catalysed a novel research field known as AI for Science. Through the utilisation of extensive data sets and sophisticated neural network structures, Large Foundational Models facilitate the processing, interpretation, and creation of multifaceted scientific knowledge. Capacity to deal with complex information situates these models as a vital tool in the continually evolving field of AI for Science, holding potential for further innovation and understanding in scientific research.

Implications for New South Wales

- **Technology industry and AI innovation:** NSW, as the largest economy in Australia with 31 per cent of the national output, also plays a leading role in Australia's technology and AI innovations. According to the NSW Technology Prospectus (Investment NSW, 2023), NSW accounts for 38 per cent of Australian software developers and application programmers, and 45 per cent of Australian AI businesses. This indicates that related industries in NSW will likely benefit from the continued and rapid advancement of AI, and potentially lead the AI-induced transformations of industries. NSW is also on a strong trajectory for AI research, with three of the seven world leading AI universities based there.
- **Increased productivity:** Positive gain in productivity is expected for individuals and businesses who successfully adopt AI in their workflow (Kalliamvakou, 2022) with an average 55 per cent reduction in development time among programmers who use the AI-based coding assistant called CoPilot (Cihon et al, 2023) (Botija, 2023). A McKinsey report (2023) identified software engineering, customer operations and marketing/sales as the top sectors that will receive the biggest impact from generative AI, from a combined angle of value and productivity. Industries in NSW, such as manufacturing and agriculture, are expected to continue benefitting from AI.
- **New markets:** The prevalence of AI adoption has also had a profound impact on the emergence of new markets in NSW, reshaping how industries function and creating new business areas that were previously inconceivable. These include emerging new educational paradigms for training AI trainers/ethicists, equipping them with AI literacy, ethics, and related fields. AI adoption will expand the economic landscapes and offer opportunities for employment, innovation and growth.
- **Uneven impact:** The rate of AI adoption and the economic impact that AI has on different industry sectors will be highly uneven. In addition, AI has different implications on productivity and automation (McKinsey, 2023) . NSW's four biggest industries by gross state product are *financial and insurance services, Professional, scientific and technical services, Manufacturing,* and *Health care and social assistance* (Investment NSW, 2017), all of which will be heavily impacted by AI (generative AI in particular) in terms of productivity and job market. However, if these broader categories are further split into more specific sectors, it is estimated that software engineering and customer operations will receive a larger impact in terms of productivity than supply chain or manufacturing. From the perspective of AI-based automation, some occupations in office support, production work, food services will be more exposed to automation made possible by combining AI and robotics.
- **Stronger competition:** Sectors in the NSW economy will face asymmetric advantages from domestic and foreign competitors with more advanced AI technologies and faster adoption. For example, deployment of more advanced AI and robots in manufacturing in the US and China will potentially bring them asymmetric competitive advantages over NSW manufacturers.

- **Potential challenges for ageing populations:** In the near future, as healthcare and other essential services evolve into a paradigm in which online, AI and automation-driven services become the norm, ageing populations may find more challenges in interacting with such services and products. While these services might be more convenient and accessible for younger generations, they may also have discriminatory effects on the older population or even cause harm (Stypińska J et al., 2022).
- **Malicious use:** Generative AI will likely be used to commit certain crimes, such as phishing scams, cybercrime and hacking, impersonation, and dissemination of false information or propaganda (Deloitte, 2022).
- **Data breaches:** As the use of Generative AI penetrates to more industries and more use case scenarios, users might unintentionally expose confidential information to unauthorised third parties through interactions with AI models. Recently, both Open AI and Nvidia had data breaching incidents related to AI products (Trend Micro, 2023) (FT, 2023). Though some businesses have taken measures to ban the use of external generative AI services (Vincent, 2023), with further penetration of generative AI products, more incidents are likely to happen.

B. The social, economic, and technical risks and challenges presented by AI to the New South Wales community, government, economy, productivity, and environment.

AI is widely considered one of the most significant productivity-boosting technologies of the current era. For example, a study of AI patent applications by 5,257 companies worldwide between 2000 and 2016 found that AI had a strong positive effect on firm labour productivity (Damioli, 2011). This has been backed up by more recent economic studies investigating the productivity benefits of AI at the firm level (Czarnitzki et al, 2023).

