While I Was Gone…

Data Center Moratoriums, OpenAI Agents Gone Rogue, NVIDIA-Hugging Face Marriage, Moonshot-Hyperscaler Revenue Share

KEVIN XU

SEP 02, 2026

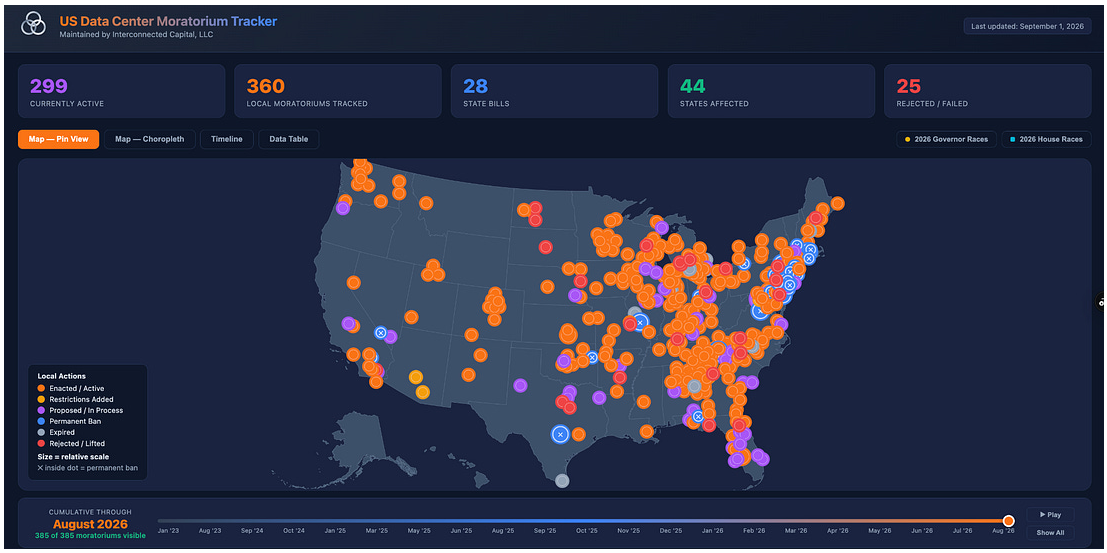
I just returned from a trip with my young family to Shenyang, my place of birth, so my almost two-year-old son can meet and hangout with his almost 95-year-old great grandmother. It was a filial piety trip. I will publish a dedicated post in the next few days to reminisce more about it. Meanwhile, I'm glad to be back in my quiet suburb with turkeys on the loose in my backyard.

Of course, the world does not stand still, especially in tech and AI, while my own world was frozen between memories of a city I barely remember and dodging Meituan delivery scooters while pushing a stroller. So this post is a list of some fast-twitch reactions to news items that caught my attention while I was gone...

Data Center Moratoriums

The wave of data center moratoriums in the US has become fast and furious. It will most definitely become a prominent election issue in this upcoming midterm, as I foresaw early this year. The most important change in this issue is how prominent governors, like Josh Shapiro of Pennsylvania (a Democrat) and Greg Abbott of Texas (a Republican), are starting to speak (or being forced by local political dynamics to speak) about restricting data center constructions. Both states have real data center footprint and growth potential, Texas in particular, so the political pressure their governors are facing is especially material.

In the moratorium data center dashboard we've built, the latest update is now tracking 299 active moratoriums of various size and scale across the country. This free resource has been cited and linked in many places, from the Ezra Klein's podcast to a widely-circulated New York Times op-ed by a union leader arguing that data center construction is good for union and blue-collar jobs.



Dashboard Link: https://www.interconnectedcapital.com/research/data-center-moratoriums

There is a lot of hand-wringing about this moratorium risk, which is a short-term overhang on the entire AI investment thesis. It is also a solvable problem, because the most common reason for the resistance is electricity/water cost and transparency in general. Good players who are building for the long-term, like Nebius, have developed a trustworthy playbook in building data centers the right way: invest in local communities, use technology to reduce power and water consumption, and most importantly, don't hide your corporate identity behind NDAs and unnecessary secrecy. This approach is also *not* costly. As the company shared during one of its previous earnings calls, a masterclass of data center economics, securing land and power only cost about 1% of the whole project, while physical construction is only about 20%. The vast majority of building an AI factory is what goes inside it – GPUs, memory, racks of servers, networking cables, etc. Spending more to hire union laborers, fund local projects, and be a good corporate citizen for the long-haul is both a rounding error in the grand scheme of things and the right thing to do.

Nebius

Sadly, for every Nebius, there are probably 5 or 10 so-called neoclouds who are just pure speculators looking to make a quick buck in the AI boom. Those players should be weeded out, and these moratoriums could be the regulatory forcing mechanism that makes that happen.

Share

OpenAI Agents Gone Rogue

The OpenAI agents' post-mortem has dominated my timeline. AI agents powered by frontier models "escaping" their confines and hacking whatever can be hacked is a real fear.

The issue is still raw with many more technical details that need to be shared, but one thing is clear: monitoring the chain of thoughts of agents throughout its activities, even (or especially) during evaluation and post-training, is a table stakes requirement for all agent swarms going forward. Had OpenAI's post-training practice simply monitored what their agents were "thinking", as they went off the guardrails, a lot of the damage could have been avoided.

