

# Sovereign Solutions:

Navigating the road  
to sovereign AI



# Table of contents

- 01. Introduction ..... 3**
- 02. Why sovereign AI? ..... 5**
- 03. The sovereign stack ..... 8**
  - 3.1 Land and power ..... 9
  - 3.2 Datacenter design and build ..... 10
  - 3.3 Silicon and servers..... 11
  - 3.4 AI at the edge ..... 12
  - 3.5 Localised AI modelling..... 13
  - 3.6 Sovereign talent ..... 14
- 04. The sovereign reality ..... 15**



01.

# Introduction

# 01.

## Introduction

Artificial intelligence has quickly become the most transformative technology the world has seen in a generation, and development, deployment and control of advanced AI systems is vital to both enterprises and governments.

As the power and utility of AI platforms has increased, so has the need to harness the potential that AI offers. It's safe to say that businesses and even nations that don't invest in the development and implementation of AI infrastructure and applications risk being left behind in a technology race that's showing no signs of slowing.

While the importance and potential of AI are clear, there is an unbalanced reliance on a relatively small number of suppliers within the AI stack, ultimately controlled by an even smaller number of nations. It's this reliance on outside entities that has paved the road towards sovereign AI.

The conceptual premise of sovereign AI assumes that an enterprise or government has complete ownership and control over an entire AI stack, with no reliance on outside entities or foreign governments. In this scenario, AI capabilities are guaranteed regardless of corporate relationships or geopolitical turmoil, and for governments that are considering the use of AI to power public sector services, that resilience is vital.

But how realistic is that vision? Can any single government or corporate entity design, build, control and operate an entire AI stack? And more importantly, what would that mean for the continued development of cutting-edge AI solutions? In practice, sovereign AI is not a binary concept, it's a highly nuanced proposition that requires careful consideration and planning.

This white paper will explore what sovereign AI really means in both absolute and practical terms, and discuss whether owning and controlling full-stack AI infrastructure is really the best option.

---

■ **Sovereign AI is not a binary concept, it's a highly nuanced proposition that requires careful consideration and planning.**

---





## 02.

# Why sovereign AI?

## 02. Why sovereign AI?

The importance and utility of artificial intelligence is undeniable, and as the technology develops and improves the impact that AI has on the world around us will increase exponentially.

In some respects, AI has become the new arms race, as governments and enterprises look to ensure their AI capabilities. But with a finite number of organisations and nations holding the keys to AI functionality, those same governments and enterprises are starting to see the necessity of controlling their own AI destinies.

Sovereign AI, as a concept, removes that dependency and gives nations and enterprises the keys to their own AI opportunities. Control and resilience are core components, ensuring that trade relations or geopolitics will have no impact on the ability to supply AI services. From trade tariffs to global military conflict, shifting geopolitical landscapes could have a direct impact on a nation or enterprise's access to AI. For governments, this is a serious consideration, where public service infrastructure could be reliant on AI functionality. Localised control of AI also ensures regulatory data compliance, with said data stored, controlled and processed within defined borders.

Similarly, a sovereign AI strategy allows regulated industries such as finance and healthcare to implement tight controls over sensitive data, keeping regulatory compliance locked down. For these types of verticals, black box AI solutions provided by external – often foreign – third-party suppliers carry a degree of risk, both in terms of compliance and continuity of service over time.

Much of the rhetoric surrounding sovereign AI focuses on risk mitigation and adopting a defensive position in a world where artificial intelligence can have significant impact on the success and growth of a business or nation. According to recent research from Accenture<sup>1</sup> where nearly 2,000 business and government leaders were surveyed, the biggest driver for considering sovereign AI was regulatory compliance at 46%. Gaining control over critical data and AI models came in next at 28%, while addressing national security was hot on its heels at 27%.

# 46%

of business and government leaders cited regulatory compliance as the key sovereign AI driver.

# 27%

felt that a sovereign AI strategy was necessary for national security.



