

Questions on notice from Senator Pocock from the Select Committee on Adopting AI

1. *From your expertise, how hard would it be for news and social media platforms to identify if a piece of content has been generated with AI?*

Several social media platforms have announced various AI 'labels' recently, including [Meta](#), [TikTok](#) and [Youtube](#), so this appears to be a capability available from tech platforms. News organisations will likely find it a lot more of a challenge, with more manual fact-checking required, particularly those with strict editorial standards.

However there are some complications with this approach. Firstly, it's complicated by having no standard criteria for what an 'AI label' represents – does it include content edited by AI, those assisted or inspired by AI (e.g. articles where ChatGPT was used in some way, like transcripts or summaries), or those that are completely and exclusively generated by AI?

It's early days for this feature, but there are already reports of [issues](#).

This is further complicated with any official arrangements AI companies might have with news organisations – like what NewsCorp recently [announced](#) with OpenAI. How would a label account for deeper partnerships like this where the AI application and content mix is done at a deeper level – at the level of training data?

It's also not clear what the 'label' is meant to signal – is it meant to signal a lack of credibility? Even if that AI use has been clearly vetted and properly edited and accounted for? For those with deep AI partnerships, does it signal 'good AI use' vs. 'bad AI use'? The label therefore could be virtually meaningless.

This approach also once again puts the onus of vetting, qualifying and sifting through credible and accurate content in the hands of the public – one of the core issues with misinformation, rather than with the platforms who are producing, facilitating and amplifying this content.

AI-identification and labelling therefore, seems like a band aid approach to a much deeper issue of AI and content in general.

2. *Given the difficulty in detecting deep fakes or AI altered images - should the Parliament look at banning the use of generative AI in election advertising and materials until this problem is resolved?*

It may be worth differentiating between clearly malicious AI content use like deep fakes – which should be banned, and generally AI altered images.

There is likely to be place for AI altered images in the future given how AI will be an enabling technology.

However, it's clear that the current regulation and frameworks specifically around election material is insufficient in tackling AI misinformation and there should be strong restrictions and guidelines around this period.

A good start could be truth in political advertising laws, applying existing restrictions placed on traditional media (e.g. advertising blackout periods in the leadup to) to digital and social media, and implementing stronger, mandatory misinformation laws (currently the only targeted misinformation law is a voluntary code of conduct which has [proven ineffectual](#)).

- 3. There has been talk about requiring watermarking to show if something has been made with AI. What do you think about this approach? Is it enforceable in Australia? Who do you think the responsibility of placing a watermark on content should sit with? The platform that generates the content of the platforms sharing it?*

This is similar to question 1 – watermarking presents numerous complications and contradictions, in terms of its meaning and generation. We need to incorporate and interrogate AI use and develop standardised editorial policies around them industry-wide.

The responsibility for identifying/flagging/labelling/culling AI enhanced or altered images should lie with all the parties involved in their development and distribution, which would be perhaps made easier if there were industry standards around their use. Certainly, the burden shouldn't fall mostly on the public who deserve and expect basic standards of accuracy, impartiality and facts from organisations.