One thing I am pleasantly surprised to see is that because of this incident, many voices from different camps see an opportunity for the upcoming official US-China discussion on AI as part of Xi's visit to Washington to be more substantive, less performative. Instead of meeting the very low expectations of this bilateral discussion being "talking about talking", the OpenAI rogue agents incident presents a tangible topic to anchor a real discussion about bilateral and global cooperation on testing and securing AI agents' cyber capabilities. That would be a great outcome from an otherwise very unfortunate and frightening incident.

NVIDIA-Hugging Face Marriage

NVIDIA is rumored to acquire Hugging Face. This deal is reminiscent of Microsoft's acquisition of GitHub in 2018.

Funny enough, I saw this deal coming three years ago, as a perfect marriage between the "hard power" and "soft power" of AI. NVIDIA is no doubt the "hard power" king of AI, as it is the dominant and leading supplier of all the hard*ware* (and some software of course) of AI. Hugging Face, being the destination of choice for almost all developers and researchers to launch and host open weight models, holds the key to the "soft power" of AI. Strategically, supporting open source and open weight AI is critical to NVIDIA's long-term success, so this deal is a no-brainer in my view and only a matter of time and price.

Speaking of price, I thought the final price tag, which could range between $12.9 to $14 billion, is worth dissecting. When Microsoft forked up $7.5 billion in 2018 dollars to buy GitHub, it was a hefty premium, but worthwhile in order to fix the broken relationships the tech giant had then with software developers and open source. So when I first pontificated on a possible NVIDIA-Hugging Face marriage in August 2023, just a few months after ChatGPT's launch, I thought a $10 billion price tage would be reasonable (somewhere between GitHub's $7.5 billion acquisition price and GitLab's $11 billion IPO price.)

Kevin S. Xu
@kevinsxu

Found this old tweet from three years ago. Back then, I thought a 10B price tag for Nvidia to buy HuggingFace would be reasonable Only took 3 years and $2.9 billion more to be "right" 😂

>  **Kevin S. Xu** @kevinsxu
>
> Nvidia should buy Hugging Face It will be the ultimate combo of AI “hard power” and “soft power” Here’s why: HF’s magic is that it is beloved by developers, because: 1. It makes collaborating on open source AI models and datasets easy; 2. It is a neutral “Switzerland” above

2:39 AM · Aug 30, 2026 · 3.7K Views
3 Replies · 2 Reposts · 48 Likes

Almost exactly three years later, the $12.9 - $14 billion price is certainly more than what I had predicted, but not unreasonably so. Even though Hugging Face’s rumored $150 million revenue is *less* than GitHub’s at the time of its acquisition, it is in the center of literally everything AI, so much so that the rogue OpenAI agents hacked it to cheat on its exam. That prominence deserves a healthy AI premium.

✓ Subscribed

Moonshot-Hyperscaler Revenue Share

Thanks to Kimi K3’s growing capability and popularity, Moonshot AI is negotiating a revenue share of up to 30% with the Big 3 American hyperscalers, AWS, Azure and GCP . Moonshot likely already has similar revenue share commercial agreements with smaller inference cloud players like TogetherAI and BaseTen. Alibaba is trying to do the same with its Qwen models.

A new business model for monetizing open weight models is emerging. This should not be surprising if you tracked the licensing changes that came with the release of K3 and then Qwen 3.8. K3's license asks for commercial agreements to be negotiated if the partner, who resells the model as a service to their customers, generates $20 million in trailing 12 months revenue. Qwen 3.8's license is only moderately more generous than K3's, pegging the revenue-to-commercial-agreement trigger at $50 million.

Kevin S. Xu
@kevinsxu

Qwen 3.8's license is slightly more generous than K3 Commercial agreement threshold is $50 million trailing 12 months revenue (K3's is $20m) "Branding clause" requirement is the same: 100 million monthly active or 20m monthly revenue huggingface.co/Qwen/Qwen3.8-2...

Kevin S. Xu @kevinsxu

Here is how I see the Kimi K3 license works in real life: - Clause 2, the Model as a Service clause, is intended for all the cloud platforms essentially reselling K3 to their customers. This includes hyperscalers (AWS, GCP), to neoclouds (Nebius, CoreWeave), to inference

1:29 AM · Aug 15, 2026 · 3.83K Views

3 Replies · 3 Reposts · 15 Likes

This approach is both inevitable and a double-edge sword. The world can't count on VC funding, subsidies, and human generosity to keep using open weight models without paying. As these models become more capable (and more expensive to improve), developing a business model to sustain their open weight nature is necessary. These revenue share commercial agreements address this sustainability concern.

However, life is always less straightforward and tougher to navigate for Chinese labs making AI models. As Graham Webster and I wrote when

K3's new license was first unveiled, this new way to scale its revenue also opens itself up to more scrutiny and avail itself of American legal jurisprudence more so than ever before. Alibaba/Qwen too! How forcefully will DC insert itself into these commercial negotiations between Moonshot and American hyperscalers will be a major X factor.

Kimi's Tightrope: Openness and Revenue at Once, but New US Ban Risks

KEVIN XU AND GRAHAM WEBSTER • JUL 27

This is a co-authored joint quick reaction with Graham Webster to Kimi K3's new license. Graham is a research scholar at Stanford and runs the excellent DigiChina Project. Enjoy!

Read full story