<sup>1</sup><https://www.accenture.com/us-en/insights/technology/sovereign-ai>



This defensive position of protecting AI capabilities and ensuring regulatory compliance is an understandable driver for sovereign AI policies and solutions, but it's also potentially short-sighted. While it's vital for governments and enterprises to safeguard their ability to deliver AI services, sovereign AI also offers significant value opportunities and the potential to build competitive advantage.

Capitalising on that competitive advantage will require senior buy-in and support in government and corporate environments, but we're already starting to see the seeds of potentially game-changing sovereign projects start to germinate. The OpenBind consortium<sup>2</sup> in the UK is building the world's largest database of information on how drugs interact with proteins. This extensive dataset will then be used to train AI models and accelerate drug discovery, placing the UK at the forefront of clinical research, and potentially saving lives in the process.

Additionally, one of the core pillars of sovereign AI is localised training and inference. At a governmental level, this means tailoring your AI services to the needs of its citizens – think local language, national and regional culture, public sector infrastructure trends, data privacy regulations, etc.

Similarly, for enterprises, tailoring AI services to the needs of their customers will deliver a better experience, generate valuable user data and insight, and potentially pave the way to that promised competitive advantage. In a sovereign AI environment, that's compliant to local data regulations, the depth of understanding that a business can extract about its customer base in that territory could be invaluable.

Government investment, funding and support is also vital for developing sovereign AI building blocks. The UK government has recently implemented a sovereign AI mission<sup>3</sup> backed by a £500m capital fund, providing financial support – between £1m and £10m available to each application – to AI businesses, while also providing access to GPU compute for organisations that need AI infrastructure to grow. The EU's AI Continent Action Plan<sup>4</sup> is also focused on growing localised AI platforms and services, with €200bn of capital committed to boost AI development in Europe. The EU plan also includes provision for up to five AI gigafactories and 19 AI factories within the territory.

<sup>2</sup><https://www.gov.uk/government/news/uk-to-become-world-leader-in-drug-discovery-as-technology-secretary-heads-for-london-tech-week>

<sup>3</sup> <https://www.sovereignai.gov.uk/>

<sup>4</sup> [https://commission.europa.eu/topics/competitiveness/ai-continent\\_en](https://commission.europa.eu/topics/competitiveness/ai-continent_en)



# 03.

# The sovereign stack

## 03.

# The sovereign stack

When we talk about full-stack AI it's easy to focus on the core compute part of the equation, but the reality is far broader, and in the context of full sovereign AI there are many aspects to consider beyond that core compute.

## Land and power

Everyone understands that datacenters form the foundation of AI, with new builds housing ever-more powerful silicon, delivering the high-performance compute for training and operating AI models. But before any concrete can be poured for a new datacenter, the land must be acquired and adequate power, along with other utilities, must be ensured.

Land and power represent the first hurdle of any sovereign AI strategy, and it's a potentially challenging one. Finding suitable locations to build large-scale datacenters isn't easy. Not only does the site need to accommodate the size of the build and have access to adequate power, but there's also the challenge of obtaining the planning permission before the project can start.

This is why governmental support can be a significant benefit at the start of the journey, especially given that so many governments are recognising the need to encourage the development of AI services within their borders. In the UK the government has implemented AI Growth Zones<sup>5</sup> to help address the land and power hurdle. AI Growth Zones are defined areas that are deemed highly suitable for AI datacenter construction. These zones have enhanced access to on-grid power at reduced cost where available, as well as the opportunity for self-build high-voltage infrastructure to expedite datacenter construction. Additionally, there will be fast-tracked planning applications within these zones, potentially removing another blocker.

Initiatives like AI Growth Zones will be key drivers on the road to sovereign AI, mitigating many of the early-stage barriers, and encouraging foundational investment. But while programmes like AI Growth



Zones try to ensure that these sites have access to enough power, there is still the question of where that power is coming from.

Many nations are heavily reliant on foreign energy, so even if the entire AI stack is owned, operated and located domestically, the government isn't truly in control of that AI infrastructure, because the power that keeps it running is not produced in-border.

