Human Technology Institute

Human Technology Institute

UTS

Prof. Nicholas Davis
Co-Director, Human Technology Institute (HTI)
Industry Professor – Emerging Technology
University of Technology Sydney
PO Box [redacted] Australia
E: [redacted]

By email: aicommittee.sen@aph.gov.au
Dr John Turner
Committee Secretary
Select Committee on
Adopting Artificial Intelligence (AI)
Parliament of Australia

Wednesday 26 June 2024

Dear Committee Secretary,

This letter responds to the Hon Senator Pocock's question taken on notice during the public hearing on 21 May 2024 of the Select Committee on Adopting Artificial Intelligence ('AI') regarding the costs involved in developing capabilities related to so-called 'sovereign foundation models'. This is a concept that encompasses large language models ('LLMs') and other AI systems that rely on foundation model architectures.[i]

I believe Australia could benefit from investment in sovereign foundation model capabilities. However, in my opinion this should not consist of a 'moonshot'-style effort to create a single, Australian foundation model or LLM. Rather, the Government should support a range of efforts to support Australia's sovereign AI capabilities, increasing Australian understanding, and flexibility around the development and use of foundation models and other AI techniques that promise significant benefits for Australia.

Any such efforts should be made with the support of rigorous, independent expert analysis on how to best realise these benefits. Such analysis should not be limited to technical or economic analysis. The development and use of foundation models, and LLM systems in particular, poses a range of profound legal questions and other risks that need to be addressed to ensure that Australian reliance on them is lawful, safe and responsible.

While this letter focuses largely on foundation models and provides evidence and examples drawn from public sources on the cost of developing LLM systems, foundation models and other deep learning-based systems are not the only class of AI models and systems that can benefit Australia and therefore deserving of national support and investment. As just one example, the UTS Human Technology Institute is at the forefront of developing innovative statistical machine learning techniques designed to help policy makers understand causality, the better to address complex social challenges such as those facing education, healthcare, energy, housing and social justice.

I suggest therefore that the Committee's report contain a recommendation relating to the importance of further analysis on how to expand Australia's sovereign AI capabilities, to the effect that "The Government should commission detailed, independent expert analysis on the desirability and feasibility of developing and extending Australia's artificial intelligence capabilities, including but not limited to foundation models."

1. Structure of and key sources for this response

I start by providing a set of relevant definitions, including a set of *characteristics* associated with sovereign foundation models and sovereign AI capabilities more broadly. I connect these to a set of *benefits* associated with sovereign foundation models. I then present three broad *approaches* for the development and deployment of foundation models, each of which can incorporate one or more of these characteristics, along with reflections on associated costs based on publicly available data from LLM development supplemented with my own analysis. I provide a summary of the primary *cost drivers* in each approach, and present two scenarios that illustrate foundation model system operating costs. I conclude with some reflections and a *core recommendation* on how the Government might want to proceed should it prioritise a flexible, highly capable Australian AI ecosystem.

While the analysis and conclusions in this letter, as well as any errors, are my own, I acknowledge the expert input I received from Professor Jeremy Howard at the University of Queensland as well as my colleagues Professor Edward Santow and Professor Adam Berry. I have also drawn in this response from the work of Dr Stefan Hajkowicz and Dr Elanor Huntingdon at CSIRO's Data61, and Professor Genevieve Bell and others from ANU.

Throughout this response, I use the terms 'foundation model' or 'LLM' in isolation to refer to the relatively compact package of code and model weights at the heart of a multimodal or language-focused foundation model. When referring to the broader, multi-layered aspects of foundation models as experienced by a user in the form of an application or subcomponent of an AI system, I use 'foundation model system' or 'LLM system'. Recognising the risks of focusing purely on foundation models at the expense of other AI approaches, I use the phrase 'AI system', relying on the OECD definition, to refer more generally to AI applications.[ii]

2. What characteristics define a 'sovereign foundation model'?

The idea of a sovereign foundation model is not a settled concept. The phrase 'sovereign foundation model', 'sovereign LLM' or 'sovereign AI' is used variously to refer to models or systems that possess one or more of the following three characteristics:

A. an AI system that is controlled by a government for its own, secure use, thus reducing the risk of foreign access or influence over its deployment (i.e. an LLM system that can be used exclusively and with complete independence by a state).

B. an AI system that has been developed or adapted for specific, nationally oriented purposes, and/or has been trained to incorporate national datasets, national values, national languages (i.e. an LLM system is optimised for sovereign purposes, and/or reflects national characteristics).

C. an AI model that has been trained 'from scratch' by or for a specific country, providing a government or other entities with awareness and oversight of training data and the full training process of the model and surrounding system components (i.e. an LLM that is the entirely the product of Australian efforts).

Characteristic *C* tends to be the strictest way of defining a 'sovereign foundation model'. However, taken broadly, the phrase can refer to an AI system with any combination of the three characteristics.

Complementing these system-level characteristics, Stefan Hajkowicz and Elanor Huntingdon, in their March 2024 CSIRO-published report on *Artificial Intelligence Foundation Models*, characterise 'sovereign capability dimensions of foundation models', as comprising:

D. The capability of nation states to build, operate and manage AI technology drawing upon data, skills, knowledge, models and computational resources within the nation or its direct jurisdictional reach.

E. The capability of government to deliver functions and services using AI technology despite changes in the ecosystem of private-sector AI suppliers. And the capability of government to guide and regulate AI development and application by the private sector.

While the terms 'sovereign foundation model' or 'sovereign LLM' are often used in the singular to refer to the desire for a country having access to a model or system possessing one or more of characteristics *A* to *C*, it is unlikely that a country would ever seek to develop or deploy merely one single model or system. Governments, businesses and universities across Australia are already making use of a wide range of foundation models and LLM systems for nationally relevant purposes, and any effort towards sovereign foundation models would tend to involve investment in a diverse set of models and systems.

