

AI Compute Sovereignty: Infrastructure Control Across Territories, Cloud Providers, and Accelerators

Zoe Jay Hawkins, Tech Policy Design Institute & Oxford Internet Institute, University of Oxford
Vili Lehdonvirta, Department of Computer Science, Aalto University & Oxford Internet Institute, University of Oxford
Boxi Wu, Oxford Internet Institute, University of Oxford

Abstract: The concept of ‘compute sovereignty’ has become a focal point in government and industry discussions on artificial intelligence (AI) governance. What is it and who has it? Based on previous literature, we propose to break these questions down to three levels: (1) how much AI compute a country has on its territory, (2) what is the nationality of the companies who own the AI compute data centres, and (3) what is the nationality of the accelerator vendors whose chips power the AI compute data centres? We examine these questions empirically through the lens of cloud computing infrastructure, focusing on nine leading public cloud providers that represent approximately 70 percent of the global market. The data is collected using a methodology previously published in Lehdonvirta, Wu, and Hawkins (2024). The findings suggest that the possession of “compute sovereignty” varies between countries depending on the level of analysis. Determining the most relevant level depends on governments’ policy aims and national contexts, and involves policy trade-offs. Policies aimed at attracting data centres to a country’s territory can enhance supply security of critical computational resources while also introducing increased consumption of energy, water and land use resources, with corresponding localised socioeconomic and environmental impacts. Regional and supply chain approaches involve different trade-offs.

1. Introduction

In a time of increasing geopolitical competition and uncertainty, the location, ownership and control of artificial intelligence (AI) technologies and their underlying material infrastructures have become strategic considerations for governments (Wang and Chen, 2018; Miller, 2022). AI compute – the specialised computational resources that are required to train and run AI models – has emerged as a sought-after commodity in a global race for AI leadership. The amount of compute required to train frontier AI models has doubled approximately every six months (Sevilla et al, 2022), driven by continued pursuit of computational scaling of AI performance (Pilz et al., 2025). For example, progress in advanced reasoning capabilities in models such as OpenAI’s o3 or DeepSeek’s R1 have relied on continued computational scaling during inference time. NVIDIA CEO Jensen Huang claimed that “the amount of computation we need as a result of agentic AI, as a result of reasoning, is easily 100 times more than we thought we needed this time last year” (Nellis & Cherney, 2025). The perceived need to secure sufficient access to AI compute for domestic firms, public sector organizations, and/or researchers has come to be referred to using such terms as “sovereign AI compute” or “compute sovereignty” (Ghahramani, 2023; Sastry et al., 2024; Lehdonvirta et al., 2024; Government of Canada, 2025). As policymakers seek to assert regulatory authority, ensure supply chain resilience, and protect strategic autonomy, ‘compute sovereignty’ is becoming a topic of interest in debates about AI governance (Sastry et al., 2024).

Sovereignty – understood broadly as a government’s supreme authority within a given territory – has a complex and contested relationship with digital technologies (Pohle & Thiel, 2020). Policies pursuing digital sovereignty have evolved significantly over the last three decades and particularly since the early Internet debates of the 1990s. Different stakeholders hold differing views on digital sovereignty; from autocratic arguments regarding the primacy of the nation state in Internet governance (as opposed to multistakeholder approaches), through to nations’ increasing pursuit of autonomy from foreign government interference (Musiani, 2024). In western democracies, digital sovereignty has recently been conceptualised through the lens of national or regional strategic autonomy in the digital realm (Blancato, 2024).

Amid rising competition to develop advanced AI systems and the corresponding scarcity of computational resources, notions of ‘digital sovereignty’ are increasing entwined with debates about the governance of AI and the emerging paradigm of compute governance (Roberts et al., 2023; Sastry et al., 2024; Lehdonvirta et al., 2024). Since 2022, almost all western democratic countries have developed policies and initiatives to increase strategic autonomy over computational resources, including a turn to industrial policies that aim to onshore the compute supply chain. In doing so, governments are assessing their dependence on corporate and state actors who own or control key nodes in the vast, global supply chain that characterises AI compute; from the owners of AI data and models, to the producers

of valuable AI accelerators and the developers and operators of the data centres that house accelerators. Through these assessments – such as the UK’s Compute Review – countries seek ‘resilience’ as well as concepts of ‘sovereignty’ (Ghahramani, 2023).

And yet, the conceptualisation, trends, and merit of the concept of ‘compute sovereignty’ have been comparatively under researched, nor analysed for their implications for the governance of AI. To contribute to the literature, this study: 1) situates the concept of ‘compute sovereignty’ in the broader scholarly context of ‘sovereignty’, ‘digital sovereignty’, and the geopolitics of AI; 2) offers a three-level framework for conceptualising compute sovereignty; and 3) provides an initial evidence base drawn from a unique census of the global AI compute infrastructures of nine leading public cloud computing providers based in the United States, Europe, and China, including Amazon Web Services (AWS), OVHcloud, and Alibaba Cloud. The term ‘public AI compute’ refers to cloud-based compute infrastructure made commercially available to users, as opposed to public sector cloud. For many AI model developers and users, AI compute is primarily accessed via the cloud computing platforms of these providers, making them a powerful empirical entry point into the phenomenon. They are not the only sources of AI compute in many countries, so the picture they provide is incomplete, but nevertheless valuable in illustrating how the question of which countries possess ‘compute sovereignty’ depends on which level of analysis we choose.

The remainder of the study proceeds as follows: a literature review of digital sovereignty and existing concepts of compute sovereignty; the methodology for conducting a global census of major providers of public AI compute; discussion of the findings’ implications for national policy considerations, international relations (IR), and global tech trends.

2. Literature Review

2.1 Digital sovereignty debates

In the fields of political science and IR, sovereignty is a complex and well-studied concept, which has evolved from 16th century notions of monarchical sovereignty to contemporary associations with democracy, the rule of law and territoriality, often emphasising the importance of sovereignty as geographically bound to a specific territory (Pohle & Thiel, 2020). Within this literature, sovereignty is discussed through the subvariants of international legal sovereignty, Westphalian sovereignty, domestic sovereignty, interdependence sovereignty, or formal and operational sovereignty (Krasner, 1999; Mueller, 2010). Applied to digital sovereignty, debates focus on a national governments’ ability to deploy and regulate technology within its borders, reflecting a broadly Westphalian notion of sovereignty.

For cyber libertarians, including prominent American proponents, such as John Barlow, the digital realm is happily beyond the reach of state power and should remain outside of jurisdictional control (Barlow, 1996; Hawkins & Weaver, 2025). For most of the last few decades, liberal democratic governments held up this more collaborative approach between governments and other stakeholders as a strength and intrinsic value of the Internet. In contrast, the importance of ‘sovereignty’ was the calling card of autocratic governments such as China and Russia seeking greater state control and pushing back on the United States’ multistakeholder vision of the Internet, where the state was just one of several stakeholders that had a role in governing the Internet (Hofmann, 2016; Raymond & DeNardis, 2015; Komaitis et al., 2024). The distinction between these positions was relatively bright and in the United States, where the notion of digital sovereignty has principally had a negative connotation (Couture & Toupin, 2019).

Since then, this relatively clear distinction has broken down. In the 2010s, the term took on new meaning following the Edward Snowden revelations, where it was adopted by European governments as they sought to reduce dependence on American digital networks to reduce risk of foreign surveillance (Hill, 2014). Democracies now invoke the concept of digital sovereignty in different ways, adding further conceptual blurriness to the term’s meaning (Musiani, 2024). As geopolitical tensions between the US and China have risen, many ‘third countries’ have found themselves considering what digital sovereignty looks like in the receipt of cloud investments from these two digital superpowers (Lehdonvirta et al, 2025). With the rise of Big Tech corporate power, the incentive for governments to assert their sovereignty has expanded from just pushing back on foreign governments but also to manage the influence of private companies (van Dijck, 2020). The perceived agency of Big Tech is such that scholars have also explored the meaning of ‘digital sovereignty’ as applicable to private corporations (Floridi, 2020; Lehdonvirta, 2022). Concurrently, the supply chain shocks of the COVID-19 pandemic intensified national conversations about dependence on international supply chains and the pursuit of resilient infrastructure. As such,

the pervasiveness and criticality of digital infrastructure to modern state, economy and society has increased, along with growing interest in policies that evoke digital sovereignty.

