Submission No 49

INQUIRY INTO ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES

Organisation:	Meta
Date Received:	14 March 2024

Meta

Meta's Submission to *NSW Parliamentary Inquiry into AI in NSW*

MARCH 2024

Executive summary

Meta welcomes the opportunity to contribute to the New South Wales (NSW) Parliament Portfolio Committee No. 1 - Premier and Finance's (Committee's) Inquiry into Artificial Intelligence in NSW (Inquiry).

Millions of Australians regularly use Meta's family of apps to share and connect with friends and family, community groups as well as small businesses and creators. We invest in policies, proactive detection technology - including artificial intelligence (AI)-powered tools - and processes and partnerships to work to ensure that our services make a positive contribution in Australia and NSW.

Whilst the public debate with respect to AI is relatively new, the work on developing this transformative technology is not. Just by way of one example, in November 2023, at Meta, we celebrated the ten-year anniversary of Meta's Fundamental AI Research (FAIR). For the past ten years FAIR has produced breakthroughs on many of the hardest problems in AI through open and responsible research – in a broad range of areas including object detection, unsupervised machine translation, and large language models – which in turn have had global, real-world impact.[1] For example, our No Language Left Behind breakthrough - a first of-its-kind AI project that open-sources models capable of delivering evaluated, high-quality translations directly between 200 languages - is helping people to access and share web content and communicate with anyone in their preferred or native languages, including low-resource languages like Asturian, Luganda, Urdu, and more.[2]

Noting the breadth of the Terms of Reference of the Inquiry, to assist the Committee, we wanted to first share some background about how Meta uses AI and our approach focused on transparency, openness and responsible innovation. This includes using AI to help to ensure a safer online environment, provide more personalised online experiences, and support innovation. We then make some general comments on key policy issues relevant to the Committee's exploration of AI. We would be happy to elaborate on these if that would be helpful to the Committee.

At Meta, we believe AI should benefit everyone – not just a handful of companies. AI innovation is inevitable and AI should be built to benefit the whole of society. Meta uses AI in a wide variety of ways as part of our content governance and integrity systems, to

[1] Meta, 'Celebrating 10 years of FAIR: A decade of advancing the state-of-the-art through open research', 30 November 2023, https://ai.meta.com/blog/fair-10-year-anniversary-open-science-meta

[2] Meta, No Language Left Behind: Driving inclusion through the power of AI translation, https://ai.meta.com/research/no-language-left-behind

optimise ads and drive sales for small businesses, and to support innovation, including in the use of large language models for socially useful purposes. Our work is guided by our Responsible Innovation Principles, including our five pillars of responsible AI that we have developed based on principles from the European Union and the OECD: privacy and security, fairness and inclusion, robustness and safety, transparency and control, and accountability and governance.[3]

Since the earliest days of Feed in 2006, Meta has used machine learning and AI to power all of our apps and services - whether it is personalised content feeds, keeping our platforms safe, or showing relevant ads. Use of AI on Meta's services has already been generating significant benefits in Australia for some time, especially among small to medium enterprises (SMEs). A recent study found that 75% of Australian SMEs report that Meta technologies enabled their business to market and sell products and services and 67% of SMEs believe their business is stronger today because of Meta technologies and apps.[4]

AI is central to our integrity systems, which are designed to protect our platform and our users, ensuring a safer experience for them. For example, we use AI to help us detect and address hate speech and other content that violates our policies. This is a big part of the reason why we have been able to cut the prevalence of hate speech on Facebook to just 0.01-0.02% (as of Q3, 2023). In other words, for every 10,000 content views, we estimate just one or two will contain hate speech.[5] As another example, we use AI to provide more age-appropriate experiences on our services.[6]

To promote greater understanding of the use of AI across our products and integrity systems, Meta has for many years invested in significant and industry leading transparency measures. With respect to content and ads ranking, we have in-product transparency tools[7] and explanations about the policies and principles that guide ranking and recommendations algorithms in our Transparency Center and Help Center.[8]

[3] Meta, 'Meta's five pillars of responsible AI that inform our work', https://ai.meta.com/responsible-ai

[4] Thoughtlab, *The Digital Journey of SMEs in Australia*, May 2023, https://thoughtlabgroup.com/the-digital-journey-of-smes-in-australia/

[5] Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', Newsroom, 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/

[6] Tech at Meta Blog, 'How Meta uses AI to better understand people's ages on our platforms', 22 June 2022 https://tech.facebook.com/artificial-intelligence/2022/6/adult-classifier/

[7] See e.g., Meta, 'More control and context in News Feed', Newsroom, 31 March 2021, https://about.fb.com/news/2021/03/more-control-and-context-in-news-feed/; Meta, 'Understand why you're seeing certain ads and how you can adjust your ad experience', 11 July 2019, https://about.fb.com/news/2019/07/understand-why-youre-seeing-ads/

[8] See e.g., Facebook Help Center, 'What are recommendations on Facebook?' https://www.facebook.com/help/1257205004624246/ ; Instagram Help Center, *'Recommendations on Instagram'* https://help.instagram.com/313829416281232/?helpref=uf_share

In addition to the use of “Classic AI” across our product and integrity systems, Meta has been investing in new generative foundation models that are enabling entirely new classes of products and experiences (“Generative AI”).[9] Innovations driven by this technology will provide enormous benefits for people and society. For example, Yale and EPFL’s Lab for Intelligent Global Health Technologies used our latest open source large language model, Llama 2 (released in July 2023)[10], to build Meditron, the world’s best performing open source large language model tailored to the medical field to help guide clinical decision-making. Meta also partnered with New York University on AI research to develop faster MRI scans. And we are partnering with Carnegie Mellon University on a project that is using AI to develop forms of renewable energy storage.[11]

By democratising access, via this open approach, to foundation language models, potential toxicity, bias, bugs and vulnerabilities can be continuously identified and mitigated in a transparent way by an open community. Advancing our efforts towards an open approach for AI has been welcomed by more than 90 global academics, policy makers and technology companies.[12]

As more recent examples of our commitment to transparency and the responsible development of AI:

- In February this year, we announced that we will be labelling AI-generated images that users post to Facebook, Instagram and Threads when we can detect industry standard indicators that they are AI-generated. If we determine that digitally created or altered image, video or audio content creates a particularly high risk of materially deceiving the public on a matter of importance, we may add a more prominent label if appropriate, so people have more information and context.[13] This follows on from research that we shared in October 2023 from our AI Research lab, FAIR, on cutting-edge invisible watermarking technology we are developing called Stable Signature, which is a new method for watermarking images created by open source generative AI.[14] These are early days for the spread of AI-generated

[9] Classic AI is known for being able to analyse large amounts of data which can be used, for example, to classify and label content (e.g., integrity models), or predict what content users will find most relevant or valuable (e.g., ranking and recommender models). Generative AI is differentiated through its ability to create new content using existing text, audio, images, or videos.
[10] Meta AI, Introducing Llama 2, http://ai.meta.com/llama
[11] Meta, ‘On AI, Progress and Vigilance Can Go Hand in Hand’, 19 January 2024, https://about.fb.com/news/2024/01/davos-ai-discussions
[12] Meta Newsroom, *Statement of Support for Meta’s Open Approach to Today’s AI,* June 2023 https://about.fb.com/news/2023/07/llama-2-statement-of-support/
[13] Meta, ‘Labeling AI-Generated Images on Facebook, Instagram and Threads’, Newsroom, 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads
[14] Meta, ‘Stable Signature: A new method for watermarking images created by open source generative AI’, Research, 6 October 2023, https://ai.meta.com/blog/stable-signature-watermarking-generative-ai/

content, and what we learn will inform industry best practices and our own approach going forward.