Furthermore, I recommend that the Committee consider characteristics *D* and *E* as higher-order goals for Australia in this domain. The drive to realise these should guide decisions as to whether and under what conditions foundation model systems with characteristics *A* to *C* are both desirable and feasible. I also note that these characteristics can and should apply to AI systems more broadly, and I recommend that the Committee consider with equal thoughtfulness and rigour the benefits, costs and risks associated with Australia possessing sovereign capabilities in other forms of AI systems beyond foundation models.

3. Why might Australia wish to invest in sovereign foundation model capability development?

Foundation models are highly flexible AI models that can be used for a wide range of use cases. In the case of LLMs, this includes analysing documents and other data and holding conversations in natural language. When developed and used safely and responsibly, these systems promise significant productivity improvements across a wide range of tasks.

The full extent of the capabilities of foundation model systems is still coming into view, as are the legal, social and economic risks that these systems bring. While technology-neutral laws apply to the deployment and use foundation model systems as they do to every other technology, important legal questions associated with foundation models, and LLMs in particular, are yet to be answered definitively by the legislature or courts. For example, there is ongoing litigation in the United States regarding the use of data protected by intellectual property and privacy laws in the training of LLMs.[iii] In other words, the precise set of duties one must fulfil to develop and use foundation model systems safely and responsibly remain the subject of public debate.

Foundation model systems are already being deployed and used within and across governments. For example, departments across States, Territories and the Federal Government are experimenting with LLMs in many forms.[iv] As with any government activity, this government use of foundation model systems is intended to be for public benefit.

Intensifying the safe and responsible use of foundation models – and deepening Australian understanding of how they work – may be beneficial for reasons other than their application in text analysis and generation. The mathematical and data science techniques that power LLMs can also be turned to myriad other positive uses, such as drug discovery. Foundation

model systems can also be used maliciously or recklessly in countless ways to harm individuals and groups and threaten the stability of Australian society.

Given these capabilities, it is reasonable to assume that nations that enjoy or build access to deep expertise in the strategic, organisational and technical aspects surrounding foundation model systems will be better placed to grasp their benefits while addressing the risks they pose.

4. What benefits could flow from developing capabilities around sovereign foundation models and other AI systems?

Implicit in characteristics A to C above is the idea that a range of benefits *may* flow from sovereign foundation model capabilities, subject to the caveat described below. The potential benefits include:

1. **Increased system and data privacy and security**: Locally developed or deployed foundation model systems can be designed to increase the security and privacy of the system itself and all related data.
2. **Increased data and system control**: Locally deployed and managed foundation model systems can be operated with greater levels of control and oversight of their data inputs, operations, computational and power usage, environmental impact etc.
3. **Reduced dependence on foreign providers:** Locally developed and/or deployed foundation model systems could reduce the risk of the Government systems or other critical infrastructure being exposed to threats of external influence or system disruption created by reliance on foreign providers.
4. **Customization for nationally determined purposes and needs**: Sovereign foundation model can be designed or adapted to address specific national challenges and needs across Australia.
5. **Incorporating Australian legal, cultural, and linguistic nuances**: Whether via development from scratch or by using fine-tuning, a locally adapted LLM system can better understand and generate text in Australia's most used languages, incorporating colloquial context and nuances that may be absent in default, foreign-trained models. A locally adapted LLM could also incorporate an understanding of Australian legal norms, Federal and State laws and regulations, which differ from foreign standards and expectations.
6. **Supporting national trust and adoption of AI systems:** The availability of Australian foundation models could increase public trust in its use and spur the safe and responsible adoption of AI more generally that results in benefits to Australians.
7. **Spurring specialised job creation and innovation**: Fostering local activity related to building, fine-tuning, operating and maintaining foundation models could lead to the creation of a new subset of locally provided data science services, with associated job growth and the possibility of exports of related IT services. Sovereign capabilities would also allow accelerated model specialization and applications in and across areas of national interest or strategic priority. This would spur innovation in both Government and industry, supporting Australian competitiveness.
8. **Advancing Australian research and knowledge in the technical, strategic and organisational aspects of foundation model development and use**: Researching new ways to create, update and deploy foundation models could create a vibrant subset of the data science education and research community in Australia, while investment in the strategic use, organisational deployment and stakeholder impact of foundation models would enhance local knowledge around how to ensure they are used safely and responsibly. This could support a range of commercial and quality of life benefits to Australians while contributing to the global body of knowledge in

artificial intelligence and related fields. It could also ensure that Australia has access to relevant experts as LLM technologies evolve over time.

Given the current state of AI awareness, adoption, and governance in Australia, and the fast-moving nature of foundation model technologies, I suggest that benefits 7 and 8 are particularly important and desirable for Government analysis.

I also note that these benefits are not unique to sovereign foundation models – all eight can apply to sovereign capabilities related to other leading AI techniques on which Australia currently relies or may rely on in the future.

5. What risks relating to sovereign foundation model investment should Australia consider?

All these benefits are desirable for Australia, and all of these can, in theory, flow from investment in activities related to sovereign foundation model capabilities, as well as investment in AI capabilities more broadly. However, there is an important caveat in realising these potential benefits: the benefits described above are not inherent to nor automatically flow from increasing technical capabilities.

Crucially, for these benefits to be realised, Australia's technical AI capabilities must be matched by both strategic capabilities related to decision making about whether, why, how and when a particular AI system is warranted, combined with a legal and regulatory environment that effectively manages the risks that can flow from AI system use across sectors. The benefits will also rely on a flourishing, well-supported education and research and development ecosystem that delivers sufficient expertise and human talent.

The need for clear laws and a correspondingly effective regulatory environment is particularly acute when it comes to sovereign foundation model capabilities. For instance, enhancing Australia's sovereign foundation model capability could easily lead to *less* data privacy and security for Australians at large, rather than more, should their widespread application prove to be unlawful, unethical and/or harmful (see benefit 1 above).

