

Governing Artificial Intelligence:

Law, Policy, and Institutions

Law 4052 // Computer Science 283 // Communication 252A // Political Science 245B/445B

Fall Quarter 2025-26

Rob Reich
Anka Reuel
Nate Persily
Sanmi Koyejo

Roberta Fischli &
Dominic Zappia
(Head TAs)

Welcome to Cybersecurity Land

1. Remote Code Execution (RCE) vulnerabilities found in PHP (open-source programming language for dynamic website building)
→ Lets an attacker run their own commands directly on a compromised web server

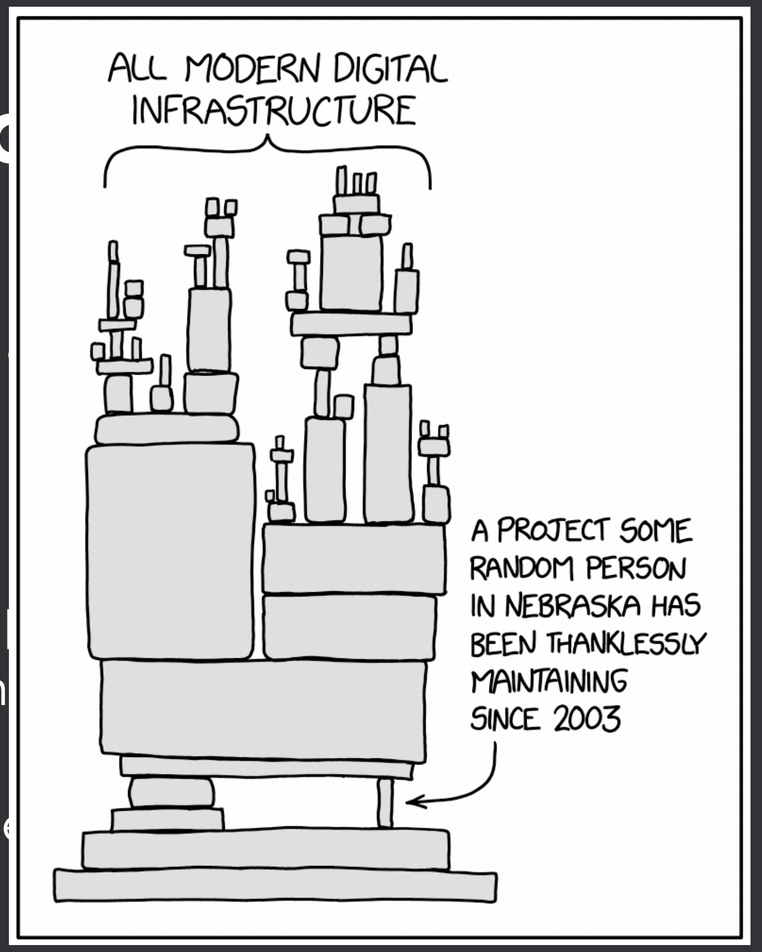
2. Private keys of Indian certificate authority eMudhra were not safe (each key should have been a 116-digit number but some only had one digit, i.e., easily guessable)

→ Certificates are digital ID cards of website (third-parties distribute them and vouch for security & identity of entity running the website)

→ Potential for Man-in-the-Middle (MITM) attack (malicious actor secretly intercepts, reads, or alters communications between two parties)

Live AI cyber case stud

- PHP vulnerability was found using GPT-5.6 Sol, Gemini 4 Argon,
- Patch available ≠ patch implemented
  - PHP team announced ahead of time that they are releasing a
    prepare but not every website has an IT department that mon
  - Patch available → easier to reverse-engineer vulnerability base
    AI (exploit in the wild already confirmed)



Source: xkcd

- Certificate compromise: no AI involved afaik, but we'll likely have many low-hanging fruit vulnerabilities in current internet infra; AI can help with broad sweeps (quantity of detected vulnerabilities goes up) and identifying more complex attack chains (more on that later!)

What you should do now

- Change passwords across websites (esp. critical things like health, finance, …) and DON’T reuse passwords across website
- Enable 2FA; at least Authenticator (not SMS 2FA!) and ideally, for critical things, use a Yubikey (~$50; get two to have a masterkey if you loose one)
- If you have your own website: update to PHP version 8.5 or larger
- Limit sensitive web traffic for a few days if possible (not everyone patches instantly)
- Run BitDefender (or similar virus/anti malware software) on your device (if malicious actors have control over server side, they can theoretically infect visitors of that site)
- Don’t approve Claude code commands if you’re not 100% sure what you’re doing and if you can’t be sure that the command only affects your local machine
- Update your browser & activate ad blockers (and check again for next few days)

October 21, 2026 Security Product

Patch the Planet: global response coordination as agents uncover critical vulnerabilities

Share

Coordinating a global response

Giving maintainers the capacity and resources to respond

Building more secure software foundations

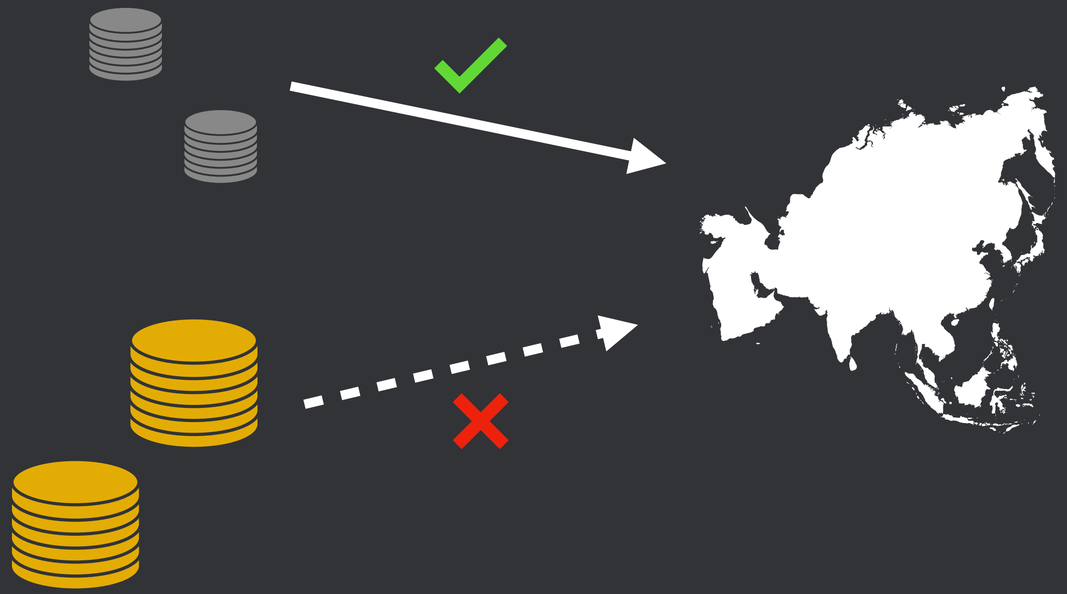
Help us improve this approach

Over the past two months, we led a coalition of maintainers, researchers, and security providers to address critical vulnerabilities in PHP, an open-source programming language used to run websites and web applications. Calif identified a critical flaw using GPT-5.6 Sol as part of our Patch the Planet initiative, with Wiz also independently discovering and reporting both vulnerabilities. PHP's maintainers led the fixes and release, with our teams and partners like Wiz, Cloudflare, and others contributing research, technical support, and coordination to prepare protections ahead of disclosure.

Hardware(-enabled) Governance Mechanisms

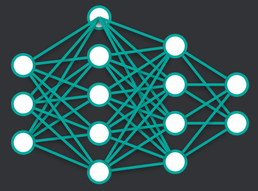
Option 1: Export Controls

Idea: Limit export of powerful chips to certain countries



What Goal Are We Trying To Achieve?

Prevent/delay training of frontier models



Delay or constrain large-scale deployment

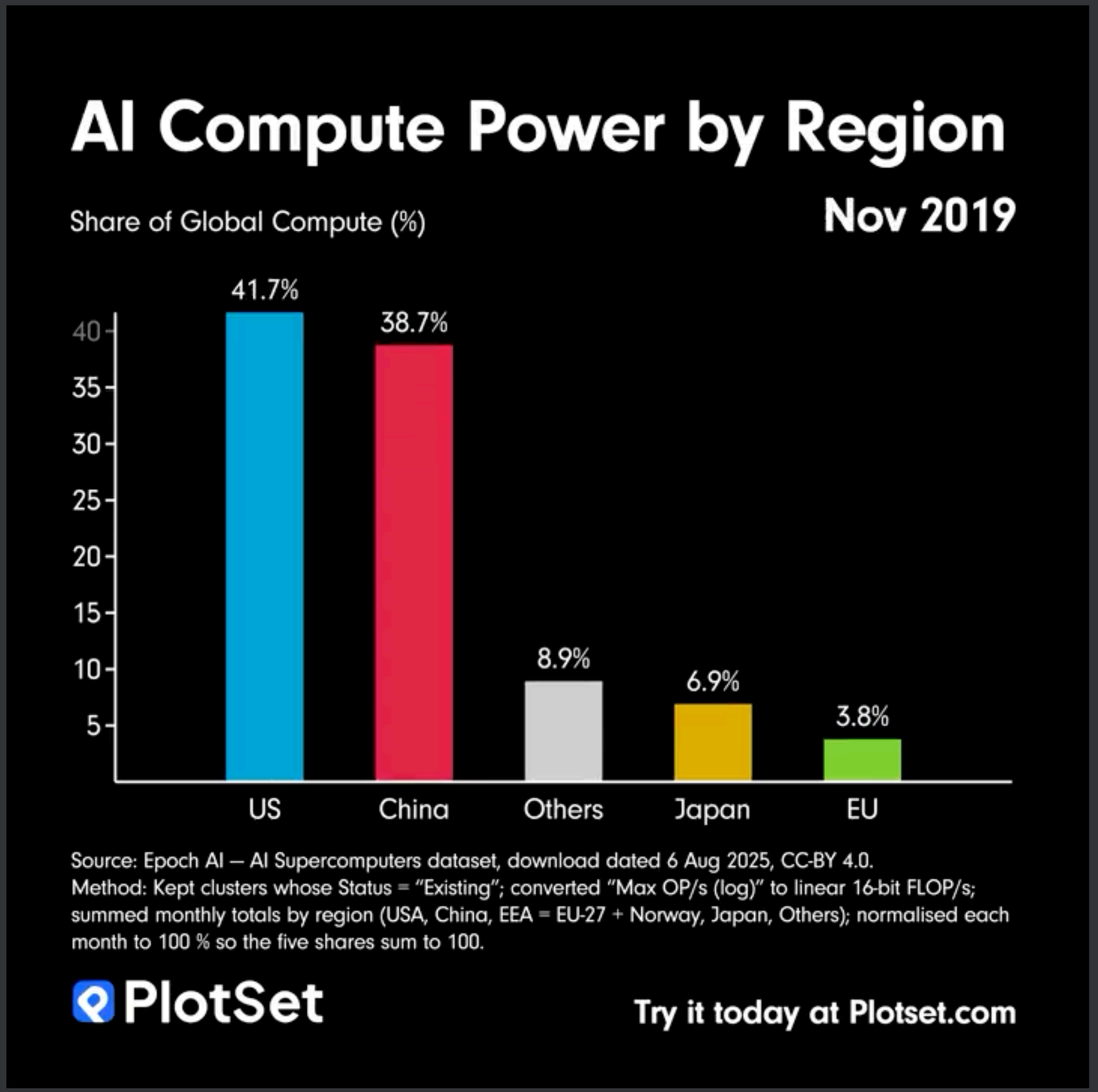


Monitor use, progress or adoption abroad



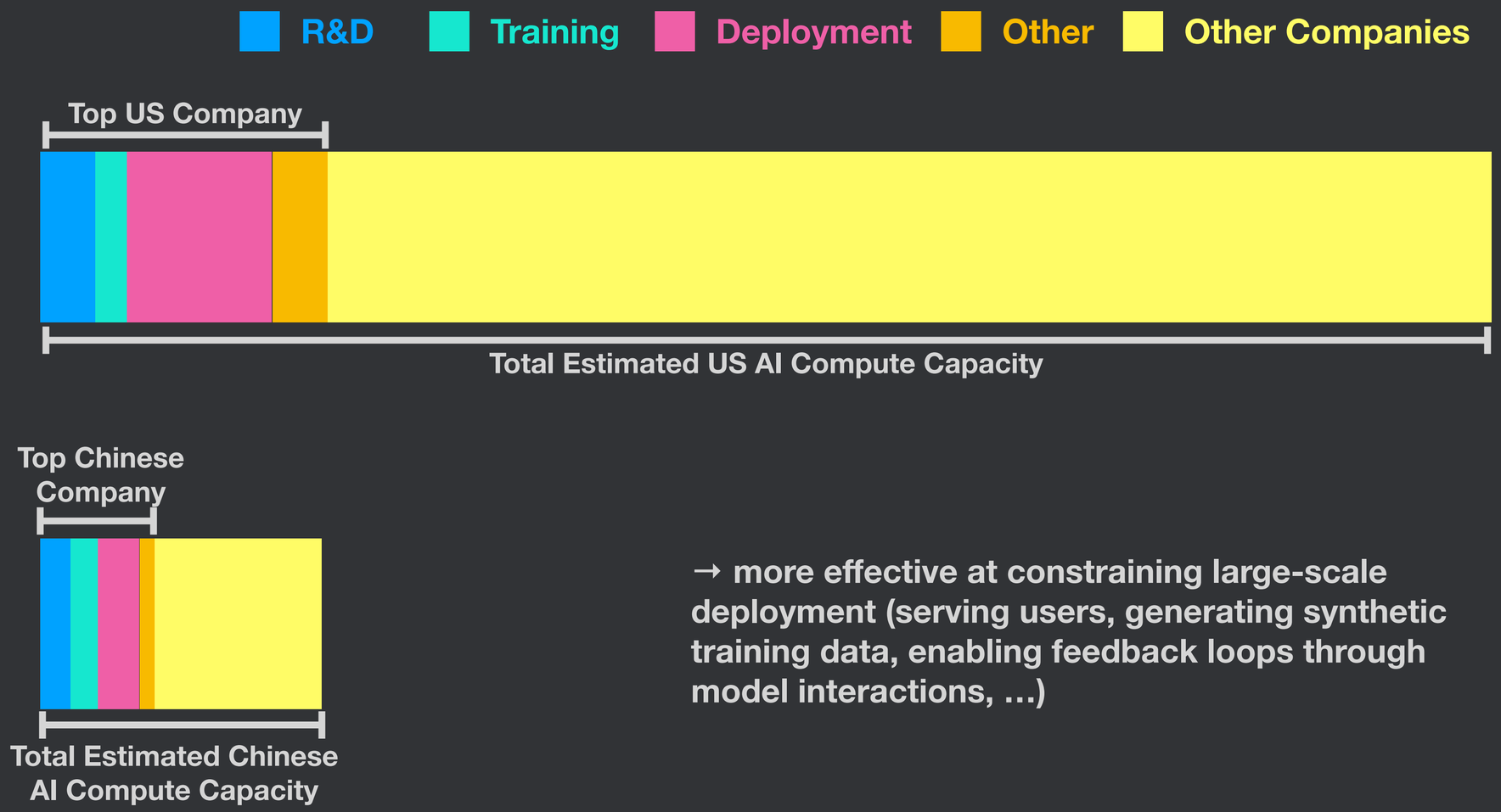
Prevent harmful use cases/models





Based on data in Pilz et al. (2025)

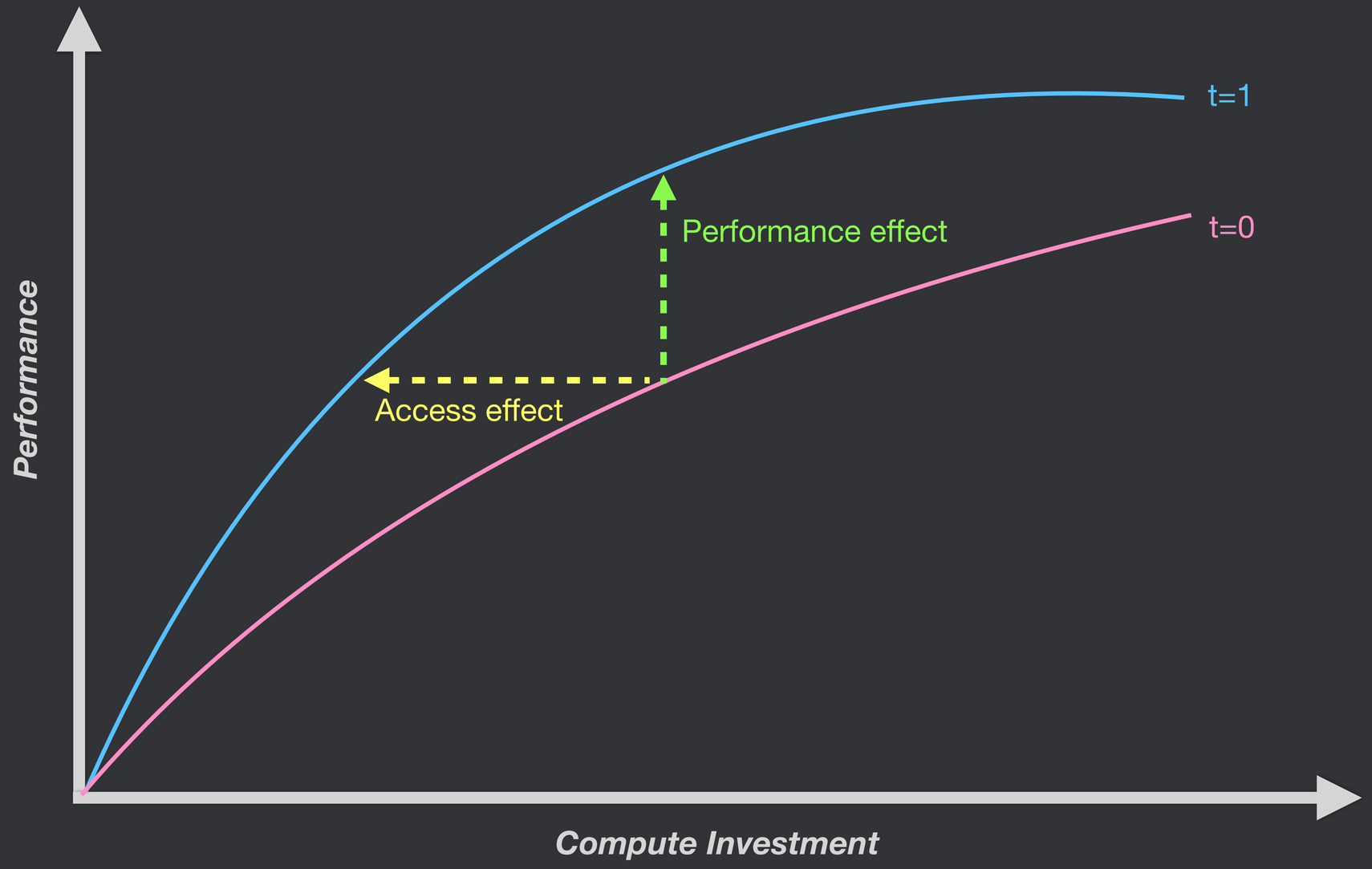
Export controls don’t automatically limit single training runs



→ more effective at constraining large-scale deployment (serving users, generating synthetic training data, enabling feedback loops through model interactions, …)

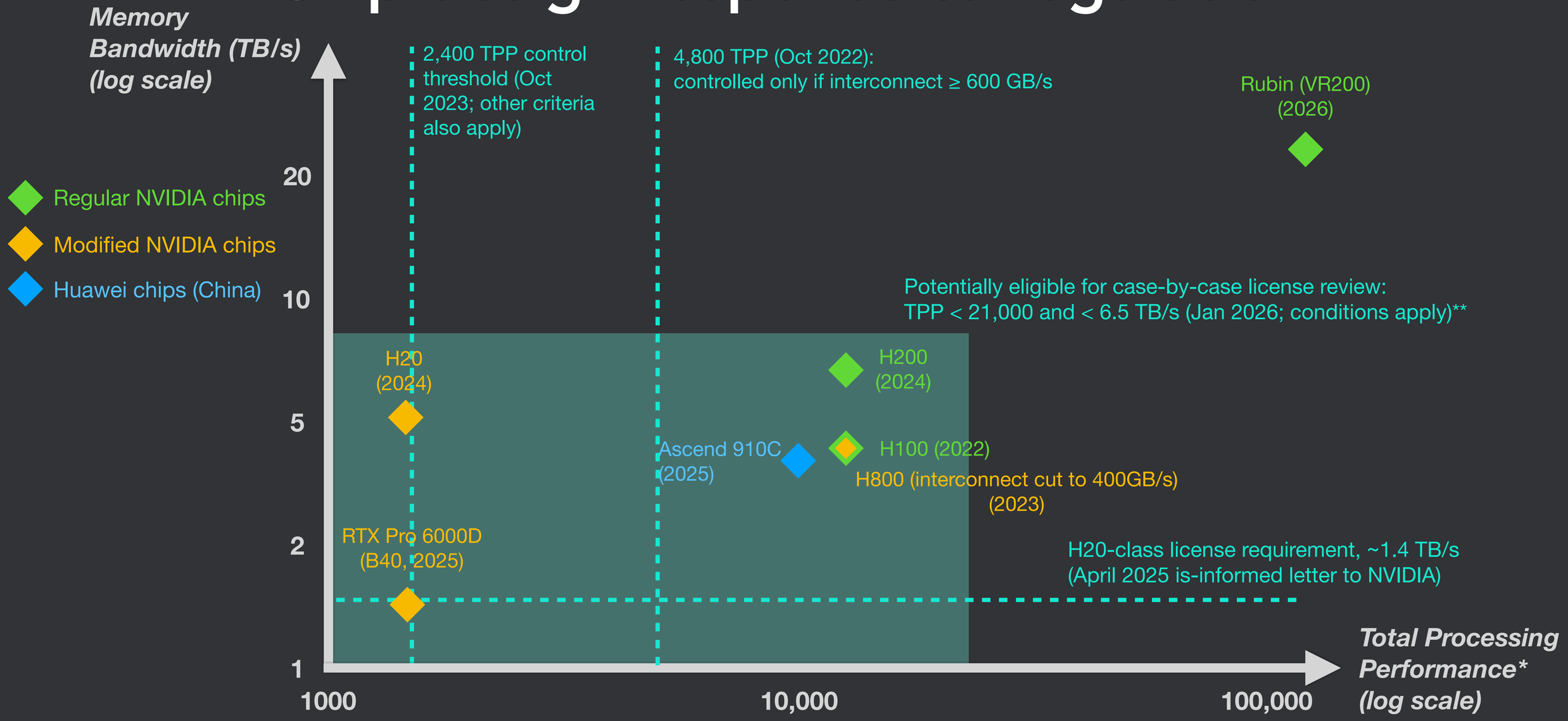
Based Pilz et al. (2025) and Heim (2025); note that these graphs are for illustrative purposes only & is based on 2025 estimates

Effects of compute on performance over time



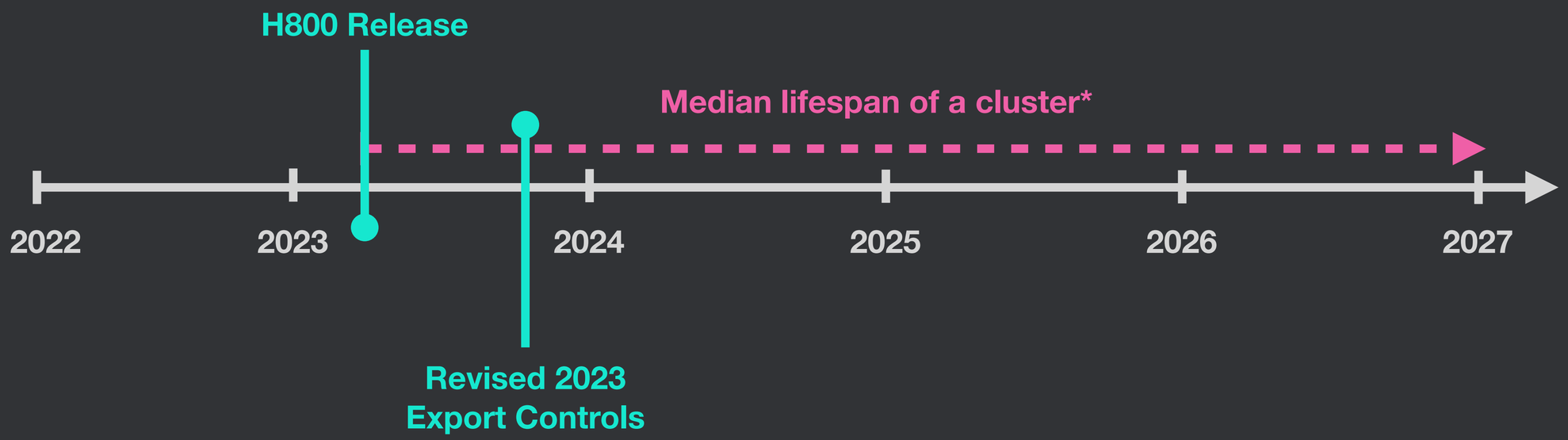
Source: Heim (2025)

Chip design responds to regulation



Source: adapted from Heim (2025); updated Oct 2026 from NVIDIA datasheets; BIS rule 91 FR 1684; *BIS measure; ** qualifying US exports to China/Macau only

Median lifespan of a cluster is 3.7 years



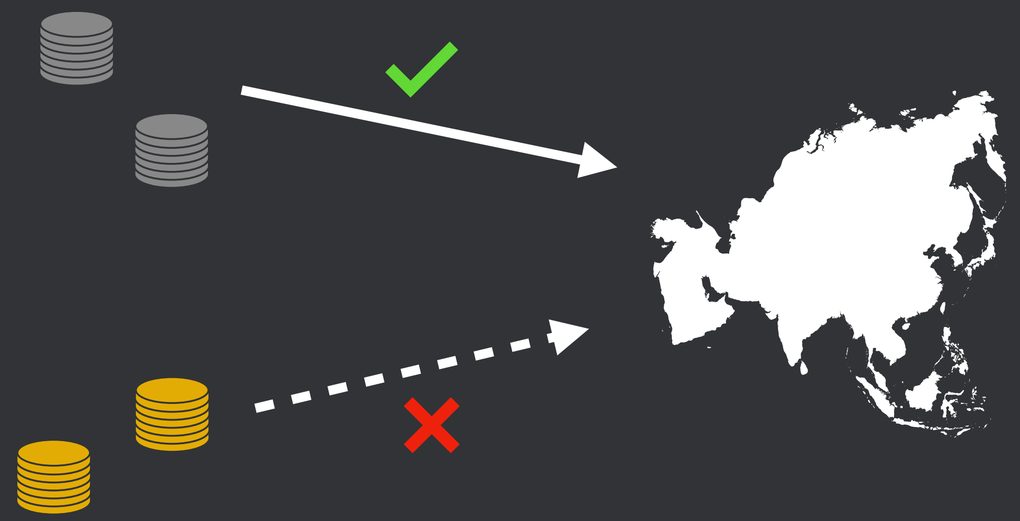
Source: EpochAI (2025); *based on 2025 estimates

