Cowell Clarke
COMMERCIAL LAWYERS

Generative AI:

Opportunities, Risks, and the Safe and Responsible implementation of the technology in Australia.

Takeaways from the Australia's Chief Scientist's Repaid Response Information Report on "Generative AI".

Australia's Chief Scientist's Rapid Response Information Report on "Generative AI" was released to the public on 1 June 2023. [i] The Report addressed the following questions:

- What are the opportunities and risks of applying large language models (LLMs) and multimodal foundation models (MFMs) learning technologies over the next two, five and ten years?
- What are some examples of strategies that have been put in place internationally by other advanced economies since the launch of models like ChatGPT to address the potential opportunities and impacts of artificial intelligence (**AI**)?

As Generative AIs are released to the public with increasingly powerful capabilities, Cowell Clarke has been monitoring the digital and technology sector for AI's potential uses and applications. The Report notes that the rapid advancements mean the full extent of the opportunities and risks are largely unknown. The Report (and associated consultation paper, which we discuss at the close of this article) is a significant milestone in the Australian Government's response to AIs and will inform thinking on how AI can be used safely and responsibly.

In this article, we consider some of the Report's important takeaways on the first question, namely, the opportunities and risks of applying AI technology.

Introduction

Generative AI describes the type of AI which can generate novel content. For most people, their first exposure to Generative AI has been through programs such as ChatGPT, developed by the American AI research company, OpenAI. Unlike conventional AI, which is largely used for analytical applications, Generative AI can create rich, user consumer-friendly experiences – for example, having human-like conversations with ChatGPT about your favourite genre of music. [ii] However, it is important to acknowledge that Generative AIs, at their core, are highly sophisticated mathematical models which merely predict what should "come next"; they do not understand fact or fiction.[iii]

The seemingly sudden proliferation of Generative AIs can be credited to recent advances in AI architecture and modelling. These advances enable current generation Generative AIs to be far more powerful than their predecessors. For example, ChatGPT-4 was trained on 100 trillion (i.e. 100,000 billion) parameters whereas its predecessor, ChatGPT 3, was trained on a mere 175 billion parameters.[iv]

AI development is hugely expensive, requires an enormous amount of resources, and what the Report describes as a technology "stack". The stack involved in AI development includes numerous sophisticated technologies such as application program interfaces (APIs), machine learning operation management, machine learning acceleration software, supercomputing and cluster based infrastructure. Every part of the "stack" requires its own expertise and technical specialisation and so the companies at the forefront of AI development must have sufficient capabilities and capital to support the "stack". From a risk perspective, each part of the stack is susceptible to intervention and so regulating the stack must be done with both a global approach, but also an individualistic approach to each part.

The Report briefly outlines six stages which form part of the lifecycle of developing Generative AI models. While it is beyond the scope of this article to explore each stage in detail, it is sufficient to note that unique and substantial risks are present at each stage.

Australia's competitiveness in the production of, or fundamental research into, AIs is questioned by the Report which notes that although Australia has capabilities in AI-related sectors (for example, in robotics and computer vision), our fundamental capacity in the development of Generative AIs is relatively weak. There is also currently a concentration of Generative AI resources and developments in a small number of multinational companies (such as Microsoft and Alphabet, Google's parent company) and although these companies are US based, the Report suggests that the concentration "poses potential risks to Australia".

With foundational matters addressed, the Report provides an overview of the opportunities and risks Generative AI has and may come to have, in Australia. While the opportunities and risks will require ongoing consideration, it is clear that there will be implications for the Australian economy now and into the future.

Opportunities

As to opportunities, the Report notes that the potential uses of Generative AIs are difficult to predict over the next two years, let alone ten. As we seem to be at the dawn of a new technology and the future applications remain largely unknown, the Report focuses on "impact spaces", as opposed to specific opportunities. The following opportunities are considered by the Report:

- **Employment opportunities**: While there is a concern Generative AI will restrict future job creation or destroy current jobs, the Report observes that automation "often complements human labour". Generative AI may create new jobs and augment existing ones by enhancing decision making skills.
- **Maximising economic benefit**: For businesses to maximise economic benefits from Generative AI, they will need to integrate the technology and derive new business models, products and services – in addition to enhancing existing productivity and processes.
- **Broad economic applications**: Generative AI has a variety of applications in a number of sectors. For example, AI technologies have been applied in the medical sector to analyse medical images, the engineering sector to optimise engineering design and the legal sector to analyse and generate documentation. There appears to be few sectors (possibly none), where Generative AIs will not have an impact.

Although the awe of the power of Generative AIs has not subsided, there is increasing concern from both the public and experts alike about the technology's risks. The Report notes that heightened concerns of the risks of AI technology can create polarising and unproductive debate. Although consensus on these issues is unlikely, active and informed conversations on the use and application of emerging technologies is paramount.

The Report identifies three categories of risk which will be realised or mitigated depending on the actions of governments, industries, developers and consumers. These categories are:

- Technical system risks, going to the model itself and its data.
- Contextual and social risks, going to the risks to human rights and values and social inequalities.[v]
- Systemic social and economic risks, going to the impact on society as a whole (for example, democratic systems and market dominance or monopolisation of Generative AI providers).

Risks

In a similar nature to opportunities, risks of Generative AI can be difficult to predict. The following risks are considered by the Report:

- **Accuracy, inaccuracies, and biases**: The accuracy and quality of Generative AIs are only as good as the models they are trained on and the models used to "predict" the next output. This can mean that the output can be entirely incorrect or misleading and without an astute awareness for this type of output, known as a "hallucination", end consumers may unknowingly adopt or rely on that output.