Nevertheless, possessing a broad-based sovereign capability around foundation models (along with other AI approaches) offers Australia *an expanded set of informed choices* and *greater power to make choices* from among decisions would otherwise be made or heavily influenced by foreign corporations and other nation states.

The extent to which any of the benefits described above is realised in fact will depend on the choices that are made in how a sovereign AI capability is established and operated. Without a sovereign capability, these choices are much more constrained or are simply exercised by others. If sovereign capabilities are focused narrowly on one subset of AI techniques, or one provider of foundation model capabilities, Australia may find itself tied too narrowly to a single 'moon shot' while missing the chance to lead and benefit from the development and use of other, emerging forms of AI that offer complementary or alternative benefits.

6. What goes into creating and using a foundational model system?

Before exploring costs, it is useful to distinguish between a few different components of a typical foundation model system that is used to create an output. Here, I draw particularly from CSIRO's report on *Artificial Intelligence Foundation Models* as well as HTI's research in this area. While I refer to foundation models in particular, the same categories generally apply to other AI approaches.

First, and most importantly, a critical input is the range of human experts who possess the requisite knowledge to develop and deploy foundation model systems appropriately.

At the strategic level, this includes knowledge of the existing challenges or time-consuming tasks where a foundation model system might be useful, which model and system attributes and capabilities would best support this use, how the system would be deployed, what individual and systemic harms might result if the system failed or was misused, how stakeholders will experience the system and its outputs, and the policies and standards such a system must meet to be trustworthy.

At the technical level, broad sovereign foundation model capabilities would require Australia to have local experts capable of developing, evaluating and selecting a relevant foundation model architecture, of identifying and collecting data for training or fine-tuning, of performing model training on appropriate infrastructure, then evaluating and improving the model to meet specific needs. These technically focused experts must be able to appreciate and incorporate an understanding of user engagement, impact assessment, regulatory guardrails and international standards, while collaborate with other stakeholders around the strategic aspects of foundation model use.

Second, developing and operating foundation models requires computing hardware and infrastructure for tasks such as base training, fine-tuning and ongoing operation. For very small and simple LLMs, this may comprise nothing more than a single graphics processing unit ('GPU') used for model training, or a laptop capable of running a locally stored model. For larger foundation models, training, fine-tuning and operating a sovereign foundation model may require access to one of Australia's high-performance computing and data (HPCD) facilities, or a dedicated data centre stocked with appropriately advanced GPUs.

Training 'from scratch' at the frontier – such as replicating the creation of the largest open-source models – would require scales of computation currently only available on the public cloud in foreign data centres. According to Professor Jeremy Howard, more than 99% of the computation time in training models is spent on training the base model.[v]

Third, all machine-learning based AI models rely on training data, and foundation models are particularly data-hungry. For example, training LLMs from scratch requires extremely large amounts of text data, comprising trillions of words. Fine-tuning requires high quality and representative text. While many 'public web' text datasets are freely available, it can be extremely costly to clean, curate and validate datasets to ensure high levels of quality. Specialised datasets may be extremely expensive to acquire and may be the subject of exclusive licensing to tech companies.

Fourth, combining expert know-how and datasets on computing infrastructure to create and/or deploy foundation models requires energy. Energy is such a critical component of computing that data centre capacity is commonly measured in MW, with Australia currently possessing approximately 1,200 to 1,500 MW of operational data centre capacity.

Foundation models are particularly energy hungry in comparison to training or operating other types of narrow AI. In the United States, energy availability is proving to be a critical bottleneck affecting data center construction and provision,[vi] and tech companies report significant challenges in securing energy supply for data centers earmarked for AI training and operation set to come online after 2026.[vii]

Fifth, safe and responsible LLM system use requires a supporting ecosystem of laws, standards, educational institutions, and regulatory bodies as well as organisational-level strategies and policies. HTI has done extensive work in this area.

7. How might Australia approach the technical aspects of sovereign LLM development?

To directly address Senator Pocock's question around the specific costs related to developing and training sovereign foundation models, in this section I consider three primary approaches that the Government and other Australian organisations could consider for developing and deploying sovereign foundation model systems. Given the increasing use of LLMs across Government, I use the development of an LLM as a convenient example for estimating and comparing costs.

These approaches differ primarily regarding the core characteristics and source of the core LLM itself. While only the third approach offers the possibility of meeting all three of characteristics A to C outlined above, it is not the only, nor necessarily the best, approach to realising the benefits of locally adapted foundation models for Australia in a sustainable and cost-effective manner.

In all three cases, regardless of how initial or 'base' training of an LLM was achieved, local adaptation can be achieved or enhanced through subsequent efforts given appropriate expertise and relevant datasets.

Two techniques are particularly common today: 'fine-tuning' and 'retrieval augmented generation' ('RAG'). Fine-tuning allows a deployer to add data to the model in ways that shift the model 'weights', thereby changing its behaviour to suit a specific need or context. By comparison, RAG does not change the model itself, but rather directs an LLM system to refer to and incorporate information contained with a pre-defined and additional set of data when preparing its answer for a user.

Fine-tuning is a powerful technique that can induce significant shifts in a model with relatively little additional investment in terms of compute. For example, in June 2024, data scientist Leonard Lin found that the Chinese Open Source LLM Qwen 2 72B had been trained and adapted using reinforcement learning such that its answers align closely to the values of the Chinese Communist Party. As a result, the base LLM refuses to answer questions considered 'sensitive' by the CCP in English, while the same questions in Chinese tend to produce an admonishment by the model.