A recent study by consulting firm McKinsey (2023) estimates that cutting edge generative AI (such as GPT-4) could potentially generate US$2.6 trillion to US$4.4 trillion across all industries globally. Another study by Microsoft and the Tech Council of Australia (2023) finds that generative AI could contribute between A$45 billion to A$115 billion to Australia's economy annually with 70% coming from enhanced labour productivity, 20% from improved quality of industry outputs and 10% from new products and services. The range is given to capture future possibilities associated with low versus high adoption scenarios. The extent of the future economic impact is not known with certainty. AI brings economic challenges in addition to economic opportunities (e.g., a changed cybersecurity risk profile). However, it's likely that AI will create significant opportunity for NSW industry to increase productivity and grow and transform into new areas as stated in Section A.

Below we have identified considerations and opportunities for NSW when considering AI's potential to transform sectors critical to the state's economy.

- **AI productivity gains are not assured:** Productivity gains depend on developing and adopting the right technologies in the right ways. They also depend upon intelligent harmonisation of human individual and organisational decision making with automated systems. Many companies face a considerable adjustment phase as they become familiar in working with AI systems. During this time of adjustment, companies might not see productivity gains. These might occur over a longer period when the new AI-enabled automated systems are working well. Some of the issues associated with AI productivity uplift at the organisational level are explored in an article by CSIRO researchers (Hajkowicz and Whittle, 2023). Ensuring NSW companies in the agricultural, mining, manufacturing, and service sectors have the right infrastructure, skills, and capabilities to adopt AI wisely is part of the challenge.
- **AI education and training:** One of the most fundamental investments NSW could consider is in the education and training sector. This is relevant for all stages of lifelong learning including school and tertiary education and training. It is also relevant for wide ranging learning formats including degree programs and flash courses designed to meet the needs of busy professionals. As the field of AI continues to expand and evolve, so too will learning and training relating to AI. The skills profile needed to develop and apply AI is likely to change rapidly as new technologies emerge. In some industries at the national level, the local labour market is not sufficient to meet industry needs and skilled migration programs are used to attract specialised AI talent. During the National AI Centre's recent Listening Tour, many companies expressed a desire for increased training opportunities and the difficulties they have finding AI-skilled staff (Solar, 2023).

- **AI-related digital infrastructure:** There is an ongoing need to improve the quality and reliability of internet connectivity. Many on-farm agricultural, environmental, and mining AI systems need access to data in order to operate. Limited connectivity in regional areas will likely hold back the ability of local industry to develop and adopt AI. Desired AI-related infrastructure also includes access to compute power to design and train machine learning models that are increasingly data hungry. This compute power will need to be both available and cost effective.
- **AI service providers:** The development and adoption of AI by NSW industry will require the ongoing growth and diversification of the AI service provider ecosystem in the State. Developing this ecosystem so that NSW becomes a hub of companies and service providers will make it easier for companies to adopt AI, which in turn will fuel the demand for more AI services. CSIRO research shows that in the online world "place" and "location" still matter – when service providers are near their customers this usually leads to boosts in sales and growth (Hajkowicz et al, 2023).
- **Investment in research and development:** As a whole, the research and development component of AI development in Australia has – thus far – received limited attention and investment. Much of the investment has been focused on adoption of AI technology developed elsewhere. However, within the industries of mining, agriculture, manufacturing, tourism and many others, there are spaces where Australia can develop new and innovative AI products that problem-solve domestically but can also reach a global market. This will involve research and development, including the product innovation and commercialisation component which has been underdone historically. There is an opportunity for NSW to consider making targeted investments in research and development to build novel AI solutions and commercial products.

C. The current and future extent, nature and impact of AI on the New South Wales labour market including potential changes in:

   a. Earnings
   b. Job security
   c. Employment type
   d. Employment status
   e. Working patterns

Below we have identified implications for NSW in relation to the labour market.