- In December 2023, we launched Purple Llama, an umbrella project featuring open trust and safety tools and evaluations meant to level the playing field for developers to responsibly and safely deploy generative AI models and experiences in accordance with best practices shared in our Responsible Use Guide.[15] This is a major step towards enabling community collaboration and standardising the development and usage of trust and safety tools for generative AI development.
- In September 2023, we started to roll out new AI features across our apps, such as AI stickers and, in the US, Meta AI in beta, an advanced conversational assistant available on WhatsApp, Messenger, and Instagram, which can give real-time information and generate photorealistic images from users' text prompts in seconds to share with friends. We are rolling out our new AIs slowly and have built in safeguards, such as visible and invisible markers on Meta-AI generated images.[16]

In Australia and internationally, significant discussions are taking place regarding the complexities and nuances of this technology within the context of other reviews of AI, including on appropriate governance frameworks.

Against this background and with respect to the Terms of Reference for the Inquiry, we encourage the Committee to consider the following when exploring the opportunities and challenges of AI in the NSW context:

- Consider how AI regulation can be built upon existing legislation that already impacts AI, without creating tension with existing Federal or State obligations
- Adopt a framework, when assessing Generative AI research models, that breaks out the policy issues that this new technology may present into three areas – research model training data, evaluation of user inputs and model outputs – to allow proportionate identification of potential policy responses at the State versus Federal level
- Use definitions that strike the right balance between precision and flexibility and consistent with international definitions such as that adopted by the OECD Expert Group on AI
- Ensure AI regulation is principle-based and adopts a pro-innovation, risk-based approach, focused on the uses of the technology and not the technology specifically

[15] Meta, 'Announcing Purple Llama: Towards open trust and safety in the new world of generative AI', 7 December 2023, https://ai.meta.com/blog/purple-llama-open-trust-safety-generative-ai/

[16] Meta, 'Introducing New AI Experiences Across Our Family of Apps and Devices', 27 September 2023, https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/; Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', Newsroom, 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads

- Encourage open innovation and competition so that AI benefits everyone – not just a handful of companies - and is built by an AI research community to benefit the whole-of-society
- Design any AI regulation as a product of collaboration amongst multiple stakeholders, including the Commonwealth Government, and benchmarked against many of Australia's regional allies such as Japan, the US and Singapore

We are still at the very early stages of AI technology, and there is an exciting opportunity for policymakers in NSW to work with the global community working on AI to strive towards the innovations that will help to solve our greatest challenges - locally and internationally. With respect to regulation for the frontier of this innovation, the key focus must be to develop regulations that are broad and flexible enough to adapt to future technologies while not overly restrictive to the point of suppressing valuable and beneficial innovations in, and uses of, AI technology.

For this reason, we need collaborative policymaking - including between the Commonwealth and State governments - to ensure an appropriately balanced approach.

As part of considering how to mitigate risks from AI, it is also important to recognise the extent to which AI is already widely deployed within industry especially as part of any discussions about what, if any, new regulatory frameworks may be needed. A review of the existing uses of AI by the industry will assist in identifying the role of AI to comply with regulatory obligations and community expectations in Australia and what adjustments to existing laws may be necessary to address public policy concerns identified from the broader training and deployment of AI systems.

We welcome the opportunity to provide more details about all of these in our submission below.

The responsible use of AI for safety, innovation and economic benefit

Using AI to ensure a safer online environment

We use AI to help ensure a safer online environment for users on our platforms and more broadly. Below are some of the beneficial use cases for AI at Meta.

Combating harmful content and behaviour

Billions of people around the world use Meta's services every day. Hence, detecting and combatting harmful content and behaviour at scale is a significant challenge. AI technology provides opportunities to detect harmful content before people need to see it.

While human review continues to play an important role in relation to reviewing certain types of harmful content, AI will be a more effective approach in many instances. For example, AI can moderate content at a scale beyond what humans can achieve, and it also lessens the need for human reviewers in some instances where we want to avoid humans needing to be exposed to the content (for example, in relation to child sexual abuse material).

In the last five or so years, we have had a strong focus on using AI to help enforce our Community Standards,[17] which are the rules that set out what people can or cannot do on Facebook and Instagram. Our ability to use AI to detect and action harmful content proactively has been improving over time.

Our work to combat hate speech online provides an instructive case study. Hate speech is traditionally one of the most challenging types of online content to proactively detect because it is so context-dependent. Five years ago, the volume of hate speech we removed was lower than other categories of harmful content, which meant a high degree of human reporting, review and assessment as needed. When we first started releasing our transparency report in 2017, we removed 1.8 million pieces of hate speech globally, 25 percent of which was detected proactively via AI. Since then, after very significant

[17] Meta Transparency Center, *Facebook Community Standards,* https://transparency.fb.com/en-gb/policies/community-standards/

investments in AI, our proactive detection of hate speech has increased significantly. In Q3, 2023, we removed 9.6 million pieces of hate speech, 94.8% of which was detected proactively via AI.

We have also significantly cut the prevalence of hate speech content within the last few years (from 0.10 to 0.11 per cent in Q3 2020, down to 0.01-0.02% in Q3, 2023).[18] Prevalence measures the number of views of violating content, divided by the estimated number of total content views on Facebook or Instagram.[19]

We continue to invest in this space, as harmful content continues to evolve - whether through events or by people looking for new ways to evade our systems - and it is crucial for AI systems to evolve alongside it.

This includes working with researchers and experts to try and optimise AI. For example, we have run detection challenges relating to specific types of harmful content like deepfakes[20] and hateful memes.[21]

Our ranking algorithms are also used to reduce the distribution of content that does not violate our Community Standards but is otherwise problematic. This includes clickbait, unoriginal news stories, and posts deemed false by one of the 90 independent fact checking organisations around the world who review content in more than 60 languages. (We outline this in more detail in our discussion of our Content Distribution Guidelines below.)