Option 1: Export Controls

Constrain deployment

Constrain training

Idea: Limit export of powerful chips to certain countries



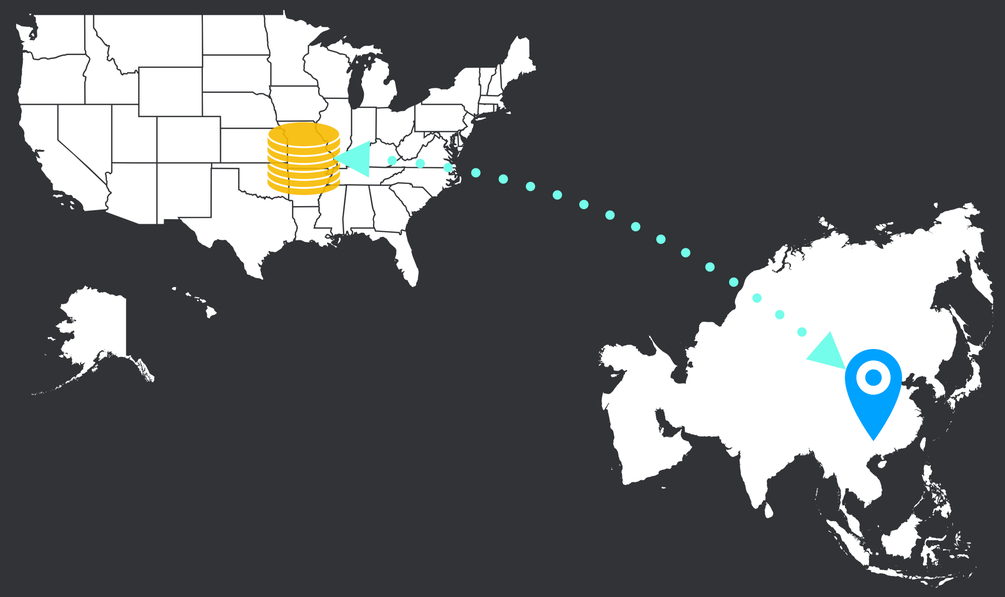
- Reduce total ecosystem capacity
- Somewhat effective at preventing cont'd model improvements (adoption/ usage feedback, preference data, …)

- Control of smuggling/use once shipped
- Incentivizes abroad chip development
- Incentivizes circumvention
- Time lag
- Doesn't constrain remote access

Option 2: Know-Your-Customer (KYC)

Constrain deployment | **Monitoring** | Constrain training

Idea: Require cloud providers to verify customers' identities and restrict access to advanced compute for unauthorized users



- Partially addresses remote access that physical chip export controls may not prevent

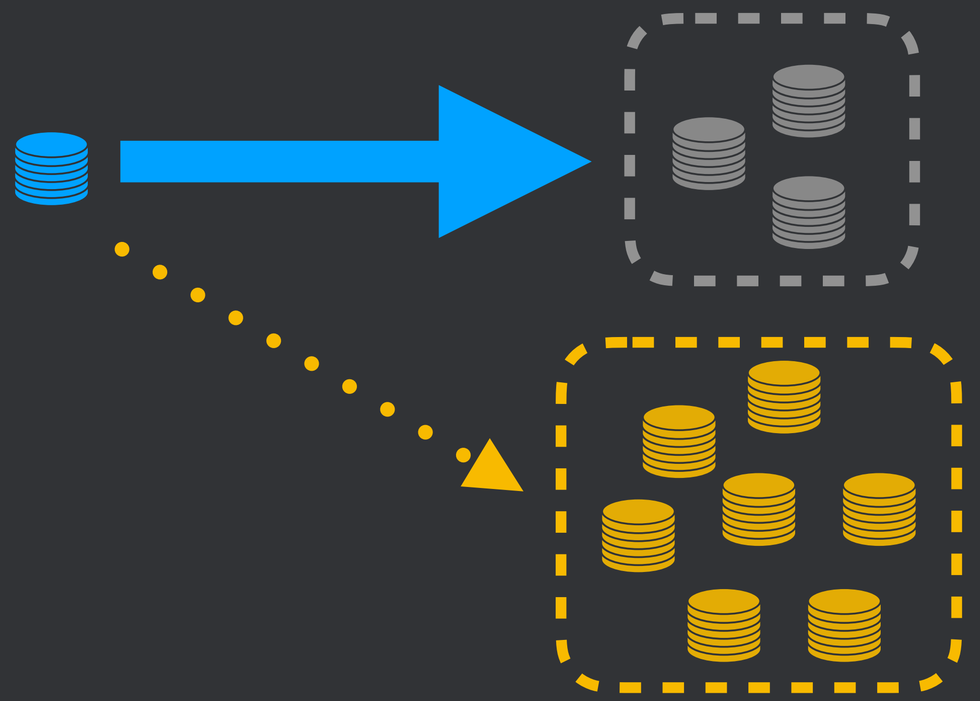
- Identity verification can be circumvented through intermediaries, resellers, and false identities

*January 2026 BIS rule includes KYC requirements to prevent unauthorized remote access to certain exported chips; broader cloud KYC requirements remain unresolved

Option 3: Bandwidth Bottleneck

Constrain training

Idea: High-bandwidth communication is limited to a certain number of chips to constrain the ability to assemble sufficiently large, fast-interconnected clusters (rather than aggregate compute)



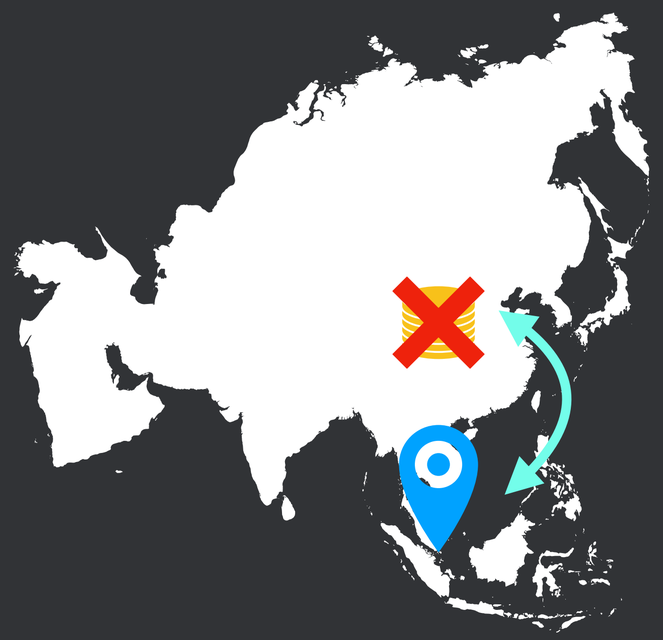
- Directly targets frontier model training
- Doesn’t restrict moderate computing

- Relies on continuation of current training paradigm → more compute means more capability
- Could be spoofed with sophisticated packaging

Option 4: Geofencing

Constrain deployment | **Monitoring** | Constrain training

Idea: Shut down chip if verification fails or chip is likely in prohibited country



- Low-overhead, automatic enforcement
- Supports intelligence on illicit flows

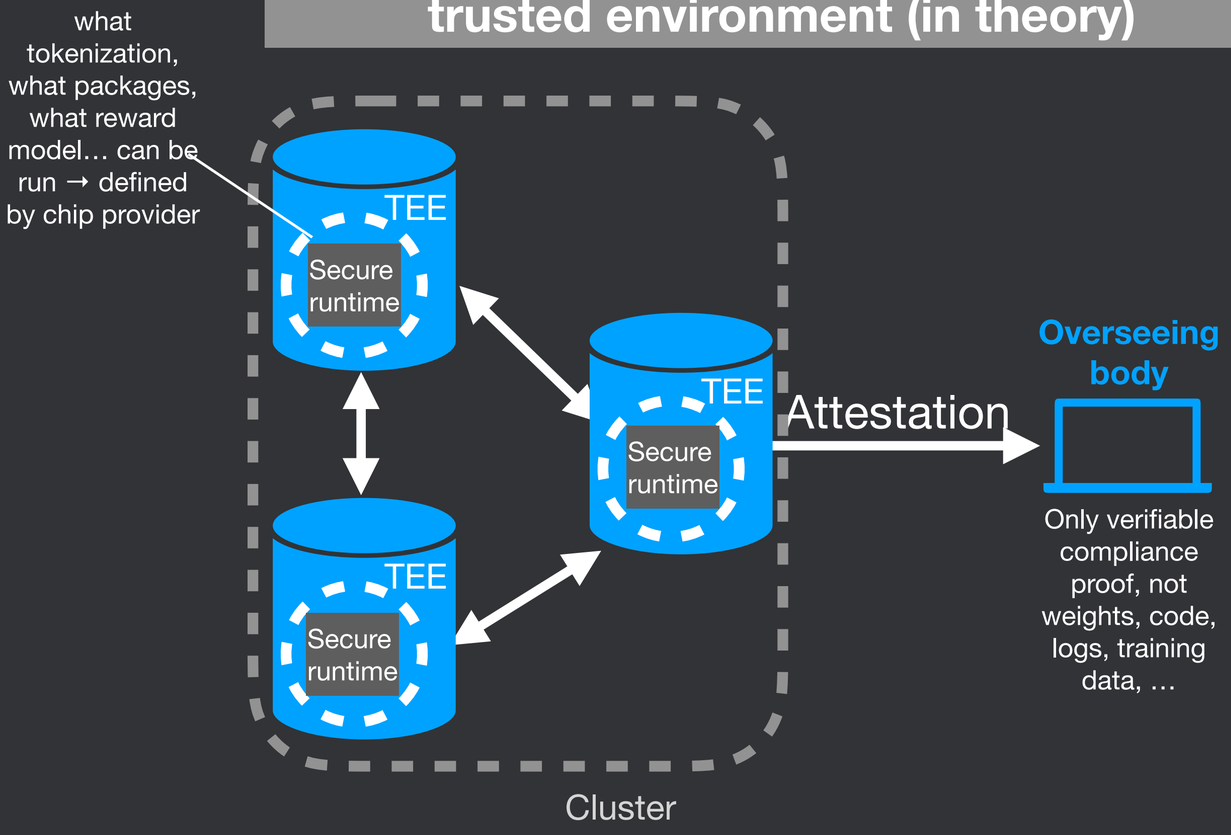
- Strong spoofing resistance is hard
- Doesn’t control how compute is used or who uses it
- Usability concerns for legitimate buyers
- Doesn’t constrain remote access

Option 5: Trusted Execution Environments

Prevent use cases

Monitoring

Idea: Control *how* compute is used in trusted environment (in theory)



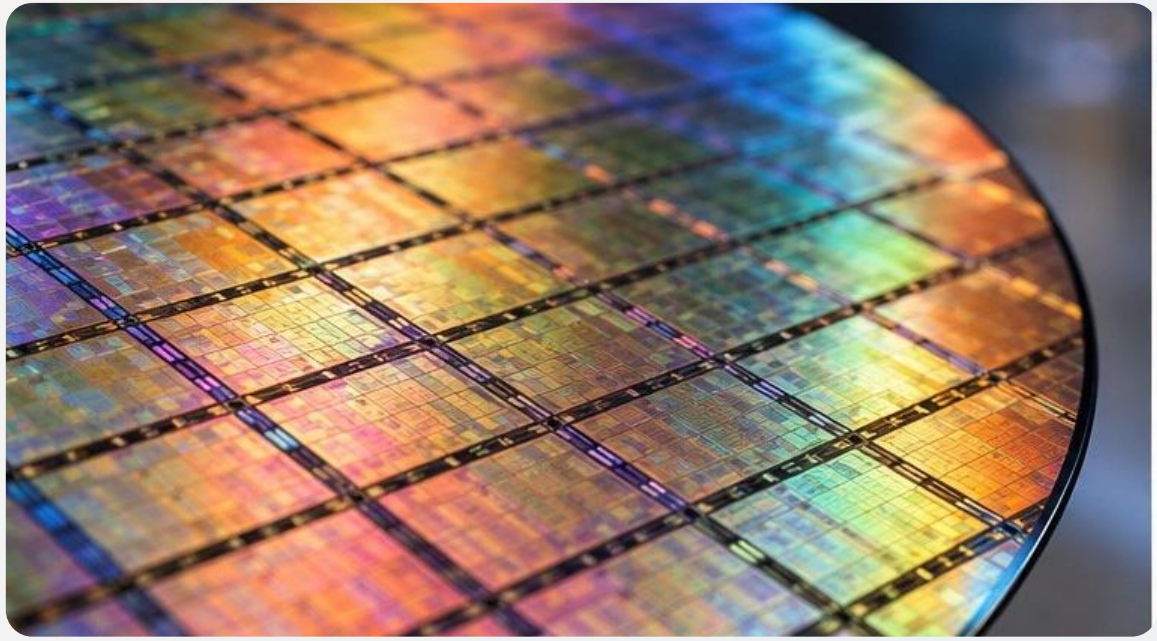
- More nuanced governance of problematic *uses* of compute
- Could be used for more crude governance, too (e.g., restrict FLOPS)

- Relatively new application area for TEEs, many open research questions
- Requires a good understanding of the “profile” of an unwanted workload
- Not possible with current chips

These approaches can be mixed-and-matched.

Core question to ask: What are we trying to achieve and what instruments accomplish that goal?

AI Infrastructure





Engineering, Geopolitics, Economics, and the Environment

Semiconductor supply chains · Data centers · Energy, climate, water and minerals

Nate Persily | Stanford Law School | October 2026

AI Infrastructure: Engineering, Economics, Geopolitics, Environment

Who makes it? Where? Who captures the value? Who bears the costs?

- Semiconductors: design, machines, fabs, packaging, critical materials.
- Data centers: accelerators, networking, construction, electricity, cooling.
- Economics and geopolitics: investment, market power, export controls and Taiwan.
- Environment and governance: power, climate, water, land and politics.





Sources: Semiconductor Industry Association / BCG · International Energy Agency

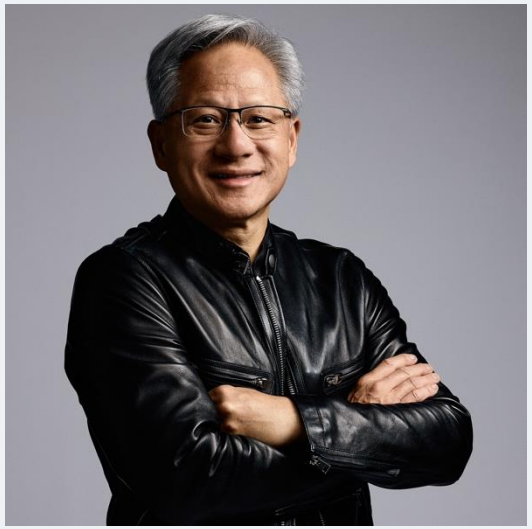
Pause to contemplate the miracle of modern semiconductors



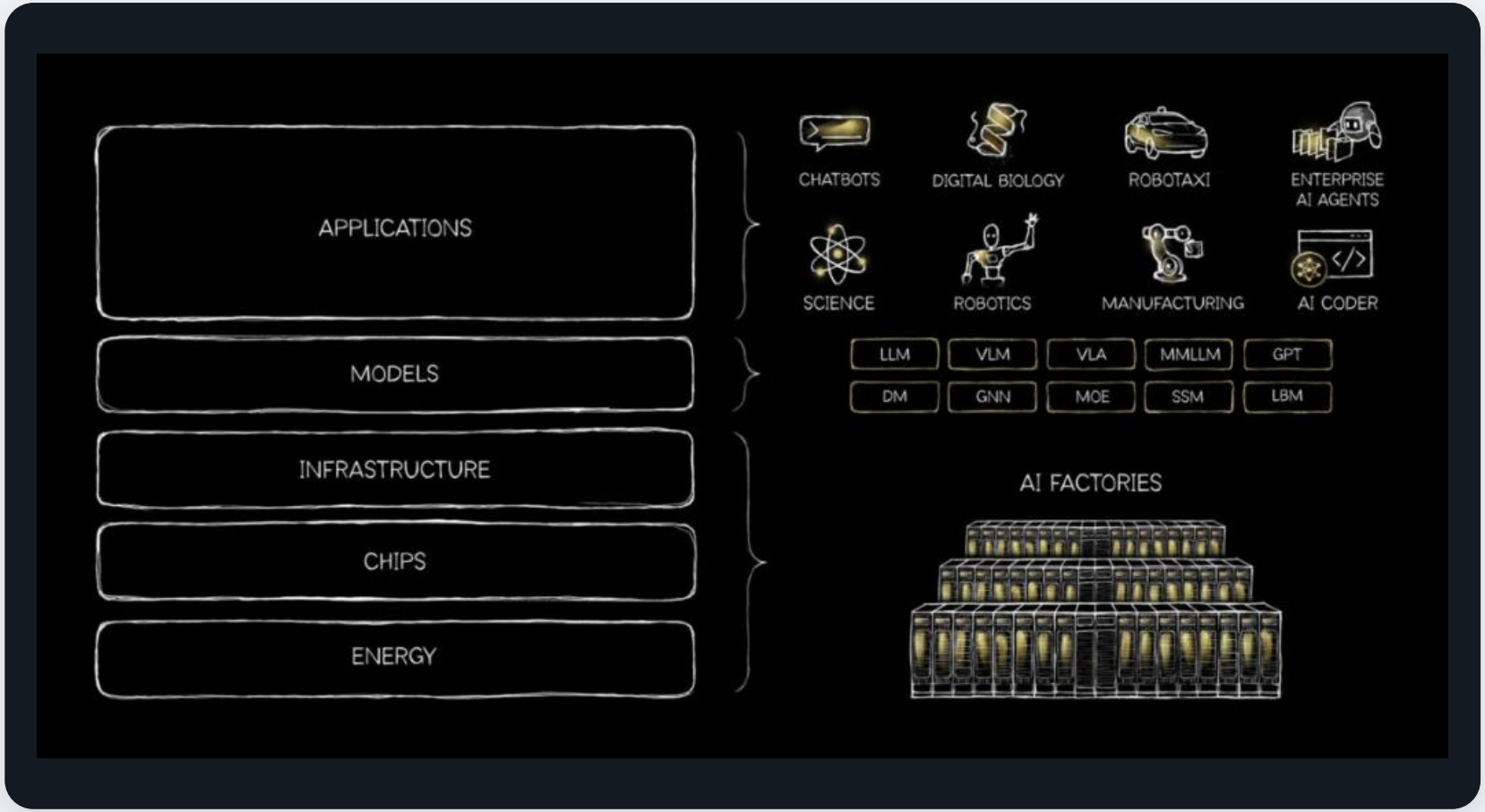
> *“If you stop to think about it, we've essentially found a way to make sand think. It's miraculous.”*

Demis Hassabis · Google DeepMind

AI as a five-layer cake



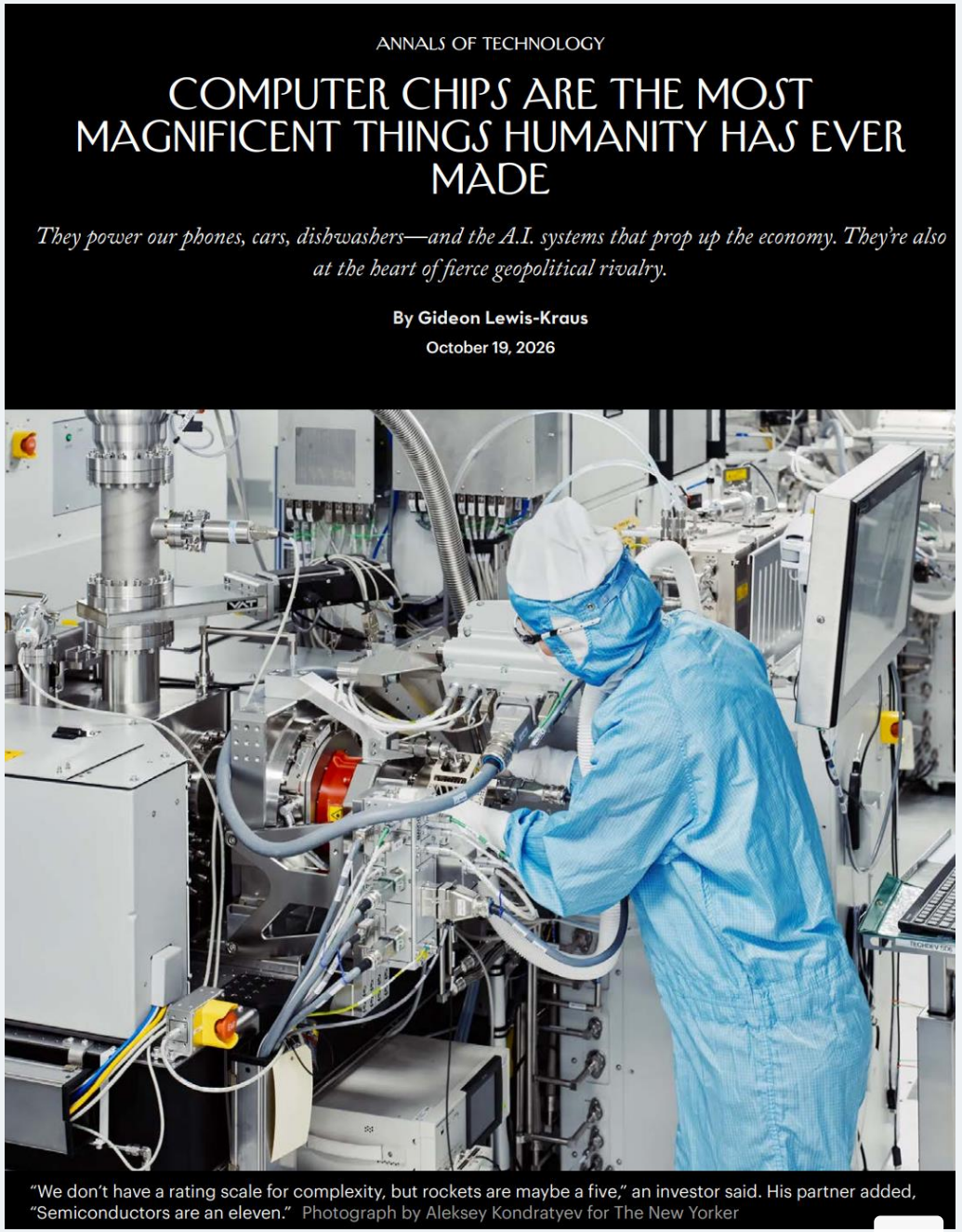
Applications rely on models; models require infrastructure, semiconductors, and electricity.



Chip manufacturing is an ecosystem, not just a factory

- An advanced semiconductor contains billions of transistors acting as tiny electrical switches.
- Manufacturing depends on silicon purification, EDA software, photoresists, masks, precision tools, inspection and packaging.
- No single company or country supplies every indispensable part of the process.
- Mass production requires highly disciplined manufacturing: contamination or tiny defects can destroy working chips.

Source: Gideon Lewis-Kraus, The New Yorker, October 19, 2026.



Before chips: the minerals and materials behind AI

Every AI accelerator and every data center begins with mining, refining and highly specialized industrial inputs.

INSIDE THE CHIP

- Silicon: ultra-pure wafers
- Copper & tungsten: connections
- Hafnium, cobalt, tantalum: some advanced process layers
- Gallium / germanium: selected photonics and power devices

INSIDE THE DATA CENTER

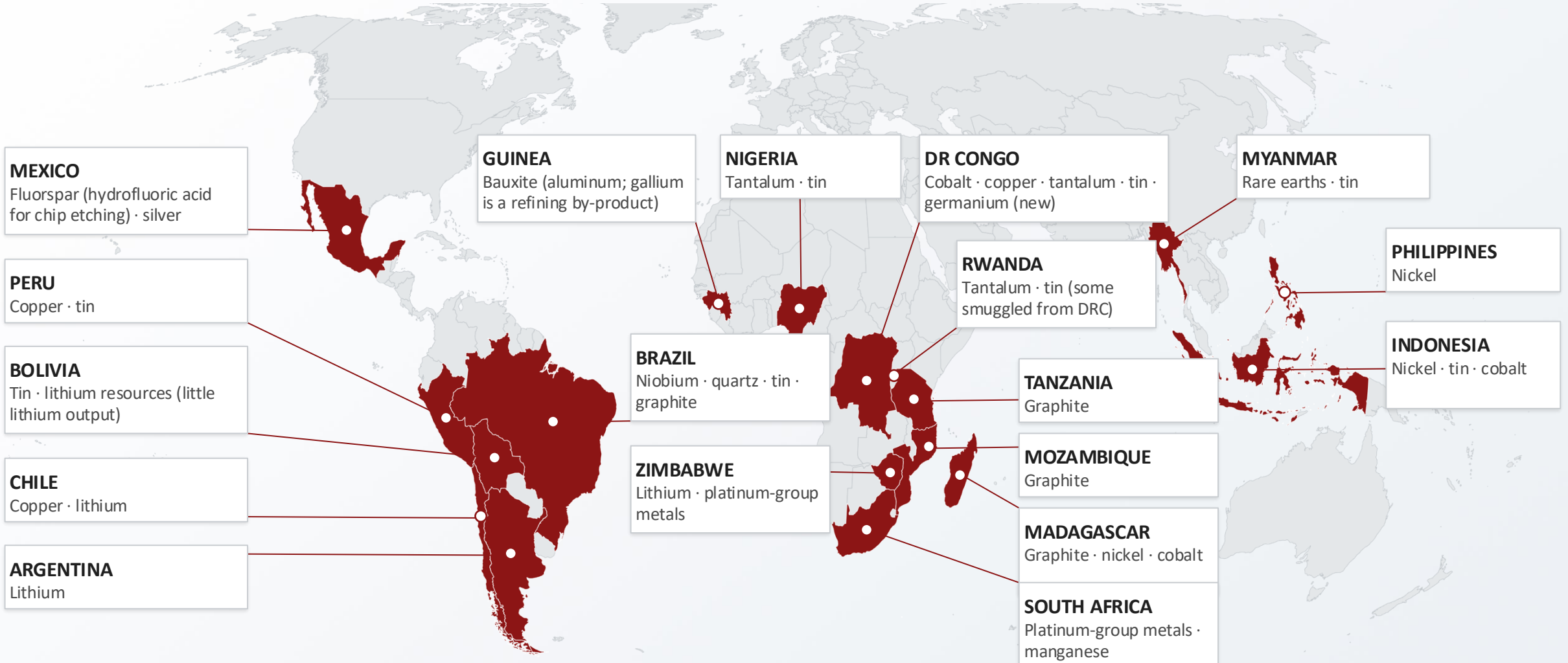
- Copper and aluminum: power conductors
- Steel, aluminum, concrete: buildings / racks
- Lithium, nickel, graphite: some batteries
- Rare-earth magnets: motors, fans and generators

WHO PROCESSES IT?

- Purity and qualification are chokepoints
- China prominent in several refining chains
- Japan / Europe important for specialty chemicals
- Mine location ≠ final processing location

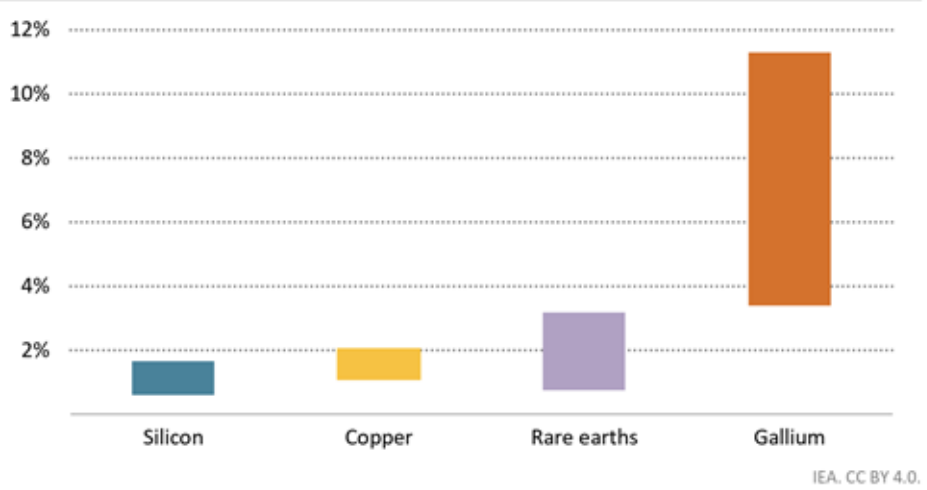
Sources: USGS Mineral Commodity Summaries (2026); IEA Global Critical Minerals Outlook (2025); chipsupplychain.org

The Global South in the semiconductor supply chain



Illustrative extraction locations, not a ranking of output. Mining only — not refining, purification, or fabrication. Several minerals (lithium, nickel, graphite, cobalt, manganese) mainly serve batteries, power, and data-center infrastructure rather than chips themselves. China, the dominant producer of rare earths, graphite, gallium, and germanium, is not shown.
Sources: USGS, Mineral Commodity Summaries 2026; IEA, Global Critical Minerals Outlook 2025; Global Witness (2026); Umicore (2024). Map: Natural Earth.