2.2 Compute Sovereignty

Debates about ‘digital sovereignty’ and its meaning in academic and policy discourse have most recently been impacted by the huge interest in AI. Many governments see AI as a potential source of national economic, military, and/or cultural advantage (Wang and Chen 2018; Miller 2022), and have consequently adopted policies aimed at securing sufficient access to AI compute resources for local industries and researchers (Vipra and West 2023). The term ‘compute sovereignty’ and adjacent terms such as ‘sovereign AI compute’ have emerged as something of a subcategory of digital sovereignty to describe this perceived need to secure access to AI compute resources (Ghahramani, 2023; Sastry et al., 2024; Lehdonvirta et al., 2024; Government of Canada, 2025).

Compute sovereignty has been pursued through industrial policies to onshore components of the compute supply chain. The *US CHIPS Act (2022)* authorised 39 billion USD in grants and loans, alongside tax credits and government research and development investments into US chip manufacturing, while the *European Chips Act (2023)* proposes to invest 43 billion euros of public funds up to 2030 to ensure Europe’s ‘strategic autonomy’ in the supply of critical chips including a Chips Fund to support European chip production including the onshoring of foundries (Shivakumar et al., 2024). These policies are often reliant on multi-billion dollar compute investments from hyperscale cloud providers who in practice deploy the AI accelerator chips in their data centres and make them accessible to developers and users. In 2025 alone, Microsoft plans to spend \$80 billion USD on global AI data centre buildouts (Smith, 2025). These investments are welcomed by governments who, through initiatives such as the *UK’s AI Opportunities Act (2025)*, propose access to land and cheaper energy, as well as fast-tracked development permits for data centre developers.

Other approaches to compute sovereignty have involved the development of public sector compute resources. Both the *UK Compute Review (2023)* and the *Canadian Sovereign AI Compute Strategy (2025)* propose investments into public sector supercomputing infrastructure to bolster domestic compute availability. This trend is likely to continue, with forecasts suggesting that worldwide ‘sovereign cloud’ spending will reach more than \$250 billion in 2027 (IDC, 2023). This global rush for ‘sovereign AI’ has been described by scholars as ‘the 21st century’s arms race’ (Pratschke, 2025).

Like the conceptual blurriness of digital sovereignty, compute sovereignty has different meanings for different actors. Defining it is complicated by the complexities of the AI compute supply chain, which is highly international and characterised by significant market concentration and vertical integration at different steps (Vipra & West, 2023; Beaumier & Cartwright, 2024). In very broad terms, the supply chain consists of companies that consume AI compute to develop and/or deploy AI systems; companies that own data centres that supply that AI compute; and companies that design and fabricate AI accelerators housed in those data centres (Sastry et al., 2024). Several of these links are characterised by intense political and economic interdependence which governments can weaponise as a geoeconomic tool to attempt to exert influence over other countries (Farrell & Newman, 2019; Beaumier & Cartwright, 2024; Lehdonvirta et al., 2025).

In the following sections, we propose to distinguish between three different understandings or ‘levels’ of AI compute sovereignty discussed in previous academic literature and policy documents: compute sovereignty as jurisdiction over (1) data centres, (2) companies that own the data centres, and (3) companies that provide the AI chips that power the data centres.

2.3 Compute sovereignty as territorial jurisdiction over AI compute data centres

The broadest interpretation of compute sovereignty is having AI compute resources physically situated within the borders of the government’s territory, referred to as ‘in-country compute’ (OECD, 2025). This conception is based on the Westphalian sovereign state model which links a government’s supreme authority to govern with a defined physical territory (Krasner, 2001). Governments that exercise territorial jurisdiction over physical compute infrastructures have the opportunity to regulate them in a way that other governments cannot (Sastry et al., 2024; Ferrari, 2024). Having territorial jurisdiction over AI compute data centres empowers governments for instance to enforce Know-Your-Customer checks, audits, or other controls at the point at which AI models and data sets enter a data centre for training or deployment (Sastry et al., 2024). For this reason, for some governments there is a

motivation to increase the territorial nexus with foreign firms providing AI compute cloud computing services. Countries attempt to shore up their territorial claims over in-country compute through both inducement (investment incentives) and regulation (such as data localisation laws).

In practice, multinational corporations also embrace this territorial interpretation of 'sovereignty'. Adopting this definition means that they can offer 'sovereign' digital services to foreign governments simply by situating sufficient proportions of their infrastructure, business functions, staff, and/or other assets within the host country's geographic borders. Microsoft offers its *Microsoft Cloud for Sovereignty program* which 'provide the necessary control without trading off cloud value' (Microsoft, 2024). The AWS Digital Sovereignty Pledge offers a 'commitment to offering AWS customers the most advanced set of sovereignty controls and features available in the cloud' (AWS, 2022). Similarly, Huawei's National Government Cloud Solution claims that its 'on-premise deployment ensures digital sovereignty' in China (Huawei, n.d).

2.4 Compute sovereignty as the nationality of the AI compute cloud providers

For some policy makers and local competitors to foreign cloud providers, AI compute being located within one's territory provides insufficient sovereign control over compute resources (Michels, 2025). In this view, the *nationality of the cloud provider* managing the provision of AI compute from the data centres within one's territory is consequential. This ownership dynamic can be framed as determining both a '*freedom to*' and '*freedom from*' perspective. Firstly, *possessing* capability and capacity to provide computational resources through locally domiciled organisations is seen as a national advantage for economic and security reasons – a form of technological self-sufficiency towards 'innovation power' (Schmidt, 2023). Secondly, and just as importantly, this approach is perceived to offer the benefit of *preventing* a foreign government's digital control over critical infrastructure. In a post-Snowden world, some are not comforted by AI compute facilities being located within their own jurisdiction if the cloud provider is a foreign company.

In this case, the interests of large corporations and their home governments are assumed by some to be aligned to the extent that corporate expansion represents a form of 'digital colonialism' where global reliance on their digital infrastructure can be weaponized by their home government (Kwet, 2019; Farrell & Newman, 2019). Even for those who do not suspect intentional state-corporate partnership, there is a remaining concern that a corporation may be forced to comply with foreign government direction. This could relate to a home government's opportunity to engage in surveillance, where "a provider who is subject to foreign jurisdiction might disclose data to a foreign government without a customer's knowledge or permission" (Michels, 2025). Others point to the long arm of US legislation - including the *Foreign Intelligence Surveillance Act 1978* and *Clarifying Lawful Overseas Use of Data Act 2018* - as posing risks of jurisdictional reach into US companies. These concerns are reflected in 'trusted vendor' frameworks, adopted by governments to mitigate the perceived risks of Chinese government interference in Chinese owned digital infrastructure such as 5G networks (Botton & Lee-Makiyama, 2018). Another way of pursuing immunity from foreign laws is the European Commission's decision to preclude majority foreign-owned corporations from receiving the highest certification under the bloc's cloud security standard, modelled after France's SecNumCloud.

In practice, compute sovereignty based on provider nationality is a challenging goal for most countries, particularly those already dependent on foreign compute providers. Given the intensely concentrated global cloud market - dominated by six hyperscale cloud providers (US corporations AWS, Google Cloud and Microsoft Azure, and Chinese corporations Alibaba Cloud, Huawei Cloud and Tencent Cloud) – cultivating locally-owned competitors with the capital and scale to meaningfully compete the global giants is no small feat (Lehdonvirta et al., 2024, 2025). As a result, in practical terms, most countries must engage in strategic trade-offs between digital infrastructure investments from US or Chinese cloud providers (Lehdonvirta et al., 2025).

While the global distribution of AI compute resources discussed in this study are a product of the investment patterns of private corporations, these corporate decisions are increasingly informed and constrained by government legislation and policy narratives around the world that frame access to - and governance of - AI compute resources as a *national security issue*, and the race for AI dominance between the US and China as a an *international security issue*. In this context, we can borrow from IR theories of great power competition, to view the global distribution of US versus Chinese owned AI compute cloud providers as resulting in individual countries being in a position of either 'hedging' or 'aligning' their compute interests in the context of US-China competition. In IR theory, 'hedging' refers to a mixed strategy deployed by a country to avoid 'picking sides' between competing powers

(Goh, 2005, 2007). According to hedging theory, countries that do not want to choose between the US and China ‘hedge’ relations and refuse a specific alignment (Goh, 2005). Much of the literature focuses on the strategies of countries in the Asia-Pacific and their relations with the US or China (Ciorciari and Haacke, 2019). In contrast, some countries pick sides by aligning with a powerful state, either to oppose the threatening rise of another (‘balancing’) (Waltz, 1979), or to join a threatening power to avoid the costs of opposing it (‘bandwagoning’) (Walt, 1987). Regardless of intent, both result in a country ‘aligning’ with a particular powerful state. These categories of ‘hedging’ and ‘aligning’ can be applied as a metaphor to describe third countries’ situations when it comes to the nationality of the cloud providers that operate AI compute infrastructure on their territories—are they of a single foreign nationality (aligning) or of multiple (hedging).