Promoting age-appropriate experiences online

Protecting our users - particularly young people - is of paramount importance to us in providing our services. Understanding how old someone is underpins these efforts, but it is not an easy task. Finding new and better ways to understand people's ages online is an industry wide challenge. For large-scale companies like Meta, AI is one of the best tools we have to help us tackle these types of challenges at scale.

Over the past decade, in consultation with experts in adolescent development, psychology and mental health, we have developed more than 30 tools and resources to protect young people from harm and create safe, age-appropriate and private

[18] Meta, *Community Standards Enforcement Report,* Transparency Center, https://transparency.fb.com/data/community-standards-enforcement

[19] Meta, *Prevalence*, https://transparency.fb.com/en-gb/policies/improving/prevalence-metric/

[20] Meta AI, 'Creating a dataset and a challenge for deepfakes', *Meta AI blog*, 5 September 2019, https://ai.facebook.com/blog/deepfake-detection-challenge/?utm_source=hp

[21] Meta AI, 'Hateful memes challenge and dataset for research on harmful multimodal content', 12 May 2020, https://ai.facebook.com/blog/hateful-memes-challenge-and-data-set

experiences for teens on our apps.[22] This includes automatically placing all teens into the most restrictive content control settings on Instagram and Facebook and hiding results in Instagram search related to suicide, self-harm and eating disorders.[23]

These controls put a number of default protections in place for those under the age of 16 (or under 18 in certain countries). They also help to empower young people to make the right choices about their experience online, and the information they want to see and share. However, people do not always share their correct age online, and we have seen in practice that misrepresentation of age is a common problem across the industry.

To address this, in June 2022, we shared details about an AI model we have developed to help detect whether someone is a teen or an adult.[24] The job of our adult classifier is to help determine whether someone is an adult (18 and over) or a teen (13–17). The role of our adult classifier is important because, for example, correctly categorising adults is important not only because it allows them to access services and features that are appropriate for them, but also because it helps mitigate risks and child safety issues that could arise on platforms where adults and teens are both present. We do not allow adults to message teens that do not follow them, for example.

Our adult classifier has significantly improved our ability to provide age-appropriate experiences to the people who use our services, but there is room to improve on this work. We are continuously testing new types of signals that might improve our ability to detect whether someone is a teen or adult. Our goal is to expand the use of our AI more widely across Meta technologies and in more countries globally.

Providing more personalised online experiences

There is a surplus of information and content online. Consequently, it can be a major challenge for individuals to easily find the people, information and experiences that are useful, meaningful and enjoyable for them.

For services like Facebook and Instagram, personalisation is at the heart of the experience. People use our services to connect with family and friends they know, to find communities that they would like to be a part of, and to pursue their interests. We are

[22] See, for example, Meta, 'Giving young people a safer, more private experience on Instagram', Newsroom, 27 July 2021, https://about.fb.com/news/2021/07/instagram-safe-and-private-for-young-people/; Meta, 'Protecting Teens and Their Privacy on Facebook and Instagram', Newsroom, 21 November 2022, https://about.fb.com/news/2022/11/protecting-teens-and-their-privacy-on-facebook-and-instagram/; Meta, 'Giving Teens and Parents More Ways to Manage Their Time on Our Apps', Newsroom, 27 June 2023

[23] Meta, 'New Protections to Give Teens More Age-Appropriate Experiences on Our Apps', Newsroom, 9 January 2024, https://about.fb.com/news/2024/01/teen-protections-age-appropriate-experiences-on-our-apps/

[24] Tech at Meta Blog, 'How Meta uses AI to better understand people's ages on our platforms', 22 June 2022, https://tech.facebook.com/artificial-intelligence/2022/6/adult-classifier/

transparent about how we use AI to make recommendations for people or content that our users may want to engage with.

One of the ways that people connect with friends, family and other accounts that they follow is via a “Feed”.

Historically, these feeds showed content in chronological order. However, as more people started using our services, more content was shared and it was impossible for people to see all of the content that was shared, much less the content that they cared about. Instagram, for example, launched in 2010 with a chronological feed but by 2016, people were missing 70 per cent of all their posts in Feed, including almost half of posts from their close connections. So we developed and introduced a Feed that ranked posts based on what people cared about most.[25]

We provide this personalised experience via AI. Our ranking algorithms use thousands of signals to rank posts for each person’s Feed with this goal in mind.[26] As a result, each person’s Feed is highly personalised and specific to them. Our ranking system personalises the content for over a billion people and aims to show each of them content we hope is most valuable to them, every time they come to Facebook or Instagram.

The goal is to make sure people see what they will find most meaningful - not to keep people glued to their smartphone for hours on end.

One way we measure whether something creates long-term value for a person is to ask them. For example, we survey people[27] to ask how meaningful they found an interaction or whether a post was worth their time, so that our system reflects what people enjoy and find meaningful.[28] Then we can take each prediction into account for a person based on what people tell us (via surveys) is worth their time.

However, AI does not just bring benefits in terms of convenience, ease or helping people discover new online content; it also brings significant economic benefits.

[25] See, for example, A Mosseri, ‘Instagram Ranking Explained’, *I*31 May 2023, https://about.instagram.com/blog/announcements/instagram-ranking-explained/

[26] A Lada, M Wang, ‘How does News Feed predict what you want to see?’, *Meta Newsroom,* 26 January 2021, https://about.fb.com/news/2021/01/how-does-news-feed-predict-what-you-want-to-see/

[27] R Sethuraman, ‘Using surveys to make News Feed more personal’, *Meta Newsroom,* 16 May 2019, https://about.fb.com/news/2019/05/more-personalized-experiences/

[28] Meta, How users help shape Facebook, *Meta Newsroom,* 13 July 2018, https://about.fb.com/news/2018/07/how-users-help-shape-facebook/ ; A Gupta, Incorporating more feedback into News Feed ranking, Newsroom, 22 April 2021, https://about.fb.com/news/2021/04/incorporating-more-feedback-into-news-feed-ranking/

Many Australian businesses, especially small businesses benefit from using personalised advertising because it is more efficient and allows them to better reach the right consumer for their business and compete with larger established businesses.

Even just a few years ago, effective advertising was simply not an option for many Australian small businesses: either because it was too expensive (for example, a commercial on free-to-air TV) or too inefficient (for example, newspaper ads which would only be relevant to a subset of a newspaper's readers).

Innovation in advertising (in particular, personalised advertising) has transformed and improved the options available to small businesses for effective advertising.

Firstly, personalised advertising has driven down the cost of advertising overall. According to the Progressive Policy Institute, the share of GDP that is spent on advertising in Australia has dropped 26 per cent from 1991-2000 to 2010-2018. And globally, internet advertising has dropped in price by 42 per cent from 2010 to 2019 (at the same time that other forms of advertising increased in price), due to innovation and advancements in targeting that have made advertising more efficient.[29] These developments are good for advertisers like small businesses and the benefits flow through to consumers, since lower advertising costs means lower prices for the items they buy.