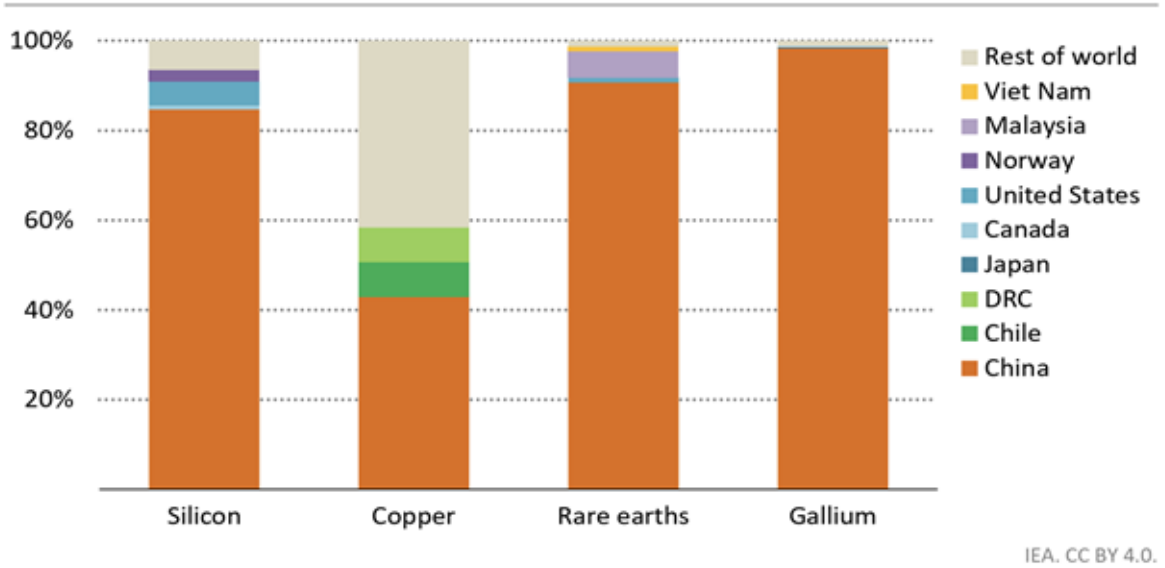
Figure 5.4 ▷ Demand for critical minerals required to meet the growth in data centre capacity in 2030 as a share of their total demand in 2024



Data centre growth to 2030 will have varied impacts on mineral demand; the share of total demand is small for traditional metals, but security of the mineral supply will still be critical

Note: The bands for each mineral represent the estimated range of their demand from AI data centres in 2030 as a share of their total demand in 2024.

Figure 5.5 ▷ Geographical concentration of the supply of selected refined critical minerals needed for data centre expansion, 2024



The supply of many refined minerals essential to the construction of data centres is highly concentrated in a handful of regions, making supply chains vulnerable to disruptions

Note: DRC = Democratic Republic of the Congo.

International Energy Agency (IEA), Energy and AI (2025),

Critical minerals ≠ rare earth elements

These terms describe different classes of material—and their exact function in the supply chain matters.

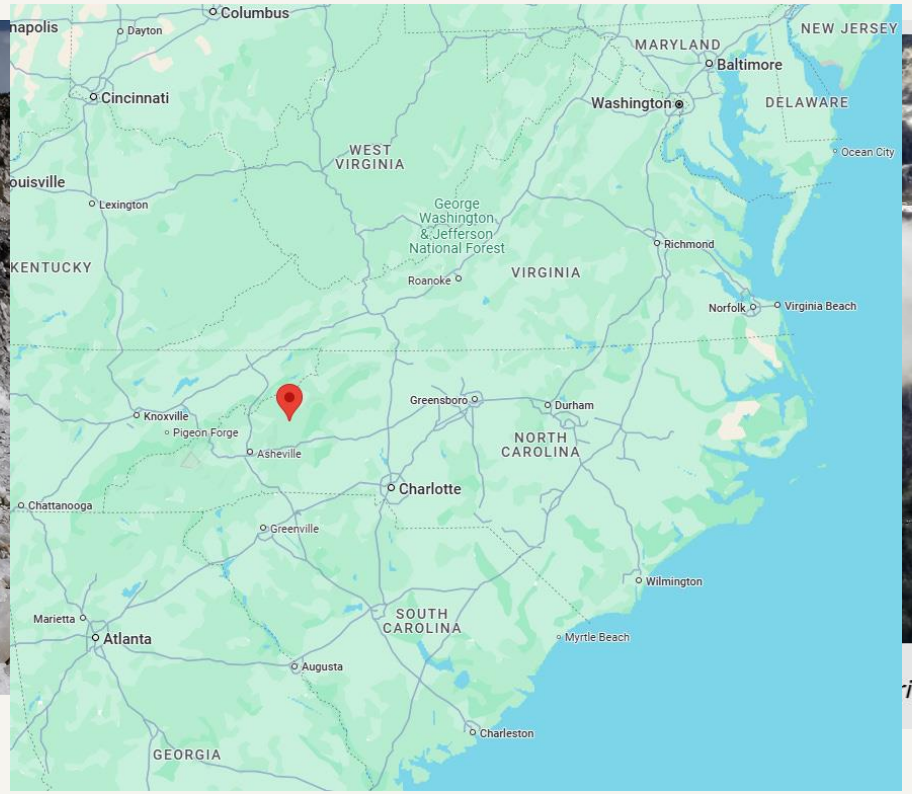
- Silicon and copper are essential to wafers and interconnects; neither is a rare earth.
- Gallium and germanium are important in compound semiconductors, electronics, optics and related equipment—not the main silicon logic substrate.
- Rare earth elements are significant for magnets, motors and components across industrial systems.
- Neon and other process gases, photoresists and specialty chemicals create different chokepoints.

Silicon + copper	Wafers / interconnects	Not rare earths
Gallium + germanium	Compounds / optics / parts	Not rare earths
Neon + resists	Process gas / lithography	Not minerals in same sense
Rare earths	Magnets / industrial equipment	17-element family

Sources: U.S. Geological Survey · Semiconductor Industry Association / BCG

70-90% of the high purity quartz to make silicon ingots comes from….





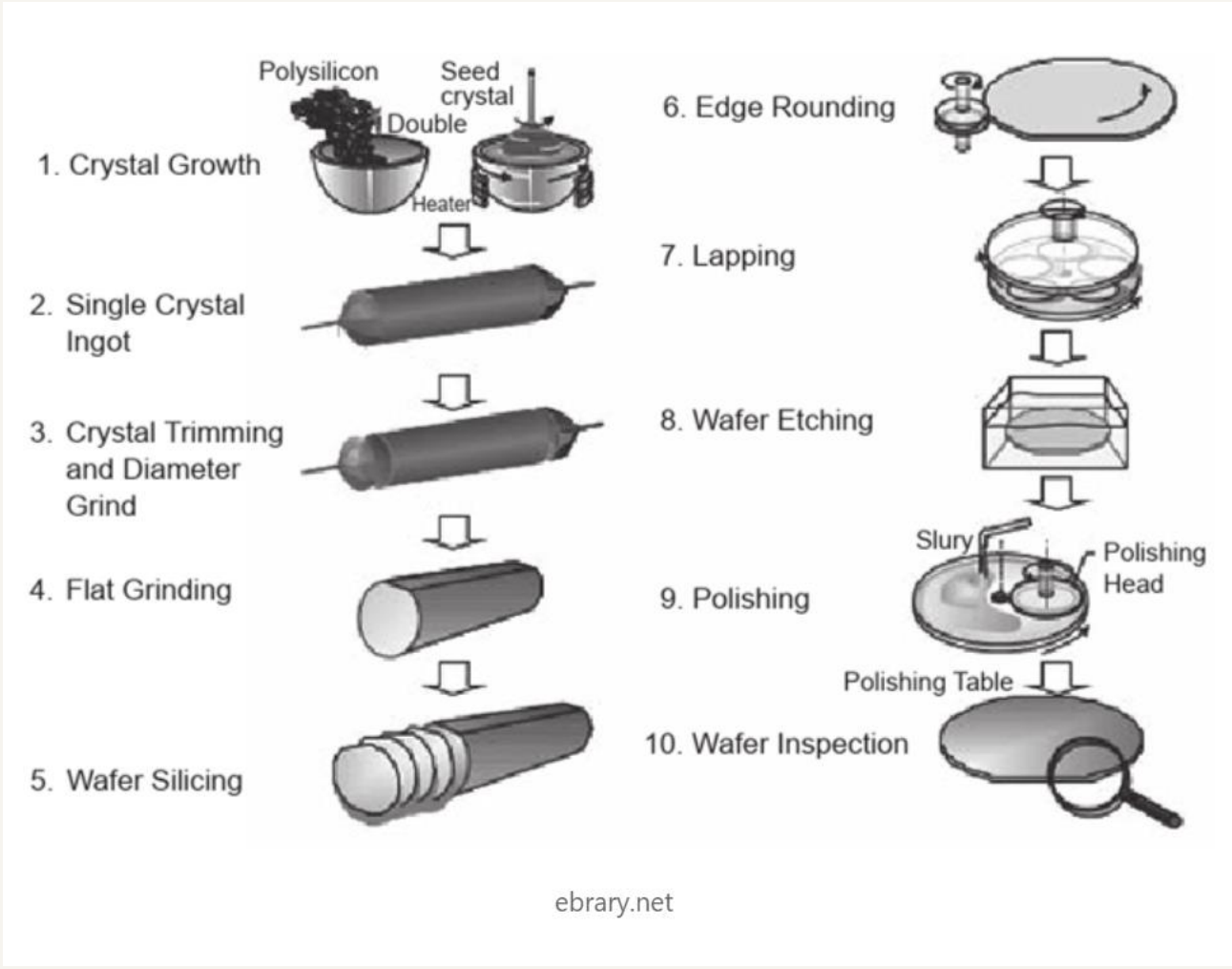


rity quartz mines in North Carolina)

Spruce Pine, North Carolina!

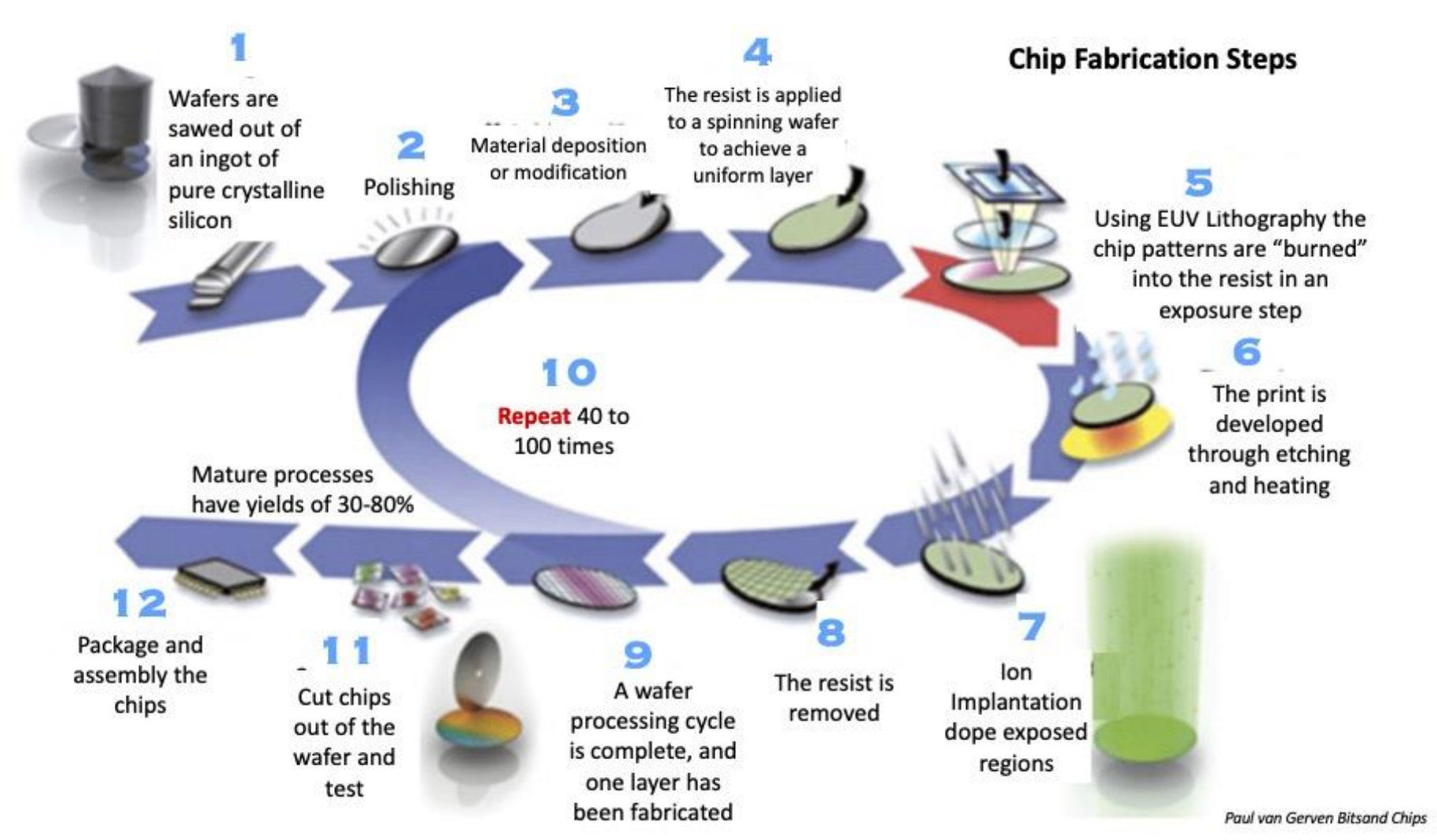
It all starts with Silicon







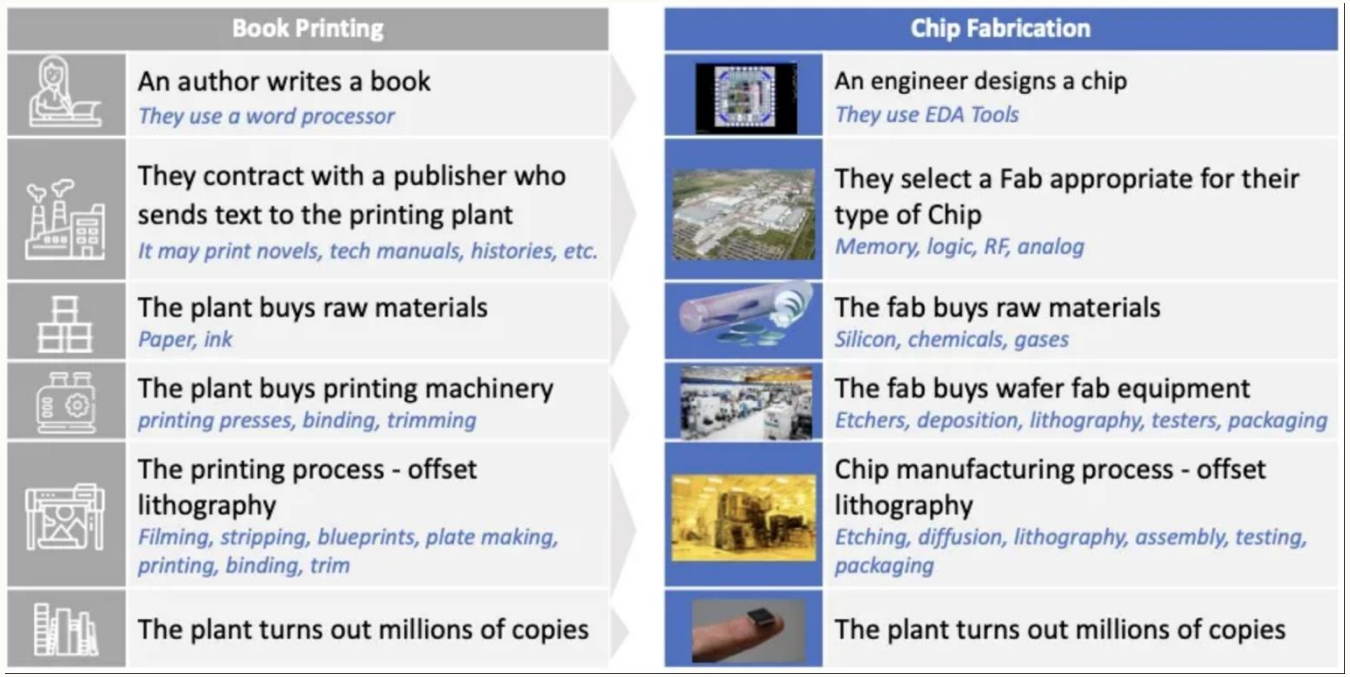
From wafers to chips...



Credit: Prof. Steve Blank -- https://steveblank.com/2022/01/25/the-semiconductor-ecosystem/

A foundry is a printing plant—operating near the atomic scale

The printing analogy clarifies what NVIDIA designs, what TSMC builds, and why tools and materials matter.



Design software : word processor · foundry : printing plant · masks, chemicals and tools : ink and presses

Source: Steve Blank, The Semiconductor Ecosystem (2022).

ASML: the monopoly hidden inside every leading-edge fab

ASML, based in Veldhoven, Netherlands, is the only commercial manufacturer of EUV lithography systems.

- EUV light at 13.5 nm helps print extremely small patterns on silicon wafers.
- ASML's scanner depends on Carl ZEISS SMT (Germany), TRUMPF laser technology (Germany) and many other suppliers.
- TSMC, Samsung and Intel buy these systems; the U.S. and allies can restrict exports and servicing.

ASML's overview of how microchips are made.

ASML short process video ↗

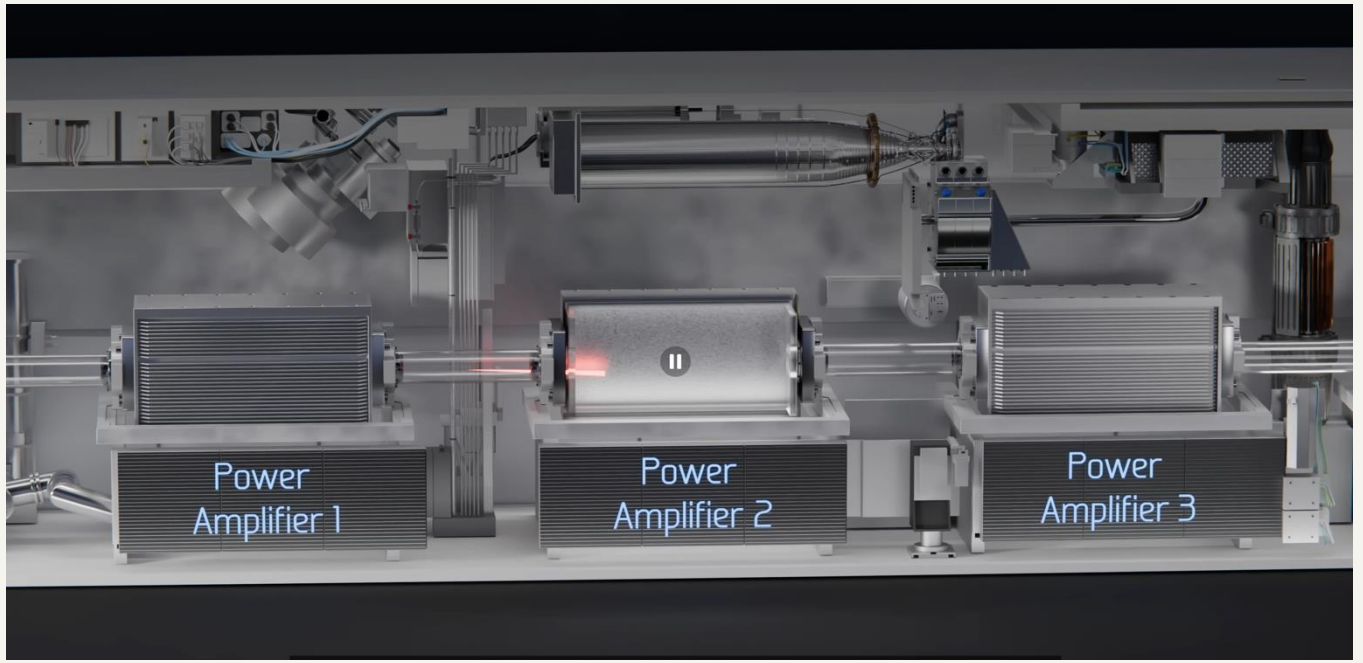


EUV: 13.5 nm light; hundreds of subsystems; commercial sole supplier

Sources: ASML · NIST

The Amazing Invention of Ultraviolet Lithography



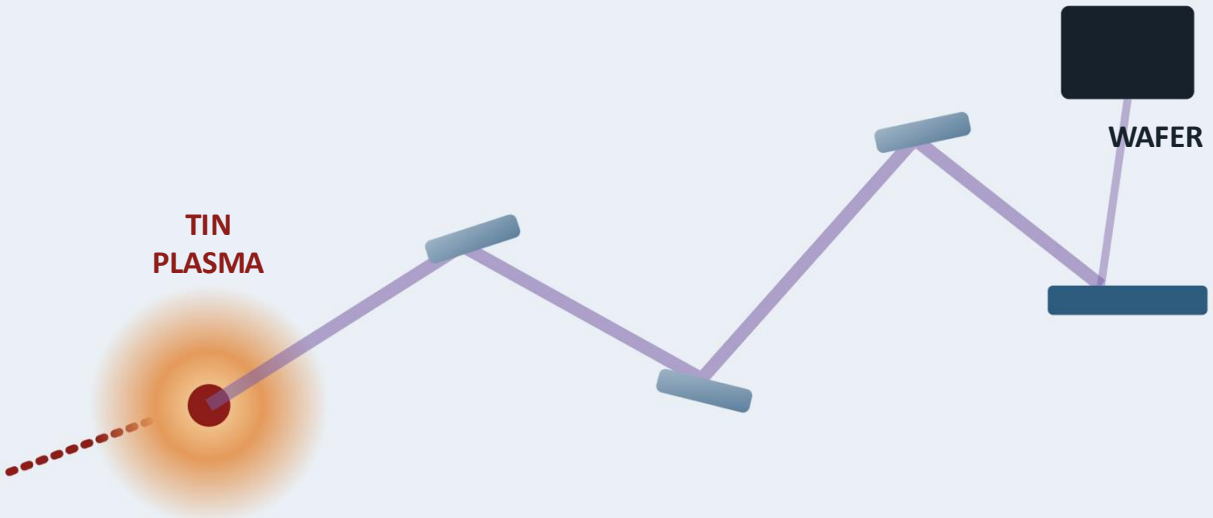


https://youtu.be/B2482h_TNwg?si=F-AxQ48ddcIKC1BI

Gideon Lewis-Kraus, "Computer Chips Are the Most Magnificent Things Humanity Has Ever Made," The New Yorker (Oct. 19, 2026)

Inside EUV: a laser, tin plasma and precision mirrors

In vacuum, 13.5-nm ultraviolet light bounces through mirrors rather than glass lenses.



ASML

System integrator: Netherlands

ZEISS SMT

Optical mirrors: Germany

Laser + source

TRUMPF (Germany); ASML (U.S.)

Visual adaptation: Chip Supply Chain, Lithography; ASML and ZEISS technical materials

Nine astonishing facts about semiconductor fabrication

99.999999999%

Purity of electronic-grade polysilicon used to grow silicon wafers

208 BILLION

Transistors in NVIDIA Blackwell: one GPU composed of two connected dies

220,000° C

Tin plasma temperature in the ASML EUV light source; approximately 40 times hotter than the Sun's surface

60,000 / SEC

Laser-struck molten tin droplets generate EUV light used to print circuit patterns

0.1 MM

Maximum surface unevenness if a ZEISS EUV mirror were enlarged to the size of Germany

60 PICOMETERS

Accuracy of wafer-stage position sensors, measuring 20,000 times per second

7 x GRAVITY

Acceleration of the wafer stage that moves wafers beneath projected patterns

$350-400M

Reported approximate price of a High-NA EUV lithography system (not every ASML tool)

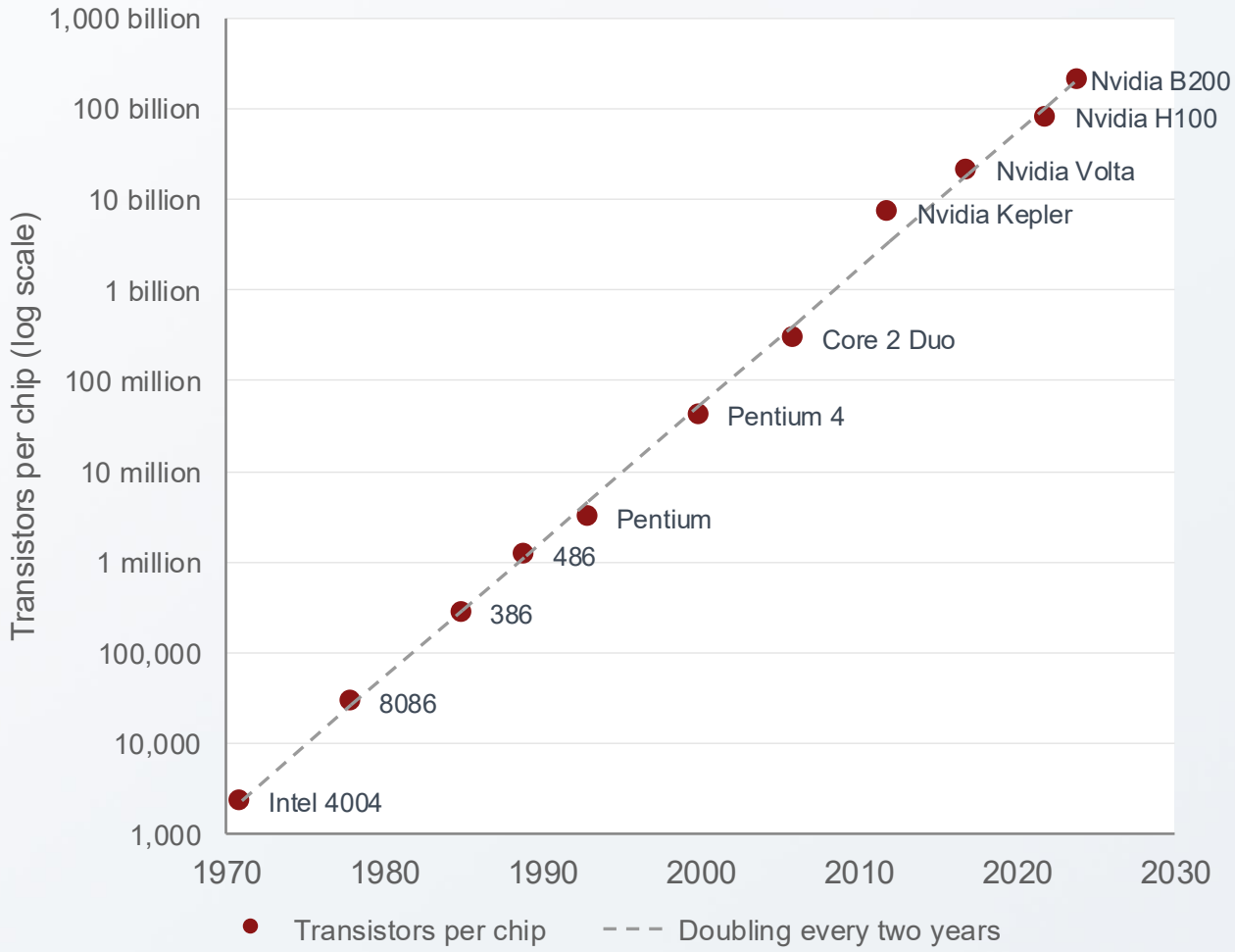
5,100

Suppliers in ASML's global network in 2025, including parts and services

Sources: Tokuyama; NVIDIA; ZEISS; ASML (technical materials and 2025 annual report). Picometer = one trillionth of a meter.

