Korea's Trillion-Dollar Sovereign AI Investment: Nvidia Wins, Hynix Loses

Korea hosts a Squid Games, National AI Tournament, the best non-Chinese open source model gets eliminated, why Nvidia needs open source, implications for Hynix and Samsung

MAX KAN, RAY WANG, AND DYLAN PATEL

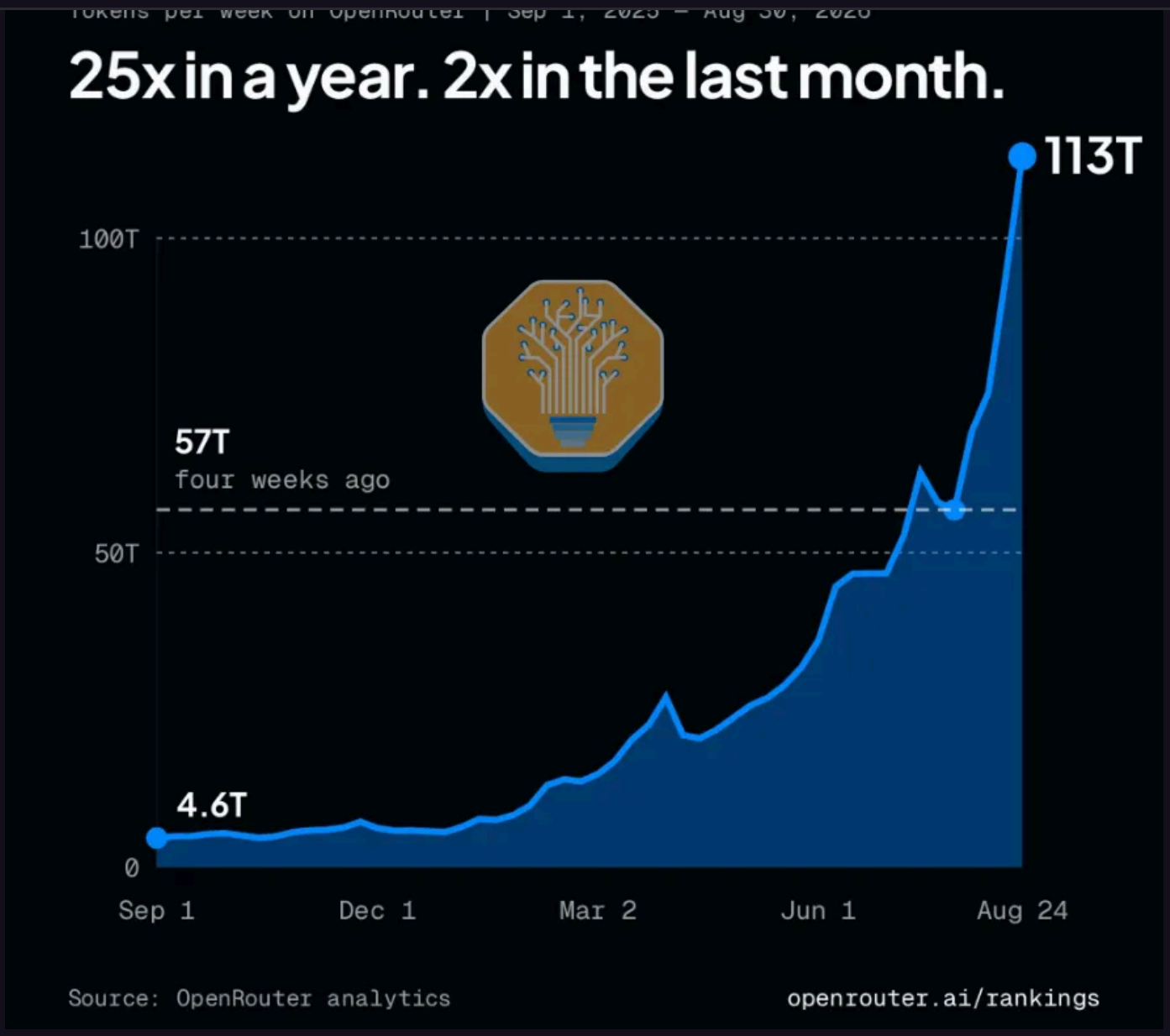
SEP 02, 2026 • PAID

54 2 Share

Every day, businesses and governments around the world are becoming increasingly reliant on America's frontier models. Startup CEOs already can't imagine running their companies without AI, and it won't be long until the same is true for every other organization in the world.