2.5 Compute sovereignty as AI compute accelerator vendor nationality

The third, and most granular conception of compute sovereignty centres on the nationality of the providers of AI accelerators. Even if data centres are located within a country’s territorial jurisdiction, and owned by a company headquartered in that country, their ability to continue to provide compute relies on the supply of advanced AI accelerators. In this framing, compute sovereignty is impacted by the major suppliers of this hardware: the US chip design firm NVIDIA, and US and Chinese hyperscale cloud providers.

Accelerators are made from semiconductor materials which make the supply of accelerators heavily dependent on the semiconductor supply chain. This supply chain is highly concentrated and consists of many steps where control could be exercised; to simplify, the main steps are design, fabrication and distribution, with key processes distributed across centralised nodes (Sastry et al., 2024; Thadani & Allen, 2023). The US design firm NVIDIA designs 80-95% of chips with Intel and AMD being marginal competitors, while 90% of semiconductor fabrication occurs in Taiwan by the Taiwan Semiconductor Company (TSMC) which is in turn entirely reliant on lithography equipment supplied by the Dutch ASML company (Miller, 2022). Finally, chips are distributed via NVIDIA to compute providers such as AWS, Google and Microsoft Azure who make up over 70% of the global cloud compute market. Although chip fabrication is largely outsourced to foundries such as TSMC, the vendors retain licensing rights and often embed export restrictions into their supply chain contracts. Moreover, TSMC does not have design capabilities, meaning that the US, as the home-state of leading vendor NVIDIA, has notable control over the supply chain (Beaumier & Cartwright, 2024). However, this supply chain has long been entwined with geopolitical tensions and concerns about the insecurity of this concentrated global supply chain (noting the fragility of the ‘Silicon Shield’ argument that Taiwan’s role in the semiconductor supply chain elevates its importance to the US and deters Chinese aggression (Weil et al., 2024)). The major US and Chinese hyperscale cloud providers have all increased investments into in-house accelerators in an attempt to compete and diversify the market.

The emergence of in-house accelerators has been intensified by the geopolitical tensions by export controls seeking to shape the global accelerator market by restricting China’s access to US accelerators. The US Government’s *CHIPS Act (2022)* prevents NVIDIA from selling its most advanced accelerators to China with the intention to prevent cloud providers and other data centre operators in China from building significant AI compute resources (Allen, 2024). While the policy appears to have been successful in diminishing Chinese developers’ access to AI compute in the near term, it has incentivised Chinese technology firms and the government to invest into sovereign AI accelerator design and fabrication capabilities. The leading public cloud provider Huawei, recently partnered with SMIC to develop an advanced semiconductor (a 7nm chip) with capabilities that would otherwise have been prohibited within US export controls (Gupta et al., 2024). Alibaba and Tencent, have likewise also announced their own AI accelerator chip development programmes.

As with hyperscale public cloud computing, it is hardly realistic for most countries to move to autarky and rapidly develop a domestic AI accelerator industry to secure ‘compute sovereignty’ for themselves (Beaumier & Cartwright, 2024). Instead, any active attempts by governments to ensure supply security and avoid dependence on this level would revolve around ‘friendshoring’ (Maihold, 2022) or choosing which foreign AI accelerator chip vendors to depend on. As on the previous level, we can conceptualize countries’ situation in this regard as either ‘hedging’ or ‘aligning’, depending on whether they rely on chips from multiple nationalities or just one.

3. Methodology: Census of Public Cloud Regions

Our methodology builds on prior research mapping the political geography of US and Chinese hyperscale public cloud providers through the lens of data centres and AI accelerators (Lehdonvirta et al., 2024, 2025). In this study, we conducted a census of US, Chinese and European public cloud providers' cloud region (defined below) and the most powerful AI-relevant accelerators available in each region.

Following the census, we analysed the geographic distribution of regions where customers can access US or Chinese AI accelerators. Our census comprised three leading cloud providers in each of the US, Chinese and European markets considering both representation from major economic markets, as well as the relative market share of each provider. The US, Chinese and European markets capture approximately 78% of global spending on cloud infrastructure (Zhang, 2024), while US and Chinese providers represent the majority of the global cloud market (Lehdonvirta et al., 2025). Although European cloud providers are not significant global players they have a non-negligible role in their region, particularly as European countries pursue compute sovereignty policies that aim to support European providers. The census therefore included AWS, Google and Microsoft (US providers), Alibaba, Huawei and Tencent (Chinese providers), and Exoscale, Hetzner and OVHcloud (European providers) (Table 1).

Table 1. Public cloud providers included in the Census

Provider Nationality	Name of cloud provider	Name of company	Year cloud service established
US	Amazon Web Services	Amazon	2006
	Google Cloud Platform	Google	2008
	Microsoft Azure	Microsoft	2010
China	Alibaba Cloud	Alibaba Group	2009
	Huawei Cloud	Huawei Technologies	2017
	Tencent Cloud	Tencent Holdings	2013
Europe	Exoscale	Telekom Austria	2011
	Hetzner	Hetzner Online GmbH	2018
	OVHcloud	OVH Groupe SA	2010

Source: Cloud providers' websites

Using publicly available data accessed via cloud providers' websites and customer interfaces we collect a list of each provider's public cloud regions as of November 2024. A 'cloud region' is an organisational designation for a cluster of interconnected data centres called 'availability zones', which are located in a specific geographic location and named after a nearby city. A developer wishing to purchase on-demand or subscription-based AI compute would therefore choose the region where their desired type of accelerator is accessible and physically hosted. The chosen coverage of accelerators is based on those considered the most powerful and relevant in the training and inferencing of common AI models in 2024 (Table 2). The accelerators covered are those made available to the public via hyperscale cloud provider's platforms, which comprise NVIDIA accelerators as well as provider's in-house accelerators. This covers a significant portion of the total AI accelerator market because NVIDIA chips make up roughly 80-95% of the total chip market with US hyperscale cloud providers alone make up 40% of NVIDIA's revenue (Sastry et al., 2024).

We have categorised these accelerators into those relevant for both frontier model training as well as inferencing, and those suited predominantly to inferencing. When choosing accelerators, developers will consider the computational needs of the project, as well as cost and availability. Training and inferencing accelerators ('training-relevant') are typically the most power and expensive GPUs (i.e. NVIDIA's A100 or H100) required to train advanced models because they are more computationally intensive (demanding high compute bandwidth). GPUs like the NVIDIA H100 are typically purchased by labs that can afford them, for their training abilities, although they are still considered the optimal GPU for both training and inferencing (Patel & Nishball, 2023). In

contrast, accelerators that are typically used for inferencing (‘inferencing-relevant’), like NVIDIA’s V100, are often more memory intensive for higher throughput (demanding more memory bandwidth).

As smaller, more efficient models enter the market, such as DeepSeek’s R1 or OpenAI’s o3-mini, the distinction between training-relevant compute and inferencing-relevant compute has been drawn into question. However, researchers argue that the current paradigm of AI development is still focused on large-scale frontier models; with many smaller models distilled from larger ones and reliant on them for data generation, evaluation and continued improvement (Pilz et al., 2025). Therefore, the purpose of our categorisation is to separate today’s most powerful accelerators that are best suited for intensive workloads to train cutting edge frontier models, from other accelerators which – while theoretically possible to use for certain training - are fundamentally less powerful. A summary of all variables included in the Census is shown in Table 3.

Table 2. AI accelerators included in the Census and their classification as of 2025

Provider nationality	Provider name	AI Accelerator	Training or inferencing	Year of release
US	NVIDIA	H100	Training-relevant	2023
US	NVIDIA	A100	Training-relevant	2020
US	NVIDIA	V100	Inference-relevant	2017
US	NVIDIA	H800	Inference-relevant	2023
US	NVIDIA	A800	Inference-relevant	2023
US	Google	Trillium	Training-relevant	2024
US	Amazon	Trainium	Training-relevant	2023
US	Amazon	Inferentia	Inference-relevant	2018
US	Microsoft	Maia 100	Training-relevant	2023
Chinese	Alibaba	Hanguang 800	Inference-relevant	2019
Chinese	Huawei	ASCEND 310	Inference-relevant	2018

Source: Details on AI accelerators (Dilmegani, 2024).