Investment in renewable energy with battery energy storage systems (BESS) would be the ideal foundation for a sovereign AI stack, delivering a more sustainable AI solution from the ground up. But whatever form it takes, there must be adequate power generation within the territory to facilitate any sovereign AI plan.

<sup>5</sup> <https://www.gov.uk/government/publications/delivering-ai-growth-zones/delivering-ai-growth-zones>

## Datacenter design and build

With land and power secured, the next step in the journey is designing and building the AI datacenters. Ironically, one of the most valuable tools for designing AI datacenters is AI itself. AI powered digital twin<sup>6</sup> technology allows designers to simulate datacenter builds, predicting operational parameters over time, and ensuring that the right build decisions are made before the costly process of actually constructing anything.

Modern, modular datacenter designs have significantly accelerated build and deployment schedules, while also facilitating scalability over time as needs evolve. Traditional datacenter builds can take between 24 – 36 months from breaking ground to deployment, but with AI developing and evolving at such pace, that's simply not viable anymore. Modular and prefabricated designs can be deployed in 9 – 12 months<sup>7</sup>, ensuring that installations are not out of date before they even switch on.

Modular scalability is also key, allowing datacenters to be built at a defined output level for launch, but with the flexibility and elasticity to add capacity – and improved density – at a later date, increasing performance to align with growing needs over time. The ideal scenario for governments would be working with a local modular datacenter provider, ensuring that the ongoing modular supply chain remains sovereign as the infrastructure evolves.

The hyperscale datacenters tasked with training and executing highly-complex AI models require copious amounts of processing power, and the more processing power that can be packed into the datacenter, the more valuable it will be, especially for AI. To facilitate this need, compute density has reached a level where traditional air cooling is no longer viable, meaning that advanced liquid-cooling solutions must be baked into hyperscale datacenter designs. Consequently, datacenter developers with experience and expertise in liquid-cooling are ideally positioned to design, build and deploy these cutting edge, modular hyperscale facilities.



**24 – 36 months**  
Traditional datacenter

vs.

**9 – 12 months**  
Modular datacenter

<sup>6</sup> <https://www.nvidia.com/en-gb/glossary/digital-twin/>

<sup>7</sup> <https://submer.com/design-build/>

## Silicon and servers

AI requires incredible amounts of silicon. While the AI conversation generally revolves around GPUs, that's only one part of that silicon equation. Besides the GPUs, there's also intense need for CPUs, memory and storage, while high-speed interconnect technology is also crucial, ensuring that all those components are communicating and operating as quickly and efficiently as possible.

The latest generations of super-chips place all that compute, memory and interconnect technology on-die, allowing for unified memory models and incredibly fast allocation and access between GPU and CPU workloads. All this is made possible by advanced fabrication processes – the latest NVIDIA Vera Rubin<sup>8</sup> platform is based on a 3nm process, delivering new levels of high-density compute.

That cutting edge silicon must be built into servers, and those servers housed in racks. The design and construction of those servers and racks is important because all that high-density compute generates significant heat, so cooling must be a core consideration from the outset. Also, for hyperscale datacenters, it's not just on-chip compute density that matters, it's compute density throughout the entire facility – the volume of servers and racks

that each datacenter module can house will dictate overall performance.