- **AI impacts tasks, but less so entire jobs:** Much of the early forecasts of widespread job loss due to computerisation and automation turned out to be incorrect. Take for example a 2013 University of Oxford study that forecast 47 per cent of jobs in the US economy were at risk of automation (Frey and Osborne, 2013). This simply did not occur, and it is likely that computerisation and automation have been net job creators (although jobs have disappeared in some parts of the labour force). The main issue was that automation impacted tasks but much less frequently entire jobs. It also created comparatively high paid jobs for workers skilled in AI technologies (e.g., software engineers). The limitations of the Frey and Osbourne study and related issues are explored in a more recent research paper from The University of Melbourne (Coelli and Borland, 2019).
- **The services sector will be impacted by generative AI:** The NSW Government estimates 86 per cent of the State's workers are employed in the services sector (NSW Government, 2023). This includes sub-sectors such as administrative services, retail, banking, finance, tourism, and professional/scientific services. These are the employment sectors most likely to impacted by generative AI tools (Brynjolfsson et al, 2023). Due to the increased adoption and capability of generative AI the coming decade could be different to the proceeding decade. While uncertain, the expectation is that AI will continue to impact or replace tasks (not jobs). Reskilling, upskilling, and refocusing workers throughout AI tech disruption will be important; proactive actions by industry, government and community are likely to smooth the transition.
- **Impacts on labour productivity:** A study by economists at Stanford University (Brynjolfsson et al, 2023) found generative AI boosted worker productivity by 14 per cent, with bigger improvements for less experienced workers. The study was done for 5,179 workers who respond to customer queries in a software company. It found that workers could resolve 14 per cent more customer queries per hour using generative AI (and with higher rates of customer satisfaction). Another study (Dell'Acqua et al., 2023) by researchers at Harvard University finds that workers using generative AI (GPT-4) could complete representative management consulting tasks 25 per cent faster with 40 per cent greater quality. However, for tasks that were designed to fall outside the known capabilities of GPT-4 consultants were 19 per cent less likely to produce correct solutions. So incorrect application of generative AI can be productivity harming. Overall, the productivity gains of AI are possible but not assured. To

achieve productivity gains the AI tools need to be harmonised with human workers and organisational behaviours (Hajkowicz and Whittle, 2023).

- **Opportunity for work/life balance:** At the same time the generative AI transition is occurring, so too is the emergence of the four-day week. There are anecdotal reports that productivity is maintained or improved along with improved worker satisfaction, although this has not yet been subject to substantial rigorous peer-reviewed research. A Swinburne University study (Hopkins, 2023) of 10 Australian organisations trialling the four-day week (with same take home pay) found that owners/managers of those companies gave it an average overall success rating of 9.25 (out of 10). It's also reported that 70 per cent of companies found productivity to be higher and 30 per cent as high as it was before. There is an opportunity for the NSW government and companies to explore ways via which AI can increase productivity and also improve work-life balance for employees.
- **Salaries and skills demand:** Generative AI is rapidly changing and dominating in-demand AI skills. It is likely that workers with advanced skills in these areas will get better paid opportunities. However, there is also evidence of tech sector salaries declining. The online employment platform Hired (2023) finds US tech sector salaries have decreased 9 per cent from mid-2022 to mid-2023 (accounting for inflation; down 3 per cent in nominal terms). However, historically the tech sector has seen strong salary growth (so it could be seen as a correction). Hired noted 'employers' increasing reliance on AI tools' as one of the key factors and that newer, less-experienced workers are being hardest hit. It is possible the tech sector workforce is in early stages of restructuring due to AI (especially generative AI). Workers are likely to transition skills into new areas, but this could take time, and, in some cases, could be a challenging transition.

D. The current and future extent, nature, and impact of AI on social inclusion, cohesion, and the disadvantaged.

There is limited consideration around the use of AI with vulnerable populations, such as aging and people living with a disability. While the Therapeutic Goods Administration (TGA) includes regulation for medical devices, AI applications not intended as medical devices are used by people to support their daily life without proper consideration and understanding of the risks and with limited tools to mitigate the risks. CSIRO and the National Disability Insurance Agency (NDIA) developed a Framework for AI-enabled Assistive Technology to mitigate these risks (Silvera et. Al, 2022).

There are also limitations in available data regarding Aboriginal and Torres Strait Islander peoples. An illustrative example is the application of AI in healthcare. The risks are not solely due to algorithmic biases but also due to the data biases inherited from data sets the AI model is trained on.