Table 3. Census variables (unit of analysis: cloud regions)

Variables	Levels
Provider	9 (AWS...Tencent)
Country	46 (Argentina...United States)
H100 available	2 (Yes/No)
A100 available	2 (Yes/No)
V100 available	2 (Yes/No)
H800 available	2 (Yes/No)
A800 available	2 (Yes/No)
Trillium available	2 (Yes/No)
Trainium available	2 (Yes/No)
Maia 100 available	2 (Yes/No)
Hanguang 800 available	2 (Yes/No)
ASCEND 310 available	2 (Yes/No)
N of cloud regions	225

By focusing our census on public cloud providers, we focus on sovereignty as it relates to compute which is publicly available through purchase or subscription via a provider’s cloud platform. It is worth clarifying that this compute is ‘public’ because of its availability to the public and not because of any government affiliation (i.e. the public in public house, rather than in public sector) (Lehdonvirta et al., 2024). As such, the study excludes private compute which is purchased by corporate or state actors for internal research and development infrastructure; public sector compute deriving from state-owned capabilities or private sector compute used for in-house AI development.

Additionally, our focus on public cloud providers limits the data available for the census. Particularly regarding Chinese cloud providers and accelerators, data is limited to what is available via customer cloud interfaces. When collecting accelerator data for the Census in 2024, there was a notable delay between the announcement and later launch of accelerators, perhaps pointing towards the broader insecurities of the semiconductor supply chain.

4. Findings

4.1 Compute sovereignty as territorial jurisdiction over AI compute data centres

A major way in which “compute sovereignty” has been understood is having data centres running AI accelerators situated within a country’s territorial jurisdiction. Under this definition, which countries appear “compute sovereign” in the public cloud AI compute census data set?

The census discovered 225 cloud regions situated on the territories of 43 different countries. A cloud region is in simple terms a cluster of large, interconnected data centres situated roughly in the same area. Of these 225 cloud regions, 132 were found to be equipped with AI accelerators. These 132 accelerator-enabled cloud regions were situated in 33 different countries. In other words, a total of 33 countries had some amount of public cloud AI compute capacity situated within their territorial jurisdictions. Most countries in the world had none. These public cloud AI compute host countries are predominantly located in North America, Europe, Middle East, and Asia (**Table 4**). In South America, only Brazil hosts public cloud AI compute on its territory, while in Africa only South Africa does so.

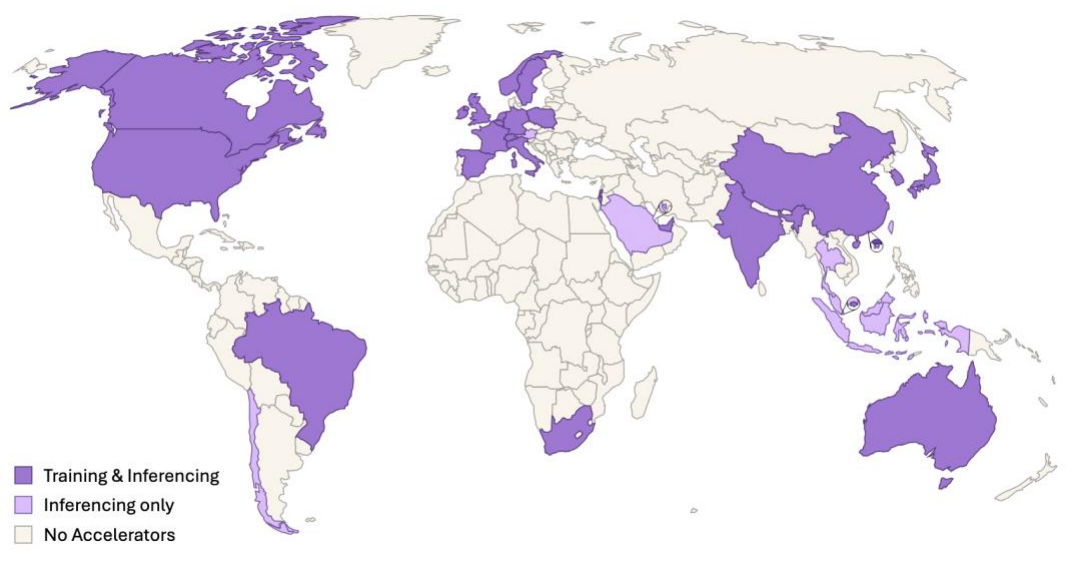
A further distinction may be made between cloud regions that are equipped with accelerators that are considered powerful enough for both training and inferencing or just inferencing new frontier AI models. The census revealed that a total of 24 of those countries hosted public cloud AI *training*-relevant compute, while 9 of the 33 countries hosted *only inference*-relevant AI compute. The remaining countries in the world have no public cloud AI compute according to our census of nine leading providers (**Table 4**). Individual regions can host multiple chip types, therefore the count of regions with training-relevant chips and regions with inference-relevant chips are not mutually exclusive i.e. some regions will be counted in both tallies. For example, Korea has a total of 4 compute regions. Of these 4, 3 are regions with training-relevant chips while 2 are regions with inference-relevant chips i.e. one of those regions contains both training-relevant and inference-relevant chips.

Table 4. Public cloud AI compute regions by host country and classification

Country	Number of compute regions	Training-relevant regions	Inference-relevant regions
<i>Countries with training-relevant public cloud AI compute on their territories</i>			
Australia	4	3	1
Belgium*	1	1	0
Brazil	2	1	1
Canada	5	3	2
China	22	4	22
France*	5	3	3
Germany*	7	4	4
India	5	3	3
Ireland*	3	2	2
Israel	2	2	0
Italy*	4	3	1
Japan	4	4	1
Korea	4	3	2
Netherlands*	2	2	2

Norway	1	1	0
Poland*	2	1	1
Singapore	6	2	5
South Africa	4	2	2
Spain*	1	1	0
Sweden*	2	2	1
Switzerland	2	2	0
UAE	3	3	1
UK	5	3	3
USA	26	22	13
*EU27 total	27	19	14
Global Total	122	77	70
<i>Countries with only inferencing-relevant public cloud AI compute on their territories</i>			
Austria*	1	0	1
Bahrain	1	0	1
Chile	1	0	1
Hong Kong	2	0	2
Indonesia	1	0	1
Malaysia	1	0	1
Saudi Arabia	1	0	1
Taiwan	1	0	1
Thailand	1	0	1
*EU27 total	1	0	1
Global Total	10	0	10

Figure 1. In-country public cloud AI compute availability by country



4.2 Compute sovereignty as AI compute provider nationality

The second way in which ‘compute sovereignty’ has been understood is by reference to the nationality of the cloud providers and other firms and organisations that operate AI data centres. The highest level of compute sovereignty from this point of view is achieved when data centres are not only situated within the country’s territory but also operated by domestically owned cloud provider firms. Under this definition, only the US, China, and a handful of European countries qualify as compute sovereign in the census data set. This result partly reflects the fact that the global market for the provision of public AI compute is dominated by a small number of large providers, which are the subject of our census. However, smaller, local cloud providers not captured by the census may nevertheless make meaningful contributions to their home markets’ AI compute capacity. Therefore, we cannot use the census data to make definitive conclusions about every country’s compute sovereignty status in regards to the existence of national compute providers.

That being said, while some countries do have their own strong and growing market of locally owned public AI compute providers, the hyperscalers remain dominant across the global cloud market (Lehdonvirta et al., 2024, 2025). Therefore, the ownership and nationality of the foreign companies that provide this AI compute, is a relevant consideration for governments when thinking about compute sovereignty. In lieu of AI compute made available by domestically owned cloud providers, whether the oversight and guarantee of these services is aligned to one foreign power, or split between multiple, has implications for a nation’s strategic calculus. Using our census, we compare the presence of foreign providers in each country to ask: are countries hedging their international digital infrastructure dependency between providers from more than one foreign country, or aligning with providers from a single foreign country only?

In total, 18 countries relied on providers from just one of either China or the US, while 12 countries hosted providers of multiple nationalities (**Table 5**). The US and China themselves are not sorted into these categories but included in the table for context and completeness. Austria is the only country that has no foreign AI compute, relying solely on European providers.

