



Submission: Parliament of South Australia Select Committee on Artificial Intelligence

This submission is made by the Campaign for AI Safety to the Select Committee on Artificial Intelligence appointed by the Parliament of South Australia. Our submission is largely informational in response to the terms of reference. Please refer to [our submission](#) to the Australian Government's consultation on AI regulation for detailed policy recommendations to mitigate the risks of AI.

The Campaign for AI Safety is a not-for-profit association established in Australia with members in Australia and other countries. We are concerned about the dangers AI poses to people and advocate for a stop on the advancement of certain AI capabilities. We also advocate for regulation that promotes and mandates ethical AI. We are not affiliated with any political group.

Table of contents

The current state of AI development, deployment and application across various sectors, with a particular focus on the economic, social and ethical implications for South Australia.....	3
Economic implications for South Australia.....	3
Social and ethical implications.....	5
Issues surrounding the use of AI in the commission of criminal offences.....	7
Proposed government actions.....	8
The challenges and opportunities of AI in relation to privacy, data security, and the ethical use of AI, including the risk of bias in AI decision making.....	9
Stronger liability rules to protect South Australians.....	9
The potential for South Australia to develop a competitive advantage in AI, including through the development of a strong AI research and development sector, the attraction of AI investment, and the training and retention of AI talent; and.....	11
South Australia has a strong AI research and development sector.....	11
Developing a competitive advantage in AI safety research.....	11
Any other related matter.....	13
Implications of AI for copyright law and livelihoods of creative professionals.....	13
Policy recommendations.....	13

The current state of AI development, deployment and application across various sectors, with a particular focus on the economic, social and ethical implications for South Australia

The current state of AI development is dominated by a handful of large technology corporations (Microsoft, Google, Meta etc) locked in a race¹ to be the first to build and release the most powerful and smartest AI models. The implications for South Australia (and the rest of the world) are uncertain: the technology has been described as having the potential to radically transform the whole of society and the economy for the better and solve wicked problems like climate change. At the same time, leading AI scientists are concerned about ethics, gross misuse and catastrophic uncontrollability of very advanced systems². This uncertainty has led governments around the world, including Australia, to take action to regulate AI in different ways.

Economic implications for South Australia

AI promises significant economic growth and productivity gains...

AI is often viewed as an enabler of productivity and economic growth, at a time where OECD countries including Australia experience the 'productivity paradox' (historically low productivity growth despite technological advancement³). Accenture forecasts that AI could double annual global economic growth rates by 2035 due to increased labour productivity, 'intelligent automation', and the diffusion effects of innovation (i.e. creation of new sectors and new revenue streams)⁴. PwC estimates global GDP may increase by up to 14% by 2030 as a result of AI⁵.

The Productivity Commission's most recent 5-year productivity inquiry report states the productivity gains can be significant, "from robot-assisted warehouses that automate online order fulfilment and reduce accidents, to AI-enabled IoT sensors installed in smart cities that allow real-time optimisation of infrastructure, energy and service use and maintenance notification"⁶. It found that AI uptake in Australia lags behind other OECD countries and businesses in the knowledge services and information media and telecommunications industries were more likely to be using AI⁷. We were not able to obtain a breakdown of this data by state.

...however this could come at the expense of jobs, wealth redistribution and market competition

¹ "[These two AI models claim to be better than ChatGPT. Here's what we know](#)", Sabrina Ortiz, Associate Editor on ZDNET (27 June 2023); "[Google DeepMind CEO Demis Hassabis Says Its Next Algorithm Will Eclipse ChatGPT](#)", Will Knight, WIRED (26 June 2023).

² "[Pause Giant AI Experiments: An Open Letter](#)", Future of Life Institute (published on 22 March 2023, signed by 27,000+ technologists and other individuals); "[The Godfather of AI Quits Google and Warns of Danger Ahead](#)", New York Times (1 May 2023).

³ "[Innovation-productivity paradox: Implications for Regional Policy](#)", Thanos Fragkandreas and Birkbeck, OECD-EC High-Level Expert Workshop series on Productivity Policy for Places (3 and 5 March 2021).

⁴ "[Why Artificial Intelligence is the Future of Growth - Accenture](#)", Mark Purdy and Paul Daugherty, Accenture (2016).

⁵ "[The macroeconomic impact of artificial intelligence](#)", PwC (February 2018).