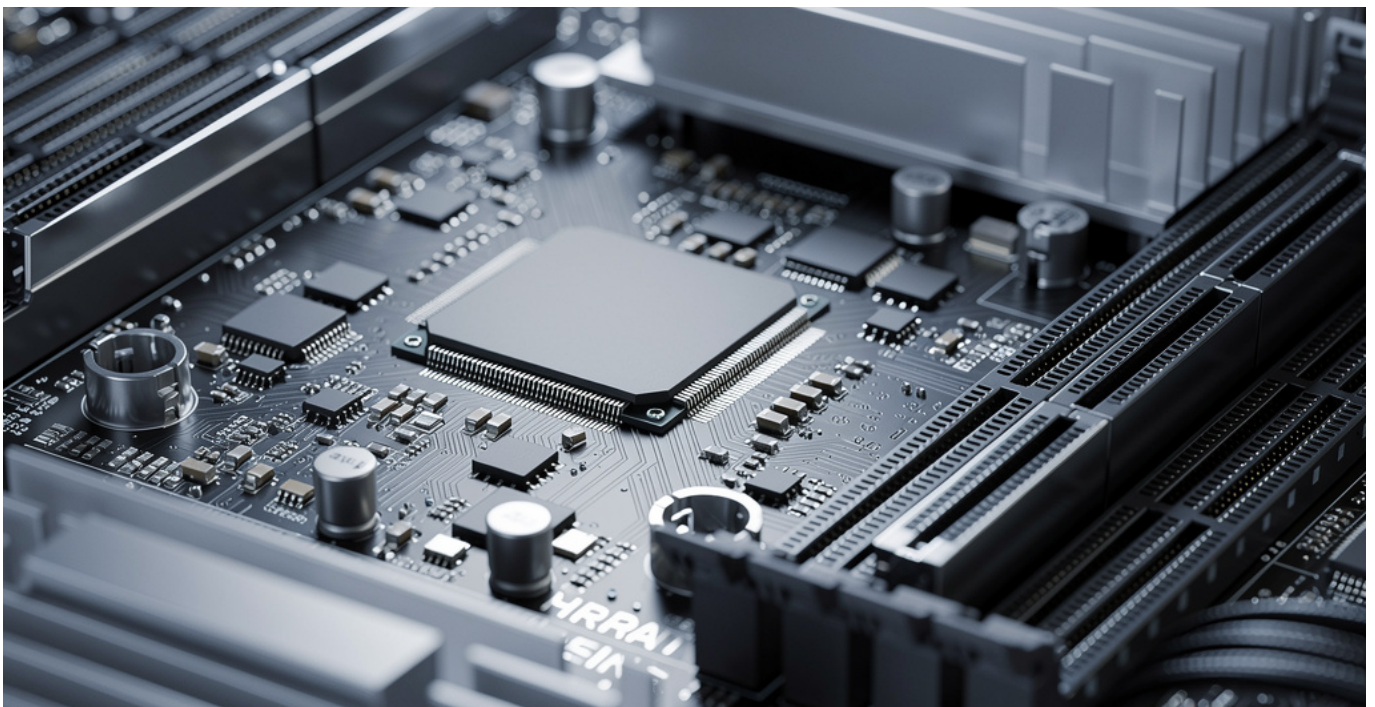
Direct liquid cooling (DLC) marries the need for cooling efficiency in a high-density compute environment, with the need for overall compute density across the entire datacenter – more chips, more servers, more racks, more performance.

Any sovereign AI strategy must account for the hardware that will populate hyperscale datacenters, ensuring that said hardware is owned outright and operated independently of any outside entities.

---

■ **The latest NVIDIA Vera Rubin platform is based on a 3nm process, delivering new levels of high-density compute.**

---



<sup>8</sup> <https://www.nvidia.com/en-gb/data-center/technologies/rubin/>

## AI at the edge

While hyperscale datacenters with their powerful high-density compute capabilities are vital for developing, training and running large language models, when it comes to inference, distributed low-latency compute is key. Deploying compute nodes at the edge – close to where the data is captured – allows inference workloads to be executed with minimal delay. In a sovereign AI environment, where public service infrastructure could rely on AI applications and platforms, low-latency processing of local data will be crucial.

Edge compute also provides the opportunity to develop hyper-localised models and applications. Whereas LLMs will always be developed and trained in large-scale cloud-based datacenters, AI factories at the edge can utilise small language models trained specifically on hyper-local needs, to run applications and services tailored to local users.