At the same time, it's become abundantly clear that access to frontier models is at the mercy of Anthropic, OpenAI, and the United States government. Fable 5 was temporarily banned by the USG, and GPT 5.6 and Astra were similarly delayed. Both models have cyber, bio, and other safety safeguards that, though well-intentioned, often prevent good users from completing harmless tasks. Given recent security incidents and general worries about increasingly powerful AI, it is extremely likely that frontier model usage will only become more restricted from here. In fact, we believe it's plausible **OpenAI/Anthropic will eventually stop offering API access entirely for their most capable models.**

Open source seems like the obvious solution to all these dependency concerns, but it's far from a silver bullet. First, all the "open source" licenses are starting to become increasingly restrictive. As just one example, any "model as a service business" making over $20M a year must negotiate a separate agreement with Moonshot to serve Kimi K3. Second, just because a lab open sources their models today doesn't guarantee they'll continue doing so in the future. Imagine if in 2028, you're stuck with 2027 level intelligence for some extremely high-value use case because no relevant model is open source and everything has to stay on-prem for security reasons. It's kind of like having your Fable request downgraded to Opus except 100x worse.



Open source token volumes have increased significantly over the past month.
Source: OpenRouter

The only way to fully address these concerns is to pretrain your own model that runs on your own GPUs. Until now, this has largely been a question for companies: do the potential benefits justify the enormous investment required? **Soon, however, this same calculation will confront every major nation state.**

Today, we'll do a deep dive on South Korea's sovereign AI efforts. Beyond being home to two of the most important companies in the AI supply chain, Korea also has a long tradition of technological self-reliance (as any foreigner that's had to use Naver Maps knows) and is currently the clear leader in sovereign AI.

Hynix shareholders.

Korean Squid Games - National AI Tournament: attempting to build a domestic frontier model

In June 2025, the Korean government announced “독자 AI 파운데이션 모델”, or the “Independent AI Foundation Model” project. As the name suggests, the goal is to develop a model that Korean organizations can train, modify, and operate without depending on a foreign AI lab.

Perhaps the most interesting thing about the project is its structure. Rather than selecting a single national champion upfront, the government is hosting a tournament. All participants are provided subsidies for the three pillars of AI—compute, data, and researchers—and evaluated every 6 months. At each stage, losers are eliminated and have their resources reallocated towards the winners.

The competition began with 15 consortiums. Ten passed the initial document review, and five were selected in August 2025: Naver Cloud, LG AI Research, SK Telecom, NC AI, and Upstage. Most readers are likely entirely unaware of any of these companies’ AI efforts, but some of them are surprisingly credible. SKT, LG, and Naver all pre-trained LLMs pre-ChatGPT for example.

The exact subsidies varied by team. For the first round, the government rented ~3000 H100 equivalents from SKT and Naver and distributed them to the other 3 contestants. They also spent ~$45M USD on data. Most of this was paid to Korean companies for things like books, news articles, and video broadcasts, which—along with some Korean government records—were shared among all the contestants. However, they also gave each company ~$2M to buy data themselves. Lastly for researchers, the government offered each company ~$1.4M to try recruiting overseas talent, but only Upstage took them up on the offer.

Surviving teams will be given additional resources as the tournament progresses, but the total government budget of ~$350M is still a rounding error compared to US labs. However, as we’ll explain later in this article, this is just a small portion of Korea’s planned AI investment. Additionally, their results so far show that training a decent model from scratch may be cheaper than most think.

government would give them additional resources to scale up their models throughout 2027.

However, there was some unexpected drama at the end of round 1!

The 5 teams were judged 40% based on benchmark scores, 35% on expert review, and 25% on user testing. NC finished last, which means the other 4 teams should've moved on, but the government decided to disqualify Naver as well. This is because they used vision and audio encoders from Alibaba's Qwen model family.

In Naver's defense, these are separate neural networks whose outputs are then fed into the main LLM, and the Korean government was originally unclear about what contestants could vs couldn't use from open source. It was only after the submissions were made that the government decided foreign architectures were allowed, but even auxiliary components must be trained and developed after initializing the weights.

Naver maintained that they'd already developed their own encoders and could replace the Qwen pieces at any time, but the government was predictably stringent and refused to reinstate them.

Now short 1 team, the Korean government decided to host a supplemental competition to find a new 4th contestant and ultimately selected Motif Technologies, a neolab that spun out of Moreh, a Korean AI infra software company—in February 2025.

All 4 companies open sourced their newest models in July and the results for the most recent round were announced on August 18th. But before we spoil the results, let's take a look at the models.

So how good are their models?

Here's an overview of the 4 models:

Source: SemiAnalysis

And here's how they compare on the benchmarks:

Source: SemiAnalysis Tokenomics Model

more impressive.