⁶ "[Volume 4 - 5-year Productivity Inquiry: Australia's data and digital dividend](#)", Australian Government Productivity Commission (7 February 2023): Report no. 100, page iv.

⁷ "[Volume 4 - 5-year Productivity Inquiry: Australia's data and digital dividend](#)", Australian Government Productivity Commission (7 February 2023): Report no. 100, page 13.

The promise of significant economic benefits for South Australians could come at the expense of impacts on the labour market and redistributive effects of AI uptake. An analysis of US wage bills has found a sharp slowdown in wage bill growth in the most recent 30 years than the previous four decades due to an acceleration of automation and a deceleration in the creation of new tasks⁸. Analysis commissioned by the UK Government is inconclusive about the net impact of AI on employment⁹. The World Economic Forum's May 2023 Future of Jobs survey globally found respondents expect structural job growth of 69 million jobs and a decline of 83 million jobs. This corresponds to a net decrease of 14 million jobs, or 2% of current employment¹⁰.

South Australian workers who perform tasks that use AI and are hard to replace (inelastic labour supply) will likely benefit more from the adoption of these technologies than those in the opposite scenario (those working in tasks that can be performed by AI). It is likely that the number of South Australians in the favourable scenario are few as they are likely to be very highly skilled, globally sought after workers. It is possible that increased adoption of AI and automation could lead to greater economic inequality (job loss and lower wages for workers and monopoly rents and profits for owners of capital).

Encouraging South Australian businesses to use South Australian-developed technology might keep some of the economic wealth generated from these technologies within the state and country, which indirectly benefits Australian workers and the rest of society.

If the jobs displacement impact is large and laid-off workers are unable to find work in other parts of the economy, the South Australian Government may need to recalibrate its taxation and welfare policies in order to maintain South Australians' standards of living. The costs of this may or may not be greater than the benefits of widespread adoption of AI.

There is also concern that AI uptake could make the Big Tech companies (Google, Microsoft, Meta etc) more powerful than they currently are as they also dominate the AI industry. The Competition and Markets Authority, the UK's regulator for competition and consumer protection, shares this concern¹¹. AI developers such as Anthropic reportedly want to "automate large portions of the economy"¹². There is empirical evidence that AI investment is associated with increased market concentration, and higher AI adoption and larger gains from AI investments for larger companies¹³. AI can exhibit significant economies of scope and be used across multiple markets, this could lead to these companies becoming de facto conglomerates. They could command high prices, lower the quality of goods or services, cause consumer detriment and make entry by new businesses very difficult, reducing competition in Australia and elsewhere. The ACCC's investigation into digital markets has found that Big Tech companies leverage market power into downstream markets using their online platform and giving preferential treatment to third party suppliers¹⁴.

⁸ ["Automation and New Tasks: How Technology Displaces and Reinstates Labor - American Economic Association"](#), Daron Acemoglu and Pascual Restrepo (2019): Journal Of Economic Perspectives, Vol. 33, No. 2, pp. 3-30.

⁹ ["The potential impact of AI on UK employment and the demand for skills"](#), Department for Science, Innovation and Technology and Department for Business, Energy & Industrial Strategy, GOV.UK (8 October 2021).

¹⁰ ["Future of Jobs Report 2023"](#), World Economic Forum (May 2023).

¹¹ ["CMA response to DCMS pro-innovation approach for regulating AI"](#), presented to Parliament by the Secretary of State for Digital, Culture, Media and Sport by Command of Her Majesty (Updated 20 July 2022).

¹² ["Anthropic's \\$5B, 4-year plan to take on OpenAI"](#), Kyle Wiggers, Devin Coldewey, Manish Singh, TechCrunch (7 April 2023).

¹³ ["Artificial Intelligence, Firm Growth, and Product Innovation"](#), Tania Babina, Anastassia Fedyk, Alex Xi He and James Hodson (Updated on 3 February 2023): Journal of Financial Economics (JFE), Forthcoming.

¹⁴ ["Digital Platforms Inquiry – Final Report"](#), Australian Competition and Consumer Commission (26 July 2019) and ["Antitrust: EC opens formal investigation against Amazon"](#), European Commission (17 July 2019).

Social and ethical implications

There is agreement in the plethora of government-issued frameworks and guiding principles¹⁵ that the main ethical challenges of AI are transparency, fairness, explainability, accountability and safety.

