Vietnamese Overseas Initiative for Conscience Empowerment
AUSTRALIA

P.O. BOX 8 DALLAS VIC 3047
AUSTRALIA
ABN 15 625 886 946
Email: contact@voiceaustralia.org

Response by VOICE Australia to Discussion Paper Safe and Responsible AI in Australia

21 June 2023

We thank the Department of Industry, Science and Resources for the above Discussion Paper ("Paper") and the opportunity to respond.

VOICE Australia is a small NGO focussed on civil society and refugee matters. We do not claim expertise in AI, and we provide our inputs as concerned citizens.

Australia should participate in international standards-making (Ref: Question 2 in Paper)

There will be international efforts to standardise certain aspects of AI - for example, to ensure that AI-produced contents are watermarked as such. We believe that the Australian government should play an active part in such international efforts.

Protect consumers of "free" AI systems (Ref: Question 2 in Paper)

Australia's consumer protection law was designed mainly for when money changes hand. The attention-based economy, such as Facebook or TikTok, requires no purchases. We realise that it is quite challenging to incorporate attention-based economy concepts into consumer protection law, but as consumers concerned about the potential upcoming huge impacts of AI, we hope that the government considers this question: *How could Australia's consumer protection law protect consumers for products or services requiring no monetary payments?*

Protect Australians from AI systems originating from China (Ref: Questions 2 and 14 in Paper)

Why should AI systems from China deserve more caution?

First, China is an AI powerhouse. Second, it openly (on domestic-facing media) says that its strategic goal is to change the world order. Third, China's actions match these words. It has a long-standing track record of weaponising many things - not least hostage diplomacy involving Australians. Relevantly, it systematically weaponises software - for decades, it has mobilised its software workers to attack the world, such as stealing industrial secrets. From this to mobilising its

AI companies to meet its strategic goal is only a small step. Given that the military-civil fusion concept is embedded in its law, and that government control of private industry is as strict as the government wants it, the fact that China's AI systems are produced by its private industry does not mean Chinese AI systems are free from government influence. It is the opposite. It means the ruling party can use private industry as a force multiplier in its efforts to change the world order.

We recall that before Citizen Lab made it public, the Zoom conferencing app quietly routed Australian-to-Australian calls into China, thus such calls could be intercepted on its servers. Citizen Lab wrote *"Because Zoom* [claims to, but] *does not implement true end-to-end encryption, they have the theoretical ability to decrypt and monitor Zoom calls"*. Either Zoom did it on Beijing's request, or Zoom did it in anticipation of being requested. Both do not bode well for Chinese AI systems.

Use the risk-based approach with China's AI systems

AI systems can have benefits and risks much bigger than conferencing apps, therefore AI systems from China are in a high-risk category and deserve special attention.

A cautionary tale is TikTok. It was allowed in because Chinese apps were by default allowed in, like apps from any other country. It quickly grew roots in our society, and removing it now would involve trade-offs that would not have been involved if it were not allowed in in the first place.

As China does not allow certain overseas apps, it expects that the world does the same to it.

Self-motivated AI systems

Currently, ChatGPT and the like know many answers but they do not answer until asked. Likewise, self-driving cars do not start until given a command. But this is likely to change. Consider future AI systems that come up with their own internal motivations. Producing the self-motivation capability is probably less difficult than producing the AI capability itself. Imagine a self-driving car that does not like a particular spot because last time it was there a drunk person climbed onto its roof. For the AI system's developers, solving the problem of producing a motivation (wanting to find another spot to park) deals with only the internal states of the system. Solving the problem of dealing with cats crossing roads, and myriad other problems, deals with both internal states and external complexities, therefore the former may be easier. Thus, it is likely that self-motivated AI systems will surface in the near future.

Motivated AI systems will impact society greatly, and it is difficult to predict their behaviors. Even developers of such systems cannot predict what would happen, but some might not care much. Therefore regulation would be necessary.

Australia cannot predict how self-motivated AI systems will arise, but it can have a say. Regulation can say that companies based in Australia planning to develop self-motivated AI systems must first

inform regulators, and AI systems from abroad capable of self-motivation must first demonstrate that their systems are safe before getting regulatory approval.

Definition of AI: Some implicit limitations should be removed (Ref: Question 1 in Paper)

The Paper's definition of AI, *"Artificial intelligence (AI) refers to an engineered system that generates predictive outputs"*, imposes on itself 2 limitations: That it is engineered, and that it generates predictive outputs.

If the word "engineered" implicitly means "engineered by humans", then the definition will be quickly become outdated. Soon, almost certainly there will be AI systems engineered by AI itself. AI is already writing simple code. Some AI systems are already matured by setting up adversarial AIs which duel until they produce improved versions of themselves.

As to predictive output, it might not be long before an AI system does more than just generating predictive outputs. For example, an AI that stops you in the street and asks you survey questions would generate predictive outputs, but it would do more than that. If regulation covers only the predictive-output capabilities of the system then it fails to regulate the other capabilities.

If the "*human*-engineered" and "predictive outputs *(only)*" limitations are removed, would regulation work?

We think the answer is Yes.

Removing the implicit "*human*-engineered" assumption should be simple. After all, consumer protection laws intended for hand-made pottery need few modifications for factory-produced pottery.

Removing the predictive-outputs-*only* limitation requires more thought. If an AI system does more than just generating predictive outputs, but if that additional capability is not known at the time of designing regulation, then how can regulation deal with it?

There are 2 basic options. Option 1 is to explicitly restrict regulation to predictive-outputs capability only, then in future when that other capability is known, review regulation to catch up with it. Option 2 is to remove the predictive-outputs-*only* limitation now. When the new capability materialises, if regulation designed for predictive-outputs capability happens to still be applicable to it, then do nothing. If regulation produces unintended negative effects, then review regulation. We think Option 2 better protects our society.

-End-

P.O. BOX 8 DALLAS VIC 3047
AUSTRALIA
ABN 15 625 886 946
Email: contact@voiceaustralia.org