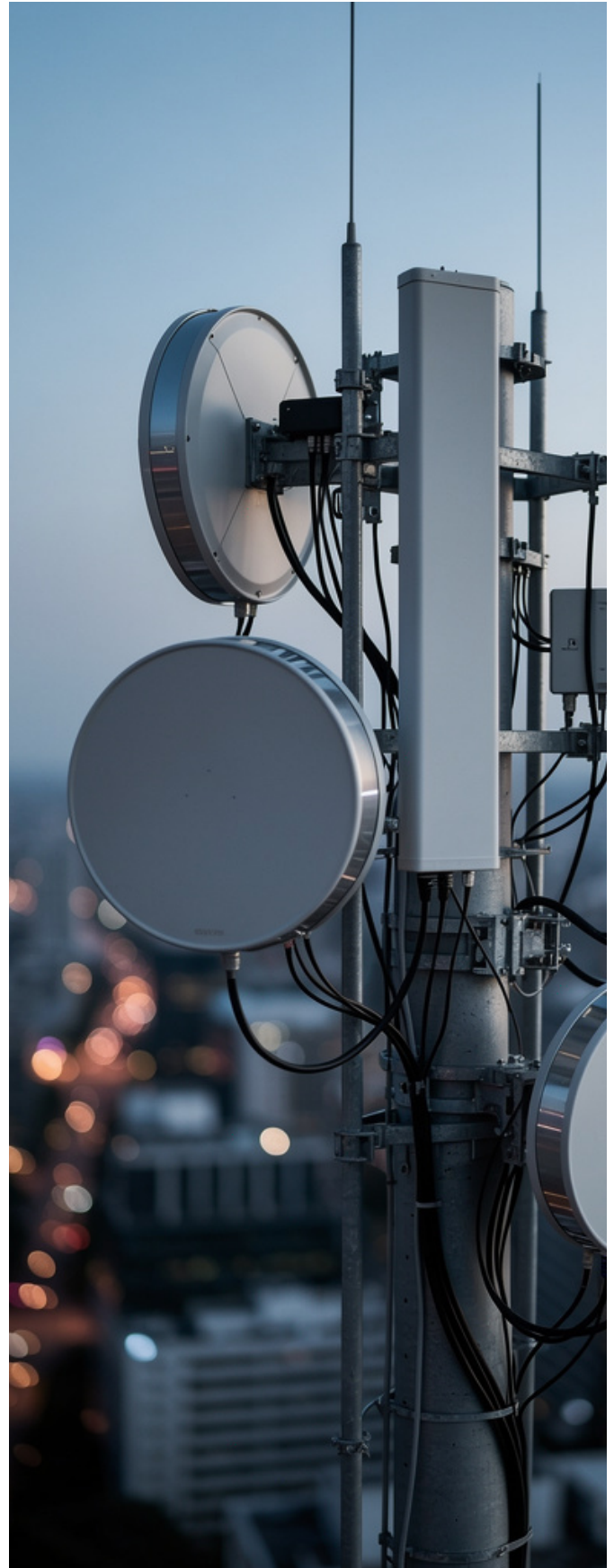
---

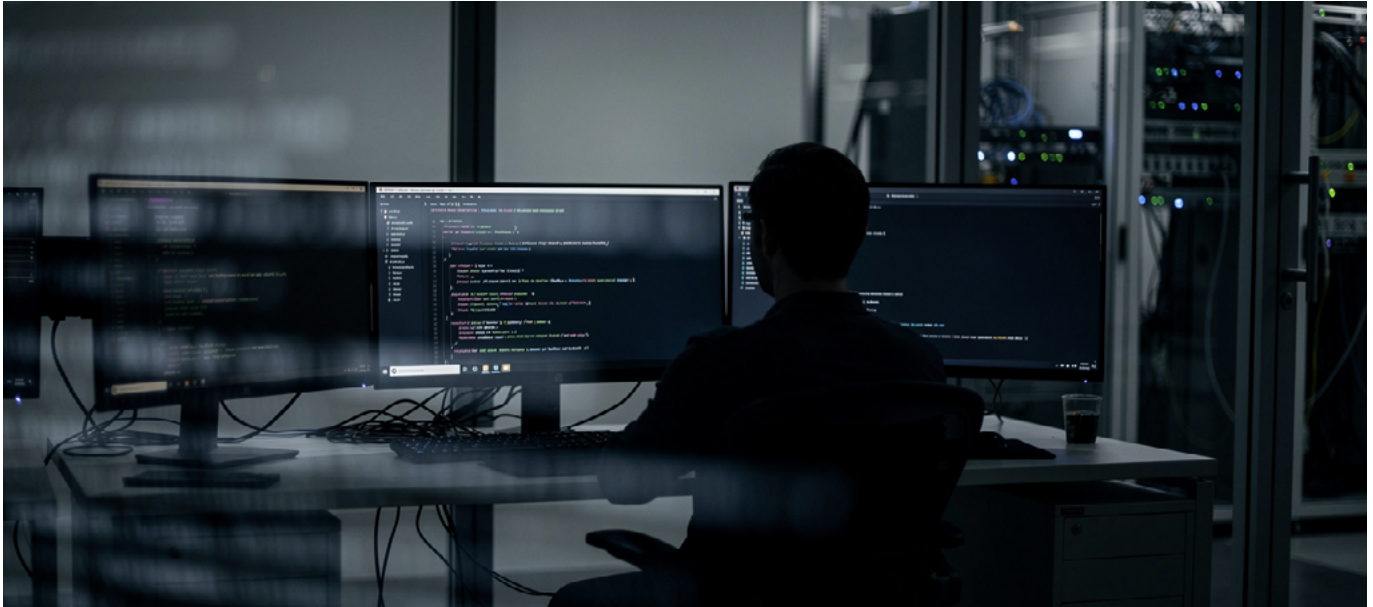
■ **AI factories at the edge can utilise small language models trained specifically on hyper-local needs.**

---

Sovereign edge compute can also be deployed far quicker than hyperscale datacenters, augmenting existing telecommunications infrastructure with AI servers housed in immersion cooling tanks providing low-maintenance, high-performance inference compute directly connected to 5G networks and fibre-optic broadband hubs.

And with edge compute nodes processing data at the point of collection and returning the results to the user, there are no issues with data sovereignty regulations, since that data is barely leaving the point of origin, and therefore never leaving jurisdictional borders.





## Localised AI modelling

Any AI application is only as good as the model it's based on, and any model is only as good as the training it has undergone. A sovereign AI solution needs an LLM that has been trained locally, making it more aware of the needs of the users that will be interacting with it. This includes local languages, dialects, customs, geography, transport infrastructure, etc. Not to mention local laws, regulatory requirements, taxes, national statistics, and so on.

Using an 'off-the-shelf' cloud-based LLM for localised applications and services at best won't deliver a good user experience, and at worst, could present compliance violations with local laws and regulations. This is why localised LLM development and training is a key component of a sovereign AI stack, and why governments are investing so much time, money and expertise into it.