This behaviour is far from set in stone, however. Professor Jeremy Howard reports that a subsequent model called Dolphin 2.9.2 Qwen2 72B was produced by fine-tuning Qwen 2 for approximately 0.03% of its original training time, at an estimated cost of USD$2,000. Compared to the original Qwen 2 model, the fine-tuned version displays completely different behaviour when posed the same 'sensitive' questions, being entirely willing to answer questions and providing very different perspectives on the same issues. Professor Howard asserts that "In practice, the behavior of all models can be entirely changed with just a few hours of fine tuning on a single, modestly sized computer."

Approaches 1 and 2 below are already both common today; indeed, the UTS Human Technology Institute relies on both for its internal use of and experiments with LLM systems.

Approach 1: Licensing access to proprietary LLM from an overseas tech partner

There are currently no major Australian companies offering an LLM with capabilities close to the frontier of current LLM development and research. The Australian Government may therefore choose to partner with an overseas provider in a commercial arrangement to access one or more leading, proprietary LLMs, such as GPT-4 from OpenAI, Claude 3 from Anthropic, or Gemini from Google.

The Government can license these LLMs for use in a variety of forms, including services embedded in commercial, cloud-based applications (e.g. an LLM operating within a word processing application), capabilities embedded in or used in locally developed applications via Application Programming Interfaces (APIs), or as standalone services. Where possible and desirable, the Government could 'fine-tune' selected models or use other methods to adapt LLMs and create systems that are nationally aligned for specific uses. The Government would likely seek to arrange with the provider to operate a closed-source LLM on Australian soil in a secure data centre, and carefully manage the security of fine-tuning processes, prompt inputs and system outputs.

Costs

The Government would pay the foreign provider to license and use this model. While such costs are open to negotiation with providers, costs are usually charged based on system throughput, with a price being charged for increments of chunks of input and output data (known as 'tokens'). This price tends to be inclusive of all infrastructure and energy costs.

For example, the latest multimodal model from OpenAI, GPT-4o, is currently offered at a retail cost of USD$5 for 1 million input (or prompt) tokens, and USD$15 for 1 million output (or answer) tokens. This pricing implies that USD$8 will provide a user with roughly 250 in-depth conversations with ChatGPT.[viii] Fine-tuning costs are additional, and it is worth noting that OpenAI currently does not allow all its models to be fine-tuned.

Advantages

Partnering with foreign providers in this way would give the Australian Government immediate access to the most advanced LLMs in the market without the significant initial investment of time and resources required to develop its own LLM.

According to the 2024 AI Index Report from the Stanford Institute for Human-Centered Artificial Intelligence, closed models currently significantly outperform open ones on all benchmarks, with a median performance advantage of 24%. For Agent-based behaviour, where an LLM is asked to perform multi-step tasks that require reasoning and decision-making, the advantage of closed-source or proprietary models is 318%.

Such an arrangement is also infrastructure efficient. Responsibility for both the physical hardware and supporting software required to run the LLM is managed by the partner, removing the need for Australian organisations to develop, manage and continuously upgrade the computing infrastructure for training, fine-tuning and operation of the LLM systems. Scaling up or extending use is therefore easy to achieve.

Disadvantages

Given the proprietary nature and overseas source of the underlying models in this approach, the Government would be dependent on a foreign entity and its guarantees in relation to key aspects of the LLM and supporting infrastructure. This includes the model architecture, the content of the original training data, and – possibly – the security of the environment in which the system is housed and used, should this also be provided by the technology partner. Moreover, such a partnership would be seen as an implicit endorsement of the organisation and may result in accepting liability for risks related to intellectual property, other harms or ethical breaches that occurred in developing the model.

Relying on proprietary models also creates dependencies on the continuing existence of the provider, and their willingness to provide ongoing support and model upgrades. It also restricts Australian choices and control on the future evolution of the underlying models.

At very large scales of use, the operational costs of such a system, when priced by throughput, could be more expensive than other approaches.

Approach 2: Adapting and deploying open-source models on nationally managed infrastructure

As a second approach, the Australian Government could choose to adapt to local needs one or more open-source LLMs, creating a series of sovereign LLMs that take advantage of the extensive expert work already undertaken by others in data preparation and training.

While the terms and conditions governing the use of open-source LLMs can differ greatly, an LLM designated as "open-source" generally means that its code and model weights are available for scrutiny, and that a license allows for public use, modification, and distribution.

Llama 3 (from Meta in the USA), Falcon 180B (from the UAE's Technology Innovation Institute), Qwen 2 72B (from Alibaba in China), and Mixtral-8x7B (from Mistral.ai in France) and other open-source LLMs are able to be freely inspected, modified and deployed within the terms of their relevant licenses and acceptable use policies.

As with option 1, Australian experts would very likely fine-tune these models to create locally adapted LLM instances for specific, national purposes and to embed Australian characteristics. The extent of the fine-tuning would depend on the tasks the LLM system was intended to perform, the amount and quality of data being used to fine-tune the LLM, and time and resources (particularly the computational power) being invested to do so. Should the licensing arrangement permit, this could result in a highly customised, sovereign LLM provided at a fraction of the original training costs.

Costs

While there is normally no licence fee associated with the use of open-source models, open source LLM systems must be deployed on infrastructure, and there are costs associated with the resulting computational load for operating the model. Furthermore, not all models badged as open source are truly open and able to be used for all purposes. For any given model, the Government may need to navigate a range of license restrictions linked to intended use.

Fine-tuning the model also consumes computational resources, though such fine-tuning is usually a fraction of the original training time. For example, as mentioned above, Jeremy Howard estimates that fine-tuning Qwen 2 to create the Dolphin model with dramatically different behaviour to the base model would have incurred computational costs in the order of a few thousand dollars.[ix]

Advantages

Fine tuning a suitable base model can significantly alter the behaviour of an LLM, effectively crafting it into a locally adapted system in terms of its behaviour and interaction with users. This allows for low-cost alignment of an LLM system with national values, and the opportunity to incorporate additional, national data, into the model weights.

Because the code and weights of open-source models are open, technical experts have greater insight into the system than with proprietary models. Many providers of open-source models also give insight into the data used for training, providing further transparency around the scale, types, lawfulness and ethical nature of training input.