Table 5. Public cloud AI compute infrastructure by host country and provider nationality (highest foreign provider values bolded)

Table 5. Public cloud AI compute infrastructure by host country and provider nationality (highest foreign provider values bolded)			
Country	Foreign providers’ cloud regions		
	US providers	Chinese providers	European providers
<i>China & US</i>			
China	2	20	0
USA	24	2	0
<i>Countries hosting cloud regions of providers from multiple foreign countries (hedging)</i>			
Canada	4	0	1
France*	3	1	1
Germany*	4	2	1
Hong Kong	1	1	0
India	4	1	0
Ireland*	2	1	0
Poland	1	0	1
Korea	3	1	0
Singapore	3	3	0

South Africa	3	1	0
UAE	2	1	0
UK	4	0	1
EU27* Total	10	4	3
Global Total	34	12	5
<i>Countries hosting cloud regions of providers from a single foreign country (aligning)</i>			
Australia	4	0	0
Bahrain	1	0	0
Belgium*	1	0	0
Brazil	2	0	0
Chile	0	1	0
Indonesia	0	1	0
Israel	2	0	0
Italy*	4	0	0
Japan	4	0	0
Malaysia	0	1	0
Netherlands*	2	0	0
Norway	1	0	0
Saudi Arabia	0	1	0
Spain*	1	0	0
Sweden*	2	0	0
Switzerland	2	0	0
Taiwan	1	0	0
Thailand	0	1	0
EU27* Total	10	0	0
Global Total	27	5	0
<i>Countries hosting no cloud regions of providers a foreign country</i>			
Austria	0	0	1

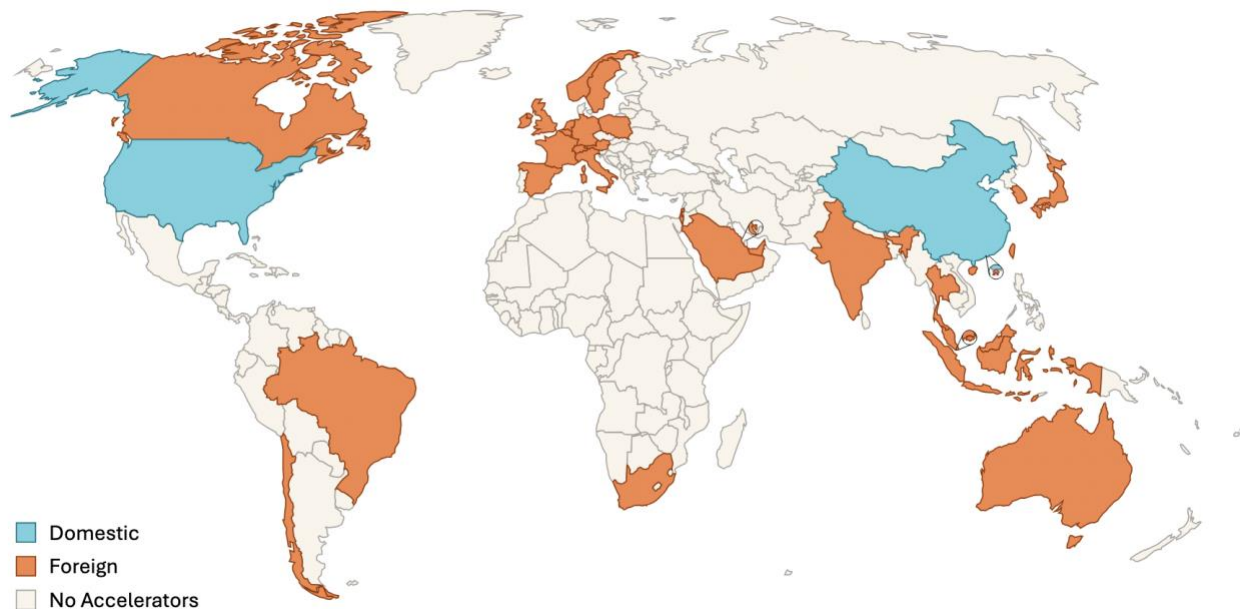
4.3 Compute sovereignty as AI accelerator chip vendor nationality

The third way in which “compute sovereignty” has been understood is by reference to control over AI accelerator chip supply chains. Governments that are able to exercise control over chip supply chains are able to secure continued supply for their own domestic providers and data centres, whilst potentially denying access to rivals’. Which countries have “compute sovereignty” under this definition?

The US company NVIDIA are by far the largest seller of AI accelerators to cloud providers by market share, comprising 80-95% of the total chip market (Sastry et al., 2024). But over the past two years, some Chinese hyperscale cloud providers have introduced their own in-house AI accelerator chip models. Of the 132 accelerator-enabled regions in the census, only six offer Huawei Ascend accelerators, with three regions in China and three regions in France, Singapore and the UAE. This means 95.5% of accelerator-enabled regions in the census are powered by US-owned accelerators. These results reflect US, and particularly NVIDIA’s, immense market concentration. As a result, in the census, only the US and China host AI accelerators sourced from domestic chip vendors. They are therefore also the only countries hosting public AI compute data centres operated by domestic vendors equipped with at least some domestically designed chips. Outside of France, Singapore and the UAE, the

remaining AI compute host countries' accelerators were all sourced from vendors of a single foreign nationality, namely the United States. In the terminology adopted earlier, the vast majority of countries were 'aligned' with the US on this level of compute sovereignty.

Figure 2. AI accelerator vendor nationality by country: domestic vs. foreign



5. Discussion

Our analysis of data on public cloud AI compute reveals that a country's 'compute sovereignty' status varies depending on which of multiple understandings or 'levels' of compute sovereignty is used. This section discusses the findings of each level of analysis in turn and draws implications to future research and policy.

5.1 Territorial Sovereignty: Do you have (or want) AI compute in-country?

Territorial compute sovereignty, or the possession of 'in-country compute', is the most widely held type of compute sovereignty around the world, as judged through the lens of public cloud AI compute, a large and important segment of AI compute. Even still, only a minority of countries worldwide (33) qualified as having at *least* inference-relevant AI compute within their jurisdiction, and only 24 of those countries *also* have training-relevant AI compute within their jurisdiction. That is about 17% and 12% of UN member states respectively. This concentration of in-country access in such a small number of places reveals the global disparity in distribution of this strategic resource around the world between the 'compute haves and have nots' (Lehdonvirta et al., 2024).

However, should every government want AI compute on its territory? As outlined in the literature review, the possible strategic benefits that motivate the 'compute sovereignty' discourse include security of supply to what is perceived to be a critical resource, and also the ability to use control over that resource to regulate AI applications and possibly also to deny access to strategic rivals (Beaumier & Cartwright, 2024). Besides any possible strategic benefits that 'compute sovereignty' brings, governments also see economic benefits in large digital infrastructures. Building data centres housing AI compute brings with it tax revenue and local employment, initially for construction, followed later by management and operation of the data centre (TechUK, 2024).

That being said, there is also increasing attention on the possible downsides of hosting AI data centres within a jurisdiction. The economic benefits of data centres, usually deriving from economic impact assessments conducted by Big Tech corporations or data centre industry bodies, have been challenged by researchers and civil

society advocacy groups. Data centres require high upfront capital investment despite having low labour intensity, and thus has a relatively limited value add compared to other digital industries. Data centre developers may be given tax breaks or subsidies from local governments, which may offset their tax revenue contributions (Whiton, 2023). AI data centres consume large amounts of electricity, causing upward pressure on electricity prices for alternative industrial users and consumers especially in places where electricity supply is fairly inelastic (Judge, 2022). Data centres can also require vast quantities of water to operate, with disproportionate extractive dynamics for geographies experiencing ongoing droughts (Monserrate, 2022).

From a purely economic perspective, investing into developing national data centre industries may only pay off for countries that have resources that make them particularly competitive in this industry, such as cooler climates and high-capacity electricity infrastructure serving low-cost, low-carbon power, and access to plentiful cooling water. Moreover, community and civil society groups can dispute data centre developments that impact on their local areas (Rone, 2022, 2024). Such political costs can make data centre promotion politically unattractive especially for local governments responsible for zoning but not for foreign policy or national security. The pursuit of territorial AI compute sovereignty thus involves trade-offs between competing priorities. As a result, the correct policy to pursue will depend on the specific strategic circumstances of each country.