Implications for South Australians about AI include bias and discrimination (e.g. political bias in ChatGPT¹⁶), infringements on privacy and individual rights (e.g. bias in risk assessments in criminal sentencing for minority groups¹⁷), physical and mental harms (e.g. use of chatbot by a charity provided harmful advice to people with eating disorders¹⁸ and children using AI-generated “deep fake” pornography to bully peers¹⁹), and dangerous capabilities of AI (e.g. use in cyberattacks²⁰, warfare²¹ and more). An ANU poll conducted in April 2023 found that only 20 per cent of Australians were more excited than concerned about AI²².

These implications have the potential to severely harm South Australians’ lives. We make [detailed recommendations to mitigate these risks](#) in our submission to the Commonwealth’s AI regulation consultation in July 2023.

Proposed actions for ethical use of AI by South Australian government agencies

Government agencies have an important role in ensuring the safe and responsible use of AI in decision making and service delivery. When AI is misused, it can cause irreparable physical and mental damage including lost lives as demonstrated by the Commonwealth’s debt assessment and recovery program which wrongly recovered debt payments using automated decision making²³. We note the Government is leading the nation in trialling a generative AI chatbot to enhance student learning and has stated it will prioritise safety²⁴.

We recommend the South Australian Government increase transparency and accountability in administrative decision-making by:

- notifying an individual where the government uses AI in administrative decision making
- giving individuals the opportunity to request reasons for AI-informed administrative decisions including a technical explanation

¹⁵ For example: “[Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI](#)” secured by the White House in July 2023, [draft EU AI Act](#), [Australia’s AI Ethics Principles](#), Ontario Government’s [Artificial Intelligence \(AI\) Transparency Guidelines and Principles](#).

¹⁶ “[The politics of AI: ChatGPT and political bias](#)”, Jeremy Baum and John Villasenor, Brookings (8 May 2023).

¹⁷ “[Machine Bias](#)”, Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica (23 May 2016).

¹⁸ “[US eating disorder helpline takes down AI chatbot over harmful advice](#)”, Lauren Aratani, The Guardian (1 Jun 2023).

¹⁹ “[AI being used for child sex abuse images in regulation-free zone](#)”, Nick Bonyhady, The Australian Financial Review (15 August 2023).

²⁰ “[WormGPT: New AI Tool Allows Cybercriminals to Launch Sophisticated Cyber Attacks](#)”, The Hacker News (15 Jul 2023): A new generative AI cybercrime tool called WormGPT can launch sophisticated phishing and business email compromise attacks.

²¹ “[A Military Drone With A Mind Of Its Own Was Used In Combat. U.N. Says](#)”, Joe Hernandez (1 June 2021).

²² “[Views of Australians towards science and AI ANU Centre for Social Research and Methods](#)”, Professor Nicholas Biddle (10 July 2023).

²³ “[Report of the Royal Commission into the Robodebt Scheme](#)”, Royal Commission into the Robodebt Scheme (Updated on 11 July 2023): The Royal Commission into the Robodebt Scheme has concluded. Commissioner Catherine Holmes AC SC presented the Report of the Royal Commission into the Robodebt Scheme to the Governor-General, His Excellency, General the Honourable David Hurley AC DSC (Retd) on 7 July 2023. It was tabled on 7 July 2023.

²⁴ “[Nation-leading trial in SA schools to focus on the safe use of AI](#)”, Government of South Australia, Department for Education (5 July 2023).

- creating an independent merits review function for all AI-informed administrative decisions²⁵.

We also propose updating existing ICT acceptable use policies given the advent of generative AI. Some specific changes to the policies should be implemented and can be put to use without delay are:

1. Public servants and contractors **must not enter private or sensitive information into generative AI tools** such as Chat-GPT, Bard, DALL-E, etc. because the information is often transferred overseas and may be used for model training purposes (i.e. be permanently incorporated into AI models under the control of foreign actors).
2. In the writing of policy documents, drafting legislation or other forms of legal writing, public servants and contractors should use AI software with transparent training datasets. This is due to the possibility that the biases²⁶²⁷ in the training data can sway the thinking of the writers as they use autocomplete functionality.
3. Public servants should be made familiar with the pitfalls of existing AI technologies, such as “hallucinations”²⁸.
4. **AI tools that are based on deep learning** (including most generative AI) are non-transparent black boxes and therefore **must not be used for any form of ADM**.
5. Public servants should adhere to *Australia’s AI Ethics Principles*.