Deployers also enjoy greater levels of control and flexibility over how open-source models are adapted, housed, managed and operated, including greater visibility over aspects such as energy consumption and infrastructure security.[x] Open-source models provide the flexibility of being deployed locally, in enterprise or co-located data centres, or via public cloud services.

For certain scales of use (e.g. entirely local use, or use at a very large scale), running fine-tuned open-source models be more cost-effective than either licensing a commercial model or developing an LLM from scratch. This is because the costs of deploying the model at scale will, over time, approach the infrastructure and energy costs of the system. For example, recent analysis indicates that running Meta's latest open source model Llama-3-70B is 8 times cheaper for input tokens and 5 times cheaper for output tokens when compared to OpenAI's top-of the line GPT-4 on the same infrastructure.[xi]

Disadvantages

The primary disadvantage is that, as described above, open source LLMs currently lag in capability compared to the leading proprietary models. This is particularly true in terms of the multi-modal or agent-like capabilities.

Approach 3: Develop an LLM 'from scratch' in Australia

The third approach is for Australia to train, develop and deploy its own series of 'home grown' LLMs from scratch. This would involve one or more Australian organisations – potentially with significant support from the Government – investing substantial amounts in research, talent acquisition, computational resources, data curation and time.

First, an organisation would first need to strategically align with the Government and other potential users of a home-grown sovereign LLM (or LLM system) regarding its purpose and key characteristics. For example, is the LLM intended to be a very general and flexible model, or focused on a set of specific tasks or use cases? Are there particular characteristics or capabilities that are desirable?

Second, the organisation would need to develop a team to design, gather the data and train the model. Depending on the availability of experienced researchers in this field – who are in global demand – it could take considerable time to recruit and assemble such a team. Data acquisition for foundation model training is a challenging and expensive task that poses particular challenges to public sector entities.[xii]

Third, the organisation would have to secure the appropriate computing infrastructure at the necessary scale. The UAE's Falcon 180B model relied on AWS public cloud infrastructure for training over a time period of more than 2 months, using approximately 4,096 specialised A100 GPUs.[xiii] Access to this kind of specialised hardware in the form of advanced and increasingly scarce GPUs is required to train large models in reasonable timeframes. The operational and economic risks that relate to an ongoing need for data centre capacity and GPUs is a significant enough issue that Microsoft mentioned these twice in its 2023 Annual Report.[xiv]

Fourth, the organisation would have to evaluate, test and fine-tune the model to ensure it complies with Australian expectations and use policies.

Costs

The cost of such an effort would be highly dependent on the size and complexity of both the intended base LLM and overall desired LLM system. While companies generally do not provide the full economic details of LLM training, we can look at publicly available data to estimate the range of costs involved in training an LLM system in Australia from scratch.

When it comes to the most capable current LLM systems, we can look for example to OpenAI's CEO, Sam Altman, who indicated that training GPT-4 cost over USD$100 million.[xv] At the University of Technology Sydney's Vice-Chancellor's Annual Democracy Forum on 19 June 2024, Meredith Whittaker, President of Signal, indicated that the current cost of each training run for the latest generation of large language models was approximately USD$100 million, noting that training an LLM would involve multiple runs.

These order-of-magnitude estimates for the most capable and recent proprietary LLMs are supported by Stanford University's AI Index Report 2024, which estimates that training Google's Gemini Ultra and OpenAI's GPT-4, were, in USD, respectively around $191 million and $78 million. This report also notes that training costs are rising over time as model size and complexity grows.[xvi] For this reason, start-ups focused on training extremely large, closed source LLMs from scratch tend to quickly partner with cloud infrastructure providers, as France's Mistral.ai has done in partnering with Microsoft to access its Azure AI infrastructure.[xvii]

Other LLMs are cheaper to train. The Stanford report estimates that training Meta's Llama 2 model cost approximately USD $4m. Others estimate that the cost of training the UAE's Falcon 180B LLM on more than 4,000 Nvidia 100 A100 GPUs was approximately USD $14m,[xviii] while my own calculations based on current GPU pricing suggests that the training cost of Meta's Llama 3 70B on Nvidia H100 GPUs was approximately USD $17m.[xix]

This can be compared with the cost of investing in the hardware directly. Assuming that supply was available, an organisation seeking to purchase 4,000 Nvidia A100 80GB GPUs to replicate the training of the UAE's Falcon 180B LLM would have to budget for approximately AUD $100 million, or USD $67 million for the purchase of the GPUs alone. The energy cost of training equivalent to the Falcon 180B LLM would add another AUD $644,000 or USD $430,000 to the costs.[xx]

In addition to these costs, acquiring the necessary human talent for home grown foundation model training is expensive. There is significant global competition for data scientists with experience in this domain.

Advantages

This approach would undoubtedly provide the greatest level of control over model's architecture, data, and training approach.

Perhaps most importantly, however, this approach would foster significant innovation and capacity building within Australia related to the development and use of LLMs.

It is hard to estimate the benefit of knowledge spillovers that would be the result of Australia investing in homegrown LLM development. Indeed, the extent of these may rely heavily on the form of incentives. However, investment in the form of support for R&D related to LLMs would undoubtedly help increase Australia's economic complexity[xxi] and drive innovation opportunities for Australia.

Disadvantages

The first disadvantage of this approach is uncertainty as to outcome. There is no guarantee that such an effort would result in a model that approaches the capabilities of the most advanced proprietary systems. During the period in which Australia invested time and money into training a system from scratch, it is highly probable that Australia's allies, strategic competitors and technology partners will produce significantly more advanced LLM systems, leaving the country with a 'sovereign LLM' that is already outdated. For reference, OpenAI has received investment to date of over USD $13 billion, at least an order of magnitude above and five years in advance of what Australian sovereign efforts might achieve.