On Artificial Analysis' Intelligence Index, Motif 3 is 10 points ahead of Upstage and by far the best Korean model. What's even more notable, however, is the fact that it's comfortably ahead of both Inkling and Nemotron 3 Ultra—the two best American open source models today.

Source: Artificial Analysis, SemiAnalysis

Motif is a sub 30 person startup that released their first pre-trained model, which was only 2.6B total parameters, last June. They've raised just $17M and have access to a mere 768 B200s (< 2MW).

In contrast, Thinking Machines is the hottest neolab around, raised $2B at $12B post for their seed, and signed a 1GW+ compute deal with Nvidia which was accompanied by additional funding. As for Nvidia itself, the resources at its disposal hardly need mentioning.

Source: SemiAnalysis

Put differently, **a tiny Korean startup you've likely never heard of trained—from scratch and on a shoestring budge—the best non-Chinese open source model in the world**.

The point here is not to diss Thinking Machines or Nvidia. Training open source models is neither of their core businesses. Instead, we want to highlight that training a near-SOTA open source model from scratch is likely less resource intensive than most think. There's no substitute for focus, and the total compute cost to train Motif 3, including experimentation, was just ~$15M USD at today's prices. This is obviously well within budget for every major nation state.

Questionable judgement from the Korean government

We've previously written in-depth about the many issues with benchmarks, but they still tend to be directionally correct. **A 10+ point gap on something like the Artificial Analysis Intelligence Index has almost always corresponded to a step change improvement in model capability** (e.g. GPT 4o to o1 or Opus 4.8 to Fable 5).

With this in mind, we were shocked to see that **Motif was the one company eliminated in the latest round of competition.** Maybe you could argue that the model is benchmaxxed and thus didn't deserve to be first, but finishing dead last is baffling.

Remember that competitors were scored on three categories: 40% benchmarks, 35% expert review, and 25% user testing. Although Motif scored the best on benchmarks, they ranked last for the other two.

Source: SemiAnalysis

The exact methodology for these last two sections is frustratingly opaque. For user testing, we have no idea what tasks people actually used the model for or what criteria they considered to judge the models. Additionally, none of the testing was blind, so it's possible users had some bias towards the more famous companies.

The expert review section is even more questionable. Essentially, 10 external experts graded all four models on things like "development strategy", "future plans", and "ecosystem impact". As for what these words actually mean, your guess is as good as ours, but it seems clear that some categories disproportionately benefit large companies like LG and SKT. We also heard that the expert reviewers were concerned Motif's technology could be transferred overseas via foreign investors. This is quite puzzling, as one of the requirements of the competition is that all of the technical components must be open sourced.

We think it's a real shame Motif will no longer get any government support now that they've been eliminated from the competition. The whole point of a sovereign AI program is to rally your domestic AI talent to build a high quality model. Motif demonstrated that they're capable of exactly this, and the government should be utilizing its reach and resources to help them with things like "future plans" and "ecosystem impact"—not punish them for it.

Now, **Motif may be forced to relocate out of Korea in order to acquire the capital and compute required to continue their research**. This is perhaps the most ironic possible result for the surprise standout of your national AI tournament. Hopefully other countries can do better as they pursue their own sovereign AI efforts.

With that being said, South Korea's ultimate AI champion will be very well resourced. In July, they announced a monster $919B investment to build 8.4GWs by 2029 and 18.4GWs by 2035. For the first phase, SK Group, GS Group, and Naver will be building 5GW, 2.4GW, and 1GW respectively. SK Group is additionally responsible for the remaining 10GWs for phase 2.

We've already identified 3 active sites for phase 1 totaling 4.4GWs. For full details including the exact location, MW ramp, and power source, see our Datacenter Model.

Naver's 1GW site in Sejong, South Korea. Source: SemiAnalysis Datacenter Model

Although these timelines are aggressive, we believe they are achievable. The regulatory environment is favorable, and President Lee Jae-myung is explicitly framing AI infra as a legacy-defining project akin to Korea's broadband push in the 1990s. Throughout the 2000s and 2010s, Korea had uniquely cheap and fast internet, which is what enabled their local search, e-commerce, and esports industries.

have the optionality to dramatically scale up their sovereign AI efforts if necessary, and this domestic compute ramp is something more countries outside of the US and China will likely copy.

Why Nvidia needs open source and sovereign AI

Jensen recently made his Twitter debut with a manifesto on the importance of open source AI. It was later co-signed by basically every relevant AI company except Anthropic. Here's an excerpt:

> "Our AI leadership will be judged not by one frontier AI model, but by whether the United States builds a strong, open ecosystem that diffuses into every sector. This is essential for creating opportunities for innovation and prosperity across the country. It requires expanding access to AI, encouraging competition, robust application layers, and giving Americans greater control over the technology they rely on. Open weight models—AI models that anyone can download, inspect, modify, and run on their own infrastructure—are an important part of that foundation because they make advanced AI more accessible, adaptable, and widely available."