Secondly, it has made advertising much more effective. There is a much greater level of transparency and measurement for advertisers' return on investment when using personalised advertising compared to other forms of advertising.

Personalised advertising has become even more important for Australian small businesses as they recover from the COVID-19 pandemic and associated economic crises. A 2021 report by Deloitte found that 82 per cent of Australian small businesses reported using free, ad-supported Meta apps to help them start their business.[30] It also found that 71 per cent of Australian small businesses that use personalised advertising reported that it is important for the success of their business. Particularly over the past few years, personalised advertising has helped businesses target new customers as they have needed to pivot away from bricks-and-mortar operations during the pandemic, and then pivot back to support the economic recovery.

[29] M Mandel, *The Declining Price of Advertising: Policy Implications,* https://www.progressivepolicy.org/issues/regulatory-reform/the-declining-price-of-advertising-policy-implications-2/

[30] Deloitte, 'Dynamic Markets Report: Australia - unlocking small business innovation and growth through the personalised economy', Meta Australia blog, October 2021, https://australia.fb.com/economic-empowerment/

Consumers also benefit from personalised advertising because they receive advertisements that are more relevant and tailored to their interests. Personalised advertising enables them to discover relevant content (like new brands, new travel destinations or new communities of interest) and find products and services that are more likely to be meaningful and engaging to them.

Further evidence of the benefit of AI-driven advertising is found in research that shows that users prefer personalised advertising to non-targeted advertising: research found that *"the high personalization ad was clearly preferred to the low personalization ad"* by participants in the research, and those users would "*rather share their clicking behaviour and receive behavioural targeted and therefore relevant ads, than random ads"*.[31] The UK Centre for Data Ethics and Innovation described it as: "*[p]eople do not want targeting to be stopped*" and that most people see *"the convenience of online targeting as a desirable feature of using the internet"*.[32]

We provide more detail in the next sections on how we preserve the value that both people and businesses get out of personalised advertising, while respecting privacy and empowering people to control their information online.

Supporting innovation

The AI innovations that companies like Meta invest in will, as with many technological innovations, provide exciting additional benefits for users. Take for example, the diversity and inclusion benefits that will result from our work on projects relating to language translation[33] and preservation[34]. Nearly half the world's population - billions of people - are not able to access online content in their preferred language. Today's machine translation systems are improving rapidly, but they still rely heavily on learning from large amounts of textual data, so they do not generally work well for low-resource languages, i.e., languages that lack training data, and for languages that do not have a standardised

[31] M Walrave, K Poels, M Antheunis, E Van den Broeck and G van Noort, *Like or Dislike? Adolescents Responses to Personalized Social Network Site Advertising*, Journal of Marketing Communications, Vol. 24, No. 6, 2018, pp. 607, 609, available at: https://www.tandfonline.com/doi/abs/10.1080/13527266.2016.1182938?journalCode=rjmc20; see also, NS Sahni, CS Wheeler, and C Pradeep, 'Personalization in Email Marketing: The Role of Noninformative Advertising Content,' Marketing Science, Vol. 37. No. 2, 2018, pp. 241, available at: https://pubsonline.informs.org/doi/10.1287/mksc.2017.1066)

[32] Centre for Data Ethics and Innovation, *Review of online targeting: Final report and recommendations*, February 2020, pp. 6, 48, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/864167/CDEJ7836-Review-of-Online-Targeting-05022020.pdf.

[33] Meta, 'Inside the Lab: Building for the Metaverse with AI', *Newsroom*, 23 February 2022, https://about.fb.com/news/2022/02/inside-the-lab-building-for-the-metaverse-with-ai/

[34] Meta, 'Preserving the World's Language Diversity Through AI', Newsroom, 22 May 2023, https://about.fb.com/news/2023/05/ai-massively-multilingual-speech-technology/

writing system. This problem will be known acutely by First Nations people in Australia and the immediate region.

As one example, in May 2023, as part of our long-term effort to build language and machine translation (MT) tools that will include most of the world's languages, we announced a series of AI models - our Massively Multilingual Speech (MMS) AI research models - that could make it easier for people to access information and use devices in their preferred language. MMS models expand text-to-speech and speech-to-text technology from around 100 languages to more than 1,100 — more than 10 times as many as before — and can also identify more than 4,000 spoken languages, 40 times more than before. There are also many use cases for speech technology that can be used in a person's preferred language and can understand everyone's voice. We are open-sourcing our models and code so that others in the research community can build on our work and help preserve the world's languages and bring the world closer together.[35]

We can also see the benefits of AI that can be quickly adapted to support public policy goals, such as public health. The use of AI-driven forecasting models during the COVID pandemic provides an example. From April 2020, we created and shared high-quality, localised COVID-19 forecasting models using AI technology to help healthcare providers and emergency responders determine how best to plan and allocate their resources in their particular area. This helped researchers, public health experts, and organisations better understand the spread of COVID-19 given the number of coronavirus cases changed quickly in different communities around the world. We also open-sourced the entire stack of COVID-19 forecasting models so that response teams, governments, and researchers could use them to further help their communities.

Finally, last year Meta released Llama 2 – the next generation of our open source large language model.[36] Large language models — natural language processing (NLP) systems with more than 100 billion parameters — have transformed NLP and AI research over the last few years. Trained on a massive and varied volume of text, they show new capabilities to generate creative text, solve basic maths problems, answer reading comprehension questions, and more. Llama 2 is free for research and commercial use.

Meta has put exploratory research, open source, and collaboration with academic and industry partners at the heart of our AI efforts for over a decade. We have seen first-hand

[35] Meta, 'Preserving the World's Language Diversity Through AI', Newsroom, 22 May 2023, https://about.fb.com/news/2023/05/ai-massively-multilingual-speech-technology/

[36] Meta Newsroom, *Meta and Microsoft Introduce the Next Generation of Llama,* July 2023 https://about.fb.com/news/2023/07/llama-2/

how innovation in the open can lead to technologies that benefit more people. Dozens of large language models have already been released and are driving progress by developers and researchers. They are being used by businesses as core ingredients for new generative AI-powered experiences. We have already mentioned its use to create Meditron. As another example, Japanese startup Elyza has developed a Japanese large language model based on Llama 2.[37] We can envisage other use cases for Llama 2 such as credit card companies using it to improve anomaly detection and fraud analysis, medical professionals making more accurate diagnoses by identifying anomalies in medical images, and businesses using it for organisational tasks.

