Question 6- Hansard Page 13

CHAIR: Obviously, it is clear to us all that there are economic opportunities for AI but it is also obvious that there are severe and possibly existential risks. I think we can walk and chew gum at the same time by not only encouraging local industry and skills and training but also making sure we are putting strong safeguards in place to protect us from the very real dangers. Of course, that is a developing safeguard, so I'm very interested if there is any additional evidence you want to give us or take on notice the time frames of likely safeguards that could be implemented. If you could take that on notice that would be a great deal of assistance to the committee.

Ms Solar: We can take it on notice and share some of the best practices and thoughts that we have for what can help industry.

CHAIR: And certainly our safeguards regarding the challenges to the broader community. I think there were some questions from senators regarding both protection of democracy and fake news. Obviously, there is criminality, which has also been raised during the hearing. All of those matters are of extreme interest to us and the general public, obviously. The walking and chewing gum mean we want to know how fast you can walk and how quickly you can chew. There is a great need.

Dr Hajkowicz: I will take that on notice.

Answer

Safeguards can be soft/hard law approaches or technical means. For soft/hard laws to come into existence, the timeline is up to the agencies and organisations devising them. For technical means, including those to support hard/soft laws, the implementation timeline depends on the nature of the soft/hard laws, the context of deployment, and the thresholds/level of risk tolerance and organisational maturity. Due to the technical challenges of frontier AI models, technical safeguards are still an active area of research. For instance, real-time safety-critical systems requiring high accuracy, or high-stakes AI detection for large-scale misinformation and AI-generated content, lack effective safeguards as human verification is not scalable for such contexts. CSIRO is actively conducting research in AI safety and guardrails in collaboration with international partners in this area.