Question 5 – Hansard page 13

Senator GHOSH: I'll just put two questions on notice quickly because I'm aware of the time. There was a reference to the 125 foundational models you are aware of. Could I ask you to take on notice what is the proportion of those models that are open-source models? Also do we know the extent of automation that exists within those models? If all of them use automation, then is there a breakdown on the extent of automation in the different types of models?

Dr Hajkowicz: In terms of open source, I would have to take that on notice. I don't know what percentage would be open source. There is probably more than 125 by now; that was six months ago, which is recent in this. The other issue of the extent of automation: these are foundation models that are trained using deep learning techniques.

CHAIR: If you could take it on notice, it might be of assistance to the committee because we are mindful of the time. That would be of assistance to us getting through the day and will give us the opportunity to ask further questions of other witnesses.

Answer:

There have been two prior studies which catalogue global AI foundation model development and give us insights to the share which might be open source.

One of these studies was led by Xavier Amatriain (2024) who leads AI product development at Google. This study identifies 77 foundation models since the year 2017. This paper lists 33 of AI foundation models (43%) as having an open-source licence, with the other models having a commercial licence, unknown or other licence status.

Another study was commissioned by the German Federal Ministry for Economic Affairs and Climate Action and done by the German Academy for Artificial Intelligence (AKI, 2023). This study found 125 foundation models. Whilst not explicitly identifying the number that are open access, it notes that the lack of open access models is a barrier for 58% of companies surveyed.

As for the question of automation, AI foundation models and all types of AI can be used to partially or fully automate increasingly wide-ranging and complex tasks performed by humans. These models are created using machine learning techniques. The AI model creates a capability for task automation, but the AI model itself does not use automation. There are no varying levels of “automation” within an AI foundation model. There are, however, different levels (quantities) of training data volumes and computational power used within AI foundation models. Different foundation models also use different techniques of machine learning.

It is important to note, however, that not all AI foundation models are equal. Whilst a significant number of models are open source, there is not an even spread of model capability in terms of safety guardrails and overall model capability. The more useful and capable models with better or safer guardrails tend to be commercially valuable and are provided under commercial licences.

References

Amatriain X et al (2024) Transformer models: an introduction and catalog. arXiv:2302.07730. DOI: 10.48550/arXiv.2302.07730.

AKI (2023) Large AI Models for Germany – Feasibility Study 2023. A report by Akademie für Künstliche Intelligenz AKI gGmbH at German AI Association for German Federal Ministry of Economics and Climate Protection, Berlin.