We suggest the South Australian Government publish guidance on how to use AI in the public service for government agencies and develop onboarding materials **for new staff and annual refresher training** and the code of ethics training²⁹. The Government should consider sanctions (suspension, reduction in classification, reassignment of duties, termination of employment, etc) for breaches of this guidance.

²⁵ These recommendations are adapted from those in the Australian Human Rights Commission’s [Human Rights and Technology Final Report 2021](#).

²⁶ “[The politics of AI: ChatGPT and political bias](#)”, Jeremy Baum, John Villasenor (8 May 2023).

²⁷ “[Political Bias in Large Language Models](#)”, Lucas Gover (17 May 2023): The Commons: Puget Sound Journal of Politics: Vol. 4: Iss. 1, Article 2.

²⁸ “[Hallucination \(artificial intelligence\)](#)”, Wikipedia (16 July 2023).

²⁹ “[Code of Ethics for the South Australian Public Sector](#)”, Government of South Australia, Office of the Commissioner for Public Sector Employment (Updated 7 October 2021).

Issues surrounding the use of AI in the commission of criminal offences

We refrain from providing comprehensive commentary and suggesting policy recommendations in relation to criminal law, detection and enforcement. We note that issues and research surrounding the use of AI has focussed on ethical and social implications including how to regulate in a civil context but there is comparatively little analysis on controlling AI for criminal use.

It is well demonstrated that AI technology can be reorientated to facilitate criminal acts. We focus on a few examples below but the list seems endless: theft and fraud, forgery and impersonation, sexual offences, trafficking of drugs by autonomous means, harassment etc.

1. Hacking and cyberattacks

Attempts have been made to develop AI capable of hacking, such as WormGPT³⁰, a generative tool that can launch sophisticated phishing and business email compromise attacks. Cybersecurity researchers have demonstrated a variety of potentially malicious use cases. For example, a team at Clarity Ltd, a cybersecurity business, used ChatGPT to win a hacking tournament³¹. In March 2023, EuroPol (the law enforcement agency of the European Union) published explorations of how ChatGPT is able to facilitate a significant number of criminal activities, ranging from helping criminals to stay anonymous to specific crimes including terrorism and child sexual exploitation³².

2. Deception, disinformation and social manipulation

The Alignment Research Center found that GPT-4 could pretend to be a blind person to hire a human via an online job ad to pass the CAPTCHA test so that it could access the internet³³. Graphika, a research company that studies disinformation, has uncovered 'deepfake' video technology and AI-generated images in pro-China campaigns disseminated through social media to influence and manipulate people's views³⁴.

3. Bioterrorism

A MIT experiment found that non-experts could use large language models (LLMs such as ChatGPT and Google's Bard) to identify, acquire and release viruses that could cause widespread harm. In one hour, MIT non-scientist students were given detailed instructions on how to engineer four potential pandemic pathogens and potential mistakes to avoid, and the chatbots named suppliers that were unlikely to verify orders³⁵.

³⁰ "[WormGPT: What to know about ChatGPT's malicious cousin](#)", Charlie Osborne, ZDNET (20 July 2023).

³¹ "[ChatGPT Helped Win a Hackathon](#)", Kim S. Nash, WSJ PRO (20 March 2023).

³² "[ChatGPT: The impact of Large Language Models on Law Enforcement](#)", European Union Agency for Law Enforcement Cooperation, Tech Watch Flash (Updated on 17 April 2023).

³³ "[Update on ARC's recent eval efforts](#)", ARC Evals (Updated on 18 March 2023).

³⁴ "[Deepfake It Till You Make It](#)", GRAPHIKA (7 February 2023).

³⁵ "[Chatbots allow people with no lab training to create pandemic viruses, study finds](#)", Matthias Bastian, The Decoder by DEEP CONTENT (18 June 2023).

The scientists who created Chemcrow, a GPT-4 powered tool that can execute common chemical tasks across areas such as drug and materials design and synthesis, acknowledge it can be repurposed for harmful applications, such as designing chemical weapons³⁶.

In 2022, researchers tweaked an existing biochemical research AI product³⁷ to reward toxicity: it produced molecules that could be deadlier than existing biochemical weapons.

We are also very concerned by the potential for dangerous and emerging capabilities in very advanced AI systems³⁸. Examples include:

- Weapons acquisition
- Long-horizon planning
- Situational awareness
- Persuasion and manipulation
- Self-proliferation.