Second, this approach is inherently significantly more costly than the previous two thanks to the cost of the base model training.

The third disadvantage of this approach is that the high costs involved in training means that the Government may be tempted to "pick a winner" and award significant funding to a single entity tasked with the creation of a home-grown LLM trained from scratch. Rather than promoting knowledge spillovers, this could result in the centralisation of LLM-related talent across Australia.

8. Cost driver comparisons and scenarios for LLMs

Table 1 summarises the cost-drivers, selected costs, and advantages related to the three different approaches of LLM development referenced above. I note that costs related to expertise are highly dependent on scale and ambition of foundation model development.

Source of costs	**Approach 1:** *Proprietary model licensed from and managed with foreign partner*	**Approach 2:** *Open-source model licensed and fine-tuned and managed by Government*	**Approach 3:** *Home-grown LLM trained from scratch, fine-tuned and managed by Government*
Expertise required	Strategic teams able to understand and decide why, how and when LLMs should be used, and which models and tech partners meet needs. Technical teams able to engage with proprietary, cloud-hosted models and able to undertake fine-tuning.	Strategic teams able to understand why, how and when LLMs should be used, and which open-source models and configurations meet needs. Technical teams able to engage with open-source, self-hosed and public cloud-hosted models and able to undertake fine-tuning.	Strategic teams able to understand why, how and when LLMs should be used, and where a home-grown LLM would be of most value to meet needs. Technical teams able to manage and deliver the full range of LLM architecture choice, training, testing, and fine-tuning.
Base model training cost	N/A – model is fully pretrained	N/A – model is fully pretrained	Open-source equivalent model: Approx. AUD $25m High-end, proprietary-equivalent model: Approx. AUD $150m
Fine-tuning cost	Throughput-based, charged per million tokens. More expensive than fine-tuning an open-source model.	Priced by computation, but approx. $AUD 10k per use case	Can be considered to be included in base model training.
Infrastructure capex cost	N/A – included in operating cost	N/A – included in operating cost	If trained on own chips: $100m for purchase
Flexibility of use	High	High	Very High
Transparency of system	Low	High	Very High
Opportunity for skills development	Medium	High	Very High
Contribution to independence and resilience	Low	High	Very High

To complement this cost comparison, Table 2 below presents a simple calculation of the ongoing operating costs for each approach across two government LLM use scenarios at current benchmark prices. Note that these calculations are focused on operating costs only, and do not consider any capital costs or their amortisation related to base training or fine tuning.

Scenario A assumes that a specialised team within government relies heavily on an advanced, fine-tuned sovereign LLM. In this scenario, 100 public servants engage heavily on a daily basis with an advanced model, collectively processing 8m tokens of text, the equivalent of 11,000 pages of A4 formatted documents in size 11 Arial font.

For the proprietary model in this scenario, I have assumed the use of OpenAI's GPT-4, noting that it is a) not commonly available for fine tuning and that b) it is one of the most expensive models available. For the open source and home-grown models, I have benchmarked operating costs to the current price of a popular open source LLM, Meta's Llama 3, assuming that the home-grown model would be of the size and complexity of Llama. I note that this an extreme comparison, as running GPT-4 currently costs 40 times more than the average of Llama 3 prices per 1m tokens.[xxii]

Scenario B assumes widespread use of a sovereign LLM across 250,000 public servants, each engaging in 10 interactions per day (each a quarter of the intensity of scenario A) at a collective throughout of 5 billion tokens.

For the proprietary model in this scenario, I have assumed the use of OpenAI's GPT-4o, noting that it is also not commonly available for fine tuning. It is, however, considerably cheaper than GPT-4. As in Scenario A, for the open source and home-grown models, I have benchmarked operating costs to the current price of a popular open source LLM, Meta's Llama 3, assuming that the home-grown model would be of the size and complexity of Llama.

Simple operating cost comparison by LLM approach	**Approach 1:** *Proprietary model licensed from and managed with foreign partner*	**Approach 2:** *Open-source model licensed and fine-tuned and managed by Government*	**Approach 3:** *Home-grown LLM trained from scratch, fine-tuned and managed by Government*
Scenario A: operational costs of daily use by a specialised team of 100 experts within government[xxiii]	AUD $450 per day	AUD $11 per day	AUD $11 per day
Scenario B: operational costs daily use by 250,000 public servants[xxiv]	AUD $56,250 per day	AUD $6,750 per day	AUD $6,750 per day

9. Which mix of approaches for sovereign foundation model development and use creates the best outcomes for Australia?

The three approaches above provide useful order of magnitude estimates of a subset of costs related the technical development of LLMs with sovereign characteristics A to C. However, it would be a mistake to narrow Australia's options in this regard purely to these aspects.

Rather, I suggest the Committee consider how investment by both Government and industry can contribute to CSIRO's 'sovereign capability dimensions of foundation models' outlined as characteristics D and E above. There are:

D. The capability of nation states to build, operate and manage AI technology drawing upon data, skills, knowledge, models and computational resources within the nation or its direct jurisdictional reach.

E. The capability of government to deliver functions and services using AI technology despite changes in the ecosystem of private-sector AI suppliers. And the capability of government to guide and regulate AI development and application by the private sector.

Viewing the approaches above in light of these broader characteristics suggests that it would be a mistake for Australian businesses, non-profit organisations, academic institutions and government entities to rely solely on proprietary LLMs or other closed source foundation models from overseas providers.

At a minimum, incentivising the creation of a deep understanding of a wide range of open-source models could benefit Australia both thanks to their potential to be housed, operated and fully controlled in Australia and by creating demand for Australian experts who can assess, choose, fine-tune, deploy, manage and govern such systems.

Spurring investment in 'home grown' development and training of foundation models would also undoubtedly support the capability dimensions above. While the up-front cost for this would be far higher, it would serve to boost local research and development efforts.