It is important to remember that current Generative AIs are predictive. Based on the information patterns on which they are trained, their programming predicts what output should follow a given input. They have no separate ability to assess whether their output is true or accurate.

The Report suggests that future Generative AIs may need to cite its sources. Since the Report was finalised, there has been an instance of a lawyer in the United States who relied on ChatGPT in preparing Court submissions. Unbeknownst to the practitioner, the case citations produced by ChatGPT as authority for his client's position were completely fabricated. Our own testing has shown many instances of ChatGPT citing sources for output that are wrong or even non-existent.

Another factor which can contribute to the (in)accuracies and biases of Generative AIs includes representation bias, which is when training data is drawn from one segment of the population to the exclusion of others (for example, only training an AI model on Western literature or "male" voices). In one example, biased training data sets used for predictive policing can perpetuate existing inequalities in over policed populations.

- **Misinformation**: Generative AIs have the capacity to generate high-quality, cheap and personalised content – and this includes content which might be harmful. An example of harmful content includes deep-fakes (images or videos of people generated by Generative AIs, which are completely fake but are virtually indistinguishable from genuine content).

  The Report gives an example of a scenario where Generative AIs produce unique but high quality submissions for parliamentary inquiries, giving a false indication of public opinion.

  An example which occurred after the Report was finalised, but one which demonstrates the real world impact a generated image can have, is the posting to Twitter of a generated image of an explosion near the Pentagon which caused a dip in the stock market.

- **Human Rights Due Diligence**: Establishing responsibility for an adverse impact from an AI-enabled system is challenging. In some AIs, the inner workings are not known with precision and so it can be difficult, if not impossible, for a human to assess the reliability of the results or seek redress.

Comprehensive and ongoing risk assessments and human rights due diligence may identify risks and allow for the development and implementation of mitigation strategies. One example is a "human in the loop" approach, where a human has a part in the ultimate decision to ensure accountability and fairness. This approach, however, is not appropriate in all deployments of AIs, particularly in circumstances where the benefit of AI is dependent on efficiency at scale.

- **Data privacy, security, and sovereignty**: There is limited information on the provenance of the training data used to train some of the larger Generative AIs. The Report considers it likely that permissions were not obtained for the use of large data sets drawn from the internet and that under Australian privacy law, the use of that data might be questioned.

  A particularly important consideration for more economically developed nations is sovereignty of their data. If a Generative AI comes to be integrated with Australian public services and so, for example, it is trained on Australians' healthcare information, questions as to sovereignty over that information will arise. The Report considers a framework for the sharing and use of data in highly complex networks between public and private entities will be required. Similarly, ethical questions will also arise where patients' medical information is used to train Generative AI without the patients' consent.

- **Computing power – environmental impact and capacity**: Generative AIs require supercomputing like capabilities and this will be a significant barrier to some developing nations entering the space.

  The creation of Generative AIs requires large data sets which need to be stored in cooled data centres. These data centres have significant energy and water consumptions, and so too does computing and processing the data.

  While advancements in algorithms have improved efficiencies, hardware upgrades can create significant e-waste, leading to an increased demand for critical minerals.

The Report highlights that the opportunities, risks and applications of Generative AI in Australia are largely unknown. An era of rapid development has no doubt begun.

Concluding remarks

As Generative AIs are increasingly adopted throughout Australia, a considered regulatory framework will need to be established. As the Report recommends, to ensure the framework balances the opportunities and risks associated with AI technology, ongoing informed discussion is critical. A framework should be proactive and not reactive, capable of keeping up with industry developments, and foster deployment of Generative AIs in a safe and responsible manner – this is no easy feat.

Where to from here?

To build on the observations made in the Report, the Department of Industry, Science and Resources has opened consultation for its discussion paper on "Supporting responsible AI". The paper will consider governance mechanisms to ensure AI is developed and used safely and responsibly in Australia.

To read more about the discussion paper, you can view the Department's consultation hub here. Submissions close on 26 July 2023.

Cowell Clarke continues to monitor this space and will provide further updates on key developments in responsible AI use.

[i] Bell, G., Burgess, J., Thomas, J., and Sadiq, S. (2023, March 24). *Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs).* Australian Council of Learned Academies.

See the Chief Scientist's media release, here.

[ii] It is noted that most end consumers do not encounter the AI models themselves. Rather, end-consumers encounter the services, applications and businesses that deploy generative AIs. For example, consumers of ChatGPT interact with the generative AI through a web interface (i.e., they do not interact *directly* with the AI).

[iii] For example, a text-based generative AI is using a mathematical model to predict what word would come next in a sentence, based on previous words already selected.

[iv] Parameters are controls on a Generative AI which dictate or restrain the types of outputs which can be generated. For example, ChatGPT uses a parameter called "Temperature" which controls the creativity or randomness of the output

[v] An example of societal inequality which may be overlooked due to the rapid advancement of AI technology is a lack of availability for the technology for regional and older Australians. AI technology often require considerable internet bandwidth, power, suitable devices, and a level of computer literacy – these prerequisites are readily not available to everyone, and these segments of Australia's population may be particularly vulnerable.

For further information please contact our team:

Matthew Hawke
Associate Director
D: +61 8 8228 1130
E: mhawke@cowellclarke.com.au

Matthew Del Corso
Lawyer
D: +61 8 8228 1193
E: mdelcorso@cowellclarke.com.au