Making AI more transparent and explainable

At Meta, we believe that the people who use our products should have meaningful transparency and control around how data about them is collected and used, and that this should be explained in a way that is understandable. That's why we are:

- Being meaningfully transparent about when and how AI systems are making decisions that impact the people who use our products;
- Informing people about the controls they have over those systems;
- Making these systems are explainable and interpretable; and
- Investing in research, explainability and collaboration

At the user level

Some of the transparency measures and tools that provide people with greater insight and control over their experience include:

- *Why Am I Seeing this post?* - helps users to better understand and more easily control what they see from friends, Pages and Groups in their News Feed. Users are able to tap on posts and ads in News Feed, get context on why they are appearing (such as how their past interactions impact the ranking of posts in their News Feed), and take action to further personalise what they see.[38] This includes the ability to customise their Feed, such as switching between an algorithmically-ranked News Feed and a feed sorted chronologically with the newest posts first.[39]
- *Why Am I seeing this Ad?* - provides users with context on their ads, to help them to understand how factors like basic demographic details, interests and website

[37] Akira Oikawa, 'Generative AI should be 'open and democratized': Meta chief', *Nikkei Asia,*19 October 2023, https://asia.nikkei.com/Editor-s-Picks/Interview/Generative-AI-should-be-open-and-democratized-Meta-chief

[38] Facebook, 'What influences the order of posts in your Facebook Feed', Help Center, https://www.facebook.com/help/520348825116417; Meta, 'Why Am I Seeing This? We Have an Answer for You', Newsroom, 31 March 2019, https://about.fb.com/news/2019/03/why-am-i-seeing-this

[39] Facebook, 'More Control and Context in News Feed', *Newsroom*, https://about.fb.com/news/2021/03/more-control-and-context-in-news-feed/

visits contribute to the ads in their News Feed. We are continually improving our transparency offerings to reflect feedback we receive. In 2023, we updated this tool to provide users with clear information about the machine learning models that help determine the ads they see on Facebook and Instagram Feed.[40]

- *Ad Preferences* - allows users to adjust the ads they see while on Facebook and gives them the ability to update their ad settings to control information we can use to show their ads.[41]
- *Control what you see on Facebook and Instagram* - helps users to learn more about and control what kind of posts they may see on Facebook and Instagram, including who they see posts from.[42]
- *Content recommendation controls* - our content recommendation controls - known as "Sensitive Content Control" on Instagram and "Reduce" on Facebook – make it more difficult for people to come across potentially sensitive content or accounts in places like Search and Explore.[43]

At the system level

As well as providing transparency at the user level, we recognise that there continue to be discussions about the best ways to provide model and systems documentation that enables meaningful transparency around how these systems are trained and operate. Our transparency initiatives at system level include:

- **System Cards.** Within the Transparency Center, we share 25 system cards for Facebook and Instagram that explain how the AI systems in our products work.[44] They give information about how our AI systems rank content, some of the predictions each system makes to determine what content might be most relevant to users, as well as the controls users can use to help customise their experience.

- **Transparency Center.** The Meta Transparency Center provides a one stop-shop that contains details of our policies, enforcement and integrity insights, including in relation to the use of AI to inform ranking of content, our efforts to reduce problematic content and our AI-driven integrity efforts as part of our content

[40] Facebook, 'How does Facebook decide which ads to show me?', Help Center, https://www.facebook.com/help/562973647153813/?helpref=uf_share; Meta, 'Increasing Our Ads Transparency', Newsroom, https://about.fb.com/news/2023/02/increasing-our-ads-transparency, Newsroom, 14 February 2023

[41] Facebook, 'Your Ad preferences and how you can adjust them on Facebook', Help Center, https://about.fb.com/news/2023/02/increasing-our-ads-transparency/

[42] Facebook, 'Control what you see in Feed on Facebook', Help Center, https://www.facebook.com/help/1913802218945435/?helpref=uf_share; Instagram, 'How Instagram Feed Works', Help Center, https://help.instagram.com/1986234648360433

[43] Meta, 'Introducing Sensitive Content Control', Newsroom, 20 July 2021, https://about.fb.com/news/2021/07/introducing-sensitive-content-control; Facebook, 'Manage how content ranks in your Feed using Reduce', Help Center, https://www.facebook.com/help/543114717778091

[44] Meta Resources, *System Cards*, https://ai.meta.com/tools/system-cards/

governance. Specifically, the Center includes an overview of how Artificial intelligence (AI) systems inform the ranking of content for many experiences on Meta's products, such as viewing Facebook Feed, watching reels on Instagram or browsing Facebook Marketplace.[45] We also provide a deeper look at the types of signals and prediction models that we use in our ranking systems to reduce problematic content.[46] And finally, the Transparency Center houses our Community Standards Enforcement Report that provides data on how much harmful content we action, prevalence of harmful content, proactive detection rates as well as appealed and restored content.[47]

- **Technical research.** One of the most significant AI challenges is ensuring that AI can behave in a way that people can easily understand and be able to anticipate how others will respond to their actions. With the most widely used approach — reinforcement learning (RL), where the agents learn mainly from rewards collected during interactions with the environment — the agent typically develops its own unique behaviours and communication protocols. It might arbitrarily make decisions that are unintelligible both to humans and to other agents trained independently. This can make real world-AI collaboration difficult.

  Meta has developed a new, more flexible approach to teaching AI to cooperate and make their actions understandable to people: off-belief learning. Instead of using human labelled data, off-belief learning starts with the quest to search for a "grounded communication," where the goal is to find the most efficient way to communicate without assuming any prior conventions. To help the field of AI, we recently published a paper on our work, open-sourced the code and released a public demo where everyone can play with our model trained using off-belief learning.[48]

  We have also worked with start-ups to “lift all boats” and encourage sharing best practice about AI explainability across the industry. In April 2022, we worked with cross-industry partner Trust, Transparency and Control (TTC) Labs in a series of co-creation workshops with start-ups and the Singaporean data privacy regulator to develop a framework for AI explainability. We published this framework to share

[45] Meta Transparency Center, *Our approach to ranking explained,* June 2023, https://transparency.fb.com/features/explaining-ranking/

[46] Meta Transparency Center, *Our approach to Facebook Feed ranking,* June 2023, https://transparency.fb.com/en-gb/features/ranking-and-content/

[47] Meta Transparency Center, *Community Standards Enforcement Report,* https://transparency.fb.com/data/community-standards-enforcement/

[48] Meta AI, ‘Teaching AI to be more collaborative with humans without learning directly from them’, *Meta AI blog*, 18 April 2022, https://ai.facebook.com/blog/teaching-ai-to-be-more-collaborative-with-humans-without-learning-directly-from-them/.

our collective thinking and help to advance the debate about effective frameworks for explaining AI.[49]

We have also supported independent AI ethics research that takes local traditional knowledge and regionally diverse perspectives into account. In 2020, we invested in eight independent research projects around APAC, with recipients from Monash University and Macquarie University.[50]

Continued research and collaboration with experts can assist in supporting technical work that enables AI to be more explainable and predictable.

