Building localised LLMs for NZ: Some challenges and opportunities

Alistair Knott

Victoria University of Wellington - Te Herenga Waka
The Social Data Science Alliance Global Partnership on AI

Why would we want to build our own LLMs in NZ?

LLMs are revolutionising people’s lives and work.

- This isn’t a prediction about the future—it’s happening right now.

Why would we want to build our own LLMs in NZ?

At present, all our LLM services are provided from overseas.

- Nearly all from the US, from the big companies in Silicon Valley.

The US is not the friendly country it once was. And the big tech companies are creating a monopoly.

- What happens, if the NZ economy becomes increasingly dependent on LLMs?
- Maybe we should be building our own?

It’s a little easier to build LLMs now than it used to be.

Two years ago, OpenAI was the only game in town.

- Now the market for LLM/Gen AI provision is much more competitive.

Three recent developments have levelled the market for providers.

It’s a little easier to build LLMs now than it used to be.

1. Many LLM developers have released open-weights models.
   - Some of these are from private companies. (Meta, DeepSeek)
   - Some are from other sovereign LLMs (e.g. Switzerland’s Apertus).

A new route for sovereign LLMs is to *fine-tune* open-weights models.

2. Efficiency improvements in LLMs.
   - More efficient generators (e.g. DeepSeek, discrete diffusion) make it cheaper to train/deploy large models.

3. Agentic LLMs.
   - Research in LLMs is moving away from training big models, and towards ‘agentic’ systems, that can act sequentially/collectively.
   - The base generators here can be ‘small’ language models.

Framing the sovereign LLM question for AI worriers

Many people are worried about the disruptions coming from Gen AI.

- Especially in NZ.
- How can a sovereign LLM project be presented to worriers?

I like the framing that NZ should adopt a defensive position towards Gen AI.

- Current offerings from overseas come with many risks.
- We can take some steps to mitigate these, by regulating.
- But Gen AI is arriving, one way or another.
- If we develop our own, we may be able to better control the risks.

I'll summarise a few risks—assuming NZ's economy & culture become increasingly dependent on LLMs.

Risks of importing LLMs from overseas

1. Economic risks
   - If our LLMs come from overseas, an increasing proportion of our national budget will flow offshore.
   - If local LLMs are created (and used), money can be retained onshore. And exports are in prospect too.
     - Summary: we want to keep a slice of global AI profits in NZ.

2. Security risks
   - In some overseas LLM provisions, sensitive NZ data can go offshore.
     - That creates security risks - e.g. for government data.
   - If we're dependent on LLMs from other countries, those dependencies can be exploited.
     - E.g. for trade concessions, political concessions.
   - 'Food security'. . . 'energy security'. . . 'AI security'??

Risks of importing LLMs from overseas

3. Privacy risks
   - If LLMs run on overseas servers, personal data is also going overseas.
   - That’s of great concern to some people, and some communities.

4. Political risks
   - LLMs echo the ideology of their creators.
   - The ‘information ecosystem’ is already partly controlled by AI. (Through social media platforms.)
   - As LLMs become prevalent, new influences will arise.

If we make our own LLMs. . .

1. They can run on local servers.
   - And so keep our personal and national data secure.
2. They can be owned and operated locally.
   - And so keep AI profits onshore.
3. They can be aligned to NZ values.
   - How do to this right is of course a huge question!
4. They can be governed locally.
   - We can't govern the LLM services provided by Silicon Valley.
   - If we build our own LLMs, we can set them up as we see fit.
   - The digital / AI realm is a new part of our lives, that governments need to include!

Public vs private local LLMs?

Who should own and operate a local LLM?

I like the idea that it should be a public utility, developed and maintained by the state. (At least for democratic countries.)

- There would have to be a charter of independence of some kind.
- Public broadcasters like the BBC are an interesting model.

Public LLMs also lend themselves well to democratic methods of governance.

Democratic alignment methods for local LLMs?

I like the idea that democratic methods can be used to assemble training sets for ‘utility’ AI systems.

- For instance, the annotations used to train a hate speech classifier can be sourced from the public through opinion polls.
- Our GPAI group did some work on this in India (see GPAI, 2024).

Opinion polling methods could also be used to source judgements for LLM alignment methods.

- For instance, the preference judgements used in Chris Manning’s DPO.
- This may help provide some measure of accountability for the outputs of a ‘public’ LLM.