Proposed government actions

A combination of arresting the research and development of more powerful AI, prohibiting high risk AI that exhibits signs of dangerous capabilities and prohibiting high-risk applications (e.g. similar to the ones listed in the draft *EU AI Act*) can help mitigate the risks of AI including their use in criminal offences. We make [detailed policy recommendations](#)³⁹ on how to achieve this in our submission to the Commonwealth's AI regulation consultation. There may be scope for the South Australian Government to adopt some of these recommendations.

³⁶ "[ChemCrow: Augmenting large-language models with chemistry tools](#)", Andres M Bran, Sam Cox, Andrew D White, and Philippe Schwaller (21 Jun 2023).

³⁷ "[Dual use of artificial-intelligence-powered drug discovery](#)", Fabio Urbina, Filippa Lentzos, Cédric Invernizzi and Sean Ekins (7 March 2022): Nature Machine Intelligence volume 4, pages 189–191.

³⁸ "[Model evaluation for extreme risks](#)", Toby Shevlane, et al. (24 May 2023).

³⁹ See our response to Question 14 in our submission.

The challenges and opportunities of AI in relation to privacy, data security, and the ethical use of AI, including the risk of bias in AI decision making

We refer to the Australian Human Rights Commission's *Human Rights and Technology Final Report 2021*⁴⁰ and its submission⁴¹ to the Attorney-General's Department Privacy Act Review for a comprehensive overview of the challenges to individuals' privacy and ethical use of AI.

Stronger liability rules to protect South Australians

AI developers have little incentive to account for third-party impacts, prioritising safety and ensuring accuracy and quality in the intense competition to develop more advanced AI.

Strengthening liability rules would help drive innovation towards more carefully developed and safer AI and make it easier to obtain compensation for damages like breaches of privacy, damages caused by safety issues or, for example, discrimination in a recruitment process involving AI technology.

That means that if the rights of an individual or a customer are infringed, they should be able to sue not just the immediate party that employed the AI system (for example, a small business using AI-powered recruiting software that has a bias against minorities), but also the provider of the AI system (e.g. the recruitment software provider or the API provider such as Amazon or Microsoft) as well as the AI lab that trained and released the AI system (e.g. Anthropic or OpenAI). There exists strict rules in Australia for manufacturers for defective products and the same principles should apply to AI products.

Shine Lawyers⁴² propose the following for law reform:

- In line with current EU proposals,⁴³ all AI developers should have a presumed duty to its end-users and non-contracting third parties for the harms their products have caused:
 - This is a rebuttable duty that applies only to significant harm.
 - This affords clarity to the general public (i.e. non-contracting third parties) on the legal recourse available to them in the event a software or AI product causes them harm.
 - The liability should include the injury of pure mental harm (e.g. embarrassment, stress) and pure economic loss.
- In line with recent changes to Australia's Unfair Contract Terms (UCT) laws in the *Competition and Consumer Act 2010* and *Australian Securities and Investments*

⁴⁰ "[Human Rights and Technology Final Report 2021](#)", Australian Human Rights Commission (2021).

⁴¹ "[Safeguarding the right to privacy in Australia](#)", Australian Human Rights Commission (05 April 2023): Submission to Attorney-General's Department Privacy Act Review, Report 2022.

⁴² See their submission for in-depth analysis and policy recommendations: "[Shine's Submission to the Department of Industry, Science and Resources - Campaign for AI Safety](#)", Atanaan Ilango and Dr Benjamin Koh, Shine's Class Actions practice (25 July 2023).

⁴³ "[Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence \(AI Liability Directive\)](#)", The European Commission (2023).

Commission Act 2001, AI-related laws should specifically state that any terms within user-agreements for AI-related software that exclude liability or prevent an individual's right to participate in class actions are to be deemed UCT and voided.

Further implementation details can be found in [Shine Lawyers' submission](#) to the Commonwealth consultation on AI regulation. We believe there is scope for the South Australian Government to enact some of Shine Lawyer's proposals (for example through state fair trading laws) to better protect its citizens.

The potential for South Australia to develop a competitive advantage in AI, including through the development of a strong AI research and development sector, the attraction of AI investment, and the training and retention of AI talent; and

South Australia has a strong AI research and development sector

South Australia has established a strong AI research and development sector especially in its higher education institutions. In 2021, the University of Adelaide's Australian Institute for Machine Learning ranked second globally for computer vision research published over the period 2016-2021. We feel further government support is not needed to grow the AI sector and attract investment. But it may help mitigate the risks of the new technology.