Table 5: Snapshot of claims regarding the benefits and drawbacks of having AI compute data centres in your territorial jurisdiction

	<u>Benefits</u> <i>of AI compute territorial sovereignty</i>	<u>Drawbacks</u> <i>of AI Compute Territorial Sovereignty</i>
National economic development	Creates tax revenue, construction and minimal operation employment	Comparatively low job creation compared to resource demands. Tax revenue may be offset if tax breaks are given (Whiton, 2023)
Governance	Enables jurisdictional control over data, and compliance with national laws such as privacy, security etc (Sastry et al., 2024)	Regulatory complexity of overlapping jurisdictions where the data centres are run by a foreign cloud provider
Sustainability	Drives private sector investment in renewable energy industry	Consumes local resources including land, energy and water, which directly conflicts with domestic/household use (Monserrate, 2022)
Operational	Improved latency, bandwidth efficiency, and reliability for national AI workloads and end users	High energy, cooling and land demands may be managed less efficiently due to local constraints – it can take years to process and access energy interconnection (CBRE, 2024)
Strategy & geopolitics	If owned by foreign provider, the high capex investment and low post-build capital mobility means the host government may have “hostage leverage” over the cloud provider’s foreign investment (Vernon, 1971)	If owned by foreign provider, risk of infrastructure dependence being weaponised (Farrell & Newman, 2019)
Social License	Perception of domestic control of data and compliance with local law may increase public trust in the digital economy	If owned by foreign provider, perception of extractive industry putting pressure on finite national resources – can lead to social contestation (Rone, 2022, 2024)

An alternative to purely national ‘in-country’ AI compute sovereignty are attempts to achieve compute sovereignty through regional frameworks of trust and collective governance. For example, the European High Performance Computing Joint Undertaking (EuroHPC) is a public-private partnership established in 2018 by the European Union, participating European countries, and private partners. It represents an effort by EU member states to pool resources and coordinate the development and use of high-performance computing infrastructure across member state borders, allowing countries to benefit from shared compute capacity while retaining a degree of influence over its governance. Among other facilities it has built LUMI, a sizeable AI compute cluster located in Kajaani, Finland. Such frameworks can offer smaller or less resource-rich states access to advanced compute infrastructure without requiring them to host it domestically. They also distribute economic and environmental costs more evenly and can support common regulatory frameworks for the responsible use of AI. However, regional

approaches require high levels of political trust and institutional coordination and may still fall short of the full control that some governments seek when invoking ‘sovereignty.’ Thus, compute sovereignty can be seen not as a binary condition but as a spectrum of arrangements—ranging from fully national control to shared regional governance—each carrying distinct political, economic, and environmental implications.

5.2 Provider Sovereignty: Does it matter who owns your AI compute data centres?

While regionalism can expand the idea of jurisdiction outward, a second challenge emerges from the inward pull of foreign legal authority. When foreign hyperscalers operate AI compute within a host country, they bring with them the regulatory reach of their home governments, creating a layered or overlapping jurisdiction. This can result in conflicted compliance obligations for private sector cloud providers and diminished regulatory exclusivity for host states. For example, a China-headquartered cloud provider operating in Malaysia may be fully subject to Malaysian law, but it is also bound by Chinese legal instruments such as the *National Security Law 2018*, which allows Chinese authorities to demand access to data stored abroad by Chinese firms. This creates overlapping sovereign claims that can result in conflicting legal obligations on the cloud providers (Abraha, 2019). As cloud services are integrated into sensitive sectors such as defence, health and critical infrastructure, this multilayered jurisdiction raises concerns for some that location without control is not sovereignty.

When applying the second level of our AI compute sovereignty framework, we find that, excluding the US and China, 12 countries (or roughly one third of the 31 other countries with accelerator enabled regions) are hedging between cloud providers from the US and China, while 18 (or just under two thirds) are aligning with cloud providers from just one of these superpowers. Of those aligned countries, 5 are aligned with Chinese digital infrastructure, and 13 aligned with US digital infrastructure.

The decision between hedging or aligning between competing superpowers is a well-established strategic concept in international relations and international security studies. As with the previous level, what constitutes ‘sovereignty’ at this level is for most countries, in reality, not so black and white. If AI compute sovereignty is interpreted as control over who operates and governs the cloud infrastructure supplying high-performance compute, then sovereignty becomes a question not only of domestic ownership, but of strategic alignment with foreign providers. In the absence of domestic hyperscalers with the capability and capacity to support economy wide access to public AI compute, many governments face the strategic choice: are they hedging between superpowers with digital infrastructure from multiple foreign cloud providers; or are they aligning with the infrastructure of one superpower?

Sovereignty through AI compute alignment – relying on a single foreign ally: Some states can be seen to be aligning by selecting a single foreign provider, typically from an allied country, and deepening that relationship to ensure reliable access to high-performance AI compute. These arrangements often reflect broader political or defence alignments, where a foreign cloud provider is seen not merely as a commercial actor but as an extension of a trusted state’s technology ecosystem (Lehdonvirta et al., 2025). For example, the Australian Government has partnered with AWS to establish a Top-Secret Cloud for Australia’s national intelligence community – the government’s most sensitive data (Australian Government, 2024). Looking at the census findings, other countries including Japan, and Israel are aligning with US cloud providers, while countries including Chile, Indonesia and Saudi Arabia are aligning exclusively with Chinese cloud providers. Through this lens, a type of sovereignty can be exercised through selective interdependence with an ally perceived to be politically stable, technologically advanced and ideologically aligned (Adler-Nissen, 2008).

Sovereignty through AI compute hedging: multiple providers for resilience: Other countries' AI compute dependence is distributed across multiple foreign cloud providers. Like balancing, this approach does not offer sovereignty through self-reliance but strategic autonomy through diversification. Evelyn Goh describes this as a strategy, which is common across countries in Southeast Asia, to 'cultivate a middle position that forestalls or avoids having to choose one side at the obvious expense of the other' (Goh, 2005, 2). For example, in this census, Singapore hosts 3 AI compute regions from US cloud providers and 3 AI compute regions from Chinese cloud providers. Conceptually, this approach can be seen as a response to the weaponized interdependence theory, which warns against being held captive by one single foreign power (Farrell & Newman, 2019).

These distinct profiles – aligning and hedging – suggest that AI compute sovereignty is not a black and white one-size-fits all goal, but a strategic continuum shaped by risk appetite, trust relations, and geopolitical alignment. For many mid-size and non-aligned states, the pursuit of AI compute sovereignty may not involve full-self-sufficiency or domestic control but navigating dependencies in a way that simultaneously preserves autonomy and reduces vulnerability.

5.3 Accelerator Sovereignty: Do you care about the nationality of your AI accelerator vendor?

The most technologically foundational – and geopolitically constrained – layer of AI compute sovereignty concerns the ownership and control of the AI accelerator vendors that design and distribute the chips required for training and deployment of AI models. Like the previous level of our framework, extreme concentration in the global accelerator vendor market means that only the US and China have regions powered by accelerators from domestic vendors. However, given the comparative market lead NVIDIA has over Chinese counterparts in performance, and with Huawei's Ascend chips not found in any region outside of China availability, most countries are left without a meaningful choice but to align with NVIDIA. What does that mean for notions of AI compute sovereignty?

Despite being a foreign-owned provider, NVIDIA positions itself as a key partner to countries around the world in their pursuit of sovereignty, stating that "*Since 2019, NVIDIA's AI Nations initiative has helped countries spanning every region of the globe to build sovereign AI capabilities*" (Lee, 2024). For example, Israel and the United Kingdom continue to scale their sovereign AI compute infrastructure investments using NVIDIA hardware. In contrast, other jurisdictions are looking to diversify their reliance on NVIDIA accelerators by building up domestic semiconductor capacity. For example, the *EU Chips Act 2022* includes a €43 billion investment plan for domestic semiconductor development (European Commission, 2022). At the same time, the US is seeking to onshore aspects of TSMC operations to further concentrate US control over accelerator supply chains (US Department of Commerce, 2024). The full impacts of this policy on the market remain to be seen.

Accelerator vendor level AI compute sovereignty presents perhaps the most complex and long-term element in our three-level framework. The building of local AI compute data centres and development of domestic cloud providers takes investment and effort over many years, but the research and development, capital and talent required for chip design and fabrication, even more so. It is therefore unsurprising, then, that most states pursue AI compute vendor sovereignty through aligning with US vendor NVIDIA.

6. Conclusion

AI is considered by some governments a core axis of global strategic competition, national security and economic innovation, and is thus increasingly entwined with questions of digital sovereignty. Underpinning its geostrategic importance for governments is one foundational input - AI compute – making questions of where AI compute is located, who owns it and who governs it, central geopolitical questions. This has implications for AI governance as countries seek to exert control and independence in global and interdependent supply chains.