In September 2025 Switzerland announced its national LLM – a locally developed and trained large language model called Apertus<sup>9</sup>, designed to provide a sovereign alternative to the cloud-based, corporate LLMs. Apertus has also been designed as an open platform<sup>10</sup>, so other nations or organisations can use, modify and build on its foundation.

---

**The development of localised LLMs is not just a trend, it's a core building block for sovereign AI.**

---

The UK-LLM project<sup>11</sup>, led by University College London, Bangor University and NVIDIA, is also working towards a locally developed AI model trained on UK data in both English and Celtic languages, including Welsh, Scottish Gaelic and Irish. And as with Apertus, this will be an open platform to build upon.

The development of localised LLMs is not just a trend, it's a core building block for sovereign AI, allowing for applications and platforms to be developed specifically for the people who will be using them.

<sup>9</sup> <https://sifted.eu/articles/switzerland-national-llm>

<sup>10</sup> <https://apertvs.ai/>

<sup>11</sup> <https://www.bangor.ac.uk/news/2025-09-15-reaching-across-the-isles-uk-llm-brings-ai-to-uk-languages-with-nvidia-nemotron>

## Sovereign talent

Datacenters, infrastructure and models are all vital parts of a sovereign AI solution, but without the people that can design, deploy, develop, train and operate that AI solution – not to mention create the applications and services that run on it – it won't amount to much.