Globally, the AI industry, which is dominated by large foreign Big Tech companies, is well-marketed, having benefited from billions of dollars in government and private equity funding including funding from the Commonwealth Government.

Developing a competitive advantage in AI safety research

The South Australian Government could **provide assistance to develop a competitive advantage in technical AI safety research**, a nascent field that is attempting to solve the alignment (ensure 'values' of increasingly powerful AI are aligned with human values⁴⁴) and control problems of AI. Building knowledge in this field will have **positive spillover effects on the development of safe and responsible AI** (e.g. lower costs for new market entrants). There is currently little private incentive for the AI developers to prioritise responsible development and safety testing in the race to develop more advanced AI systems.

Government support could go towards:

- National standards institutes to work on means of quantitative assessment of AI capabilities and safety⁴⁵; and
- Free up any existing research grants earmarked for AI capability research, computer science or fundamental science so they can be used in AI safety research (or establish new grants for AI safety research).

Universities in South Australia should be required to perform safety evaluations⁴⁶ as part of grant conditions for any AI research and development research to help ensure no new risks are introduced from deployment, and not to open-source (i.e. make available publicly) dangerous new models.

Additionally, if the South Australian Government were considering support to help South Australian businesses adopt AI, we suggest the funding be directed towards:

⁴⁴ "[What is the AI alignment problem and how can it be solved?](#)", Edd Gent, New Scientist (10 May 2023).

⁴⁵ "[Strengthening U.S. AI Innovation Through an Ambitious Investment in NIST](#)", Anthropic (April 2023).

⁴⁶ "[Model evaluation for extreme risks](#)", Toby Shevlane, et al. (24 May 2023).

- Cybersecurity suppliers to review implementations of AI systems, such as compliance with emerging standards like OWASP Top 10 for Large Language Model Applications⁴⁷.
- AI ethics training businesses that can conduct community workshops for small businesses to teach best practices in compliance with the new AI regulations and principles, data protection, and related topics.
- AI ethics consulting businesses to review compliance with any new regulations and help businesses adjust to these requirements.

The funding should only go to local South Australian companies (locally owned, with local employees). It should not be spent on credits for the use of AI tools and APIs. It should preferentially be given to support small businesses in higher-risk industries (e.g. local clinics that handle patients' health records and may begin to use AI in diagnosis or management of chronic conditions).

⁴⁷ ["OWASP Top 10 for Large Language Model Applications"](#), The OWASP Foundation (2023).

Any other related matter

Implications of AI for copyright law and livelihoods of creative professionals

We would like to bring to the Select Committee's attention the issue of copyright. Copyright law is a Commonwealth responsibility but generative AI has a profound impact on the livelihoods of Australian creative professionals and publishers.

We view respect for copyright as a crucial element of avoiding AI risks in general and promoting a culture of responsible AI. Use of copyrighted materials, privacy-protected data, and confidential information in training AI models without consent is a **violation of intellectual property rights in many jurisdictions**. Copyright matters for AI safety because it can help slow down competitive race dynamics in the AI industry and making available information about the training data used in AI models will shed more light on “emergent abilities”⁴⁸ of large language models, including abstract thinking and dangerous capabilities (such as ability to suggest dangerous chemical compounds) and whether they are truly capabilities or regurgitation of training data.

We request that the South Australian Parliament, as an urgent priority, **work together with other parliaments in Australia to enact new legislation clarifying copyright in relation to training data**, ensure that Australia works on similar clarifications in its international agreements with other countries, and that mechanisms of negotiation and distribution of compensation be established in Australian jurisdictions.

Policy recommendations

Generative AI arrived largely unanticipated by the public, creative professionals, publishers and relevant stakeholders. The vast majority of creators never imagined that their works would be used for machine learning. Copyright law never considered this novel use and the ease and speed with which content could be generated with AI. This has created a climate of uncertainty for NSW creative professionals as well as businesses using AI-generated content.