Responsible innovation initiatives

We recognise, when working on innovative technologies, it is important to provide confidence that we are building AI in a way that is privacy-preserving and cognisant of how technology can be misused.[51] This is why we have developed five pillars of responsible AI that inform our work, to ensure that AI is designed and used responsibly, which are based on principles from the European Union and the OECD.[52] These are:

- *Privacy and security:* For Meta, the values of safety, privacy, and security are mutually reinforcing. AI can be used to enhance privacy. We are investing in research on privacy-preserving machine learning technology (differential privacy, federated learning, encrypted computation) and teaching people how to use it. By making our models open source, others will be able to advance research in this area too.
- *Fairness and inclusion:* We believe that AI should work well for everyone, which is why we continue to develop and scale tests and tools that aim to minimise potential bias and enable more inclusive and accessible AI. We are continuing our work to create and distribute more diverse datasets that respect privacy and represent a wide range of people and experience, to enable researchers to better evaluate the fairness and robustness of certain types of AI model. For example, we have publicly released Casual Conversations v2, a consent-based dataset for

[49] TTC Labs, *People-centric approaches to algorithmic accountability*, https://www.ttclabs.net/report/people-centric-approaches-to-algorithmic-explainability.

[50] Meta Research, 'Facebook announces award recipients of the ethics in AI research initiative for the Asia-Pacific', *Meta Research blog*, 18 June 2020, https://research.facebook.com/blog/2020/06/facebook-announces-award-recipients-of-the-ethics-in-ai-research-initiative-for-the-asia-pacific/.

[51] Meta, 'Privacy Matters: Meta's Generative AI Features', *Newsroom*, https://about.fb.com/news/2023/09/privacy-matters-metas-generative-ai-features/

[52] Meta AI, 'Facebook's five pillars of responsible AIA', *Meta AI blog*, 22 June 2021, https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/.

evaluating trained models in computer vision and audio applications by measuring their accuracy across a diverse set of ages, genders, languages/dialects, physical attributes, voice timbres, skin tones, and more.[53] Additionally, as part of our Massively Multilingual Speech (MMS) project, we have released models[54] for speech-to-text, text-to-speech, and more for 1,100+ languages. Others are now able to build on those models, improving inclusivity.

- *Robustness and safety:* As part of our commitment to building AI responsibly, we have adopted an open source approach with respect to our large language models that promotes transparency and access. Open sourcing can lead to safer products through an open community that can iteratively improve them.

  With respect to safety and privacy for our models available for businesses and developers, our Llama 2 research paper outlines Meta's approach to safety and privacy, including analysing for bias and adopting privacy protections in pre-training data, conducting model evaluations against industry safety benchmarks (e.g. truthfulness, toxicity), and working with external partners to red team our Generative AI models to test the robustness of our AI-powered integrity systems against threats.[55]

  We also have a number of on-platform Generative AI safety features, such as input and output filters - which scan inputted and created content to detect potential violations of our policies - and continue to develop new software tools for testing and improving robustness which we share with the AI research and engineering community, such as Purple Llama.[56]
- *Transparency and control:* We continue to prioritise providing greater transparency to users through measures that explain how our AI-powered products work and enable users to understand when they are engaging with AI-generated content. As well as open sourcing our models, we often provide accompanying model cards and weights, which aids transparency and reproducibility, and recently announced the labelling of AI-generated images that users post to Facebook, Instagram and Threads when we can detect industry standard indicators that they are AI-generated.[57]
- *Accountability and governance:* We have invested in our Privacy Review efforts, developed approaches and tools to improve our understanding and ability to

[53] Meta Research, 'Introducing Casual Conversations v2: A more inclusive dataset to measure fairness', 9 March 2023, https://ai.meta.com/blog/casual-conversations-v2-dataset-measure-fairness/

[54] Meta AI, Introducing speech-to-text, text-to-speech, and more for 1,100+ languages https://ai.facebook.com/blog/multilingual-model-speech-recognition

[55] Hugo Touvron, *et al*, 'Llama 2: Open Foundation and Fine-Tuned Chat Models', 18 July 2023, https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models

[56] Meta, 'Announcing Purple Llama: Towards open trust and safety in the new world of generative AI', 7 December 2023, https://ai.meta.com/blog/purple-llama-open-trust-safety-generative-ai

[57] Meta, 'Labeling AI-Generated Images on Facebook, Instagram and Threads', Newsroom, 6 February 2024, https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads

address concerns about our AI systems, and increased transparency and control around our AI products and features.

To this end, we partner with industry and government organisations. Most recently, in December 2023, Meta and IBM launched the AI Alliance, an international community of leading organisations industry, startup, academia, research and government collaborating together to advance open, safe, and responsible AI.[58]

[58] Meta, 'AI Alliance Launches as an International Community of Leading Technology Developers, Researchers, and Adopters Collaborating Together to Advance Open, Safe, Responsible AI', AI at Meta blog, 4 December 2023, https://ai.meta.com/blog/ai-alliance/

Discussion of key policy issues

Given the breadth of the Terms of Reference, we have included some general comments below on a few of the policy issues we think are relevant in relation to AI, noting that these are not comprehensive.

Existing regulatory frameworks and reviews

Companies developing AI technologies are subject to an extensive set of regulatory requirements in Australia. To name just a few, these include privacy, online safety, intellectual property and many more. In addition, there are already a number of reviews underway at a Federal and international level to ensure these existing regulatory systems continue to be fit for purpose to address any new risks posed by AI, and generative AI in particular. Against this backdrop we would encourage the Inquiry to:

- take note of these existing regulatory frameworks, as well as other ongoing reviews;
- focus its attention on issues of most relevance to NSW which are not likely to be considered already at a Federal or international level; and
- continue to work with the Federal Government to ensure that any NSW proposals are considered in light of and are consistent with Federal laws and proposed law reforms.

Existing regulation of AI in Australia

There is an extensive Federal regulatory framework in Australia applicable to AI technologies. This was recognised by the Department of Industry, Science and Resources in its *Safe and responsible AI in Australia Discussion Paper*, which noted that potential risks of AI are already regulated both by general regulations which apply across industries, and also sector-specific regulations.[59] In that Discussion Paper, in addition to sector specific regulations, the Department identified each of the following as imposing regulatory requirements relevant to mitigating against the potential harms of AI:

- data protection and privacy law
- Australian Consumer Law
- competition law
- copyright law
- corporations law

[59] Department of Industry, Science and Resources, *Safe and responsible AI in Australia Discussion paper* https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf at p 10

- online safety
- discrimination law
- administrative law
- criminal law
- the common law of tort and contract.[60]

These existing frameworks are technology neutral and principles based, and therefore well adapted to address new technologies such as AI. Accordingly, they are effective to mitigate the potential risk of harms that could arise from AI and, in particular, generative AI.