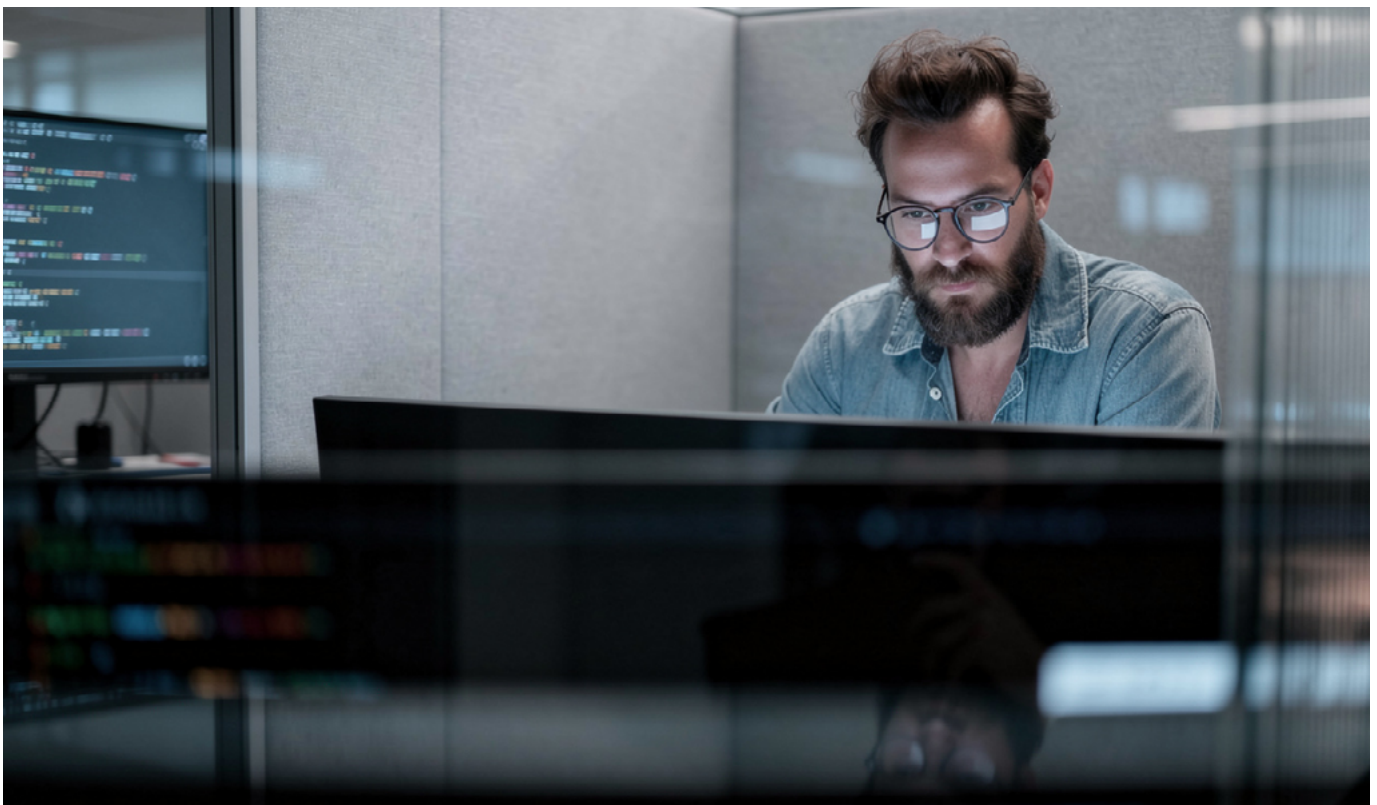
Nations and enterprises that are looking to build out sovereign AI must ensure that they have the necessary people to do so. There is no easy fix to this part of the problem, talent must be grown, developed and encouraged to stay local.