Furthermore, taking characteristics D and E as a primary lens for investing in sovereign AI capabilities suggests that Australia should be careful not to focus solely on foundation models to the exclusion of other approaches. In addition to the possibility of Australia missing out on opportunities to extend its lead in areas such as Bayesian machine learning and causal inference, the risks and legal concerns associated with foundation models and the rapid pace of AI technology development strongly suggest that Australia should develop sovereign AI capabilities that are multifaceted, broad-based and flexible.

10. What opportunities exist to support Australian sovereign AI capabilities?

I would like to conclude this response with reflections on how the Government might foster support for the development of sovereign AI capabilities in Australia in ways that maximise Australia's options and flexibility to benefit from the promises of foundation models as well as the capabilities of other, emerging AI approaches.

I suggest that the Committee and Government consider four broad areas of investment to support sovereign AI capabilities in Australia:

1. The development of an Australian AI regulatory strategy

First, I note that HTI has separately provided independent expert advice to the Government on the development of an Australian AI Regulation Strategy. Among other things, a regulation strategy of this nature would support sovereign AI capabilities by promoting consistency and regulatory certainty regarding the development and deployment of specific sub-types and use cases for AI systems, including but not limited to foundation model systems and LLMs.

2. Support for AI-dedicated High-Performance Compute and Data ('HPCD') infrastructure in the form of AI-specific data centres.

Second, the Government could consider expanding support for Australian access to flexible computational infrastructure required for efficiently training and fine-tuning AI systems.

Australia is a significant net importer of compute. Every day, Australian individuals and organisations rely on IT services that are computationally intense. While the precise deficit is challenging to estimate, the preponderance of evidence is that most of these computations are performed in data centres outside of Australia, delivering everything from cloud email services to Tik-Tok videos, to LLM answers, to GPS navigation system directions.

By my own calculations, the current market for cloud computing services in Australia reflects demand for approximately 4,000 MW of compute, and onshore data centres can supply less than half that.

Australia particularly lacks the kind of computational facilities for large scale foundation model training. For security, resilience – as well as to attract talent – Australia should consider investing in an AI-first HPCD, or dedicated and secure access to the same.

3. Support for both technical and strategic AI skills development, with a particular focus on undergraduates and early researchers

Australia is lucky to have a small set of world experts in AI resident across the country, including world-leading researchers focused on causal inference, deep learning, foundation models and their strategic, policy and legal implications.

However, to realise the full benefits and manage the risks of AI systems over time, a sovereign AI capability for Australia must consist of a significantly larger, wider, more diverse, and more inclusive interdisciplinary talent pipeline.

In particular, the Government should seek to engage and activate Australia's most promising minds across Australian universities to develop globally relevant research on the strategic and technical aspects of AI system development and use.

This could involve providing funding to qualifying universities willing to invest in the necessary mathematics and computer science courses at both the undergraduate and graduate level. These efforts should support an accelerated understanding of the mathematical underpinnings, engineering and design elements, development steps, legal and policy implications, stakeholder impacts, and implementation considerations of frontier AI techniques, including but limited to deep learning and foundation models.

The importance of diversity and inclusion in promoting AI research and skills cannot be overstated. A sovereign AI capability for Australia should reflect Australian values, draw on uniquely Australian sources of knowledge and insight, and be designed to serve and empower Australians, including First Nations communities. For example, sovereign AI capabilities could include expanded methods for respecting indigenous data sovereignty.[xxv]

The Government may choose to prioritise AI training, research and related development activities focused on ensuring the benefits of our recent AI advances reach those who could use them the most, uplifting our vulnerable and disadvantaged members of the community, while simultaneously mitigating the risks of harm that are so often seen with the introduction and embrace of new technologies.

4. Support for safe and responsible applications that can be commercialised from supported education programs

Finally, the Government should consider options for incentivising the commercialization of Australian research related to AI systems, where use or further development adds value to Australia's sovereign AI capabilities.

An important condition of such funding could be a focus on safe and responsible development and the potential for community benefit.

11. Next steps and core recommendation

Given the strategic importance of AI systems, and the potential for Australia to rely ever more heavily on AI technology, services and computational facilities owned and controlled by third parties based overseas, it is important that the Government invest soon in research that can help further illuminate the value of growing Australia's sovereign AI capabilities, and options for doing so. Such research should consist of rigorous, inter-disciplinary and independent analysis of the potential benefits, costs, enablers and legal implications of expanding Australia's sovereign AI capabilities.

I suggest therefore that the Committee's report contain a recommendation that "The Government should commission detailed, independent expert analysis on the desirability and feasibility of developing and extending Australia's artificial intelligence capabilities, including but not limited to foundation models."

Please do not hesitate to be in touch if I can assist further.

Yours sincerely,

Prof. Nicholas Davis
Co-Director, Human Technology Institute
Industry Professor – Emerging Technology
University of Technology Sydney

[i] While I do not use the term in this paper, foundation models underpin what is increasingly known as "General Purpose AI" or "General AI". See for example Explainer: What Is a Foundation Model? https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/. Accessed 26 June 2024.

[ii] Perhaps the most broadly accepted definition for an AI system is that developed by the OECD: "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en. As CSIRO points out, LLMs are a subset or variant of the broader category of foundation models, which in turn are a type of deep-learning powered AI system. https://www.csiro.au/en/research/technology-space/ai/AI-foundation-models-report. Some LLM systems may consist of multiple foundation models and other AI-drive systems working in tandem across text, image, speech and code. In their *Rapid Response Information Report*, Bell et al referred to these as 'multi-factor foundation models' or 'MFMs'. Bell, G., Burgess, J., Thomas, J., and Sadiq, S. (2023, March 24). Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs). Australian Council of Learned Academies. https://www.chiefscientist.gov.au/GenerativeAI

[iii] See for example Reed, Rachel. 'Does ChatGPT Violate New York Times' Copyrights?' Harvard Law School, https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/. Accessed 26 June 2024.