Much of the discourse around open source AI tends to wax poetic about the democratization of intelligence and decentralization of technological power. However, look under the hood and it's often just people talking their book.

No one needs open source AI more than Nvidia. As we previously explained to Tokenomics Model subscribers, selling frontier tokens at API prices is the highest ROI use case of incremental compute. Anthropic plus OpenAI are therefore set to take a larger and larger percentage of net new GWs in the future, and Nvidia is currently on track for a world where they only have 2 real customers. Sure maybe the number's 7 if you include the hyperscalers, but that's still obviously unacceptable for the world's largest company. This is doubly true when all 7 customers except SpaceX are actively trying to build their own XPUs to cannibalize Nvidia's margins.

One of the most famous competitive strategies in tech is to "commoditize your complements". The general idea is that as something becomes cheaper and more abundant, demand for everything that goes well with it increases. This explains why companies will often "give away" things for "free". Classic examples include 1) Google offering Chrome for free and open-sourcing Android, which increased access to the web and made search

In Nvidia's case, AI models and applications are the complements to its GPUs. **The long-term health of their business depends on a vibrant, diverse AI ecosystem that exists outside of just Anthropic and OpenAI.**

Their recently announced $500B MOU with the world's largest capital allocators is just one example of the clever financial engineering Nvidia can do to make this vision a reality. By making "Nvidia AI Factory Compute" an "investable asset class," the SSIs and Thinking Machines of the world have a way to leverage capital from pension funds, insurance companies, and private credit firms to buy enormous amounts of compute.

Source: SemiAnalysis

Sovereign AI is another huge opportunity for Nvidia to diversify their customer base. SK Group has already committed to building 2 GWs of Rubin as part of their 5GW buildout, and Naver has signed on for another 200MW. **We expect Jensen to continue announcing similar deals with additional countries.**

Datacenter Bet and Memory

requires sovereign infrastructure: domestic control over where the model is trained, how it is deployed, and who can access the underlying compute.

Korea's challenge is that it still lacks a scaled domestic provider of leading-edge accelerators. Rebellions and FuriosaAI are making some progress, but they remain objectively behind Nvidia and AMD in deployment scale, software maturity, and commercial availability. Hyperscaler-developed alternatives such as Google's TPUs and AWS's Trainium are not domestically controlled either.

The buildout of data centers essentially comes down to two things: capital and compute. Capital is the easier problem in our view. The Korean government has been prioritizing AI investment over the past few years, while Naver, SK Telecom, Samsung, SK Hynix, and other leading domestic tech companies are committing significant resources to data centers, AI infrastructure, and model development. Compute on the other hand, is another story. Korea can finance land, power, cooling, and buildings, but it cannot manufacture a globally competitive accelerator ecosystem just like the majority of countries in the world—or guarantee access to the latest platforms—simply by spending more money.

If Korea's objective is merely to establish a domestic AI ecosystem, a conventional supplier relationship with Nvidia may be sufficient. If it wants to become one of the world's leading sovereign AI players, however, it will need something deeper: greater roadmap visibility, close technical integration, reliable access to new compute platforms, and support for developing the surrounding domestic ecosystem.

This is why Nvidia's expanded relationships with Samsung and SK Hynix matter in this context. Samsung reportedly plans to build an Nvidia-powered AI factory using more than 50,000 GPUs, while SK Telecom's announced 2GW DSX AI factory is expected to deploy Vera Rubin systems powered by SK Hynix HBM4. While the public announcement does not disclose preferential GPU allocation or guaranteed volumes, deeper strategic alignment should strengthen Korea's ability to plan and deploy leading-edge compute—something far more difficult than simply constructing the data centers that house it.

Against this backdrop, we believe Nvidia's plans extend beyond just acquiring a large sovereign compute customer. As we previously explained to Memory Model subscribers, we believe Korea's infrastructure buildout also gives Nvidia an opportunity to deepen its relationship with SK Hynix, one of its most important memory suppliers and a key affiliate within the broader SK Group. We believe Nvidia has likely secured a long-term agreement

Expanding Beyond Memory Across the AI Infrastructure Stack Comes at a Cost

This post is for paid subscribers

Subscribe

Already a paid subscriber? Sign in

← Previous

© 2026 Dylan Patel · Privacy · Terms · Collection notice

Start your Substack | Get the app

Substack is the home for great culture