Moore's Law

- 1965: Gordon Moore predicts the number of components on a chip will double every year
  - "Cramming More Components onto Integrated Circuits," Electronics (Apr. 19, 1965)
- 1975: Moore revises it to doubling every two years
- An observation about engineering and economics, not a law of physics
  - More transistors per chip → faster, cheaper, more capable computers
- Today: slowing as transistors approach atomic scale
  - Gains now also come from stacking and combining chips and from specialized chips (GPUs)

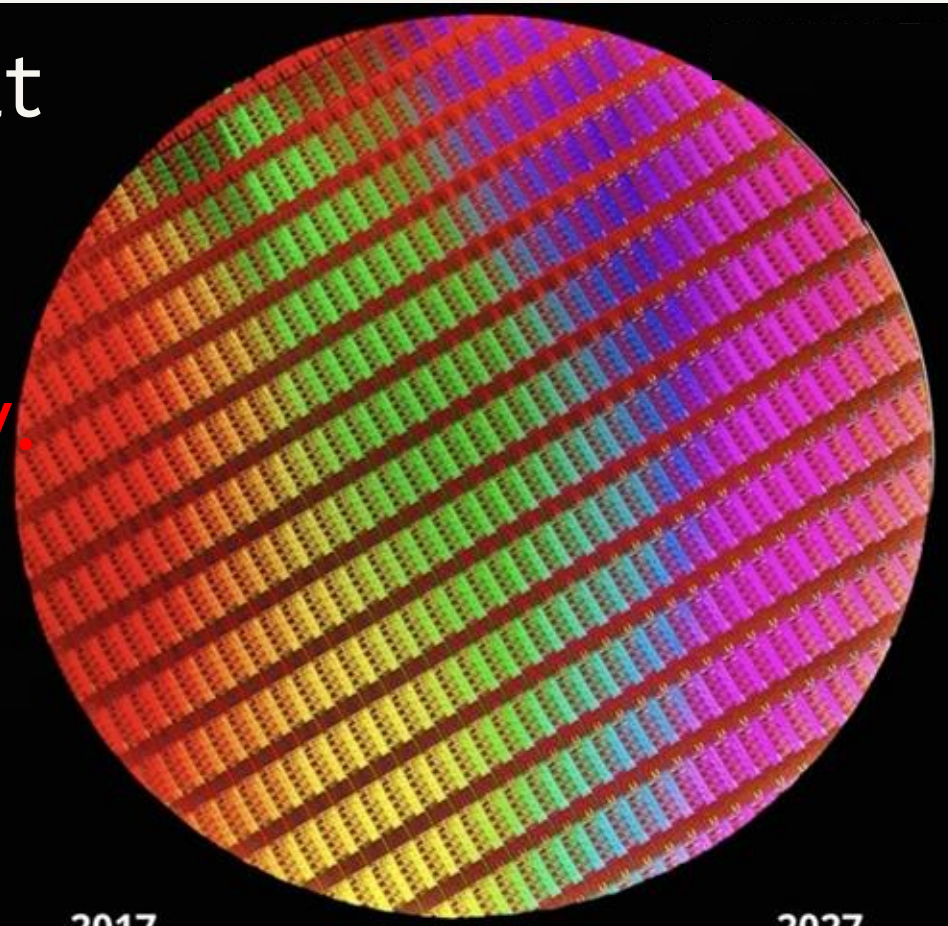


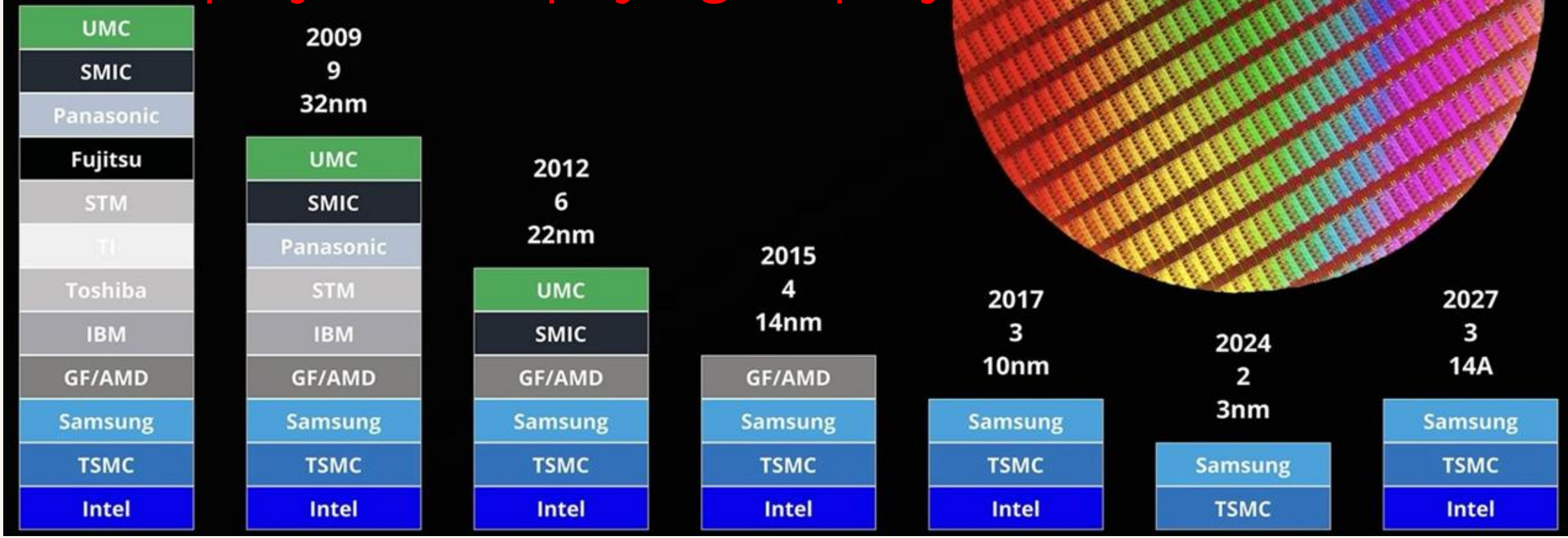
Transistor counts: manufacturer specifications. Nvidia B200 combines two dies in one package. Trend line: doubling every two years from the Intel 4004.

SECTION

Economics and Geopolitics of Chip Production

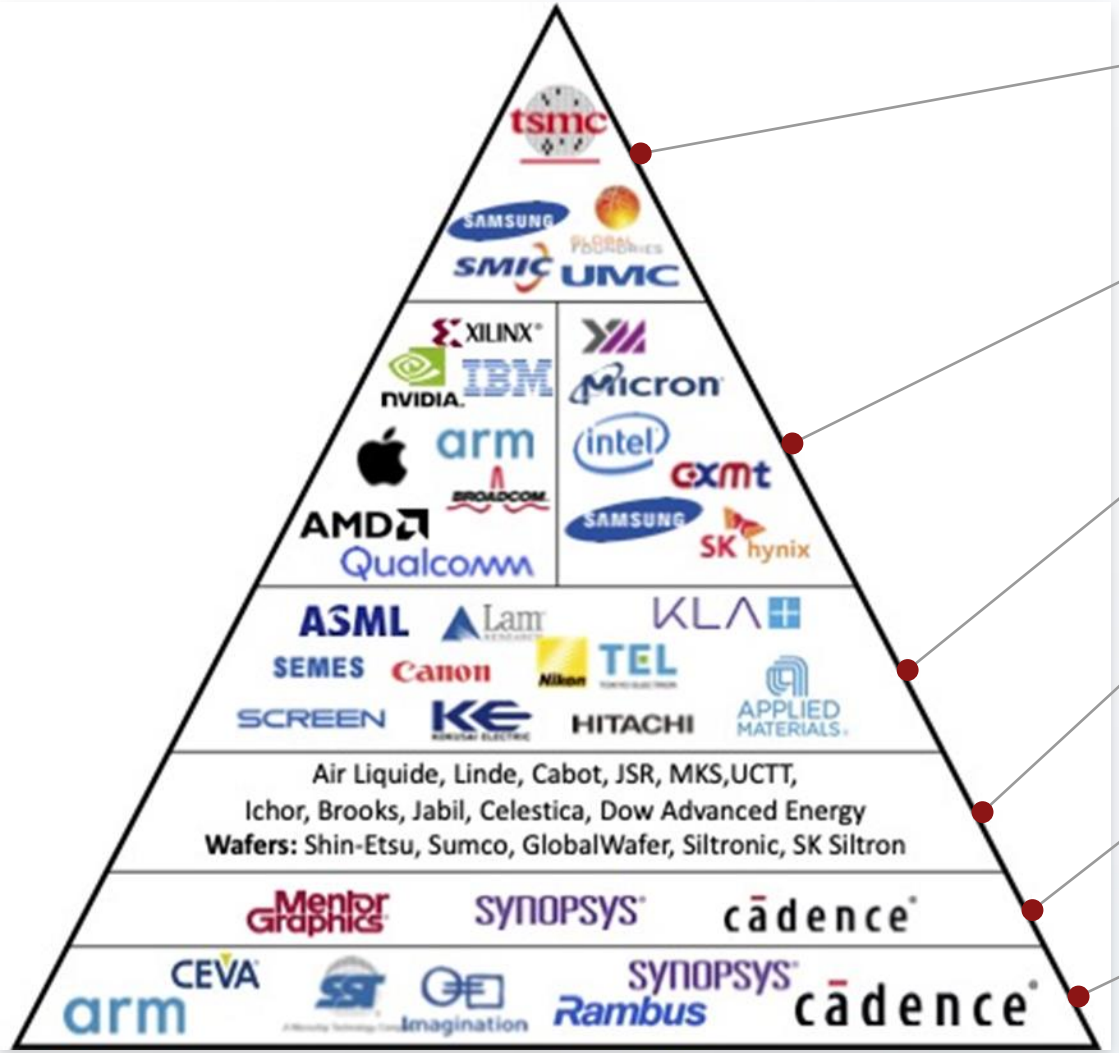
Wafer-scale fabrication at the leading edge is VERY capital intensive! Fewer players are paying to play.





Credit: Prof. Steve Blank, Deutsche Bank, IBS Consulting, Anastasi in Tech (revision)
https://steveblank.com/2022/01/25/the-semiconductor-ecosystem/
https://www.linkedin.com/posts/aginn_semiconductors-technology-ai-activity-7364126204066476033-3ASu/?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAA8i0YBkBL9PTIGLPimxOh5dnA9Jo

The Semiconductor Supply Chain



Foundries

Manufacture chips designed by other companies

Chip designers & memory makers

Left: fabless logic designers. Right: memory makers and integrated device manufacturers

Manufacturing equipment

Lithography, deposition, etch, and inspection tools

Materials, components & wafers

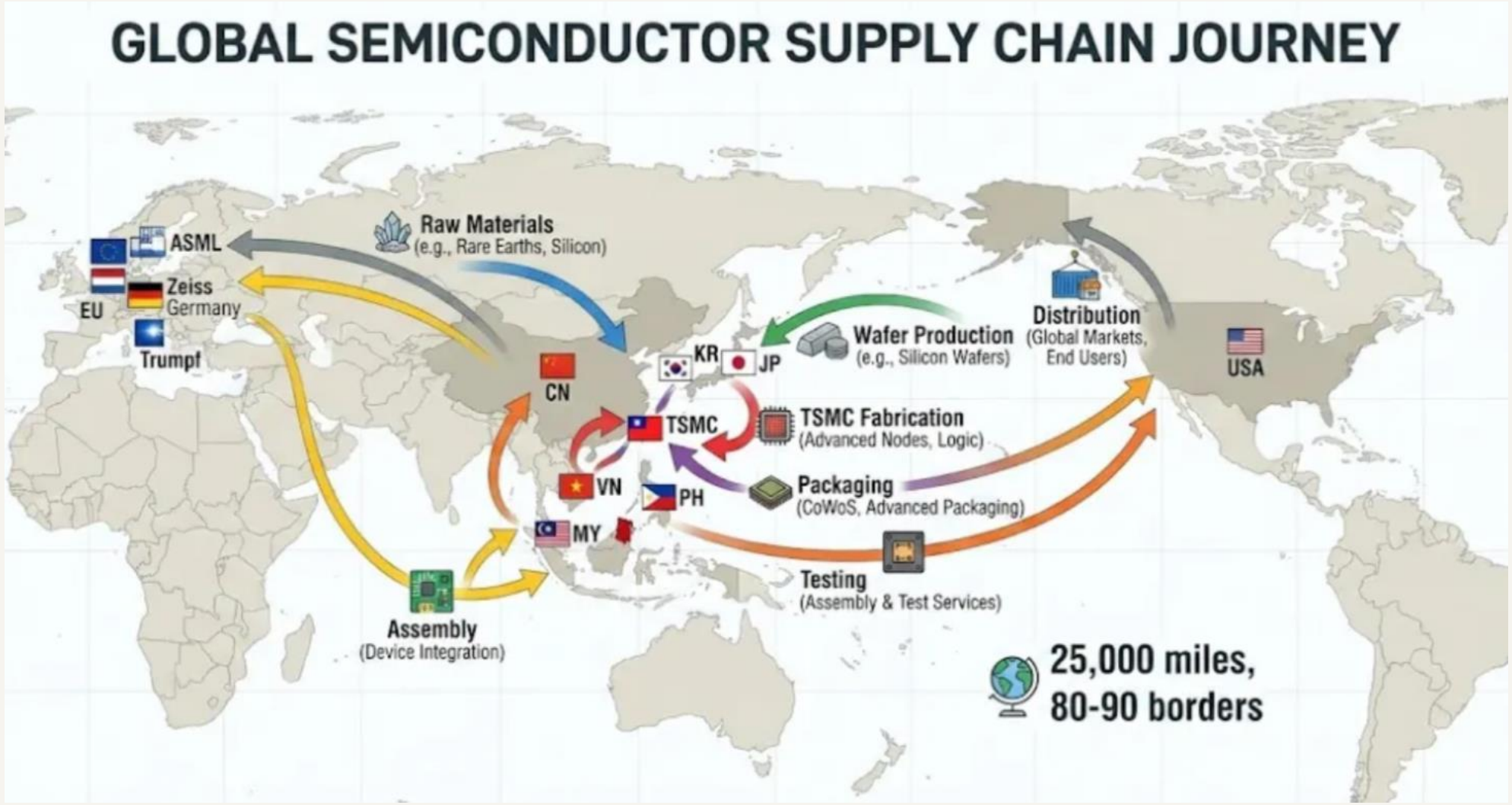
Industrial gases, chemicals, subsystems, silicon wafers

Design software (EDA)

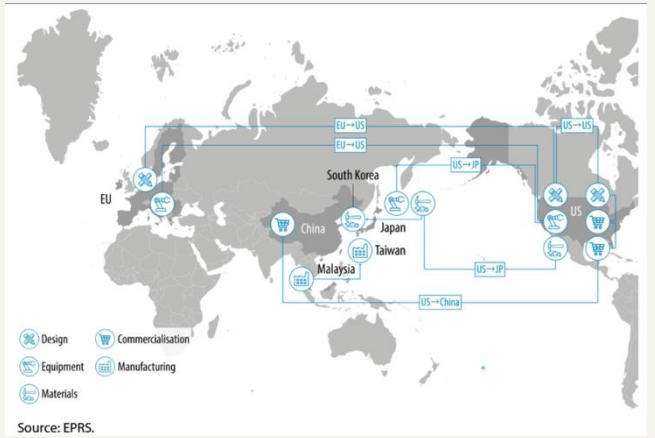
Electronic design automation tools

Semiconductor IP

Licensed, pre-designed circuit blocks


GLOBAL SEMICONDUCTOR SUPPLY CHAIN JOURNEY
ASML
EU
Zeiss
Germany
Trumpf
Raw Materials
(e.g., Rare Earths, Silicon)
CN
KR
JP
Wafer Production
(e.g., Silicon Wafers)
Distribution
(Global Markets, End Users)
USA
TSMC
TSMC Fabrication
(Advanced Nodes, Logic)
VN
PH
Packaging
(CoWoS, Advanced Packaging)
MY
Testing
(Assembly & Test Services)
Assembly
(Device Integration)
25,000 miles,
80-90 borders

Designer, foundry, equipment maker



The company that sells a GPU is often not the company that manufactures the silicon.

- Fabless designers (NVIDIA, AMD, Broadcom): architectures, circuits, packaging choices, software.
- Foundries (TSMC, Samsung, Intel Foundry): manufacture patterned wafers to specified process nodes.
- Toolmakers (ASML, Applied Materials, Lam, KLA, Tokyo Electron): sell the machines that make fabs possible.
- Memory (SK hynix, Samsung, Micron) and advanced packaging are separate critical industries.

DESIGN	NVIDIA / AMD United States
FAB EQUIPMENT	ASML + others Europe / US / Japan
FABRICATE	TSMC Taiwan
MEMORY / PACKAGE	SK hynix + TSMC Korea / Taiwan
AI SERVER	OEM + cloud global

A five-step simplification of a multi-year global ecosystem

Sources: Semiconductor Industry Association / BCG · ASML

US controls depend on allied firms; China on mineral refining

This is a network of complementary strengths, not a clean US–China economic separation.

UNITED STATES

Design software, chips, deposition & inspection

NETHERLANDS

ASML EUV / selected DUV scanners

JAPAN

Photoresists, wafers, masks, process tools

S. KOREA

HBM and other advanced memory

CHINA

Mineral refining, mature logic, local tool buildout

An export restriction only works if substitute suppliers and workarounds are difficult.

Sources: Y. Sharma, Chip Supply Chain (2026) · chipsupplychain.org

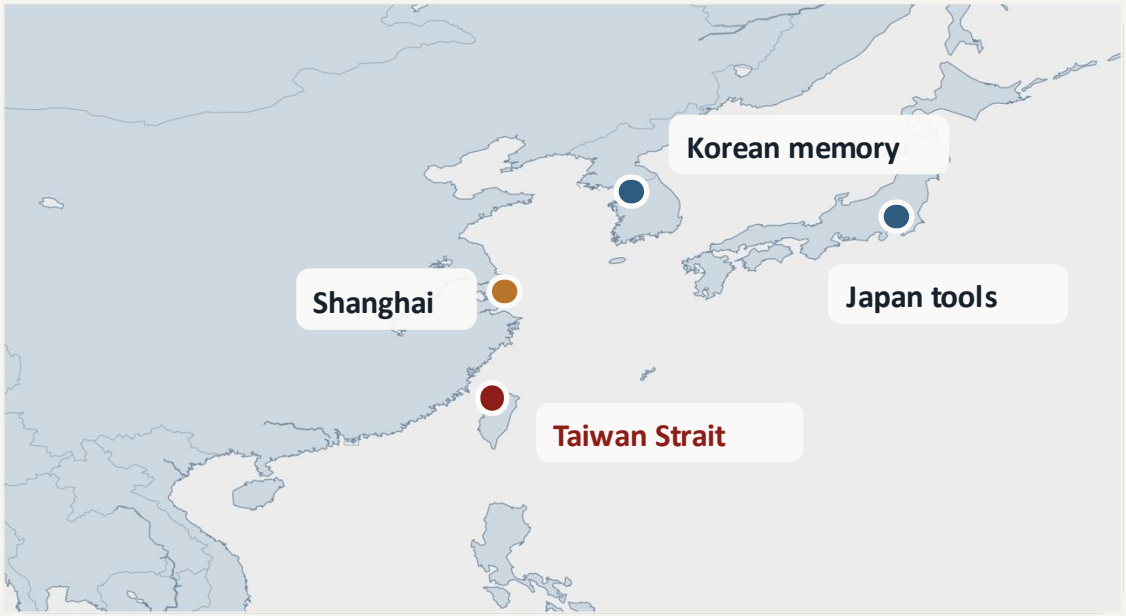
What if advanced chip exports from Taiwan stopped?

Different disruption scenarios have radically different economic and political consequences.

- Earthquake or power outage: may cause temporary shutdowns, lost wafers and recovery delays.
- Blockade: prevents inbound chemicals, spare parts and outbound finished wafers even if fabs are intact.
- Military conflict: can disrupt workers, utilities, logistics and production over an unknown duration.
- No public model can precisely predict the global loss without specifying duration and response.

$10.6T

Bloomberg Economics estimate
First-year global GDP loss in a Taiwan-war scenario



Disruption affects fabrication, packaging and freight routes.

Substitute capacity is neither immediate nor frictionless.

Sources: Semiconductor Industry Association / BCG · BCG / SIA

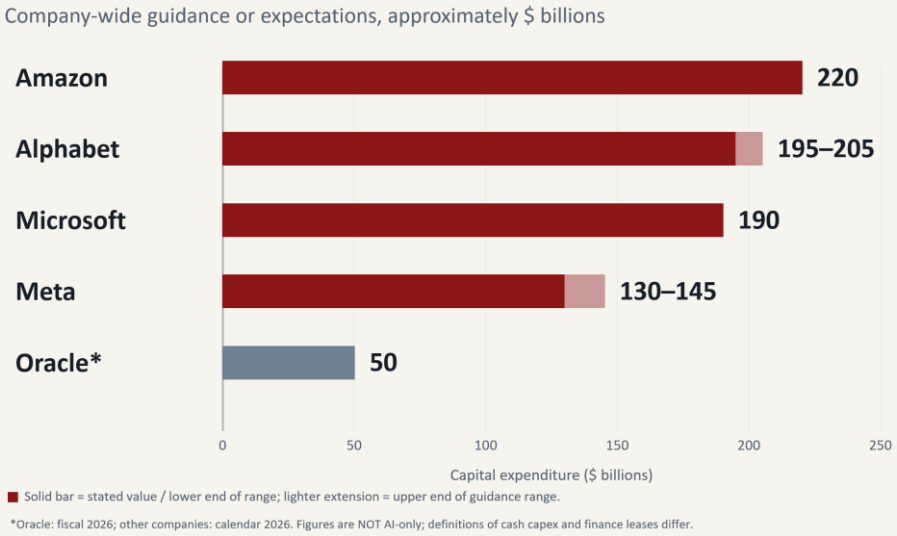
SECTION

Data Centers

The AI buildout has become a major capital-investment cycle

The evidence is large announced and realized capital spending—but not every dollar is spent on AI.

- Amazon guided to about $220 billion in 2026 cash capex; Alphabet guided to roughly $195–205 billion.
- Microsoft described roughly $190 billion calendar-2026 capex, including finance leases (April guidance).
- Meta guided $130–145 billion including finance-lease principal (July); Oracle expected $50 billion for FY2026.
- These are company-wide forecasts, not solely AI spending; their time periods and lease accounting differ.

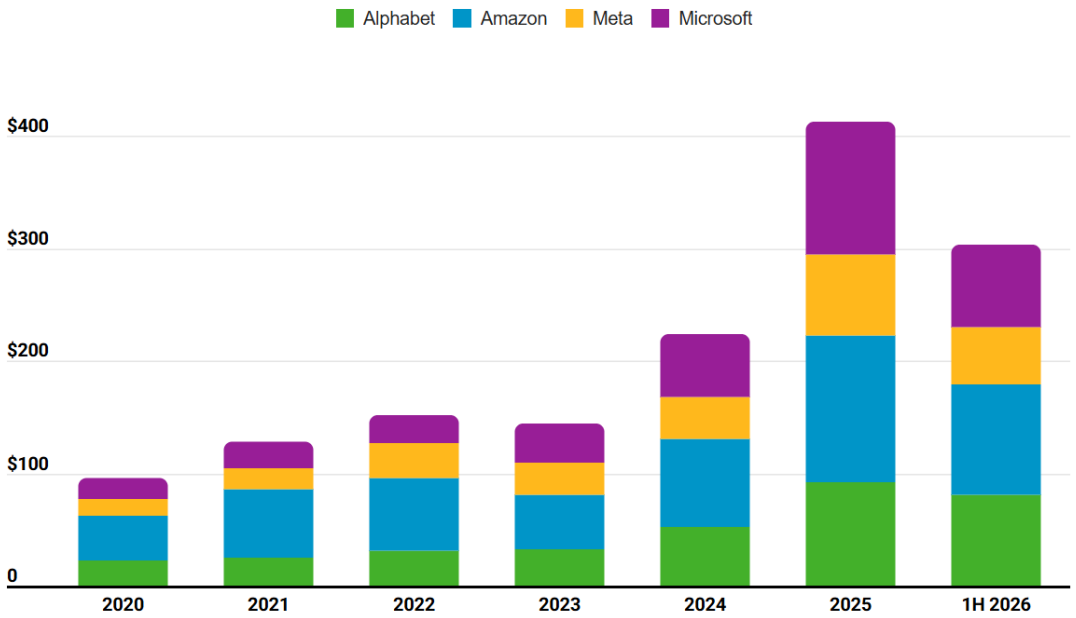


*Oracle: FY2026. Others: calendar 2026. Some ranges shown at midpoint.

Cash capex / lease treatments differ; totals are NOT AI-only.

Sources: Amazon Q2 2026 investor call / FY2026 capital-spending guidance · Alphabet 2026 CapEx range · Meta Investor Relations

AI hyperscaler capital expenditures over time



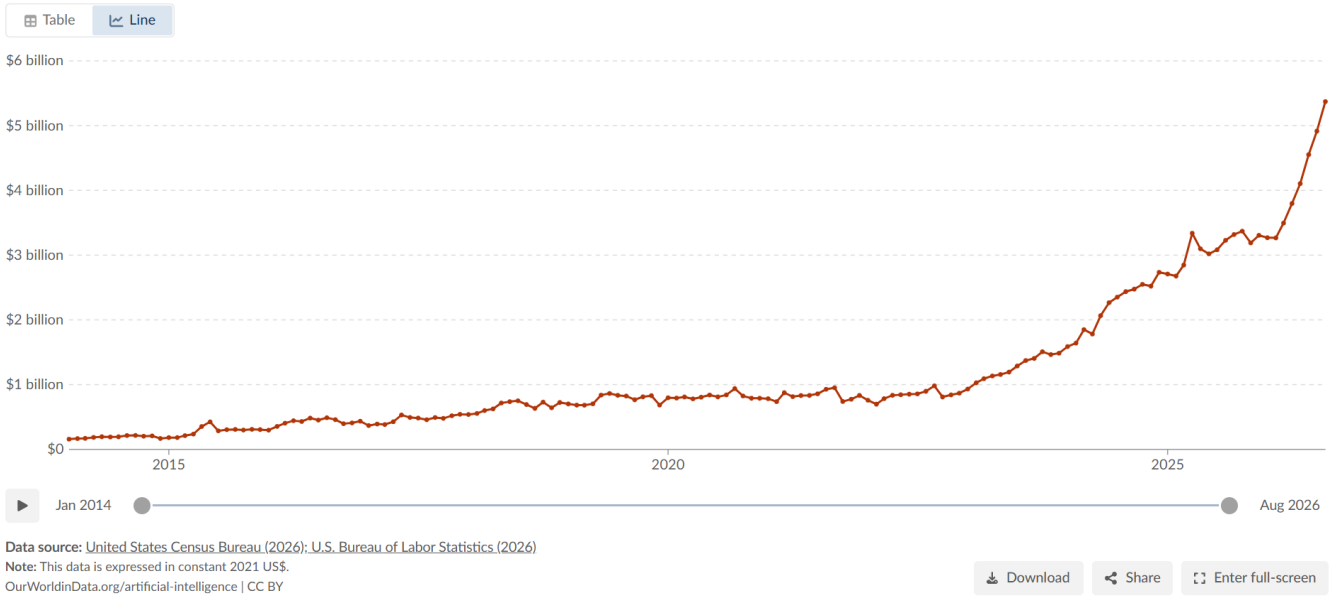
Company	2020	2021	2022	2023	2024	2025	1H 2026
Alphabet	$22.281B	$24.64B	$31.48B	$32.25B	$52.53B	$91.4B	$80.6B
Amazon	$40.14B	$61.05B	$63.64B	$48.7B	$78.6B	$131.2B	$98.4B
Meta	$15.16B	$18.69B	$31.18B	$28.1B	$37.25B	$72.5B	$50.9B
Microsoft	$17.59B	$23.21B	$24.76B	$35.20B	$55.55B	$118B	$72.9B
Combined	$95.17B	$127.59B	$151.08B	$147.22B	$228.34B	$410.2B	$302.8B

Data source: Microsoft, Meta, Amazon, and Alphabet SEC filings and earnings releases (2026). Data by calendar year.