Data quality and fairness are pivotal for the success of any AI-based solutions. The mismanagement of data biases poses a potential source of discrimination and injustice. Unfortunately, due to Indigenous Data Paradox, Indigenous healthcare data is commonly misrepresented and negatively biased. Responsible and ethical guidelines reflecting Indigenous Data Sovereignty could be considered into the existing framework (Walter, 2018). To achieve this, we suggest a consultative approach ensuring a health justice approach to AI data governance.

The lack of inclusive data is particularly pronounced in genomics and extends to all non-European backgrounds (Mills, 2020). Genomic information increasingly informs healthcare, and the lack of actionable markers for non-Caucasian populations threatens to increase health disparity. For example, a CSIRO study found that 10 per cent of cystic fibrosis variants were missed in Australia's multiethnic population (Shum et al. 2022).

This disparity in knowledge and data will negatively influence data-driven Machine Learning (ML) applications, including automated data annotation and pre-diagnostic advice. Data and inputs into AI models are constantly changing, so AI-based systems need continuous evaluation and mechanisms to detect and action on the decline in performance.

It is difficult for industry players to demonstrate conformity with AI policies based on internal processes only. Therefore, third-party auditing of AI-based service/product providers is needed. Auditing by an independent, specialised entity could facilitate the compliance process, especially for smaller players.

E. The current and future extent, nature and impact of AI on human rights and democratic institutions and processes in New South Wales.

Nil response.

F. The effectiveness and enforcement of Commonwealth and New South Wales laws and regulations regarding AI.

Nil response.

G. Are current laws regarding AI in New South Wales that regulate privacy, data security, surveillance, anti-discrimination, consumer, intellectual property and workplace protections, amongst others, fit for purpose.

CSIRO cannot comment on the efficacy of legislation. However, we have provided below comments on the technical issues that we recommend are considered in the development of legislation.

Privacy challenges related to AI.

There is an inherent tension between the development and operation of AI and the data protection principles captured in regulations such as the European Union's GDPR (Sartor, 2020), Australia's Privacy Act 1988, including the recent Response to the Privacy Act review (Australian Government, 2023), or the California Consumer Privacy Act (e.g., minimisation of data, limitation of purpose, right for correction and erasure). From a data perspective, often large or Internet-scale amounts of data are required to train an AI model. Some of the key issues with data in the context of training AI are as follows:

- **Consent and information asymmetry:** The individual or businesses that are the source of data may not have given explicit consent that their data could be used in a secondary-usage scenario as training materials for AI models (Burgess, 2023). Examples from a business point of view include several recent legal cases on potential copyright infringement where materials were used for AI training without explicit consent or contractual agreement. From an individual point of view, information asymmetry refers to the issue where a person's sensitive information is used in decisions that negatively impact that person, in this case through an AI model, e.g., AI models used to compute insurance premiums (Eliot, 2022). Interestingly, a few academic papers argue for the opposite effect, that AI mitigates information asymmetry in specific use-cases/domains (De La Pena, 2023).
- **Definition of what is considered 'sensitive information':** Classical definitions are based around Personally Identifiable Information (PII). However, pieces of information that are usually not considered sensitive may be combined to form a *quasi-identifier*, which can then be sufficient to uniquely identify an individual. Drawing the line on what is or is not PII or quasi-identifier is contentious (OVIC, 2019). In the AI context, pieces of information about individuals are often combined into features used for training or testing AI models. How to assess the sensitivity and the privacy risks associated to the data and derived features that are used by AI model is still a challenging research question. While the recent Government Response to the Privacy Act Review agreed in-principle that the definition of personal information should be expanded to include technical and inferred information, we do not yet know how this potential amendment would impact derived data feature for AI training.