This study has presented an analytical framework that allows for multi-level examination of sovereignty across the AI compute supply chain, including: 1) territorial jurisdiction over AI compute infrastructure; 2) ownership and nationality of cloud providers; and 3) ownership and nationality of AI accelerator chip vendors. Empirically, the paper presents a global census of public AI compute infrastructure of the world's three largest hyperscaler blocs – those based in the US, China, and Europe. The data reveals striking asymmetries: only 34 countries host any public AI compute; only 24 of those have access to training level compute; and most rely on cloud or chip infrastructure controlled by a small number of foreign actors. There are 18 host countries aligning with one superpower's AI cloud compute infrastructure, while 12 hedge by relying on infrastructure from both US and Chinese owned companies. All countries in the census apart from China rely on US-owned NVIDIA accelerators to

power the AI compute data centres in their territories. These insights are not just technological, they are fundamentally geopolitical.

Importantly, while this paper puts significant effort into understanding and measuring AI compute sovereignty, this should not be misconstrued as advocating for its pursuit. Territorial AI compute sovereignty presents a range of benefits and drawbacks, so this strategic question should receive its due consideration, specificity, and critical analysis from policymakers around the world.

This analysis invites several avenues for further research. First, repeating and expanding a similar census in the future will offer valuable longitudinal insights into how the global distribution of data centres and leading-edge AI accelerators unfolds, how sovereignty configurations evolve over time, and whether efforts by some states to rebalance dependencies are successful. Second, there is a need to explore how sovereignty strategies evolve in response to changing geopolitical conditions—for example, a second Trump Presidency, with its unpredictable approach to alliances, growing techno-nationalist policies, may dramatically reshape perceptions of interdependence and sovereignty, while pushing back on other countries’ attempts to regulate these US companies. Third, future work should investigate the role of these private sector actors, such as the hyperscalers, shape the legitimacy and accountability of sovereignty claims in this domain.

As Pohle & Thiel (2020) caution, digital sovereignty is not an end in itself but a political project shaped by underlying values and power dynamics. This study contributes an empirical and conceptual foundation for examining what kind of sovereignty is being asserted in the AI era, by whom, and at what cost.

Authors

Zoe Jay Hawkins is a co-founder of the Tech Policy Design Institute, an independent non-partisan think tank based in Australia. She is a Research Associate at the Oxford Internet Institute, University of Oxford, UK, and a former policy adviser to Australian cabinet ministers, government, and Amazon.

Vili Lehdonvirta is Professor of Technology Policy at the Department of Computer Science, Aalto University, Finland, and Professor at the Oxford Internet Institute, University of Oxford, UK. He leads the Digital Economic Security Lab DIESL. He is the recipient of three European Research Council grants, most recently on “The Geopolitics of Cloud Computing”. His books *Cloud Empires* and *Virtual Economies* are published by MIT Press.

Boxi Wu is a doctoral researcher at the Oxford Internet Institute, University of Oxford, UK, and member of the Digital Economic Security Lab DIESL. Their research focuses on the political and environmental impacts of AI infrastructure and has been published in *Big Data & Society*, *Philosophy & Technology*, *ACM Fairness, Accountability & Transparency* and *AAAI/ACM AI, Ethics & Society*.

Conflicts of interest

Tech Policy Design Institute has received funding from Amazon, Microsoft, and Google. No companies were involved in the funding, design, data collection, analysis, writing up, or publishing decisions of this study.

References

- Adler-Nissen R. 2008. Sovereignty games: Instrumentalizing state sovereignty in Europe and beyond. Palgrave Macmillan.
https://www.researchgate.net/publication/313861167_Sovereignty_Games_Instrumentalizing_State_Sovereignty_in_Europe_and_Beyond
- Abraha H. 2019. How compatible is the US ‘CLOUD Act’ with cloud computing? A brief analysis. *International Data Privacy Law* 9(3): 207–217. <https://academic.oup.com/idpl/article-abstract/9/3/207/5532213>
- Allen G C. 2024. Understanding the Biden administration’s updated export controls. Center for Strategic and International Studies. <https://www.csis.org/analysis/understanding-biden-administrations-updated-export-controls>
- Australian Government – Office of National Intelligence. 2024. Australian Government announces Top Secret Cloud. <https://www.oni.gov.au/news/australian-government-announces-top-secret-cloud>
- AWS. 2022. Digital Sovereignty Pledge. <https://aws.amazon.com/blogs/security/aws-digital-sovereignty-pledge-control-without-compromise/>
- Barlow, J. P. 1996. *A declaration of the independence of cyberspace*. Electronic Frontier Foundation. <https://www.eff.org/cyberspace-independence>
- Beaumier G and Cartwright M. 2024. Cross-network weaponization in the semiconductor supply chain. *International Studies Quarterly* 68(1): sqae003. <https://doi.org/10.1093/isq/sqae003>
- Blancato, F. G. 2024. The cloud sovereignty nexus: How the European Union seeks to reverse strategic dependencies in its digital ecosystem. *Policy & Internet*, 16(1), 12-32.
- Botton, N. & Lee-Makiyama, H. 2018. *5G and national security after Australia's telecom sector security review*. ECIPE Policy Brief No. 8/2018. European Centre for International Political Economy (ECIPE), Brussels. <https://www.econstor.eu/handle/10419/202509>
- CBRE, 2024 <https://www.cbre.com/insights/briefs/data-center-asking-rents-surged-as-much-as-54-percent-over-eight-months>.
- Ciorciari J D and Haacke J. 2019. Hedging in international relations: An introduction. *International Relations of the Asia-Pacific* 19(3): 367–374. <https://doi.org/10.1093/irap/lcz017>
- Dilmegani C. 2024. Top 20 AI chip makers: NVIDIA & its competitors in 2025. AIMultiple. <https://research.aimultiple.com/ai-chip-makers/>
- European Commission. 2022. Digital sovereignty: Commission proposes Chips Act to confront semiconductor shortages and strengthen Europe's technological leadership. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_729
- Farrell H and Newman A L. 2019. Weaponized interdependence: How global economic networks shape state coercion. *International Security* 44(1): 42–79. https://doi.org/10.1162/isec_a_00351
- Floridi L. 2020. The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology* 33: 369–378. <https://doi.org/10.1007/s13347-020-00423-6>
- Ferrari, F. 2024. State Roles in Platform Governance: AI’s Regulatory Geographies. *Competition & Change* 28 (2): 340–58. <https://doi.org/10.1177/10245294231218335>.
- Gaia-X. 2025. *Gaia-X strengthens European digital sovereignty at European Parliament reception*. <https://gaia-x.eu/gaia-x-strengthens-european-digital-sovereignty-at-european-parliament-reception/>

- Ghahramani, Z. 2023. *The future of compute: Report of the Review of the Independent Panel of Experts*. UK Government. <https://www.gov.uk/government/publications/future-of-compute-review/the-future-of-compute-report-of-the-review-of-independent-panel-of-experts>
- Goh, Evelyn. 2005. *Meeting the China Challenge: The US in Southeast Asian Regional Security*. Policy Studies Monograph 16. Washington DC: East-West Center. pp. 2-4
- Goh, E. 2007. Great powers and hierarchical order in Southeast Asia: Analyzing regional security strategies. *International security*, 32(3), 113-157.
- Government of Canada. 2025. *Canadian sovereign AI compute strategy*. <https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy>
- Gupta, A., Pal, A., & Reinsch, W. A. (2024). *Collateral damage: The domestic impact of U.S. semiconductor export controls*. Center for Strategic and International Studies. <https://www.csis.org/analysis/collateral-damage-domestic-impact-us-semiconductor-export-controls>
- Hofmann J. 2016. Multi-stakeholderism in Internet governance: Putting a fiction into practice. *Journal of Cyber Policy* 1(1): 29–49. <https://doi.org/10.1080/23738871.2016.1158303>
- Hill, J. F. 2014. The Growth of Data Localization Post-Snowden: Analysis and Recommendations for U.S. Policymakers and Industry Leaders. *Lawfare Research Paper Series*, 2(3), 1–41.
- Huawei. n.d. National government cloud solution. <https://e.huawei.com/en/industries/government/national-digital-infrastructure/national-cloud>
- IDC. 2023. Sovereign cloud: A policy response to the digital economy. <https://www.idc.com/getdoc.jsp?containerId=prEUR251542423>
- Judge, 2022. <https://www.datacenterdynamics.com/en/news/report-home-building-to-halt-in-west-london-due-to-data-center-power-demands/>
- Krasner S D. 2001. Rethinking the sovereign state model. *Review of International Studies* 27(1): 17–42. <https://doi.org/10.1017/S0260210500000173>
- Krasner, S. D. 1999. *Sovereignty: Organized Hypocrisy*. Princeton. <https://doi.org/10.2307/j.ctt7s9d5>
- Kwet M. 2019. Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class* 60(4): 3–26. <https://doi.org/10.1177/03063968188231702>
- Lee A. 2024. What is sovereign AI. NVIDIA Blog. <https://blogs.nvidia.com/blog/what-is-sovereign-ai/>
- Lehdonvirta, V., Wú, B., & Hawkins, Z. (2024). Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe. In *Proceedings of the AAI/ACM Conference on AI, Ethics, and Society* (Vol. 7, pp. 828-838).
- Lehdonvirta, V., Wú, B., & Hawkins, Z. (2025). Weaponized interdependence in a bipolar world: How economic forces and security interests shape the global reach of U.S. and Chinese cloud data centres, *soxArXiv*.
- Maihold, G. (2022). *A new geopolitics of supply chains: The rise of friend-shoring* (SWP Comment No. 45/2022). Stiftung Wissenschaft und Politik – SWP – Deutsches Institut für Internationale Politik und Sicherheit. <https://doi.org/10.18449/2022C45>