Monthly spending on data center construction in the United States

Our World in Data

Cost of privately-funded construction of data centers each month, including materials and labor. It excludes the cost of IT hardware like servers and storage, which can account for a large share of total investment. This data is expressed in US dollars, adjusted for inflation.



Data source: United States Census Bureau (2026); U.S. Bureau of Labor Statistics (2026)
Note: This data is expressed in constant 2021 US$.
OurWorldinData.org/artificial-intelligence | CC BY

Download Share Enter full-screen

- Chips account for 60% of total AI capex in a datacenter.
- 2026 likely total capex spend (including AI) is $745 billion.

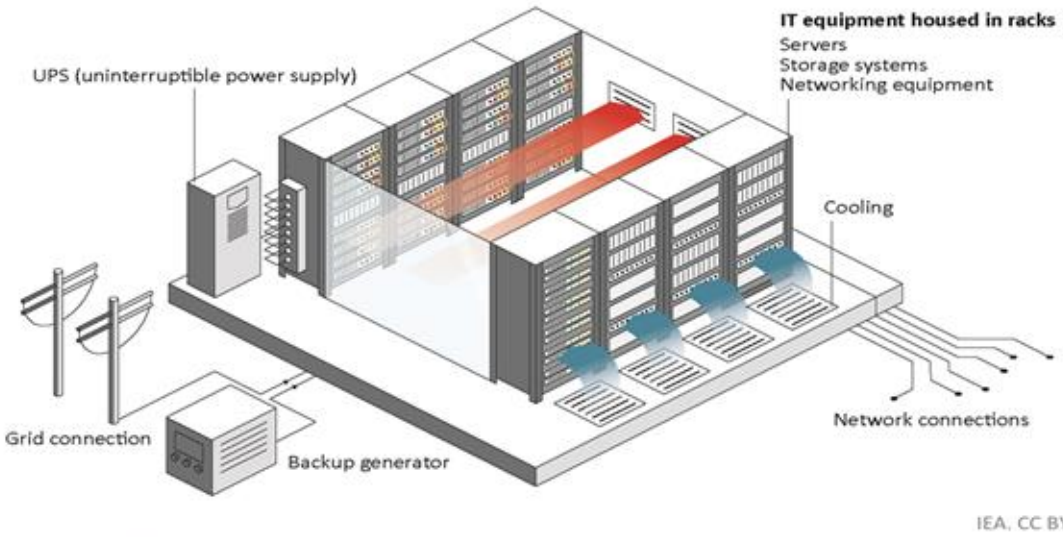
Data centers = factory for computation

It joins silicon, electricity, cooling, networks, buildings and global cloud services.

- “Hyperscaler”: a company operating very large, distributed cloud infrastructure.
- “AI cluster”: many accelerators coordinating through fast networks for training and inference.
- The facility converts purchased electricity into computation and almost entirely into heat.
- The physical location affects latency, grid emissions, water demands, legal jurisdiction and cost.

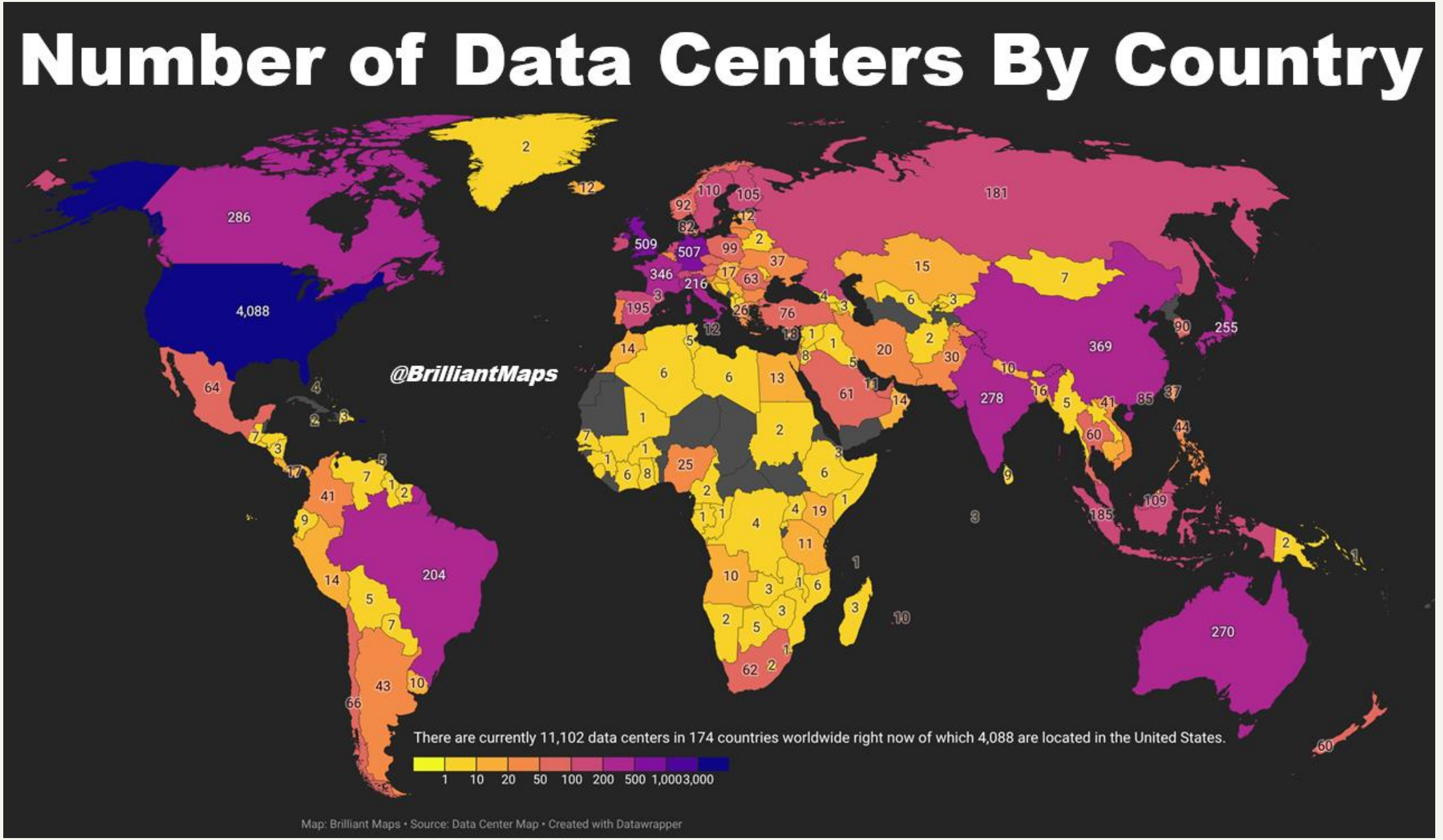


Figure 2.1 ▷ **Data centre components**



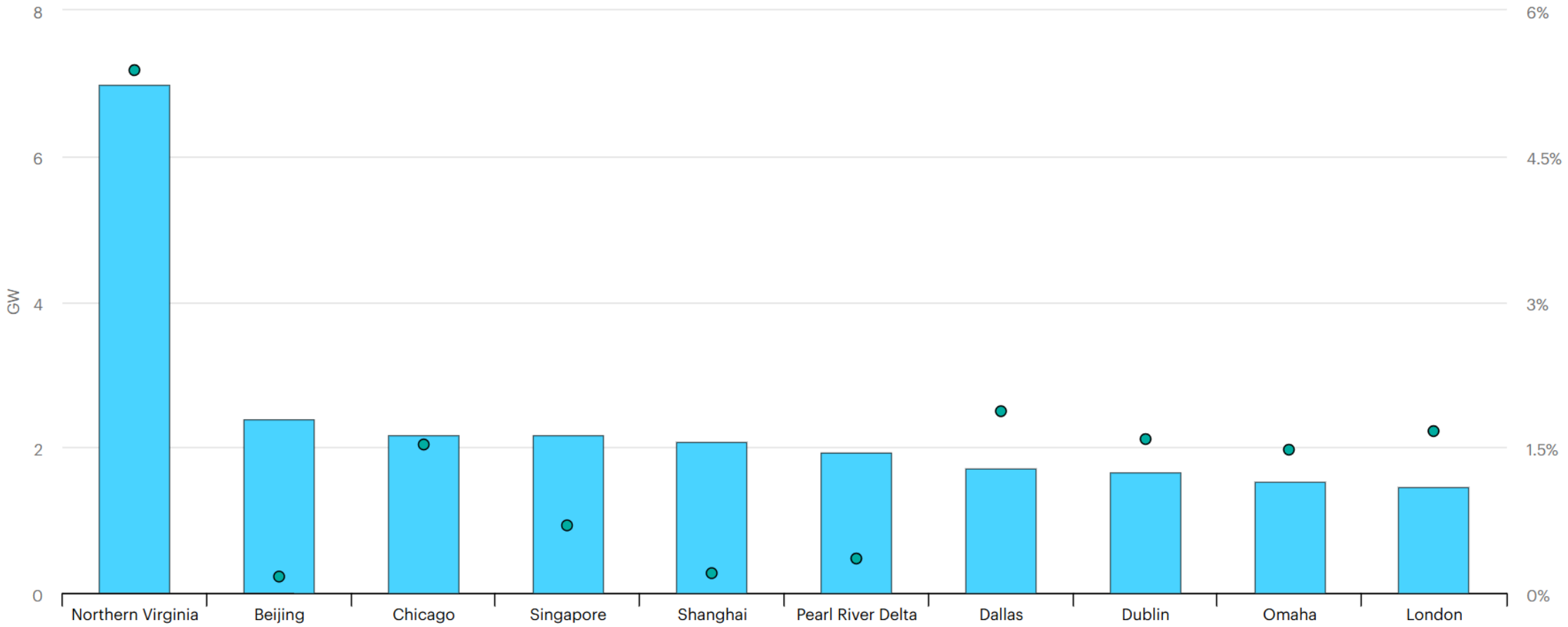
Effective design and integration of key data centre components ensure continuous operation and optimal performance

Sources: International Energy Agency · NVIDIA


Number of Data Centers By Country
@BrilliantMaps
There are currently 11,102 data centers in 174 countries worldwide right now of which 4,088 are located in the United States.
1 10 20 50 100 200 500 1,000 3,000
Map: Brilliant Maps • Source: Data Center Map • Created with Datawrapper

Top ten data centre markets by installed capacity versus share of capacity under development, 2024

Open



IEA. Licence: CC BY 4.0

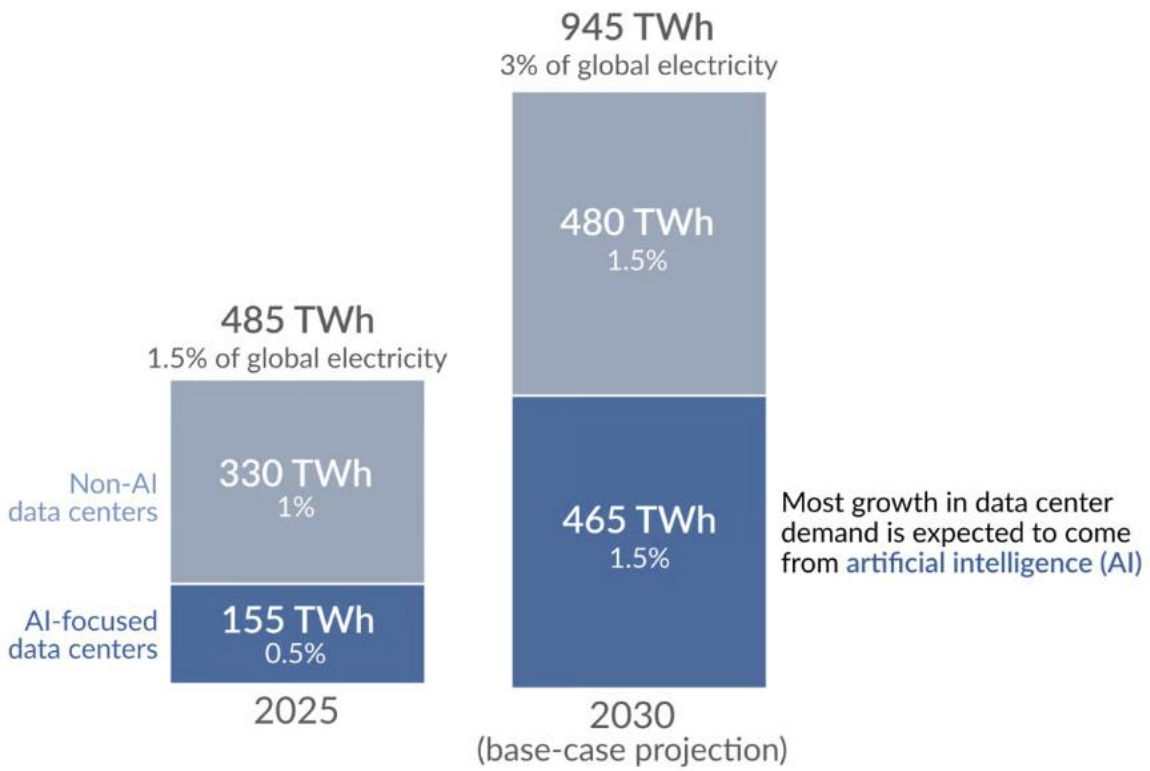
SECTION

Energy

How much of global electricity is used for data centers?

Our World in Data

Measured in terawatt-hours (TWh). Estimates for 2025; base-case projections for 2030 from the International Energy Agency (IEA).



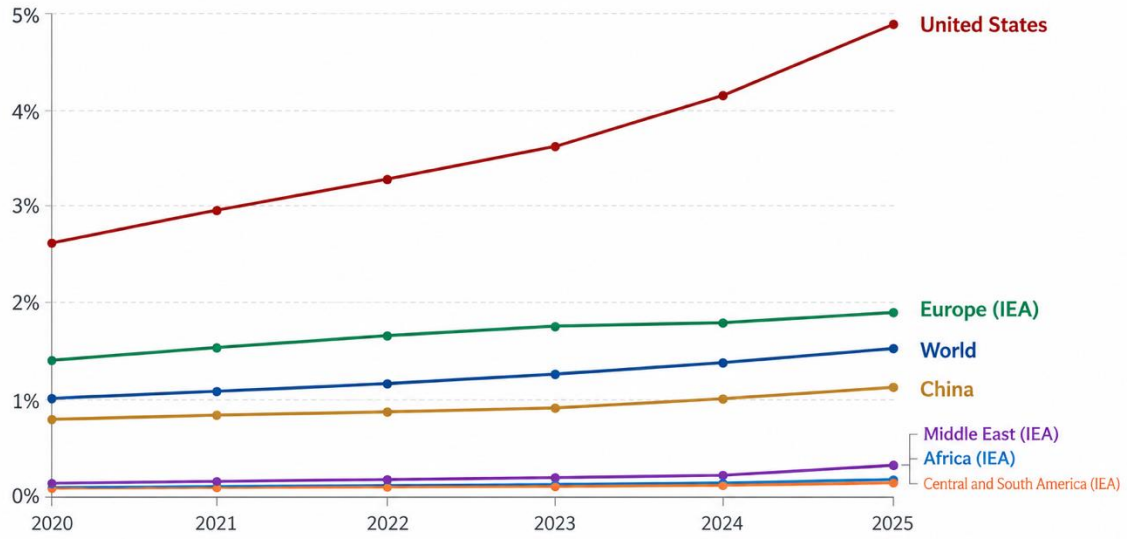
Note: Projections for 2030 are very uncertain, but shown here to reflect expectations that most data center growth will come from AI-focused ones.

Data source: IEA (2026). Key Questions on Energy and AI.

CC BY

Share of total electricity demand used by data centers, 2020 to 2025

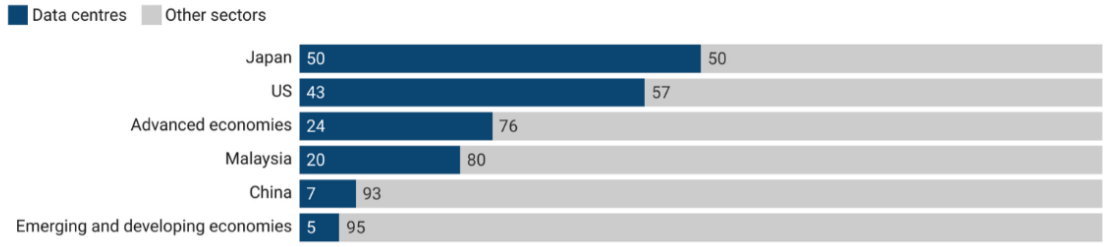
Electricity consumed by data centers given as a share of electricity generation. Data centers power a wide range of digital services, including streaming, messaging and cloud storage. Artificial intelligence (AI) focused data centers are a subset of this electricity consumption.



Data source: International Energy Agency (2025, 2026); Ember (2026).

Data centres are expected to be major sources of electricity demand growth in some countries

Data-centre share of electricity demand growth 2024-2030, %



Source: IEA

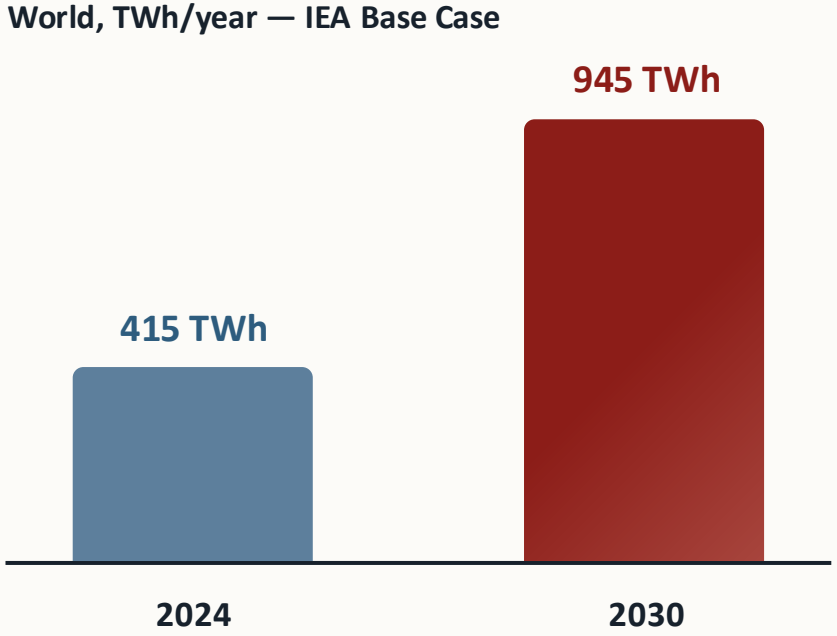
CarbonBrief CLEAR ON CLIMATE

Total Global Electricity Consumption (2025) = approximately 30,000 Terrawatt Hours

Worldwide data centers used ~415 TWh in 2024

The IEA estimates ~1.5% of global electricity consumption, with large uncertainty around AI's separate share.

- IEA "Energy and AI" (2025): 415 TWh data-center electricity use in 2024.
- IEA Base Case: ~945 TWh in 2030, more than double—but a forecast rather than an observation.
- COMPARISON – San Francisco uses 18 TWh per year
- National totals mask concentrated effects in specific grid regions.



Top 5 Largest Power Plants by Capacity

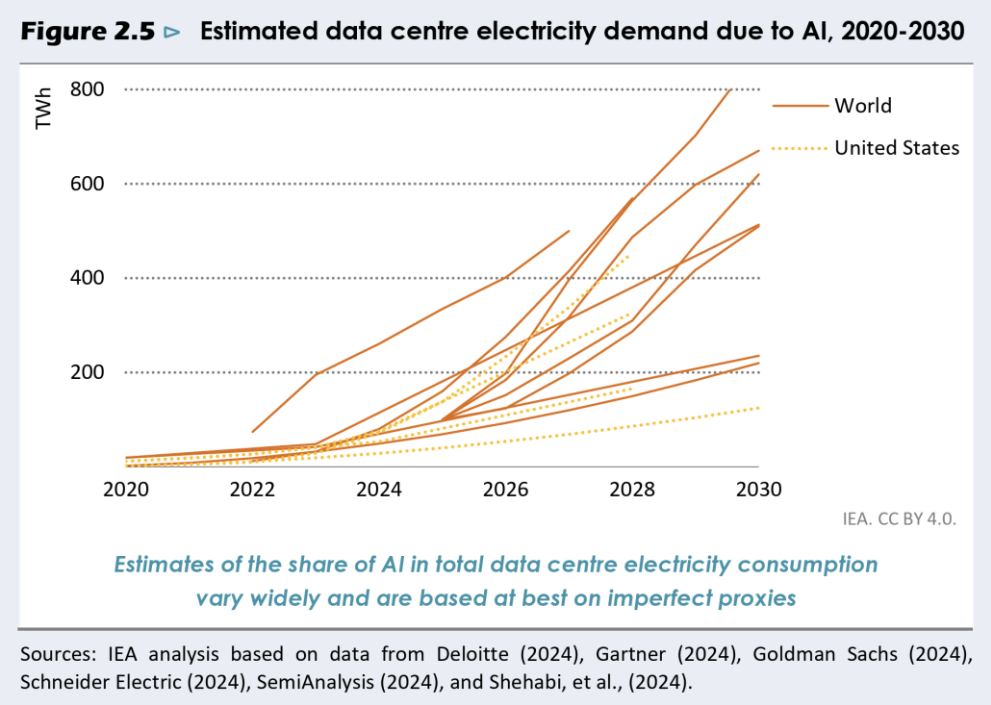
1. **Three Gorges Dam** (China) — 22,500 MW (Hydroelectric)
2. **Baihetan Dam** (China) — 16,000 MW (Hydroelectric)
3. **Itaipu Dam** (Brazil / Paraguay) — 14,000 MW (Hydroelectric)
4. **Xiluodu Dam** (China) — 13,860 MW (Hydroelectric)
5. **Belo Monte Dam** (Brazil) — 11,233 MW (Hydroelectric) Wikipedia +1

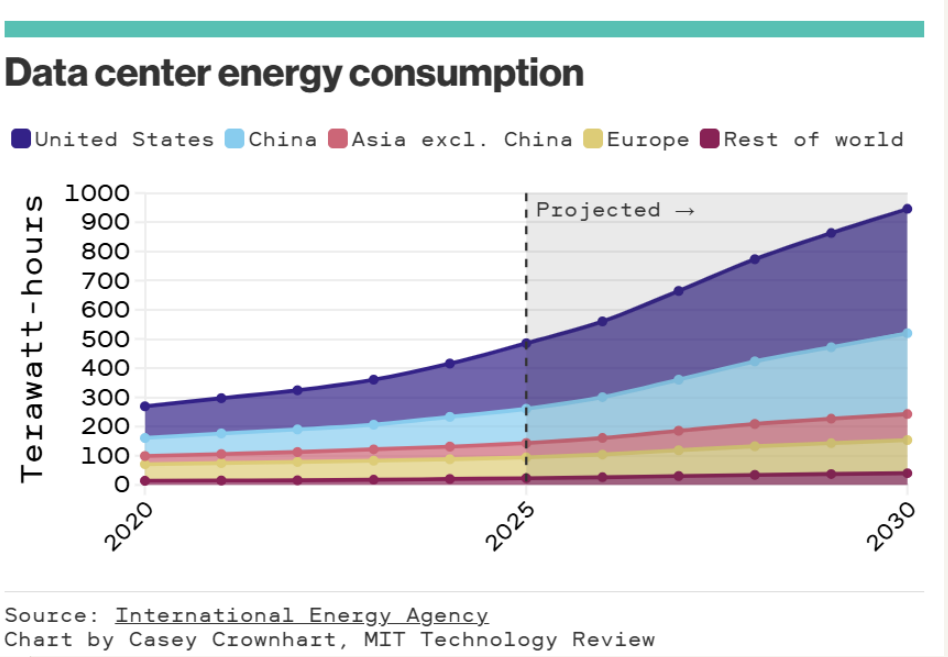
Sources: International Energy Agency

Total Global Electricity Consumption (2025) = approximately 30,000 Terrawatt Hours

Forecasts for AI electricity demand differ by several multiples

The IEA collated published estimates that use very different definitions of AI and data-center computing.





The lesson of the spread

Dotted lines: U.S. estimates.

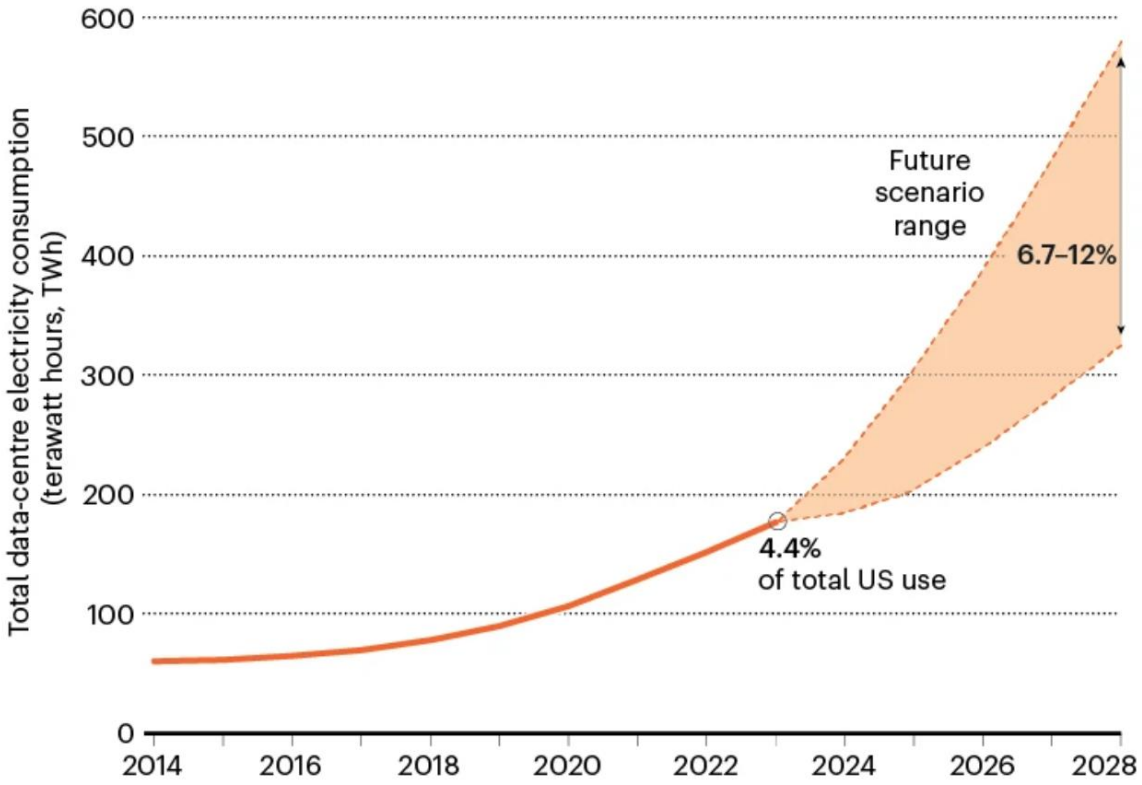
Solid lines: world estimates.