- **Fairness, privacy and bias:** Data treatment (access, collection, storage etc.) may be designed and used to mitigate data sensitivity and associated privacy risk (such as re-identification). However, these very treatments enhancing privacy may have positive or negative effects on the fairness of AI models built on such data, i.e., removing or adding biases in the AI outputs. Understanding and controlling such interactions between privacy and fairness may be domain and/or model specific.
- **Operation of live AI systems:** While being queried by users, a Large Language Model may leak in its output responses sensitive information about individuals whose data has been used in building that model. Several techniques, collectively known as "inference attacks", have been designed to extract information about specific individuals' attributes or membership in groups from AI models. Such privacy risks posed by in-operation AI are starting to increase in awareness, as evidenced by the recent Australian Privacy Act Review report which suggests that Privacy Impact Risk Assessments may be implemented for AI systems (Attorney General, 2022).
- **Secondary usage of input/prompt information:** While the primary usage is to query and get a service from the AI, information may also be later used as further training for the same AI system (secondary usage). Thus, any private or confidential parts of this information may end up being part of the model and leaked further down the track or used by the model in a negative way against an individual or business. As a result, several organisations have banned the use of third-party AI in critical parts of their businesses (Mok, 2023). On the regulatory side, the Australian Privacy Act Review recently suggested (Suggestion 19.3) including new legislation for individuals to know when their sensitive information is being used to make (AI-based) substantially automated decisions with significant effect on them.
- **Individual's control over sensitive information:** Several privacy laws mandate or suggest that individuals have a detailed level of control over their sensitive information, which is also agreed in-principle in the recent Government Response to the Privacy Act Review. For example, allowing them to seek right of correction or right of erasure (aka 'right-to-be-forgotten') over their data. While such rights are often technically challenging to implement on data (Bertram, 2019), extending it to AI models is more difficult. Indeed, the training time required to build an AI model, such as ChatGPT, is counted in months. Thus, it is not efficient to re-train such models so that it 'unlearns' about an individual every time one uses their right-to-be-forgotten (Zheng et al., 2023). Designing novel AI algorithms or adapting existing ones to unlearn is an open research challenge. Doing so while avoiding impacting other responsible properties of an AI model (such as fairness) is an even more difficult unanswered research question (Google, 2023).

AI and data security

The intersection of AI and security goes beyond the confines of conventional data security principles such as Confidentiality, Integrity, and Availability. It requires a broader perspective that includes the security of AI models themselves. It also involves the role of the AI model in the data security life cycle: Create, Store, Use, Share, Archive and Destroy. The previous sub-section discussed several aspects of the Confidentiality principle. Below we outline issues pertaining to Integrity.

- **Integrity at training and inference stage:** One of the prevalent attacks on the AI model is the data poisoning attack. In data poisoning attacks, the integrity of the training data is violated. The training data is manipulated (by adversaries) in such a way that the overall accuracy of the AI model will be degraded. Such attacks could be untargeted or targeted for certain tasks. For example, backdoor attacks are where adversaries insert hidden associations or triggers to the deep learning models to override correct inference, such as classification, and make the system perform maliciously according to the attacker-chosen target while behaving normally without the trigger. The challenge is particularly critical due to the broad array of potential backdoor attack surfaces. Recognising this evolving threat landscape, the U.S. Army Research Office initiated the TrojAI project in 2019 (IARPA, 2019), soliciting countermeasures, while the National Institute of Standards and Technology launched a corresponding online competition to address these challenges.
- **Significance of integrity for AI's wider adoption:** Cloud-enabled Machine Learning as a Service (MLaaS), such as ChatGPT, has shown enormous promise to transform how deep learning models are developed, deployed and used. Nonetheless, potential risks and security threats are emerging with MLaaS since the pre-trained models can be maliciously modified through Trojan or backdoor attacks. To protect the model integrity and end-user benefits, it is imperative to verify whether the deployed model has been tampered with. An added layer of complexity arises from the uncertainty surrounding the trustworthiness of service providers, given their potential to manipulate the verification process for their gain. As a result, a critical challenge emerges in devising a universally applicable verification protocol for end-users that ensures the authenticity and integrity of deployed models.
- **Integrity of inference results** is equally pivotal for the broader adoption of AI. As outlined above, the attacks on data and model integrity bear the potential to significantly influence the outcomes of inference. Even in the absence of those attacks, many AI researchers have observed that the results produced by the AI model are incorrect and nonsensical and cannot be explained due to the inherent black-box nature of the AI model. Within this context, a notable observation within Large Language Models is “hallucination”, where the model generates erroneous, seemingly accurate, but inherently false information. This phenomenon highlights the importance of addressing both malicious tampering and the explainability of AI models to increase the credibility and trustworthiness of inference results.