Ongoing reviews of Australia’s regulatory frameworks

Nonetheless, to address any risk that these frameworks are not well-adapted to regulating AI, there are several concurrent reviews being conducted at a Federal level. These reviews will or have already proposed amendments to existing laws, as well as introduction of new Federal laws to regulate AI technologies. A few examples include:

- Governance models for AI have been considered as part of the Department of Industry, Science and Resources’ consultations on *Positioning Australia as a leader in digital economy regulation (automated decision making and AI regulation)*[61] and *Safe and Responsible AI in Australia*.[62] Coming out of this latter review, the Commonwealth Government has proposed a new AI Safety Standard and mandatory guardrails to promote the safe design, development and deployment of AI systems.[63]
- The Australian eSafety Commissioner has specifically considered online safety risks relating to AI in a number of contexts, including the draft industry Standards for Designated Internet Services under the *Online Safety Act 2021* (OSA)*,*[64] draft amendments to the Online Safety (Basic Online Safety Expectations)

[60] Department of Industry, Science and Resources, *Safe and responsible AI in Australia Discussion paper* https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf at p 10

[61] Department of Industry, Science and Resources, ‘Positioning Australia as a leader in digital economy regulation (automated decision making and AI regulation): issues paper’, 18 March 2022, https://consult.industry.gov.au/automated-decision-making-ai-regulation-issues-paper

[62] Department of Industry, Science and Resources, ‘Safe and responsible AI in Australia consultation - Australian Government’s interim response’, 17 January 2024, https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf

[63] The Hon Ed Husic MP, ‘Action to help ensure AI is safe and responsible’, 17 January 2024, https://www.minister.industry.gov.au/ministers/husic/media-releases/action-help-ensure-ai-safe-and-responsible

[64] eSafety Commissioner, *Industry standards – public consultation*, https://www.esafety.gov.au/industry/codes/standards-consultation

Determination 2022,[65] and are scheduled for further consideration as part of the review of the OSA.

- The Commonwealth Government has agreed with proposals for amendments to the Australian Privacy Act designed specifically to address automated-decision making, and the Attorney-General's Department has committed specifically to considering the use of AI as it progresses its reforms of the Privacy Act more broadly.[66]

Global discussion regarding frameworks for AI regulation

There is also significant action at an international level to consider the best ways to regulate AI. To give a short recap – globally, in late 2023, the US Government released an *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*,[67] the Group of Seven (G7) Leaders released a *Statement on the Hiroshima AI Process* in which they, among other things, instructed relevant ministers to accelerate the process toward developing the *Hiroshima AI Process Comprehensive Policy Framework*,[68] and the UK Government hosted the AI Safety Summit resulting in the *Bletchley Declaration* to which Australia is a signatory.[69] These complement existing global frameworks, such as the OECD *Principles on Artificial Intelligence* adopted in May 2019 by OECD member countries.[70]

Regulation should be adapted to address specific harms, without undermining the benefits of AI

Australia's peers have also acknowledged that generative AI is a nascent technology at its early stages, and that a rush to regulate without taking the necessary time to determine whether there are net new, clear and actionable harms and the most effective way to

[65] Department of Infrastructure, Transport, Regional Development, Communications and the Arts, *Online Safety (Basic Online Safety Expectations) Amendment Determination 2023*, https://www.infrastructure.gov.au/have-your-say/online-safety-basic-online-safety-expectations-amendment-determination-2023

[66] Attorney General's Department, 'Government Response - Privacy Act Review Report' https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.PDF at 2 and 11.

[67] US National Archives Federal Register, 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence', Executive Order 14110, 88 FR 75191, 30 October 2023, https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

[68] G7, 'G7 Leaders' Statement on the Hiroshima AI Process', 30 October 2023, https://g7g20-documents.org/database/document/2023-g7-japan-leaders-leaders-language-g7-leaders-statement-on-the-hiroshima-ai-process

[69] UK Government, 'The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023', https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[70] *OECD, 'Artificial Intelligence: OECD Principles',* https://www.oecd.org/digital/artificial-intelligence

address those, will risk dramatically curtailing innovation in this area, and the benefits that it can bring to individuals, societies, businesses, and governments. As noted by Export Finance Australia, Goldman Sachs suggest AI adoption could lift global productivity growth by 1.5 percentage points over a 10-year period and drive a 7% (or US$7 trillion) increase in global GDP[71].

This was also recognised by the Commonwealth Government's response to the Department of Industry, Science and Resources' *Safe and Responsible AI in Australia Discussion Paper*. The Government noted:

> The potential for AI systems and applications to help improve wellbeing, quality of life and grow our economy is well known. It's been estimated that adopting AI and automation could add an additional $170 billion to $600 billion a year to Australia's GDP by 2030.[72]

We share the concerns of Australian governments and those of policy makers internationally that it is important that all technology, but particularly technology such as AI, is built and deployed responsibly. This is why we have developed five pillars of responsible AI that inform our work, to ensure that AI is designed and used responsibly,[73] as outlined in our submission above.

Principles for AI regulation

Having noted that the Inquiry is taking place amidst significant Australian and international developments in relation to AI regulation, we thought it may be helpful to set out below some general suggestions about how we think AI regulation should be approached.

The fundamental challenge is to develop regulations that are broad and flexible enough to adapt to future technologies while not overly restrictive to the point of suppressing valuable and beneficial innovations in, and uses of, AI technology. For this reason, we need collaborative policymaking to ensure an appropriately balanced approach. As we informed the Commonwealth Government in our submission to its consultation on *Supporting responsible AI*, Meta stands ready to collaborate with Australian policymakers on these important issues.[74]

[71] Export Finance Australia, *Australia—AI adoption creates benefits and challenges for businesses*, https://www.exportfinance.gov.au/resources/world-risk-developments/2023/may/australia-ai-adoption-creates-benefits-and-challenges-for-businesses/

[72] Department of Industry, Science and Resources, 'Safe and responsible AI in Australia consultation - Australian Government's interim response', 17 January 2024, https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf, p 4

[73] Meta, *Our commitment to Responsible AI*, https://ai.meta.com/responsible-ai

[74] Meta, Submission 478, Supporting responsible AI: discussion paper, https://consult.industry.gov.au/supporting-responsible-ai/submission/view/478

Many AI systems at issue are profoundly complex. There are unanswered questions about how to simultaneously achieve different policy objectives such as meaningful transparency, upholding privacy, protecting trade secrets, and encouraging innovation.

Rushing to impose onerous data, technical, and transparency legal requirements in the absence of consensus based standards and guidelines risks creating substantial risks for companies and their users.