Forecasts differ on demand, chip efficiency, model mix and what counts as “AI.”

IEA, Energy and AI (2025), Figure 2.5; original figure CC BY 4.0.

DATA CENTRES' HIGHER LOCAL IMPACT

In the United States, national electricity demand from data centres could grow from 4.4% to up to 12% by 2028. This is higher than the global projection because the United States hosts so many data centres.

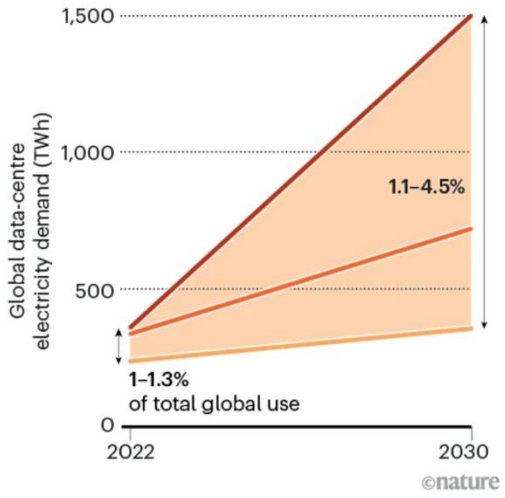


DATA CENTRES' MODEST GLOBAL IMPACT

Data centres currently use 1–1.3% of the world's electricity. But as more are built to supply artificial-intelligence models, they are expected to use up to 4.5% by 2030. The wide range shows the uncertainty in projections.

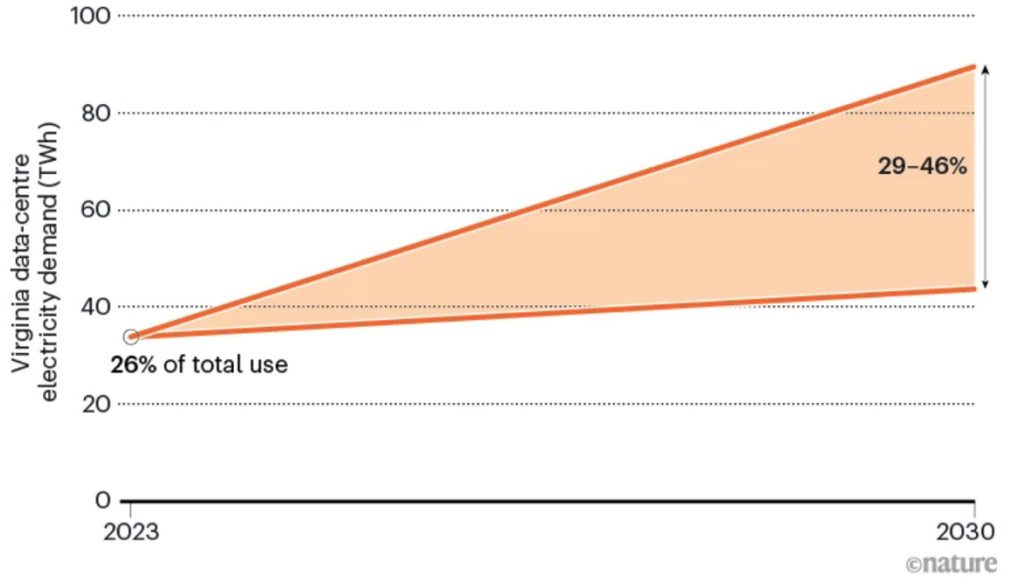
Data source
- SemiAnalysis 2024 (base case)
- IEA 2024 (high projection)
- IEA 2024 (low projection)

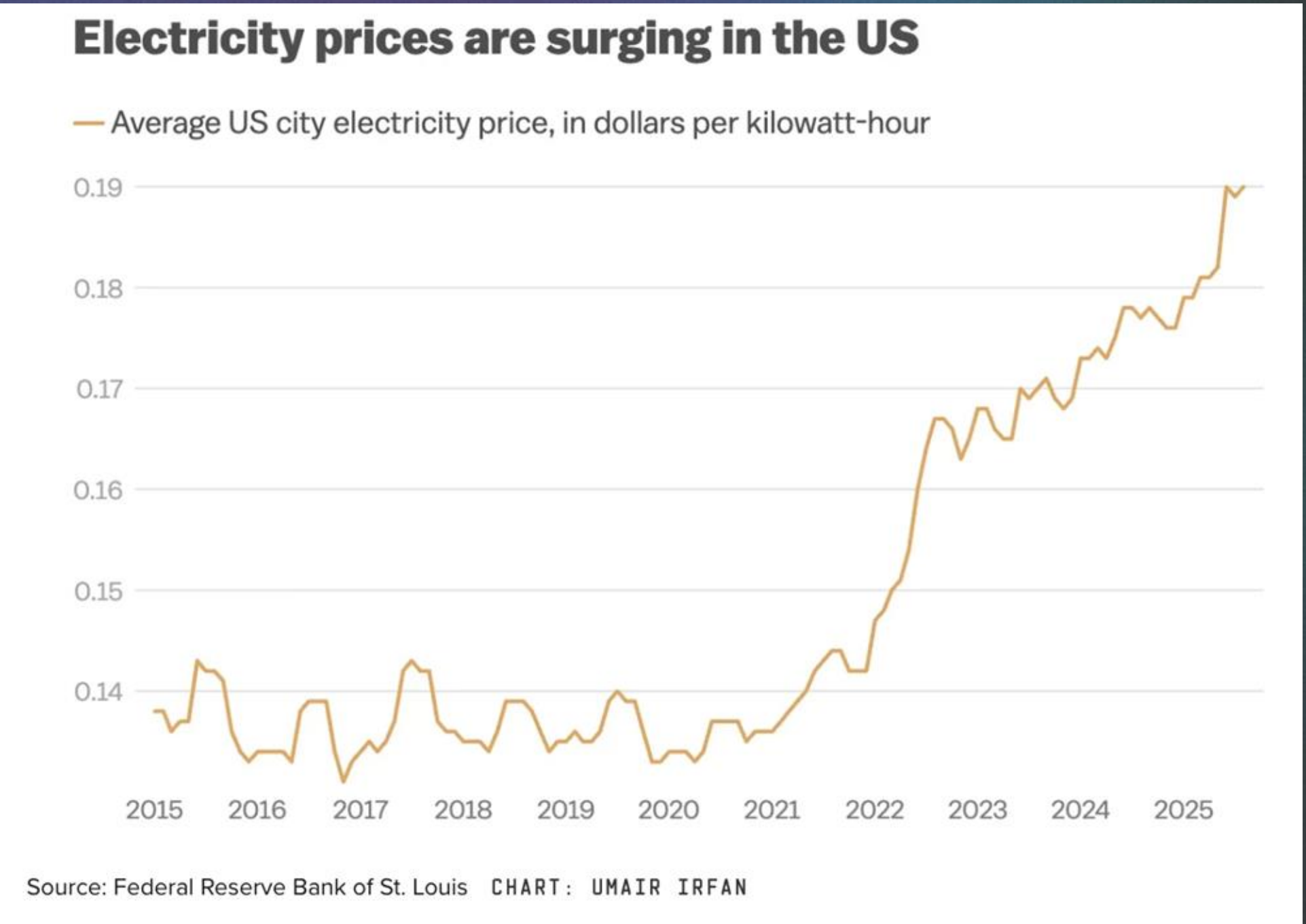
*IEA: International Energy Agency, *World Energy Outlook 2024*. The agency says it will release new figures in April 2025.



Sources: Ref. 4, SemiAnalysis

In Virginia, data centres already make up more than one-quarter of electricity use; that proportion will rise by 2030.



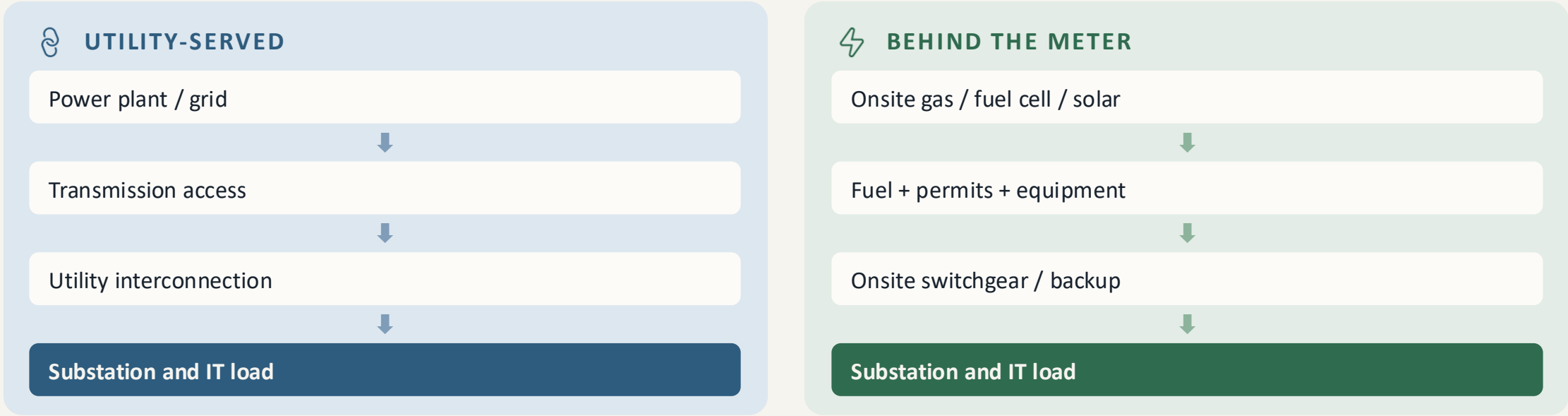


US residential electricity rates have risen more than 30 percent on average since 2020 and almost double the rate of inflation in the past year.

Political resistance is already happening.

AI data centers are increasingly seeking power outside the grid queue

On-site generation can shorten one bottleneck while creating new fuel, permitting, and emissions questions.



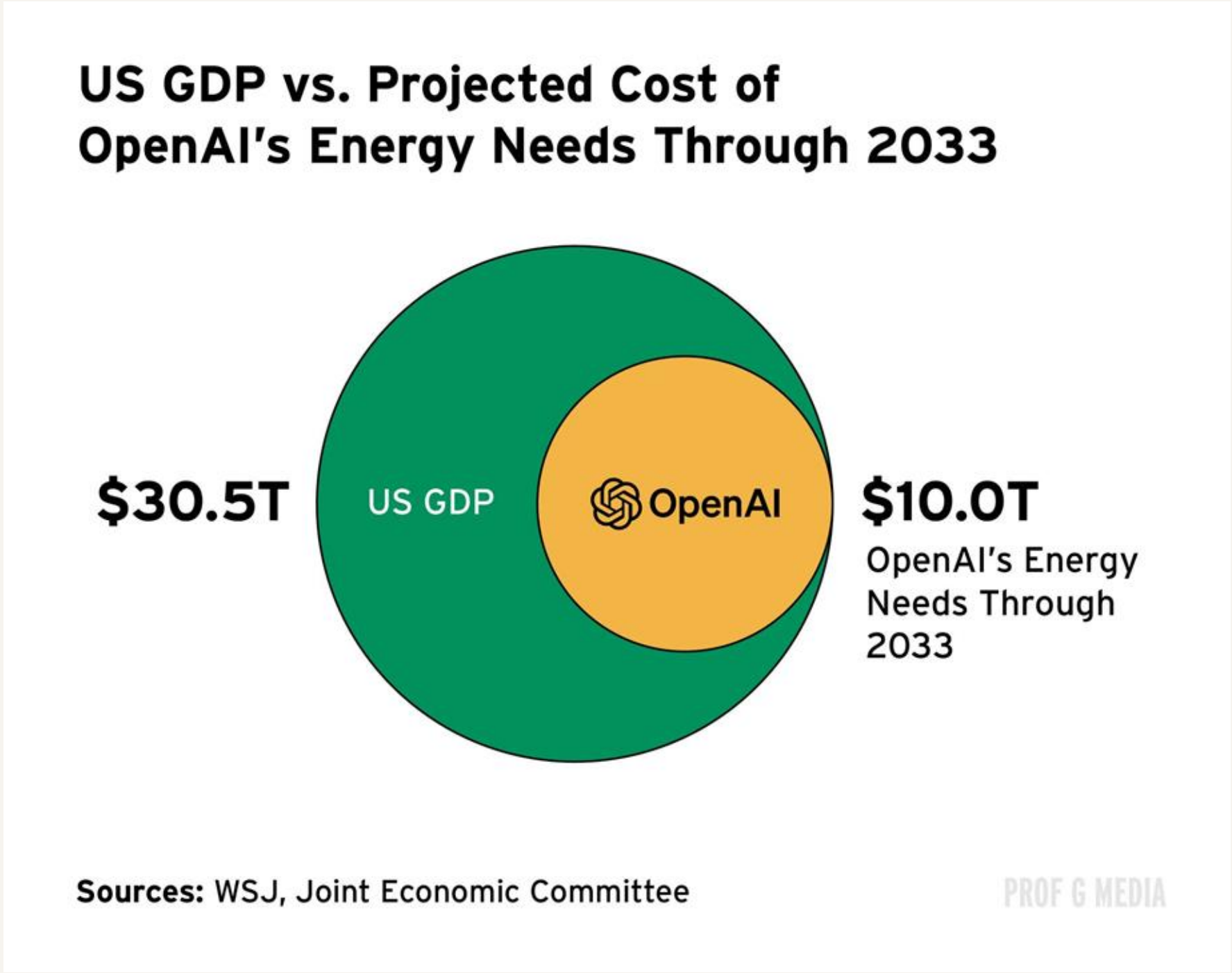
SemiAnalysis: ~75 GW of firm BTM equipment orders reported (Sept. 2026)

Sources: SemiAnalysis (June–September 2026) · Orders are not commissioned generation capacity

“Companies that have raised lots of money are at particular risk. It’s so tempting to paper over a problem with the business by spending more money instead of fixing the product or service.”

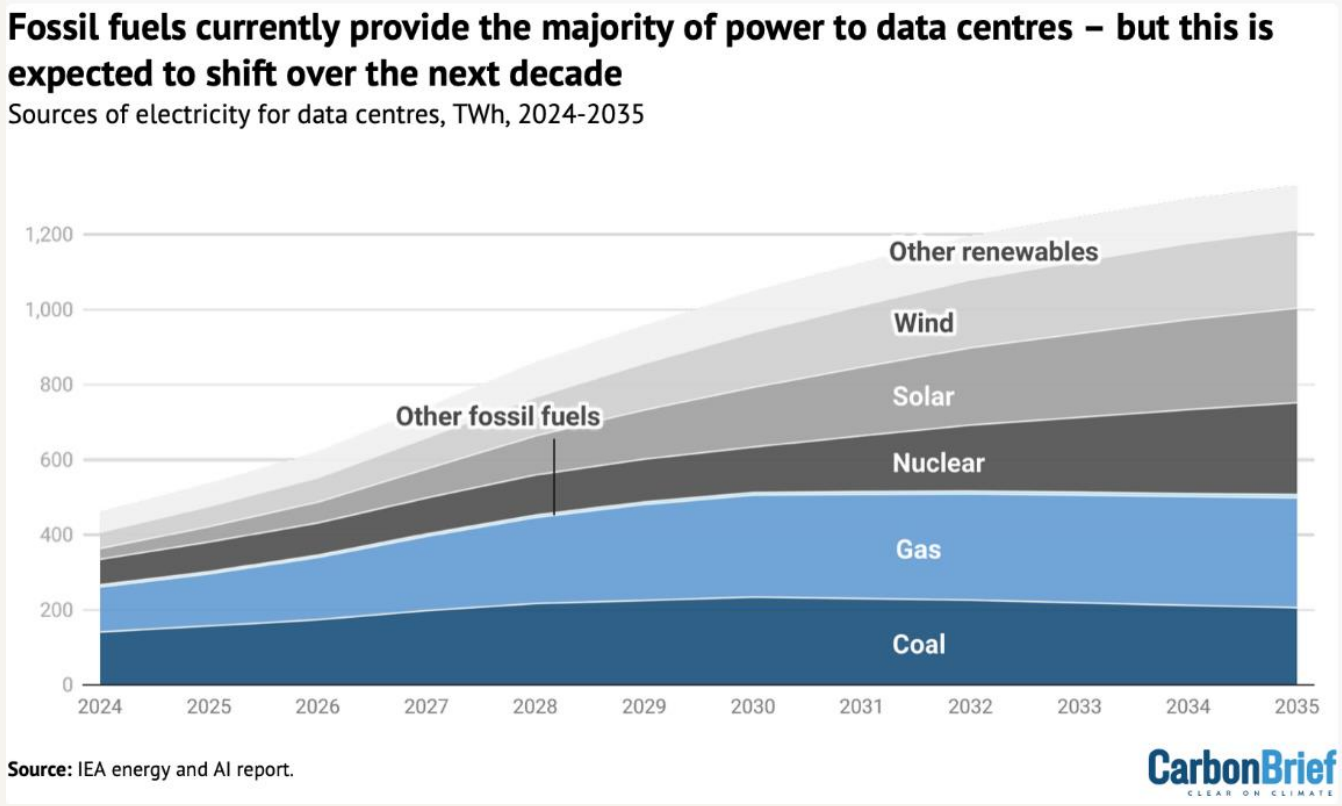
Sam Altman, 2015

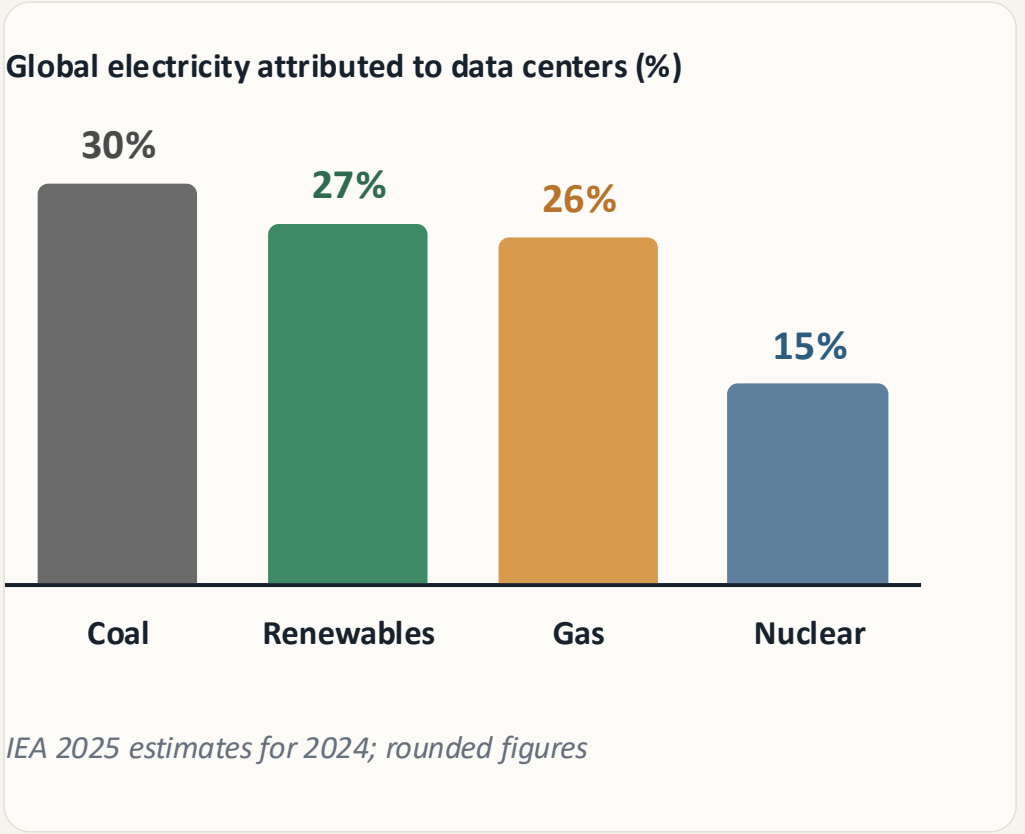
2025: Sam Altman told OpenAI employees that he wants to build out 250GW of compute power over the next eight years. For context, 250GW is about a fifth of the energy generation capacity of the entire U.S. Building that much capacity would likely cost around $10 trillion, or one-third of the entire U.S. economy last year.



What actually powers the world's data centers?

Global averages hide **dramatic regional differences** and hourly grid conditions.

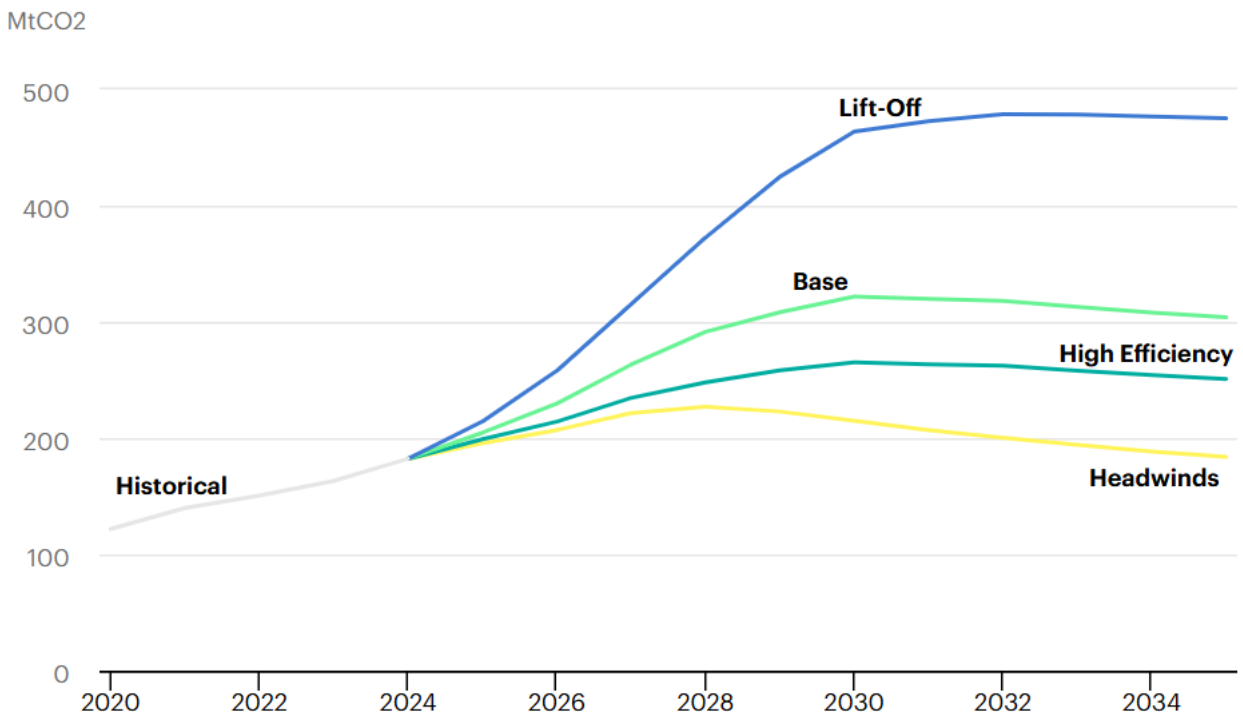




Sources: International Energy Agency

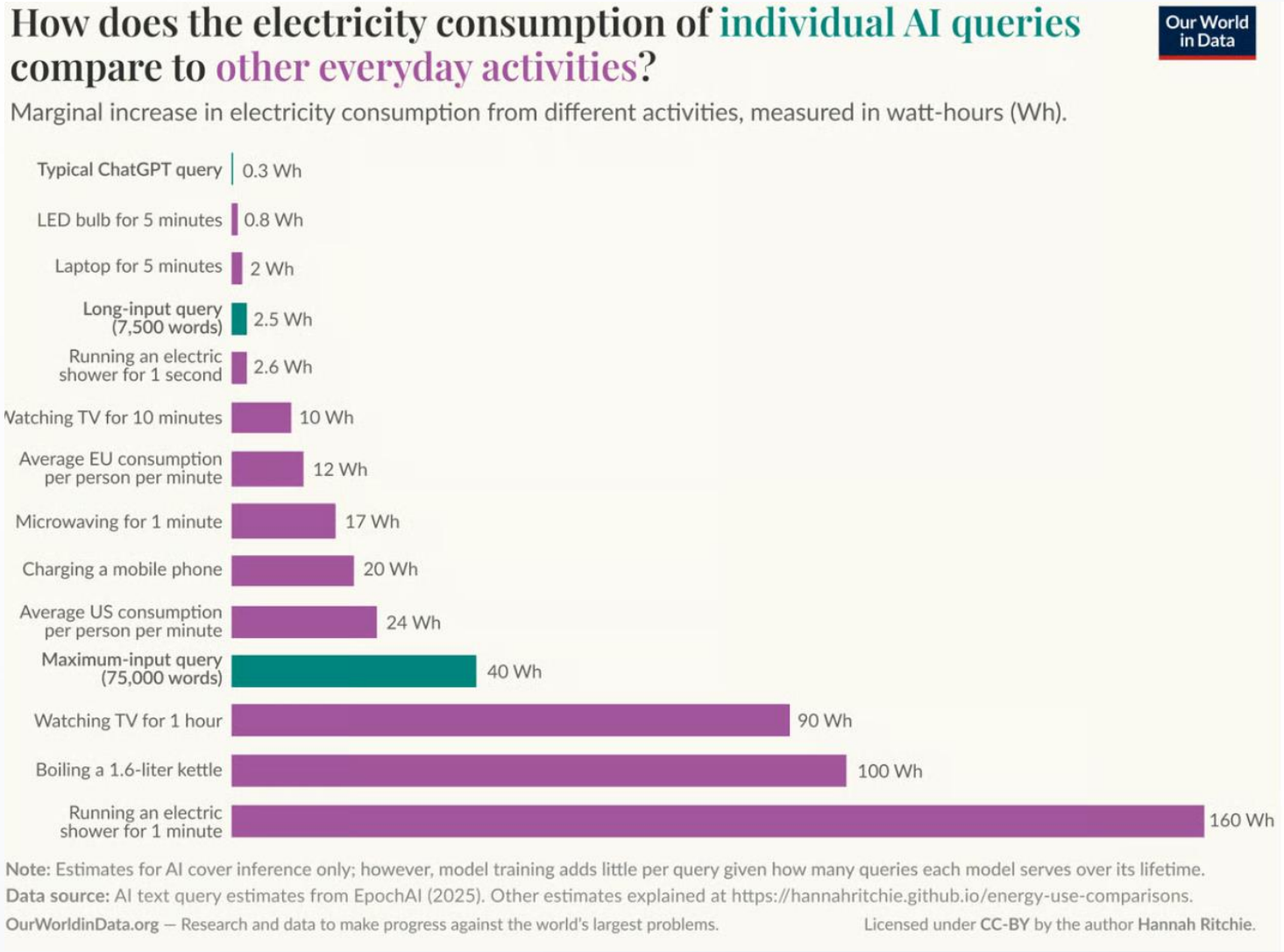
CO2 emissions associated with electricity generation for data centres by case, 2020-2035

Open



IEA. Licence: CC BY 4.0

How guilty should you feel for using AI?