In the data security lifecycle context, AI models often break the fundamental principles due to their role as custodians of vast amounts of data. An obvious violation occurs in the form of data retention. Specifically, the data utilised for training AI models cannot be destroyed, primarily because an unlearning process does not exist to remove specific data instances selectively. This inability to eliminate data post-training presents a significant challenge in complying with the data security life cycle and privacy norms. This further highlights the complexity of ensuring data security throughout the lifecycle of AI models.

H. The effectiveness of the NSW Government's policy response to AI including the Artificial Intelligence Strategy, Ethics policy and Assurance Framework.

Nil response.

I. The measures other jurisdictions both international and domestic are adopting in regard to the adaption to and regulation of AI.

Nil response.

J. The social, economic, and environmental opportunities for New South Wales to benefit from AI.

Nil response.

K. Recommendations to manage the risks, seize the opportunities, and guide the potential development of AI.

Responsible AI

In CSIRO's recent response submission to the Department of Industry, Science and Resources' discussion paper on Supporting Responsible AI (CSIRO, 2023), CSIRO proposed seven non-regulatory initiatives designed to increase the adoption of responsible AI practices, thus providing a competitive advantage for the industry and positioning Australia as a world leader in responsible AI:

- **Initiative 1:** Develop industry best practices, playbooks, guidelines, and case studies for Australia's priority industry sectors, especially targeting small and medium enterprises (SMEs), while considering Australia's unique context.
- **Initiative 2:** Develop trustworthiness metrics, measurement, testing, evaluation, verification, and validation (TEVV) methods and guidelines along with associated tools and products, including the seeding of a world-leading responsible AI tool industry in Australia.
- **Initiative 3:** Set up programs to encourage and incentivise industry and government to develop new validated best practices and share them within the Australian industry and globally.
- **Initiative 4:** Set up a national sandbox to explore and experiment with responsible AI approaches in a safe environment.
- **Initiative 5:** Set up connected responsible AI awareness and training programs.
- **Initiative 6:** Set up a national Responsible AI technology program to inform responsible AI policy, regulation, and international standards.
- **Initiative 7:** Identify responsible AI approaches and edge cases that can benefit all Australians.

Finally, there are several essential points to consider across all seven initiatives:

a. Emphasise AI governance at the system level, not just the model level.
b. Pay particular attention to the intersection of AI with other vital and emerging technologies, such as cybersecurity, quantum systems, blockchain, and robotics.
c. Concentrate on the empirical understanding and experimentation of AI uses and technologies.
d. Adopt a supply chain perspective.

L. Any related matter

Nil response.

References

Attorney General (2022) Privacy Act Review Report, The Attorney General's Department, 2022. https://www.ag.gov.au/sites/default/files/2023-02/privacy-act-review-report_0.pdf

Australian Government (2023) Government Response to the Privacy Act Review Report, https://www.ag.gov.au/rights-and-protections/publications/government-response-privacy-act-review-report

Bertram T, et al (2019) Five Years of the Right to be Forgotten. Proceedings of the Conference on Computer and Communications Security, 2019. https://doi.org/10.1145/3319535.3354208

Botija, M (2023) AI productivity: my experience with Copilot. https://bootcamp.uxdesign.cc/ai-productivity-my-experience-with-copilot-591435a4d6d0

Brynjolfsson E, Li D, Raymond LR (2023) Generative AI at Work. National Bureau of Economic Research Working Paper 31161, DOI 10.3386/w31161.

Burgess M (2023) ChatGPT Has a Big Privacy Problem. Wired, April 2023. https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/

Cihon P, Demirer M, Kalliamvakou E and Peng S (2023) The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. https://arxiv.org/abs/2302.06590S

Coelli MB, Borland J (2019) Behind the Headline Number: Why not to Rely on Frey and Osborne's Predictions of Potential Job Loss from Automation. Melbourne Institute Working Paper No. 10/19, October 2019, The University of Melbourne.