We encourage the Committee as part of any consideration of possible regulatory responses to AI to emphasise the following principles:

- **Use definitions that strike the right balance between precision and flexibility:** Any legislation should include a definition of AI that is sufficiently flexible to accommodate technical progress, but also precise enough to provide the necessary legal certainty. At the same time, however, a definition should not be too narrowly focused on a detailed and prescriptive description of the underlying technical elements of AI and machine learning because, as this is a dynamic and continuously evolving field, they will soon become outdated. We believe that legal certainty around AI developers' obligations can be achieved while still preserving the flexibility to accommodate changing needs and norms – and the ability to take full advantage of the powerful economic benefits of AI – as the technology evolves.

  We recommend adopting a definition which focuses on AI systems that learn and adapt over time because these are the capabilities that are at the core of AI, that make it different from other software applications, and that raise new and unique governance questions. Specifically, we recommend adopting a definition consistent with the definition proposed by the OECD Expert Group on AI:

  > *An AI system is a machine-based system that is capable of influencing the Environment by making recommendations, predictions or decisions for a given set of objectives. It does so by using machine and/or human-based inputs/data to: i) perceive real and/or virtual environments; ii) abstract such perceptions into models manually or automatically; and iii) use model interpretations to formulate options for outcomes.*[75]

[75] OECD Expert Group on AI, https://oecd.ai/en/ai-principles

- **Review existing regulatory frameworks:** Many of the policy concerns raised in the context of AI are already addressed by existing regulatory frameworks, particularly at the federal level. A more detailed review of existing Commonwealth and State regulatory frameworks may be helpful in assessing the extent to which they are fit-for-purpose already, are currently under review at the federal level, or may need to be adjusted at the federal and/or State level. The Inquiry may consider it appropriate to focus on State-based issues, such as the use of AI by NSW Government agencies.

- **Take account of existing and proposed Commonwealth legislation and adopt a suitable framework to identify policy concerns:** We suggest that policy concerns be broken out and considered with respect to research model training data, user inputs and model outputs. This, combined with a review of existing obligations noted above, should assist in minimising tension with existing obligations, in the event that new governance frameworks are identified as being necessary. This will also help to provide greater legal clarity, avoid duplicative regulation, and ensure a proportionate approach to any novel issues.

- **Be principle-based:** Rather than codifying inflexible rules, regulators should focus on supporting and building on ongoing efforts to establish best practices in the fields of Responsible AI. Rather than prescriptive technical requirements, AI legislation should provide opportunities for stakeholders to come together to develop and regularly update the standards and best practices for assessing, measuring, and comparing AI systems as they evolve. We refer to the OECD's AI Principles as a strong foundation on which to consider governance models.[76]

- **Take a risk based approach that is both pragmatic and evidence-based:** The development of AI standards and regulations should be underpinned by a risk based approach, focused on the most sensitive types of AI applications and sectors, such as in cases where AI may produce decisions that cause legal or similarly significant effects.

  AI is a fast-evolving field, with new techniques emerging all the time. A risk-based approach is more future-proof than approaches that focus on particular technologies or techniques, which may become obsolete within years. In contrast, the outcomes that risk-based approaches generally seek – prevent and minimise harm, ensure protections, foster innovation – are less likely to change dramatically, even as new technologies emerge.

[76] See OECD AI Policy Observatory, OECD AI Principles overview, https://oecd.ai/en/ai-principles

However, in adopting this approach it is important to carefully calibrate how to identify risk. We encourage the Committee to adopt a risk-based approach that considers the technology in context, and introduces rules in a way that is proportionate to the level of risk a situation presents.[77] This reduces the likelihood that rules are introduced unnecessarily, creating barriers to innovation and adoption of useful, low-risk AI. This type of approach tends to focus on the outcomes that one wants to achieve or prevent – the 'what' – rather than how they are arrived at. This allows companies to develop their own practices, tools, and techniques to meet expectations, in comparison to more prescriptive approaches which can impose rigid processes on business models that are not well-suited to them.

Within that risk based approach, we believe that - except for exceptional, high-risk circumstances - risk assessments should be conducted by the entities (whether private or public) acting as providers for the AI system, and cover both the potential risks as well as the potential benefits of the AI systems being built and deployed. Potential legal requirements on explainability, auditing, transparency disclosures, and data subjects' right to appeal, redress, and object should only be applicable to AI applications that pose a high risk.

- **Encourage open innovation and competition:** AI should benefit everyone - not just a handful of companies. AI innovation is inevitable and it should be built by an AI research community to benefit the whole-of-society. A specific example to help illustrate this open innovation approach is large language models. Large language models are extremely expensive to develop and train. Fostering a flourishing AI research community that enables experts from diverse disciplines to explore, challenge and innovate with cutting-edge technology depends on democratising access to the most sophisticated models, which are mostly developed by industry.

  An open innovation approach increases market contestability by spurring new market competition, creating more innovation and consumer choices. Open innovation can also facilitate new entry by providing a wide range of stakeholders with access to AI models that will allow them to innovate and compete. Open innovation also promotes sustainable economic growth by helping to close any

[77] See International Institute of Privacy Professionals, 'The case of the EU AI Act: Why we need to return to a risk-based approach', 23 March 2023, https://iapp.org/news/a/the-case-of-the-eu-ai-act-why-we-need-to-return-to-a-risk-based-approach/

gap by enabling researchers and SMEs to build on open source models, making new discoveries and building profitable businesses.

In addition, an open approach has safety benefits. With thousands of open source contributors working to make AI systems better, we can more quickly find and mitigate potential risks in systems and improve the tuning to prevent erroneous outputs. The more AI-related risks that are identified by a broad range of stakeholders - researchers, academics, policymakers, developers, other companies - the more solutions the AI community, including tech companies, will be able to find for implementing guardrails to make the technology safer.

The more access given to AI models, the more likely it is that toxicity and bias can be identified and appropriately addressed and mitigated.

- **Be a product of collaboration amongst multiple stakeholders:** the NSW Government should coordinate and collaborate with the Commonwealth and the many experts and stakeholders of the AI ecosystem and devise its regulatory strategy in conjunction with the Commonwealth and other co- or self-regulatory instruments (international AI principles, standards, ethical codes of conduct, NIST AI Risk Management Framework etc).

  Cross-ecosystem collaboration is helpful in developing rules and norms that are globally harmonised. Globally-harmonised frameworks are necessary to ensure consistent standards around the world. Such frameworks will protect people's information wherever it goes and provide predictable rules for businesses - both being essential requirements for the long-term success of the global digital economy. Additionally, it will foster a level playing field for all AI providers operating across borders.

We trust that these insights are helpful to the Committee as it undertakes this Inquiry and we welcome the opportunity to continue to collaborate with governments on delivering the many benefits of AI in Australia, whilst working to mitigate risks and addressing policy concerns.