Therefore, we recommend that Australian parliaments clarify and reinforce the copyright regime in relation to training of AI models on copyrighted materials. Specifically:

1. Third parties who wish to use copyrighted materials in training AI models must obtain **specific consent** from the owners of materials.
2. Specific consent implies that third parties **must not be allowed to coerce copyright holders** to provide such consent. For example:
 - a. Consent must not be part of terms and conditions of unrelated services (such as video or gaming distribution platforms). This way social media companies

⁴⁸ “[Emergent Abilities of Large Language Models](#)”, Jason Wei, et al. (Updated on 26 Oct 2022): Ability is defined as “emergent if it is not present in smaller models but is present in larger models”.

will not be able to deny services or features of their platforms to those who do not consent to such use of copyrighted materials.

- b. Specificity of consent means that it needs to be a separate agreement, ideally a contract with consideration, which copyright holders have an opportunity to take time to consider, review, and negotiate. It must not be a checkbox underneath a registration form on a website.
3. **Consent must be granular.** For example, copyright owners should be able to specify if they allow third parties to have technical means to generate works “in the style of” the author of the materials.
4. Copyright holders must have effective means of **negotiating and receiving compensation** for the use of their copyrighted materials (such as opt-in collective bargaining mechanisms).
5. **Data used in training AI models must be fully referenced.** References to the works used in training must be made publicly available.
6. Training models using “synthetic data” generated by other models trained on copyrighted materials must be considered equivalent to training on those copyrighted materials.
7. Legal liability in case of infringement of copyright must apply both to parties that train models and parties that use those models to generate content.
8. “Fair use” provisions in law must not apply to AI model training.
9. Materials generated completely or substantially by AI models must not be copyrightable.

These clarifications should be legislated, without waiting for courts to provide them. Appropriate legislation, if promptly enacted, will give certainty in relation to these important questions to the AI industry, the NSW business community, and copyright holders, including artists, writers, publishers, and the wider public.

We clarify the terms used in the recommendations above:

- **AI models** refer to generative AI models and systems (including constellations of models), including large language models (e.g. OpenAI’s GPT-3), multimodal models (e.g. Google Deepmind Gemini), large diffusion models (e.g. Stable Diffusion).
- **Training** includes fine-tuning.
- **Copyright** includes other substantially similar intellectual property rights.
- **Copyrighted materials** include confidential information and similarly protected content.

Case study 1: Chat-GPT powered by GPT

Chat-GPT is a famous AI-powered chatbot that uses a variety of models in the GPT series (including GPT-3.5 and GPT-4) to generate answers (or “completions”) to user inputs (“prompts”). GPT models were trained on various undisclosed sources of information.

It has been alleged that many of those sources of data are copyrighted or privacy-protected⁴⁹. Thanks to the training data, Chat-GPT is able to reproduce (albeit with some error) portions of copyrighted works and summarise them⁵⁰. Authors face diminished demand for their works when imperfect reproductions and regurgitations are accessible from OpenAI and Microsoft.

Because of the secrecy around the training data used by OpenAI⁵¹, copyright owners might not realise that their rights could be violated.

AI companies like OpenAI stand to unfairly gain from the use of works, whose creators never considered the possibility that they would be used for machine learning, do not consent to such use, are not compensated, and do not enjoy attribution of moral rights.

Case study 2: Google Deepmind's Gemini

Google Deepmind is working on a system called Gemini, which they claim will have a number of advanced capabilities⁵². It has been reported⁵³ that Gemini may be trained using materials uploaded onto Google-owned YouTube platform.

While it is not clear what data is being used for training, if Google indeed is using videos hosted on the YouTube platform, that would be a violation of users' trust and needs to be clarified in legislation as illegal because no specific consent was sought from the users who created and uploaded their videos, and no mechanism of compensation was agreed to or established for such use.

⁴⁹ [“ChatGPT Language Model Litigation”](#), Joseph Savieri Law Firm (Updated on 28 June 2023).

⁵⁰ [“OpenAI's ChatGPT may face a copyright quagmire after 'memorizing' these books”](#), Thomas Claburn, The Register (3 May 2023).

⁵¹ [“Language Models are Few-Shot Learners”](#), Tom B. Brown, et al. (22 July 2020): Only cryptic names for datasets are given in OpenAI's GPT-3 paper, such as “Books1”, “Books2”, etc.

⁵² [“Google DeepMind CEO Demis Hassabis Says Its Next Algorithm Will Eclipse ChatGPT”](#), Will Knight, WIRED (26 June 2023).

⁵³ [“Why YouTube Could Give Google an Edge in AI”](#), Jon Victor, The Information (14 June 2023).