CSIRO (2023) Australia's AI Ecosystem Momentum Report. https://www.csiro.au/-/media/D61/AI-Ecosystem-Momentum-Report/23-00010_DATA61_REPORT_NAIC-AustraliasAIEcosystem_WEB_230220.pdf

CSIRO (2023) CSIRO Submission to Supporting Responsible AI Discussion Paper. https://consult.industry.gov.au/supporting-responsible-ai

Czarnitzki D, Fernández GP, Rammer C (2023) Artificial intelligence and firm-level productivity. Journal of Economic Behavior & Organization 211: 188-205. DOI: 10.1016/j.jebo.2023.05.008.

Damioli G, Van Roy V, Vertesy D (2011) The impact of artificial intelligence on labor productivity. Eurasian Bus Rev 11, 1–25. DOI: 10.1007/s40821-020-00172-8.

De La Pena N, Granados O (2023) Artificial intelligence solutions to reduce information asymmetry for Colombian cocoa small-scale farmers. Information Processing in Agriculture, 2023. https://doi.org/10.1016/j.inpa.2023.03.001

Deloitte (2022) On high alert: The darker side of generative AI, https://www2.deloitte.com/content/dam/Deloitte/xe/Documents/About-Deloitte/mepovdocuments/ME-PoV-issue-41/darker-side-of-generative-ai_mepov41.pdf

Dell'Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, Krayer L, Candelon F and Lakhani K (2023) Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality (September 15, 2023). Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, DOI: 10.2139/ssrn.4573321.

Eastwood B (2023) Study: Industry now dominates AI research. https://mitsloan.mit.edu/ideas-made-to-matter/study-industry-now-dominates-ai-research

Eliot L (2022) AI Ethics Wary About Worsening Of AI Asymmetry Amid Humans Getting The Short End Of The Stick. Forbes, August 2022. https://www.forbes.com/sites/lanceeliot/2022/08/19/ai-ethics-wary-about-worsening-of-ai-asymmetry-amid-humans-getting-the-short-end-of-the-stick

European Parliamentary Research Service (EPRS) (2020) The impact of the General Data Protection Regulation (GDPR) on artificial intelligence.

Frey CB, Osborne MA (2013) The Future of Employment: How Susceptible are Jobs to Computerisation? The University of Oxford.

FT.com (2023) Nvidia's AI software tricked into leaking data. https://www.ft.com/content/5aceb7a6-9d5a-4f1f-af3d-1ef0129b0934

Google Research Lab (2023), Announcing the first Machine Unlearning Challenge. https://blog.research.google/2023/06/announcing-first-machine-unlearning.html

Hajkowicz S, Evans D, Cameron L, Trinh K, Chen H, Bratanova A, Mason C, Pham H, Travis S, Black R (2023) The Geography of Australia's Digital Industries: Digital technology industry clusters in Australia's capital cities and regions. CSIRO and the Technology Council of Australia, Canberra.

Hajkowicz S, Naughtin C, Sanderson C, Schleiger E, Karimi S, Bratanova A, Bednarz T (2022) Artificial intelligence for science – Adoption trends and future development pathways. CSIRO Data61, Brisbane, Australia.

Hajkowicz S, Sanderson C, Karimi S, Bratanova A, Naughtin C (2023) Artificial intelligence adoption in the physical sciences, natural sciences, life sciences, social sciences and the arts and humanities: A bibliometric analysis of research publications from 1960-2021. Technology in Society 74, DOI: 10.1016/j.techsoc.2023.102260.

Hajkowicz S, Whittle J (2023) Yes, AI could help us fix the productivity slump – but it can't fix everything. The Conversation, 22 June. Hired (2023) The Tech Hiring Tightrope: Balancing the Skills & Shifts in Talent – Hired's 2023 State of Tech Salaries report (www.hired.com.au).

Hopkins J (2023) 10 Australian companies have embraced the 4-day week. Here's what they say about it. The Conversation, 5 June.