- Michels J D. 2025. Sovereign cloud for Europe: Independent research report prepared for Broadcom. Queen Mary University of London, Centre for Commercial Law Studies.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5146122
- Microsoft. 2024. Introduction to data sovereignty. <https://learn.microsoft.com/en-us/industry/sovereignty/cloud-for-sovereignty>
- Miller, C. 2022. *Chip war: the fight for the world's most critical technology*. Simon and Schuster.
- Monserrate, S. G. 2022. *The staggering ecological impacts of computation and the cloud*. The MIT Press Reader.
<https://thereader.mitpress.mit.edu/the-staggering-ecological-impacts-of-computation-and-the-cloud/>
- Musiani, F. 2024. Understanding infrastructure as (Internet) governance. In Padovani, C., Wavre, V., Hintz, A., Goggin, G., and Iosifidis, P. (Eds.), *Global Communication Governance at the Crossroads*. Springer.
https://link.springer.com/chapter/10.1007/978-3-031-29615-3_6
- Mueller, M. L. 2010. *Networks and States: The Global Politics of Internet Governance*. MIT Press. <http://doi.org/10.7551/mitpress/9780262014595.001.0001>
- Nellis S and Cherney M. 2025. Nvidia CEO Huang says chipmaker well positioned for shift in AI. *Reuters*.
<https://www.reuters.com/technology/nvidia-ceo-huang-says-chipmaker-well-positioned-shift-ai-2025-03-19/>
- NVIDIA, N.d <https://www.nvidia.com/en-gb/data-center/tesla-v100/>
- OECD, 2025. *Forthcoming Working Paper*.
- OVH, n.d
<https://corporate.ovhcloud.com/en/company/history/#:~:text=From%202010%2C%20OVHcloud%20entered%20into,offer%20%2D%20the%20Hosted%20Private%20Cloud.>
- Patel D and Nishball D. 2023. GPU cloud economics explained – the hidden truth. SemiAnalysis.
<https://semianalysis.com/2023/12/04/gpu-cloud-economics-explained-the/>
- Pilz, K. F., Heim, L., & Brown, N. (2025, April). Increased compute efficiency and the diffusion of AI capabilities. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 26, pp. 27582-27590).
- Potkin F. 2024. Huawei aims to mass-produce newest AI chip in early 2025, despite US curbs. *Reuters*.
<https://www.reuters.com/technology/artificial-intelligence/huawei-aims-mass-produce-newest-ai-chip-early-2025-despite-us-curbs-2024-11-21/>
- Pohle, J. and Thiel, T. 2020. Digital sovereignty. *Internet Policy Review* 9, 4 (2020).
<https://doi.org/10.14763/2020.4.1532>
- Pratschke M. 2025. The global pursuit of sovereign AI is becoming the 21st century's arms race. LSE Blog.
<https://blogs.lse.ac.uk/businessreview/2025/01/27/the-global-pursuit-of-sovereign-ai-is-becoming-the-21st-century-arms-race/>
- Raymond M and DeNardis L. 2015. *Multistakeholderism: Anatomy of an inchoate global institution*. *International Theory* 7(3): 572–616. <https://doi.org/10.1017/S1752971915000081>
- Roberts H, Hine E and Floridi L. 2023. Digital sovereignty, digital expansionism, and the prospects for global AI governance. In: Timoteo M, Verri B and Nanni R (eds), *Quo vadis, sovereignty? New conceptual and regulatory boundaries in the age of digital China*. Cham: Springer.

Rone, J. 2022. The politics of data infrastructures contestation: Perspectives for future research. *Journal of Environmental Media*, 3(2), 207-214.

Rone, J. 2024. The shape of the cloud: Contesting data centre construction in North Holland. *new media & society*, 26(10), 5999-6018.

Sastry G, Heim L, Belfield H, Anderljung M, Brundage M, Hazell J, O'Keefe C, Hadfield G K, Ngo R, Pilz K, Gor G, Bluemke E, Shoker S, Egan J, Trager R F, Avin S, Weller A, Bengio Y and Coyle D. 2024. Computing power and the governance of artificial intelligence. arXiv. <https://arxiv.org/abs/2402.08797>

Sevilla J, Träger F, Vogel T, Ding Y, Kumbhar S and Perez E. 2022. Compute trends across three eras of machine learning. arXiv. <https://arxiv.org/abs/2202.05924>

Schmidt E. 2023. Innovation power: Why technology will define the future of geopolitics. Foreign Affairs. <https://www.foreignaffairs.com/united-states/eric-schmidt-innovation-power-technology-geopolitics>

Sharma, 2023 <https://datacentrenews.uk/story/hetzner-celebrates-five-years-of-helsinki-s-data-centre-park>

Shilov A. 2025. DeepSeek research suggests Huawei's Ascend 910C delivers 60% of Nvidia H100 inference performance. Tom's Hardware. <https://www.tomshardware.com/tech-industry/artificial-intelligence/deepseek-research-suggests-huaweis-ascend-910c-delivers-60-percent-nvidia-h100-inference-performance>

Shivakumar S, Bell R, Alatovic L and Padhi A. 2024. A world of CHIPS Acts: The future of U.S.-EU semiconductor collaboration. Center for Strategic and International Studies. <https://www.csis.org/analysis/world-chips-acts-future-us-eu-semiconductor-collaboration>

Smith B. 2025. The golden opportunity for American AI. Microsoft blog. <https://blogs.microsoft.com/on-the-issues/2025/01/03/the-golden-opportunity-for-american-ai/>

techUK. 2024. Foundations for the future: How data centres can supercharge UK economic growth. <https://www.techuk.org/resource/techuk-report-foundations-for-the-future-how-data-centres-can-supercharge-uk-economic-growth.html>

Thadani A and Allen G C. 2023. Mapping the semiconductor supply chain: The critical role of the Indo-Pacific region. Center for Strategic and International Studies. <https://www.csis.org/analysis/mapping-semiconductor-supply-chain-critical-role-indo-pacific-region>

U.S. Department of Commerce. 2024. *Biden-Harris Administration announces CHIPS incentives award to TSMC*. <https://www.commerce.gov/news/press-releases/2024/11/biden-harris-administration-announces-chips-incentives-award-tsmc>

Van Dijck J. 2020. Governing digital societies: Private platforms, public values. *Computer Law & Security Review* 36: 105377. <https://doi.org/10.1016/j.clsr.2019.105377>

Vipra J and West S M. 2023. Computational power and AI. AI Now Institute. <https://ainowinstitute.org/publication/policy/compute-and-ai>

Vernon, R. 1971. *Sovereignty at bay: The multinational spread of U.S. enterprises*. New York: Basic Books.

Walt S M. 1987. *The origins of alliances*. Cornell University Press.

Waltz K N. 1979. *Theory of international politics*. Addison-Wesley.

Wang, Y. and Chen, D. 2018. Rising Sino-U.S. competition in artificial intelligence. *China Quarterly of International Strategic Studies* 4, 2 (2018), 241–258. <https://doi.org/10.1142/S2377740018500148>

Weil, B., Kolmar, M., and Senn, M. 2024. The European Union, Taiwan and the Silicon Shield argument: A conceptual assessment through the lens of grand theories. *Journal of Transatlantic Studies*. <https://doi.org/10.1080/23745118.2024.2417028>

Whiton J. 2023. New data on data center subsidies, same old problems. Good Jobs First. <https://goodjobsfirst.org/new-data-on-data-center-subsidies-same-old-problems/>

Zhang M. 2024. Top 10 cloud service providers globally in 2024. Dgtl Infra. <https://dgtlinfra.com/top-cloud-service-providers/>