Summary

Thank you for reviewing this extremely important topic and for requesting public feedback.

AI has already[2], and will in future, deliver enormous benefits to society and to Australia. But, as leaders of the principal AI firms have stated, regulation is badly needed. As much as AI can be used for good, it can potentially be used for ill. Recent developments in technology (primarily social media) have shown that technology will be used for ill by certain individuals, organisations and states, if they can profit from it.

Designing regulation - in this area - that prevents harm without preventing useful innovation is very challenging. Many are saying that regulation is needed. Few are specifying what is needed.

Background

From where I stand, something went wrong with information technology in the middle of the last decade. Around 2016 it felt like the world suddenly shifted on its axis. Unexpected things started to happen, and the world became less civil. Looking back it is clear that modern information technology was responsible. Certain big technology firms were responsible for dividing nations and economic blocs, in the pursuit of corporate profits. It is clear that governments around the world lacked the regulations (and tools) to stop them and that self-regulation will not work. For decades certain individuals have been able to influence public opinion and elections via control of print and broadcast media. But now nations and societies now face a much bigger and more insidious threat – from "big tech". First from social media and second from AI.

Threats

Threats posed to the Australian public by AI and social media (so algorithms and ADM) include

- Disinformation (and misinformation, e.g. the influencing of an election)
- Misrepresentation (e.g. faking an identity)
- Personal data (storage and misuse of)
- Bias (incorrect output because of skewed data)
- Accuracy (invented "facts" or "hallucinations")
- Intellectual property (misappropriation of)
- Military (new forms of AI-driven warfare or false alerts from AI that triggers escalations and conflicts)

It is good that you have included ADM within this discussion. ADMs (and other algorithms) present the same threats – AI just increases these threats hugely.

Remedies

It's important that regulation (formal legal obligation) covers all[1] system-generated (autonomous or ADM) activity, not just AI per se. Remedies include

- Transparency

AI responses should be identified as such.

- Identity
  Trolling must be prevented by blocking user anonymity (e.g. in social media). Systems and non-human artefacts such as bots must not be allowed to impersonate humans (including in social media and emails). Accounts purporting to belong to an organisation (e.g. “Kingston Community Association”) must be verifiable and verified.
- Explanation
  Require explanation and justification for AI search and query results.
- Accuracy
  Identify hallucinations or other incorrect answers. If necessary a probability score may help.
- Military
  Ban military use of autonomous AI. Exception: permit national defence autonomous defence use. E.g. mine-clearing robots should not be regulated or should be minimally regulated.
- Sponsorship
  Identify results which have been influenced by advertising or any other sponsorship.
- Influence Operations
  Mandate mechanisms to identify and prevent state-sponsored interference (including where this is via third parties).
- Copyright
  Copyrighted materials must be protected. AI and other algorithms must not be able to circumvent intellectual property protections.
- Correction
  When an automated decision is made (from AI or another algorithm such as Robodebt) a human-driven appeal method must be provided, to permit correction of incorrect decisions. (There are stories from the US of algorithm-driven probation decisions; because of the data used these decisions may be ill-founded.)

Example

An example of a regulatory challenge is video generation using AI.

A Sydney organisation is working with film studios to overcome issues with dubbing films into other languages. The images from, say, an English language film can be (and are being) altered to synchronise with, say, a Spanish language soundtrack. This is clearly a good thing. It is consensual, it produces a superior product, it creates interesting and useful work for Australians and it helps Australia lead in the AI industry.

However the same technology in the hands of a bad actor can undermine the reputation of individuals, spread fake news and unethically influence elections. Careful drafting of legislation will be necessary to permit the former use and prohibit (and punish) the latter use.

Publishers

Part of the solution (given that misuse may originate overseas) is to hold publishers (e.g. social media companies, messaging services, AI service providers and traditional media companies that are embracing technology) responsible for deliberate fake, misleading or offensive content.

The problems with social media are caused by social media companies not suffering consequences for publishing fake, misleading or offensive content. It seems that only massive penalties and fines will prevent this behaviour.

Examples I'm thinking of include (from the world of social media)

- massacres in Myanmar (Facebook)
- massacres in north-east Africa (Facebook)
- Cambridge Analytica (Facebook)

Conclusion

If you do nothing else, I strongly urge you to consult with the following individuals or entities about the regulation that the world in general, and Australia in particular, needs

- Professor Toby Walsh (UNSW)
- Professor Sandra Wachter (University of Oxford)
- The EU[3]
- Sam Altman (OpenAI CEO)
- Demis Hassabis (DeepMind CEO).

Normally I would not advocate going to the organisation that needs to be regulated. For example there is probably no value in talking to Google or Meta (Facebook) about this. However OpenAI and DeepMind have (to date) navigated their way skilfully and have managed to maintain public trust in their products, which is a remarkable achievement. Trust is crucial not only for the development of AI but also for the continuance of civil society.

As stated in the discussion paper, it is important to build and maintain public trust in AI, in order for Australia to see the full benefits. If the use of AI results in loss of trust in government and society in Australia, as has happened in the United States, then we will see the sort of problem that occurred in Washington on 6th January 2021[6].

Notes

1. E.g. prevent future Robodebt-type issues.

2. E.g. AlphaFold from DeepMind

3. https://www.coe.int/en/web/artificial-intelligence/contact-us

4. Our national security defence activities should probably be excluded (or largely excluded) from technology regulations.

For example we will need to respond as effectively as possible to autonomous drone warfare or to covert development overseas of new chemical and biological warfare agents.

5. Our cyber security defence should probably be totally excluded from regulations.
   We need to be prepared for the worst that an unregulated “bad actor”, including nation states, can throw at us. This may include cyber offence operations.

6. https://en.wikipedia.org/wiki/January_6_United_States_Capitol_attack

I am happy to be contacted, if you would like me to clarify or expand on any points.

The answers I have given to your questions below discuss this topic from an Australian point of view, but as well as national solutions we need international solutions. International coordination in this is very important. Leadership by Australia can help steer the rest of the world in the right direction.