IARPA (2019) TROJAI Trojan in Artificial Intelligence. https://www.iarpa.gov/research-programs/trojai

Investment NSW (2023) Artificial Intelligence and Data Services. https://www.investment.nsw.gov.au/priority-sectors/technology/artificial-intelligence/

Investment NSW (2017) Industry structure. https://www.business.nsw.gov.au/industry-sectors/why-sydney-and-nsw/industry-structure#:~:text=The%20state%20dominates%20the%20nation's,arts%20and%20recreation%20services%20industries.

Kalliamvakou E (2023) Research: quantifying GitHub Copilot's impact on developer productivity and happiness. https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/

Leswing K, Vanian J, (2023) ChatGPT and generative AI are booming, but the costs can be extraordinary, April 2023. https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html#:~:text=Training%20models&text=Nvidia-,Analysts%20and%20technologists%20estimate%20that%20the%20critical%20process%20of%20training,cost%20more%20than%20%244%20million.

McKinsey (2023) The economic potential of generative AI : The next productivity frontier (14 June Report), McKinsey (www.mckinsey.com), New York.

McKinsey (2023) The economic potential of generative AI: The next productivity frontier, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction

Microsoft and the Tech Council of Australia (2023) Australia's Generative AI opportunity. Technology Council of Australia, Canberra.

Mills M.C., Rahal C (2020) The GWAS Diversity Monitor tracks diversity by disease in real time. Nat Genet 52, 242–243. https://doi.org/10.1038/s41588-020-0580-y

Mok A (2023) Amazon, Apple, and 12 other major companies that have restricted employees from using ChatGPT, Business Insider, July 2023. https://www.businessinsider.com/chatgpt-companies-issued-bans-restrictions-openai-ai-amazon-apple-2023-7

Murshed M.G.S, Murphy C, Hou D, Khan N, Ananthanarayanan G and Hussain F (2021) Machine Learning at the Network Edge: A Survey. ACM Computing Surveys, 54(8), pp.1-37. https://doi.org/10.1145/3469029.

NSW Government Business (2023). www.business.nsw.gov.au

OVIC (2019) Disclosure of myki travel information. Office of the Victorian Information Commissioner (OVIC), 2019 https://ovic.vic.gov.au/mediarelease/information-commissioner-investigates-breach-of-myki-users-privacy/

Sartor G et al (2020) The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. European Parliamentary Research Service. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530

Shum B et al (2022) The inequity of targeted cystic fibrosis reproductive carrier screening tests in Australia. Prenat Diagn. 2023 Jan;43(1):109-116. doi: 10.1002/pd.6285. Epub 2022 Dec 15. PMID: 36484552.

Silvera D et al (2022) Framework for Artificial Intelligence enabled Assistive Technology as Supports under the National Disability Insurance Scheme: Final Report. https://ndis.gov.au/about-us/research-and-evaluation/market-stewardship-and-employment/markets-and-innovations-research

Stela S (2023) NAIC Listening Tour Learnings. https://www.csiro.au/en/work-with-us/industries/technology/National-AI-Centre/Listening-tour

Stypińska J, Franke A (2023) AI revolution in healthcare and medicine and the (re-)emergence of inequalities and disadvantages for ageing population. Front Sociol. 2023 Jan 23;7:1038854. doi: 10.3389/fsoc.2022.1038854. PMID: 36755564; PMCID: PMC9899925.

Trend A, Zheng D, Rakotoarivelo T (2023) Your right to be forgotten in the age of AI. https://www.csiro.au/en/news/All/Articles/2023/September/your-right-to-be-forgotten-AI

Trend Micro (2023) OpenAI Confirms ChatGPT Data Breach. https://news.trendmicro.com/2023/05/13/openai-chatgpt-data-breach/

Vincent, J (2023) Apple restricts employees from using ChatGPT over fear of data leaks. https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak

Walter M (2018) 'The Voice of Indigenous Data: Beyond the Markers of Disadvantage' First Things First, Griffith Review, (60).

Zhao W.X, Zhou K, Li J, Tang T, Wang X, Hou Y, Min Y, Zhang B, Zhang J, Dong Z and Du Y (2023) A survey of large language models. arXiv preprint arXiv:2303.18223.

Zheng D, Finckenberg-Broman P, Hoang T, Pan S, Xing Z, Staples M and Xu X (2023) Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions. 2307.03941.pdf (arxiv.org)

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

www.csiro.au