- **YouTube or Netflix, 1 hour (HD)**
  ~0.12 kWh → 42 g CO_2
  Tied for the dirtiest single activity in the study.
- **Text-to-video generation, 6–10 seconds**
  ~0.05 kWh → 17.5 g CO_2
  Roughly the same as an hour-long Zoom call.
- **Zoom, 1 hour**
  ~0.0486 kWh → 17 g CO_2
- **Short email, no attachment**
  ~0.0133 kWh → 4.7 g CO_2
  One email is tiny, billions per day are not.
- **AI image generation, 1 image**
  ~0.003 kWh → 1 g CO_2
- **Voice assistant query (Alexa/Siri/etc.)**
  ~0.0005 kWh → 0.175 g CO_2
- **Google search or AI chatbot prompt**
  ~0.0003 kWh → 0.105 g CO_2
- **Two Gemini prompts**
  ~0.00024 kWh → 0.084 g CO_2 total (~0.042 per prompt)

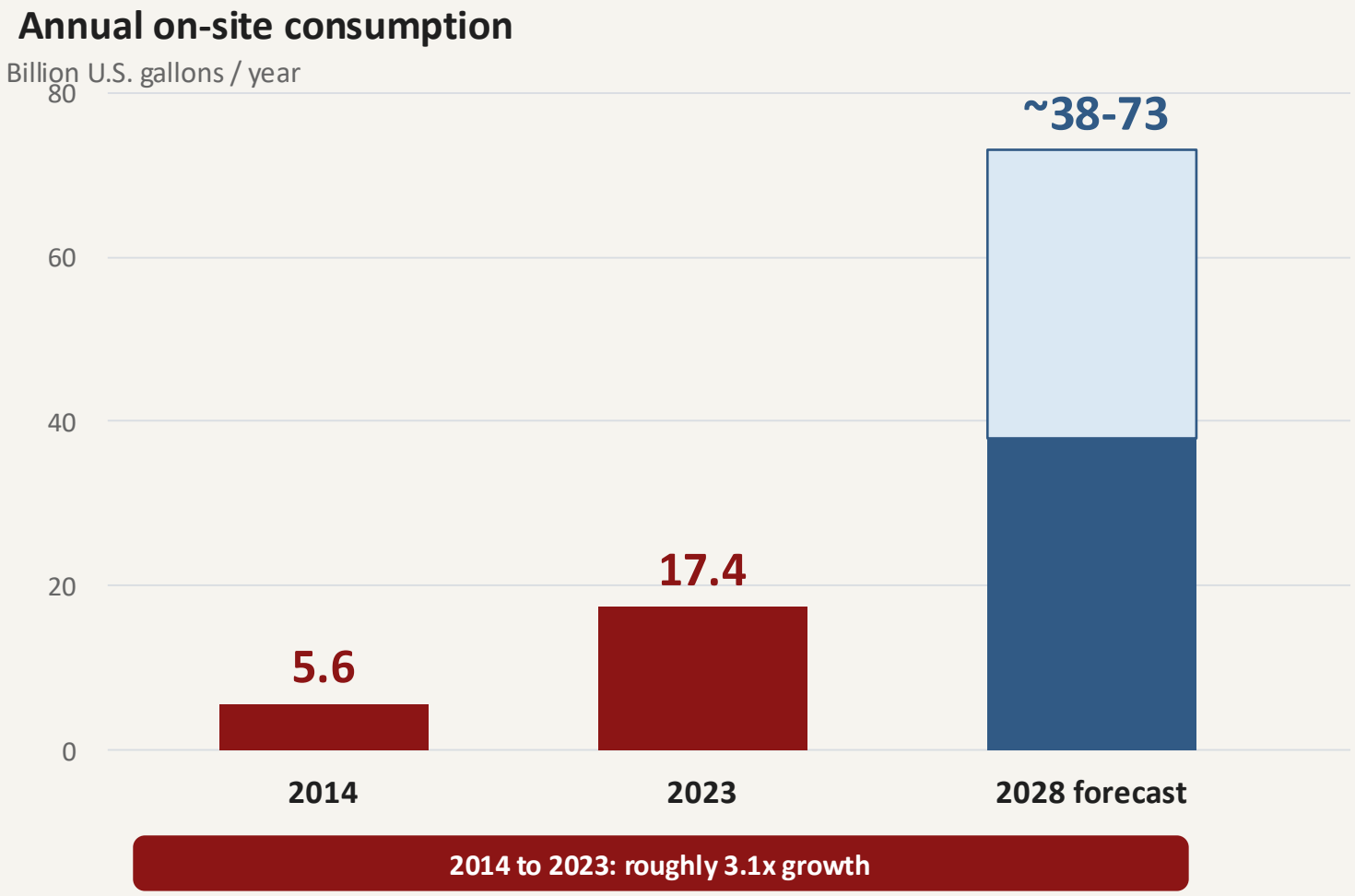
Forbes

SECTION

Water

How much water do U.S. data centers use?

Direct water consumption at facilities, all data-center workloads (not AI alone).



What the chart establishes

- 2014: 21.2 billion liters (~5.6 billion gallons).
- 2023: 66 billion liters (~17.4 billion gallons).
- 2028: roughly 38-73 billion gallons under modeled scenarios.

Forecast band is read approximately from Figure 5.9; it is not observed use.

Source: Lawrence Berkeley National Laboratory, 2024 Data Center Energy Usage Report, Figure 5.9.

Water use has four different meanings

A useful debate must specify the accounting boundary and the physical water metric.

DIRECT / ON-SITE

A data center uses cooling equipment. Evaporation from towers removes water from the local system.

INDIRECT / OFF-SITE

Electricity generation can consume water at power plants or reservoirs, often in a different watershed.

WITHDRAWAL — Water taken from a source; some may be returned.

CONSUMPTION — Water not immediately returned, largely via evaporation.

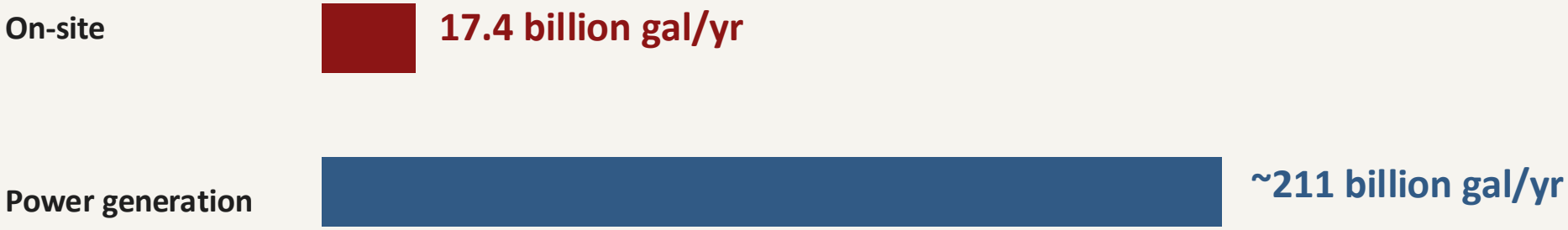
Do not confuse direct with total, or withdrawal with consumption.

Sources: LBNL (2024); Masley (2025/26); Li, Yang, Islam & Ren (2023/25).

Most of the attributed water footprint is off-site

The large number depends on how electricity-related water consumption is assigned.

LBNL model, U.S. data centers, 2023



~628M

Gallons/day total attributed in LBNL model (direct + indirect)

~200-250M

Gallons/day in Potter's rough alternative sensitivity

Why the gap?

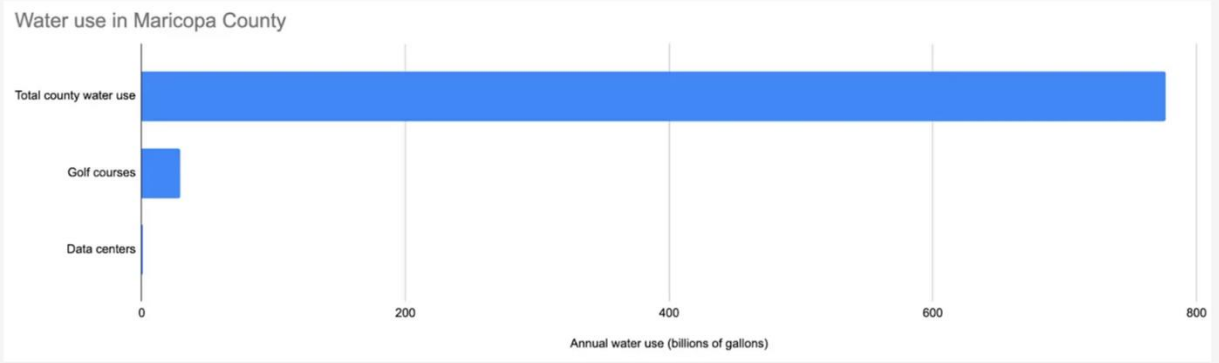
Reservoir evaporation, grid attribution and renewable-energy accounting differ.

Neither total is directly measured at every facility. Potter calls his alternative an estimate, not a full recalculation.

Sources: LBNL 2024, pp. 55-58; Brian Potter (Aug. 2025), methodological critique.

Water?


The New York Times
Their Water Taps Ran Dry When Meta Built Next Door
In the race to develop artificial intelligence, tech giants are building data centers that guzzle up water. That has led to problems for people who live nearby.


Water use in Maricopa County
Total county water use
Golf courses
Data centers
0
200
400
600
800
Annual water use (billions of gallons)

Masley: the national water footprint is small

Masley argues that headline figures lose perspective without national denominators.

~0.04%

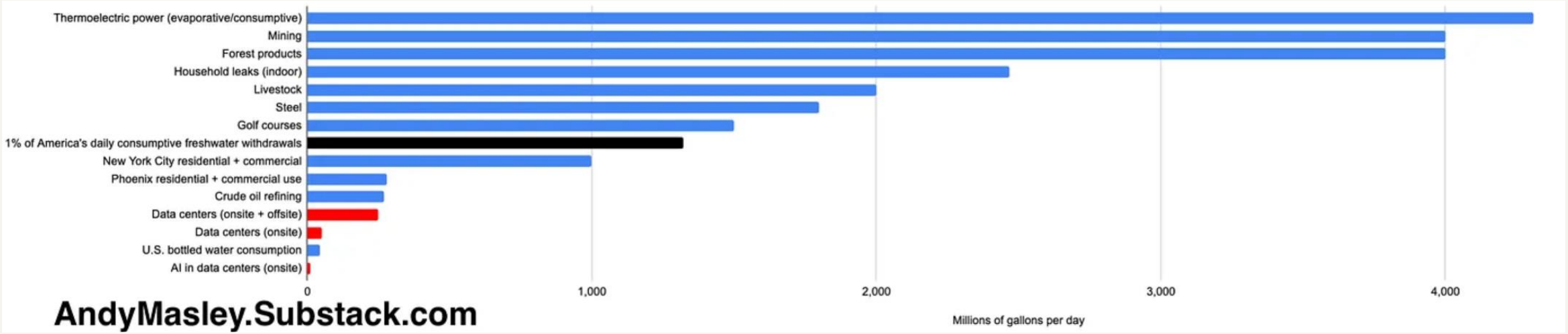
His estimate of the U.S. freshwater-consumption footprint represented by data centers' direct use (2023).

~0.008%

His estimated AI-only on-site fraction, assuming AI was 20% of data-center activity.

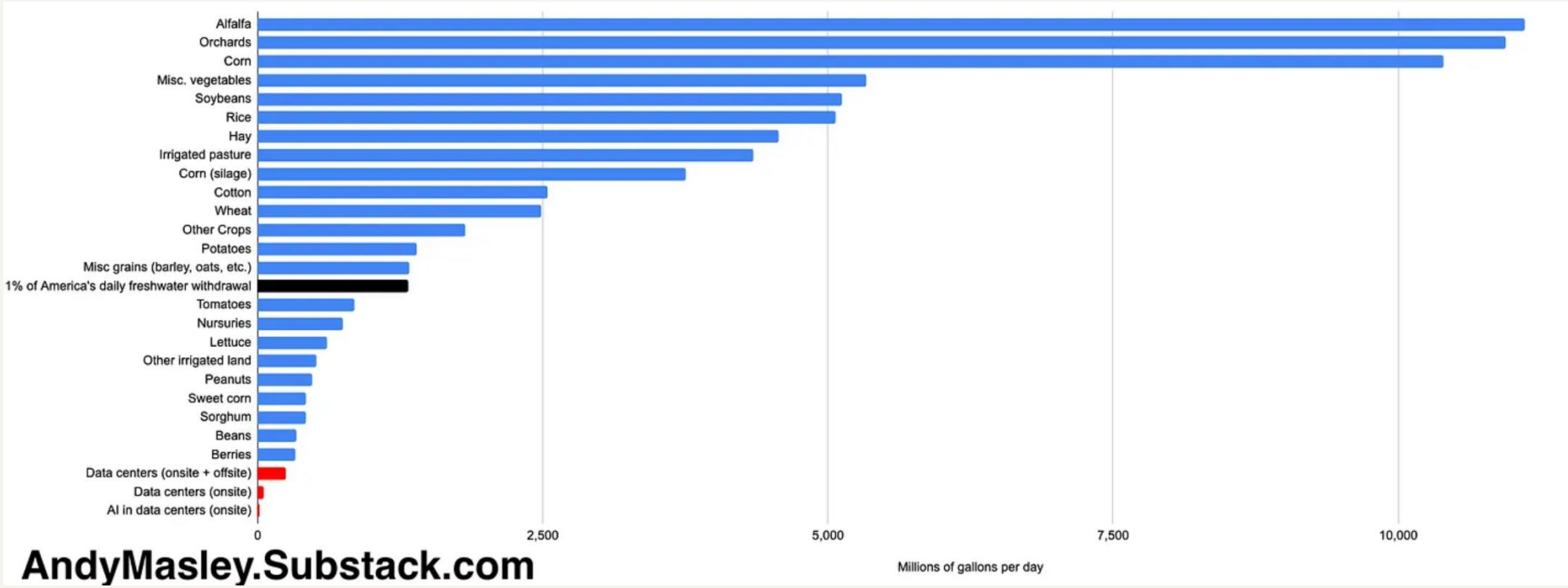
~30x

National golf-course irrigation water applied versus all data centers' direct consumption.



Sources: Andy Masley (updated Apr. 2026); LBNL (2024); GCSAA (2025).

What about water?



Data centers use far less water than irrigation

Annual U.S. volumes shown together; note that the sources measure different concepts and years.

U.S. DATA CENTERS

17.4 billion

Direct on-site consumption (2023)

U.S. GOLF COURSES

531 billion

Irrigation water applied (2024)

U.S. CROP IRRIGATION

27.6 trillion

Consumptive water use (2010-20 avg.)

Measured volumes: golf applied ~30x as much as data centers consumed directly; crop irrigation consumed ~1,600x as much.

CRITICAL QUALIFICATION: golf-course figures are applied irrigation water, not measured net consumption. Year and boundaries differ. A national ratio is not a local water-scarcity measure.

Sources: LBNL (2023); GCSAA/USGA survey (2024); USGS crop irrigation, 2010-20 average.

How much water does one AI prompt use?

Published estimates vary by model, year, task and the water boundary counted.

GOOGLE GEMINI — **0.26 mL**

OPENAI CHATGPT — **~0.32 mL**

EARLIER GPT-3 STUDY — **10-50 mL**

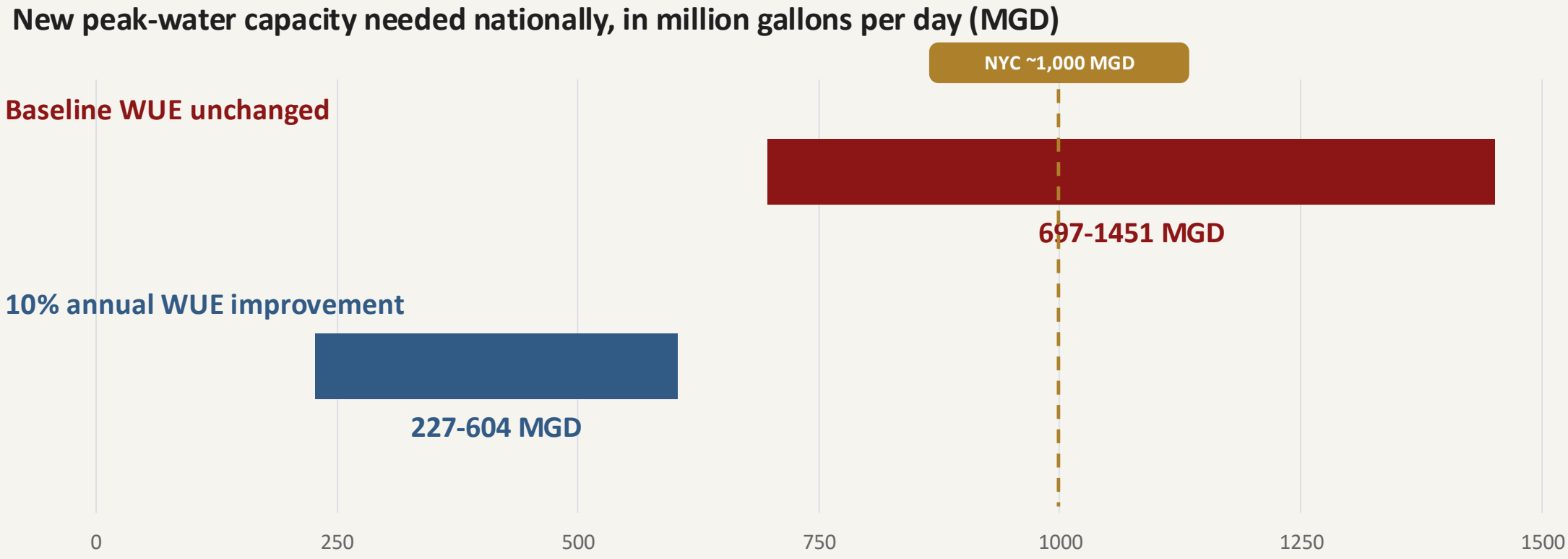
- **Leather Shoes** - 4,000,000 prompts' worth of water
- **Smartphone** - 6,400,000 prompts
- **Jeans** - 5,400,000 prompts
- **T-shirt** - 1,300,000 prompts
- **A single piece of paper** - 2550 prompts
- **A 400 page book** - 1,000,000 prompts

These are not competing measurements of the same query. Short vs. long prompts, model size, cooling, region, training allocation and indirect water change the result.

Sources: Google (2025); Sam Altman (2025); Li, Yang, Islam & Ren (2023/25).

The emerging concern is peak water capacity, not just gallons/year

A 2026 research preprint estimates additional municipal-system capacity through 2030.

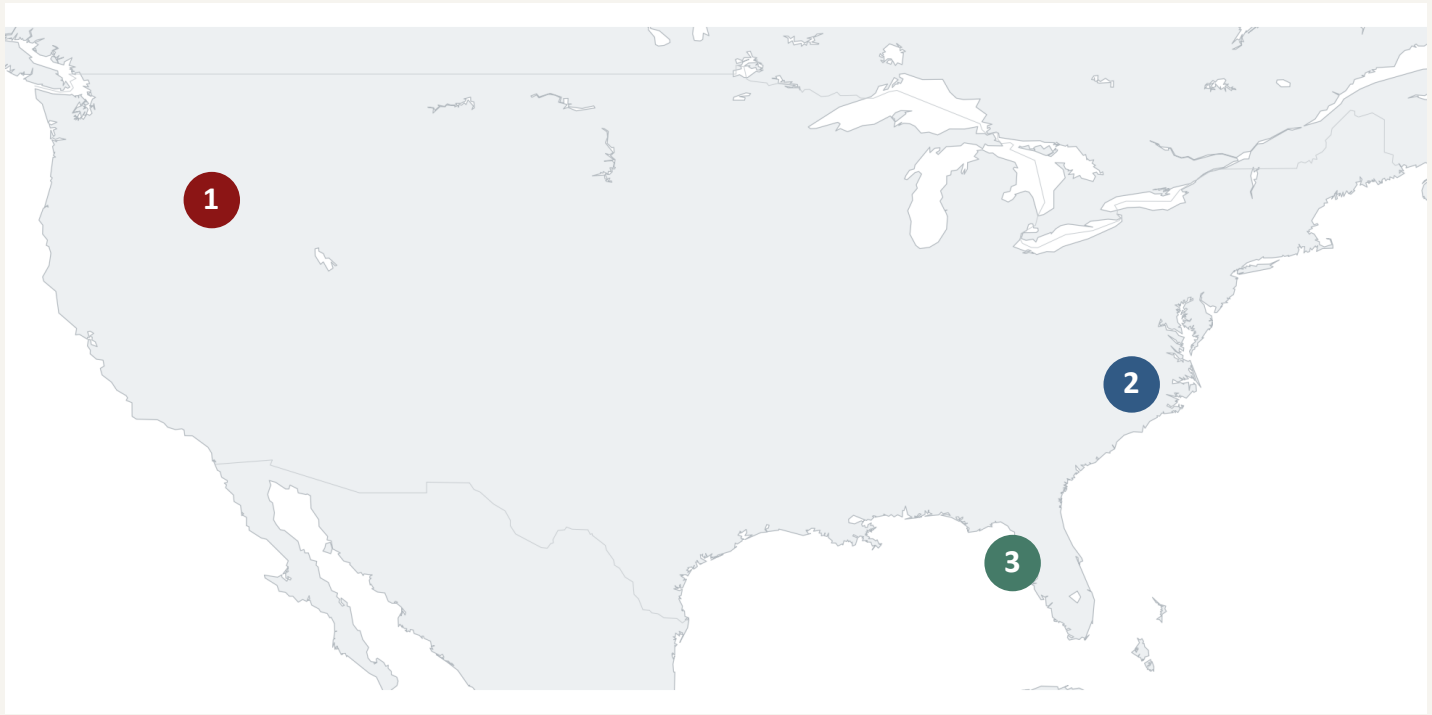


These are modeled additions to peak water-system capacity, not measured daily consumption.

Source: Han, Li, Wierman & Ren, Small Bottle, Big Pipe (March 2026 preprint).

National averages v. local demand

Three documented U.S. examples illustrate high shares, reuse and projected capacity pressure.



1 THE DALLES, OR

Google used ~434 million gallons in 2024 (~1.19 MGD). Roughly one-third of city water usage.

2 LOUDOUN, VA

Data centers ~9% of potable demand; >750 million gallons of reclaimed water delivered in 2025.

3 NEWTON COUNTY, GA

Meta ~196,000 gallons/day; 2030 plans model ~390,000-1 million gallons/day.

Actual reported use, municipal shares and projections are distinct measures. Pin locations are approximate.

Sources: City of The Dalles data via OPB (2026); Loudoun County (2026); PLOS Water (2026).

Cooling design determines how water demand appears locally

Using less on-site water often shifts cost to electricity, cooling hardware or a different source.

EVAPORATIVE COOLING

Often lowers electricity use
but evaporates local water.

- Water passes through cooling tower
- Evaporation carries away heat
- Needs makeup water + wastewater handling

DRY / CLOSED-LOOP

Can sharply reduce on-site water use
but may raise electricity demand.

- Liquid carries heat inside sealed loop
- Air-cooled chillers reject the heat
- Check regional grid water + hot-day power

Reclaimed water, watershed stress and peak demand can matter more than the headline "gallons per prompt."

Sources: Loudoun County water guidance; LBNL (2024); Ren et al. (2026); The Atlantic (2026).

SECTION

The Politics of Data Centers



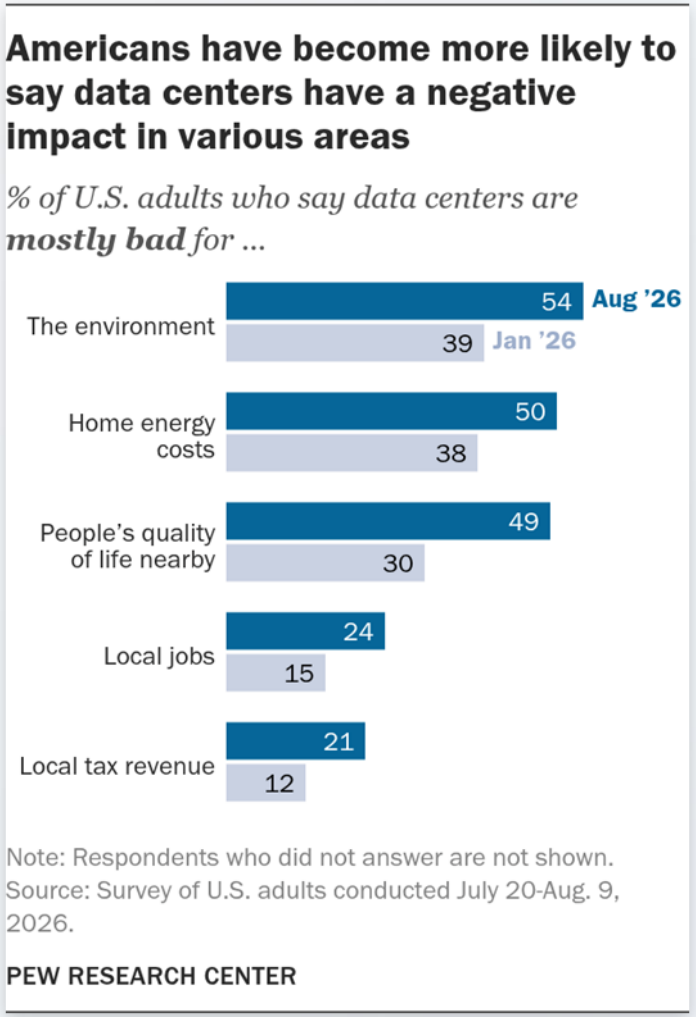
Data Centers Are Even Less Popular Than AI

- Gallup (Mar. 2026): 71% oppose an AI data center in their area
  - 48% strongly; 27% in favor
  - Nuclear power plant: 53% oppose
- Opponents cite: water and energy use, local quality of life, pollution, utility bills
- Supporters: about two-thirds cite economic benefits (jobs, tax revenue)
- Echelon (June 2026): 27% support / 62% oppose
  - Nuclear plant: 34% / 57%
  - AI data centers ranked lowest of projects tested

Source: Jones, Gallup, May 13, 2026; Echelon Insights, June 2026

Opinion on Data Centers: Every Tracking Series Up Sharply

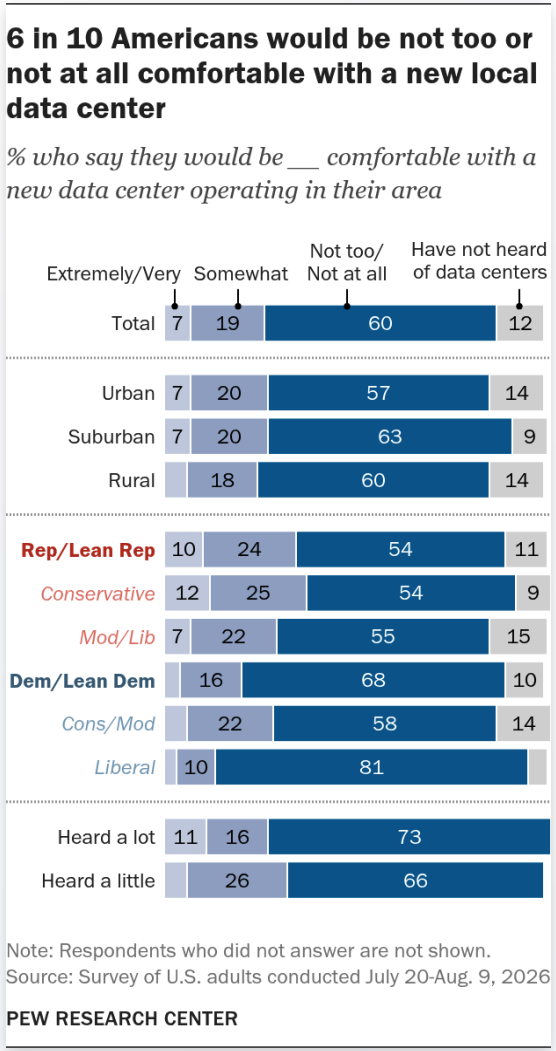
Survey (population)	Question	Earlier	Latest
Heatmap/Embold (registered voters)	Oppose a data center near them	42% (Aug. 2025)	51% (Feb.) · 70% (May) · 75% (Aug. 2026)
Change Research (registered voters)	Oppose construction near home	40% (Mar. 2025)	73–74% (Apr. 2026)
Change Research	Oppose construction in local area	26% (Mar. 2025)	65% (Apr. 2026)
Annenberg (adult citizens)	Oppose new data centers in area	49% (Feb.–Mar. 2026)	61% (June–July 2026)
Marquette (Wisconsin voters)	Costs outweigh benefits	55% (Oct. 2025)	70% (Feb. 2026)
Pew (U.S. adults)	Mostly bad for the environment	39% (Jan. 2026)	54% (July–Aug. 2026)



Question wording differs across surveys; comparisons valid only within each row.