[iv] See for example DTA. 'APS trials generative AI to explore safe and responsible use cases for government'. Digital Transformation Agency, 6 March 2024, https://www.dta.gov.au/blogs/aps-trials-generative-ai-explore-safe-and-responsible-use-cases-government

[v] Howard, Jeremy. 'Answer.AI - What Policy Makers Need to Know about AI (and What Goes Wrong If They Don't)'. Answer.AI, 17 June 2024, https://www.answer.ai/posts/2024-06-11-os-ai.html.

[vi] 'Energy Is Now the "Primary Bottleneck" for AI'. Latitude Media, https://www.latitudemedia.com/news/energy-is-now-the-primary-bottleneck-for-ai. Accessed 26 June 2024.

[vii] Moss, Sebastian. 'Meta's Mark Zuckerberg says energy constraints are holding back AI data center buildout'. Data Centre Dynamics, https://www.datacenterdynamics.com/en/news/metas-mark-zuckerberg-says-energy-constraints-are-holding-back-ai-data-center-buildout/ Accessed 26 June 2024.

[viii] Here I assume that each detailed conversation with the LLM consumes a total of 4,000 tokens, or approximately 3000 words processed, with a 3:1 ratio of input to output tokens.

[ix] Jeremy Howard (2024) Personal communication with author, Friday 21 June 2024

[x] In some cases, it may be possible to license proprietary models yet run them within locally controlled infrastructure, which would provide this visibility.

[xi] Llama 3 70B vs GPT-4: Comparison Analysis. https://www.vellum.ai/blog/llama-3-70b-vs-gpt-4-comparison-analysis. Accessed 25 June 2024.

[xii] Acquiring training data at requisite scales may be particularly complex and legally challenging for the Government to do itself. The uncertain intellectual property status of most public web data would give rise to significant legal issues.

[xiii] Falcon 180B Foundation Model from TII Is Now Available via Amazon SageMaker JumpStart | AWS Machine Learning Blog. 11 Sept. 2023, https://aws.amazon.com/blogs/machine-learning/falcon-180b-foundation-model-from-tii-is-now-available-via-amazon-sagemaker-jumpstart/.

[xiv] Microsoft 2023 Annual Report. https://www.microsoft.com/investor/reports/ar23/. Accessed 25 June 2024.

[xv] Knight, Will. 'OpenAI's CEO Says the Age of Giant AI Models Is Already Over'. Wired. www.wired.com, https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/. Accessed 24 June 2024.

[xvi] AI Index Report 2024 – Artificial Intelligence Index. https://aiindex.stanford.edu/report/. Accessed 24 June 2024. page 63.

[xvii] Boyd, Eric. 'Introducing Mistral-Large on Azure in Partnership with Mistral AI'. Microsoft Azure Blog, 26 Feb. 2024, https://azure.microsoft.com/en-us/blog/microsoft-and-mistral-ai-announce-new-partnership-to-accelerate-ai-innovation-and-introduce-mistral-large-first-on-azure/.

[xviii] 'Is Falcon 180B Really a Llama Killer?' Prompt Engineering, 8 Sept. 2023, https://promptengineering.org/is-falcon-180b-really-a-llama-killer/.

[xix] This calculation draws on data from https://huggingface.co/meta-llama/Meta-Llama-3-70B and assumes 6.4m GPU training hours using H100 chips at the lowest widely advertised cost of USD $2.69 per hour. https://lambdalabs.com/nvidia-h100-nvidia-h200-gpus#:~:text=On%2Ddemand%20HGX%20H100%20systems,only%20%242.59%2Fhr%2FGPU.

[xx] This calculation assumes 7m GPU training hours using A100 chips at their SXM full power draw of 400W at a (typical for Canberra) electricity price of 23c per kwh.

[xxi] Hidalgo, César A., and Ricardo Hausmann. 'The Building Blocks of Economic Complexity'. Proceedings of the National Academy of Sciences, vol. 106, no. 26, June 2009, pp. 10570–75. https://doi.org/10.1073/pnas.0900943106.

[xxii] For cost comparisons, see https://artificialanalysis.ai/models/gpt-4 and https://artificialanalysis.ai/models/llama-3-instruct-70b/providers

[xxiii] Use rate of 100 people having 10 interactions a day at the context limit of 8000 tokens and with 75% of tokens being input. For proprietary model, illustrative cost GPT-4 model at a cost of USD $30 for 1m input tokens, USD $60 for 1m output tokens. For open source and home-grown models, cost estimated from current costs of Llama 3 running at a combined input/output price of USD $0.9 for 1m tokens. USD-AUD exchange rate of 1.5. For simplicity, I have assumed that both the home-grown model is of the size and complexity of Llama 3, so would incur similar operational costs.

[xxiv] Use rate of 250,000 people having 10 interactions a day at an average of 2000 tokens and with 50% of tokens being input. For proprietary model, assuming use of GPT-4o model at a cost of USD $5 for 1m input tokens, USD $15 for 1m output tokens. For open source and home-grown models, cost of Llama 3 running at a combined input/output cost of USD $0.9 for 1m tokens. USD-AUD exchange rate of 1.5. For simplicity, I have assumed that both the home-grown model is of the size and complexity of Llama 3, so would incur similar operational costs.

[xxv] See for example PM&C. 'How might AI affect trust in public service delivery?' Friday 27 October 2023. https://www.pmc.gov.au/resources/long-term-insights-briefings/how-might-ai-affect-trust-public-service-delivery/ai-trust/insight1, and Walter, M and Kukutai, T (2018), Artificial Intelligence and Indigenous Data Sovereignty, input paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing" on behalf of the Australian Council of Learned Academies. https://acola.org/wp-content/uploads/2019/07/acola-ai-input-paper_indigenous-data-sovereignty_walter-kukutai.pdf