Government initiatives are key to building and growing the necessary talent pool, making education and training for these roles more attractive to the younger generation. Likewise, corporates need to highlight defined career opportunities, graduate training schemes, internships with real prospects attached, all of which can help persuade a new generation of AI-focused students and trainees, that can eventually deliver on government and enterprise sovereign AI ambitions.

---

**Nations and enterprises that are looking to build out sovereign AI must ensure that they have the necessary people to do so.**

---



■ 04.

# The sovereign reality

## 04.

## The sovereign reality

In theory, the ultimate implementation of sovereign AI is ensuring that you control every part of the stack described above – making sure that power is generated domestically, that all hardware and silicon is owned outright, that both core and edge compute infrastructure is contained and controlled within borders, that AI models are developed and trained on sovereign relevant datasets, that all data regulations are enforced, and that the necessary talent is developed and retained over time.

Not only is that a huge undertaking, but it also represents a snapshot in time. While a government or enterprise might own the silicon and hardware outright for the initial sovereign AI rollout, what happens when that hardware needs to be upgraded? Unless said government and enterprise also owns chip fabrication and hardware manufacturing facilities, they will be beholden to outside entities once again. And even if they do own and control those facilities, are they located domestically?

The reality is that no nation or organisation can build and maintain a full sovereign AI solution, and if we accept the premise that sovereign AI should be a route to competitive advantage rather than protectionism, that's not necessarily a bad thing.

A recent report from the Tony Blair Institute for Global Change – Sovereignty in the age of AI<sup>12</sup> – highlighted that “No state can dominate every layer of the AI stack. Leaders must make deliberate choices about where they want to build strength and influence.” Suggesting instead that by becoming leaders in key sections of that stack, they can build influence and become part of something bigger.

That bigger vision will ultimately drive innovation and keep pushing the boundaries of what AI is and can be capable of. Creating an isolated AI structure and relying only on the tools and talent available within that isolated structure will inevitably stifle innovation.

---

■ **No state can dominate every layer of the AI stack. Leaders must make deliberate choices about where they want to build strength and influence.**

---

While the fear of overreliance on established hyperscalers, cloud providers and LLM specialists is understandable, building stronger two-way partnerships within the ecosystem can mitigate that overreliance, while still implementing the parts of a sovereign AI infrastructure that deliver true value opportunities to nations and enterprises.

Even significant hurdles such as data privacy regulation can be addressed in a more collaborative sovereign model. Data federation and confidential computing can facilitate the broader use of datasets for training, inference and analysis without that data ever leaving its sovereign base or even being openly accessible. Essentially, the insight can be extracted from datasets without that data ever being visible to outside entities.

Not only does this allay fears around data regulation compliance, it also allows enterprises in regulated industries to share data and build more robust AI models without ever actually revealing their data to competitors.

<sup>12</sup> <https://institute.global/insights/tech-and-digitalisation/sovereignty-in-the-age-of-ai-strategic-choices-structural-dependencies>



So, what does a more realistic, collaborative sovereign AI solution look like? Ideally, it should marry the cutting-edge innovation of industry leaders with the needs of localised, defined audiences, while ensuring local regulatory compliance. A truly workable and positive sovereign AI strategy will require negotiation, agreements, partnerships, and ultimately a degree of trust.

Without this type of collaborative approach, we run the risk of AI genuinely becoming an arms race, the likes of which we haven't seen since the Cold War. We don't want a world where nations close their doors and work on the most important technology of our time in isolation for fear of being left behind. In that scenario the whole world gets left behind, robbed of the breakthroughs that collaborative frontier AI development could deliver.

Like most things, sovereign AI is not binary, it should be seen as a flexible framework for governments and enterprises to develop AI solutions that can deliver more tailored applications and platforms for defined audiences within specific regions or verticals.

Sovereign AI is an opportunity to make this technology work more effectively for specific audiences – whether that be based on geography, language, nationality or industry sector – while continuing to push the boundary of what it can do at a global level. AI is at a crossroads and we need to choose the direction we travel carefully.

**inferX**

# Real-world intelligence

[Learn more at our blog](#) →