Source: Heatmap/Embold, Aug. 2026; Change Research, 2026; Annenberg Public Policy Center, 2026; Marquette Law School Poll, Feb. 2026; Pew, Sept. 22, 2026

Local AI Data Center Construction



Overall, would you strongly favor, somewhat favor, somewhat oppose, or strongly oppose the construction of a data center in your area to support artificial intelligence, or AI, technology in the U.S.?

Group	Strongly favor	Somewhat favor	Somewhat oppose	Strongly oppose
All U.S. adults	7	20	23	48
Democrats	4	19	19	56
Independents	7	17	26	48
Republicans	9	25	24	39

Gallup telephone survey, Mar. 2–18, 2026 (n=1,000). Percentages.

Source: Jones, Gallup, May 13, 2026

Local data-center construction: additional survey evidence

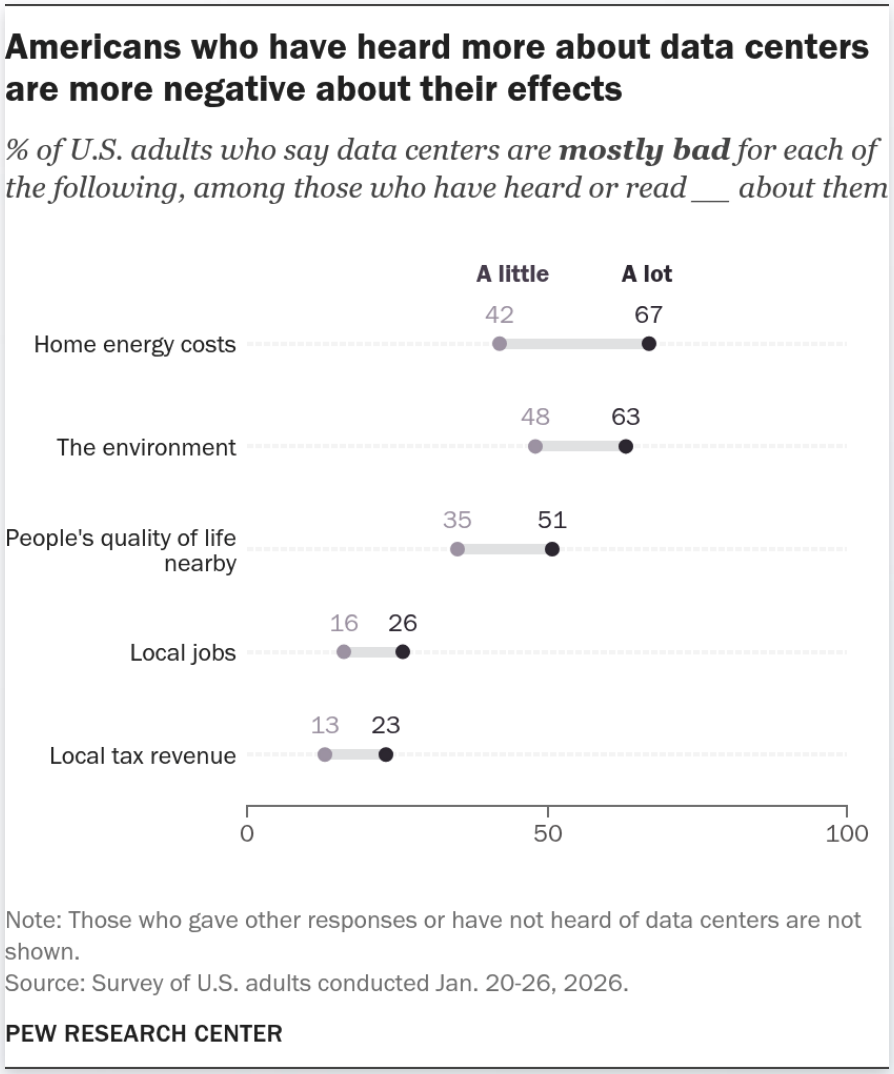
AI7. Overall, would you (oppose) or (support) the construction of new data centers in your area?

Jun.-Jul. 2026

Group	N	NET Sup.	NET Opp.	Strongly support	Somewhat support	Neither	Somewhat oppose	Strongly oppose	Not sure
Jun.-Jul. 2026	1320	**14**	**61**	5	9	25	16	44	0
Democrat	643	**10**	**69**	4	6	21	16	54	0
Independent	157	**14**	**53**	5	9	33	12	41	0
Republican	515	**19**	**54**	6	13	27	18	36	0
Non-user	253	**11**	**64**	5	7	25	13	51	0
Light user	583	**14**	**60**	4	10	26	16	44	0
Heavy user	483	**17**	**60**	7	10	23	19	41	0
18-29	213	**8**	**70**	2	5	22	16	55	0
30-49	521	**14**	**60**	6	8	25	15	45	0
50-64	292	**17**	**58**	6	11	25	16	42	0
65+	289	**17**	**57**	6	11	26	19	38	0

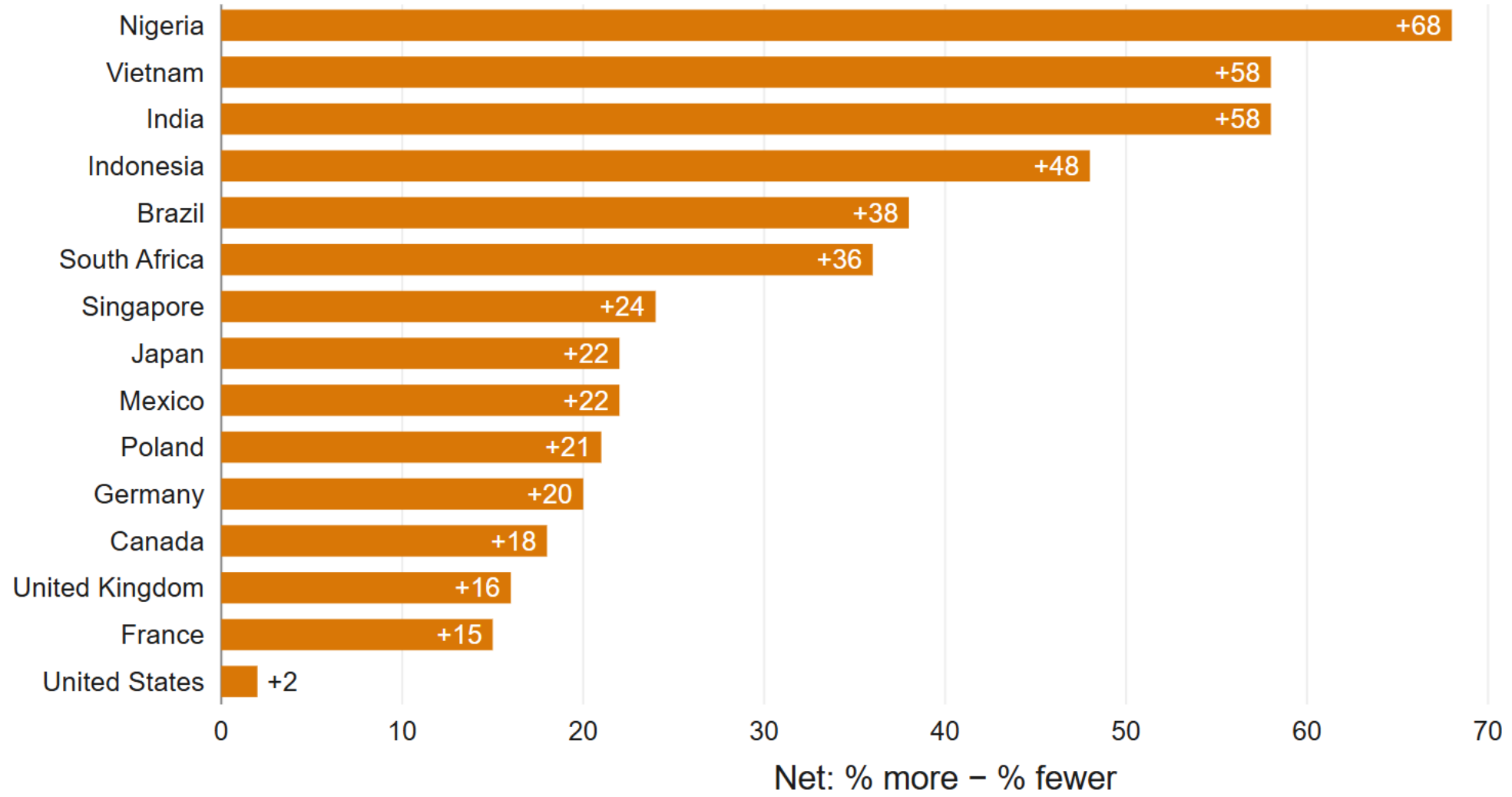
Annenberg (2026)

More Awareness, More Negative Views: To know data centers is to hate them…



Source: Pew, Mar. 12, 2026

Net support for building more data centers rather than fewer



PUBLIC FIRST
A SHGH Company

* Nigeria and India are probably higher – 81% and 73% support respectively.

Data Centers Invert the Usual Patterns

Familiarity does not soften opposition

Oppose local data center: heavy users 60% · light 60% · non-users 64% (Annenberg)

Age runs against use

Oppose: 70% under 30 v. 57% 65+ · “bad for environment”: 54% under 30 v. 26% 65+

Bipartisan, but bigger party gaps

Gallup: D 75% · I 74% · R 63% · Wisconsin Dems 56% → 85%, Reps 53% → 55% · Heatmap net: D −75, I −65, R −43

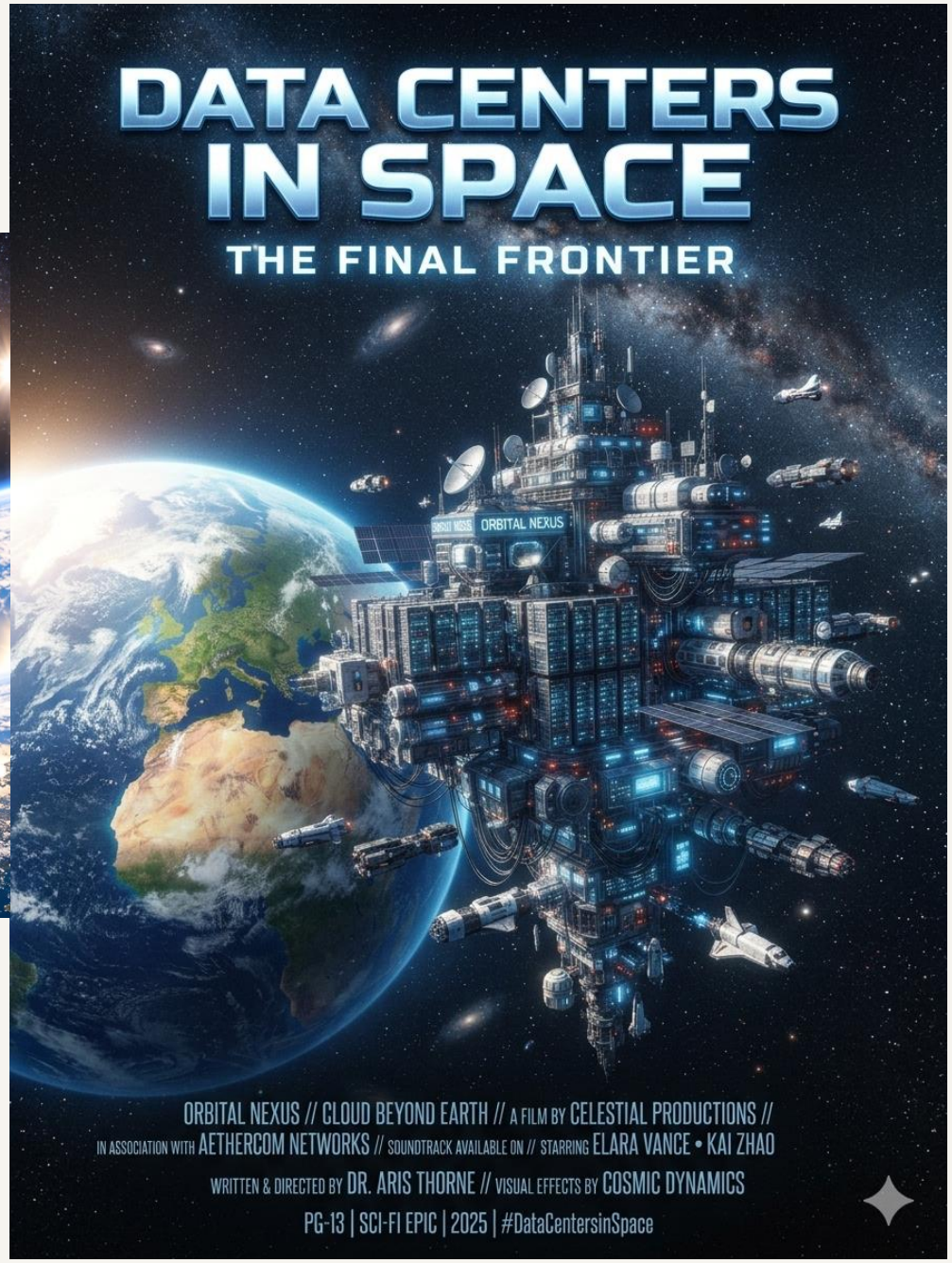
U.S. is an outlier

Public First net “more” v. “fewer” data centers: U.S. +2, India +58, Nigeria +68 (provisional)

Source: Annenberg Public Policy Center, 2026; Pew, Mar. 12, 2026; Jones, Gallup, May 13, 2026; Marquette Law School Poll, Feb. 2026; Heatmap/Embold, Aug. 2026; Public First, June 2026

Pacing the frontier….
the final frontier







Subscribe

Elon Musk

@elonmusk

Terafab.AI

Joined June 2009 >

1,419 Following **241.5M** Followers

The radical proposal: put some AI computation in orbit

Google, SpaceX and startups have explored orbital computation, but a terrestrial replacement remains speculative.

- Google's Project Suncatcher studies networks of solar-powered satellites with Tensor Processing Units.
- Google announced a prototype mission with Planet targeted for early 2027; prototypes test hypotheses, not commercial viability.
- Onboard satellite data processing already has practical uses; an orbital hyperscale AI cluster is much more demanding.
- Space platforms would require launch, maintenance, thermal control, radiation protection and communications.

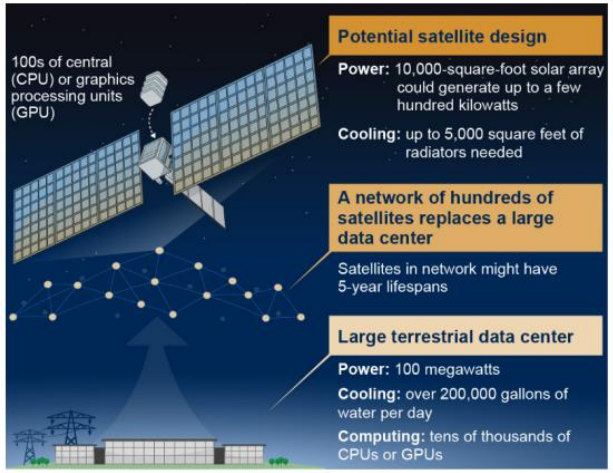
WHY THIS MATTERS

Space-based data centers would place data processing and storage systems for AI and other computing needs into satellites. This could reduce the land, electricity, and water needed for data centers on Earth. Several companies have begun development of data centers in space, but there are engineering and economic barriers to deployment.

KEY TAKEAWAYS

- Placing data centers in space could reduce the demand for resources from these facilities on Earth.
- Data centers generate excess heat, but space does not cool computing hardware efficiently. This could be a major engineering challenge.
- A significant increase in the number of satellites in orbit could be difficult to manage and cause collisions.

Figure 1. Terrestrial and Hypothetical In-Space Data Centers



Source: GAO analysis and illustration. | GAO-26-109012

These might supplement or replace the use of terrestrial data centers for energy-intensive tasks such as cloud computing services or training AI models.

GAO-26-109012 Data Centers in Space

Sources: Google Research

Space solves some problems and introduces several new ones

Sunlight is plentiful, but heat, repair, radiation and communications are difficult.

- No convective cooling in a vacuum: heat ultimately leaves by thermal radiation from large radiators.
- Radiation and debris threaten electronics; fault tolerance and shielding add mass.
- Solar power varies with orbit, orientation and eclipse; storage and large arrays add complexity.
- Launch cost, mass, high-bandwidth optical links and physical repairs become core economics.

Solar power	Arrays + eclipse + batteries
Waste heat	Vacuum → large radiators
Reliability	Radiation + debris + repair
Bandwidth	Optical links + latency

Sources: Google Research · NASA

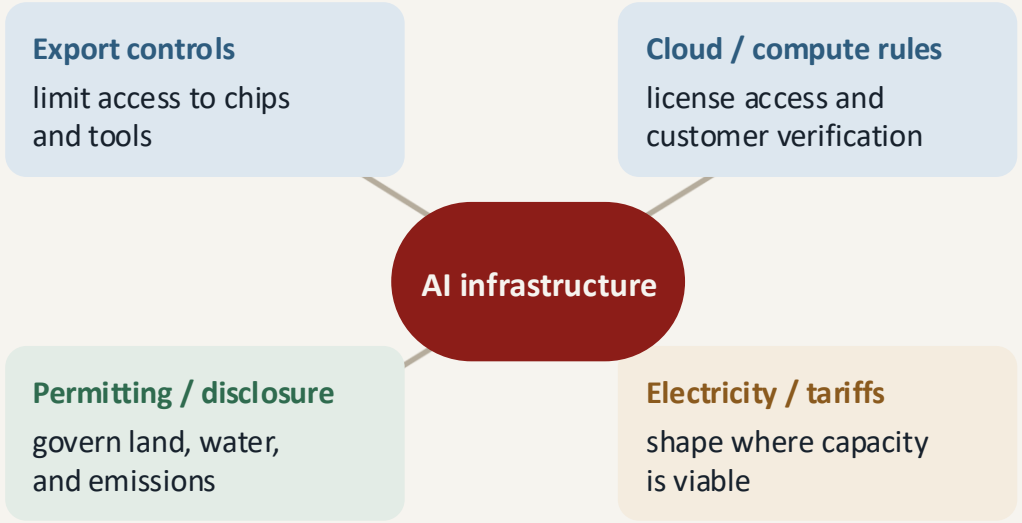
SECTION

Sovereignty

Can governments govern AI through chips, data centers and electricity?

Physical infrastructure creates identifiable entities and chokepoints—but also incentives to route around regulation.

- Export restrictions can influence the geographic availability of certain accelerators.
- Cloud licensing, customer verification and compute reporting are possible regulatory tools.
- Permitting, energy and water disclosure can govern facility impacts independent of model content.
- Limits include cloud cross-border services, open models, legacy hardware, verification and trade retaliation.



Leverage exists — but firms can reroute, internationalize, or shift to less-regulated channels.

Sources: U.S. Bureau of Industry and Security · International Energy Agency

In the AI era, sovereignty is not the ability to build everything yourself.

It is the ability to choose what you depend upon — and to change that choice when your national interests require it.



Compute Is Now an Instrument of Statecraft

Export Controls

Advanced chips and manufacturing equipment controlled under U.S. export administration rules — extending foreign policy into hardware supply chains.

Tiered Access

The 2025 diffusion framework sorted 120+ countries into tiers with differentiated access to advanced compute.

Sovereign in Geography, Not Hardware

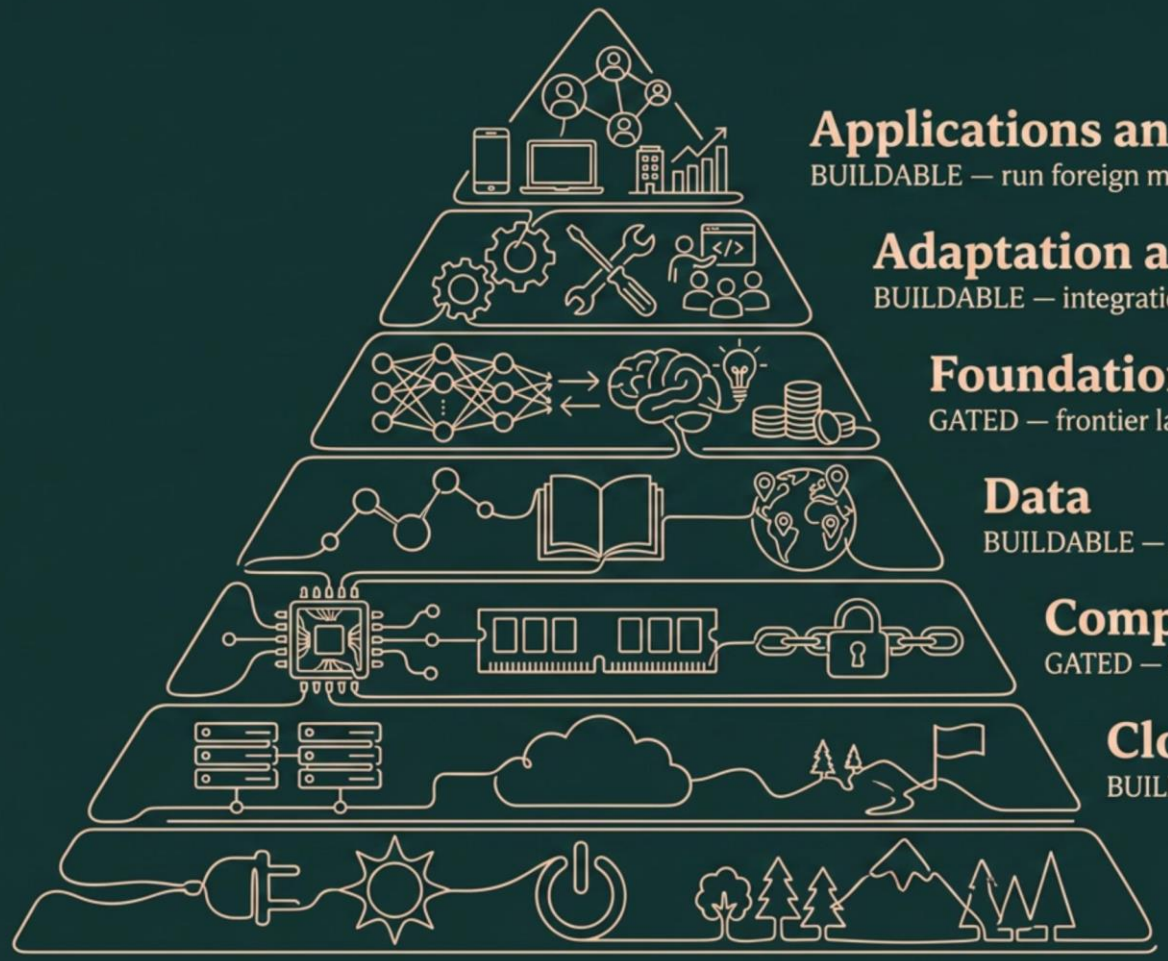
Data centers inside a country do not mean sovereignty over compute. The chips inside obey foreign rules and firmware regardless of physical location.

The Counter-Current

Open-weight models from China are now the default in much of the world. Chinese open models lead in derivative releases globally.



Sovereignty Across the AI Stack



Applications and Adoption
BUILDABLE — run foreign models across the economy

Adaptation and Tooling
BUILDABLE — integration, fine-tuning, local talent

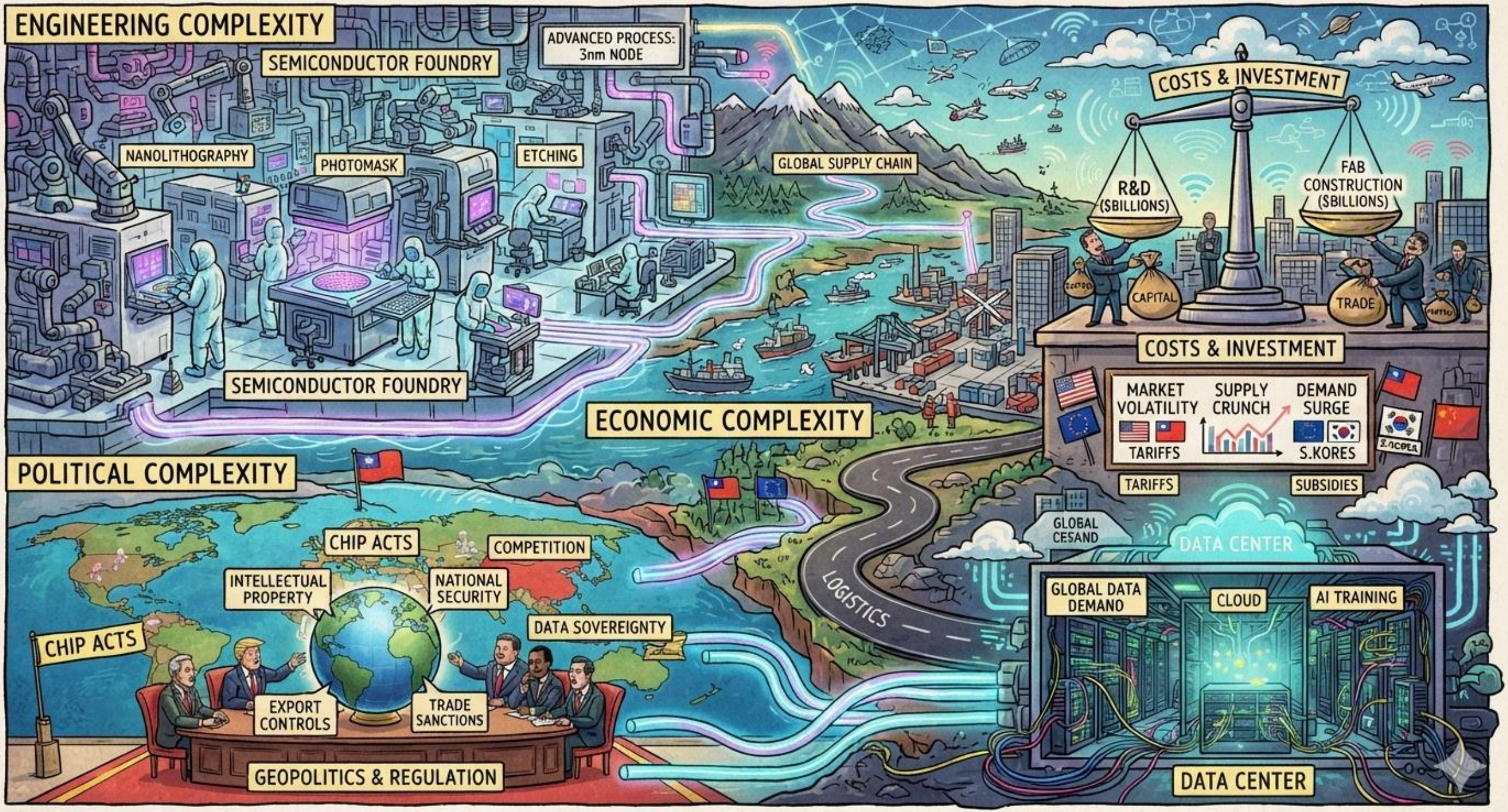
Foundation Models
GATED — frontier labs, capital, data, compute

Data
BUILDABLE — local, language and domain datasets

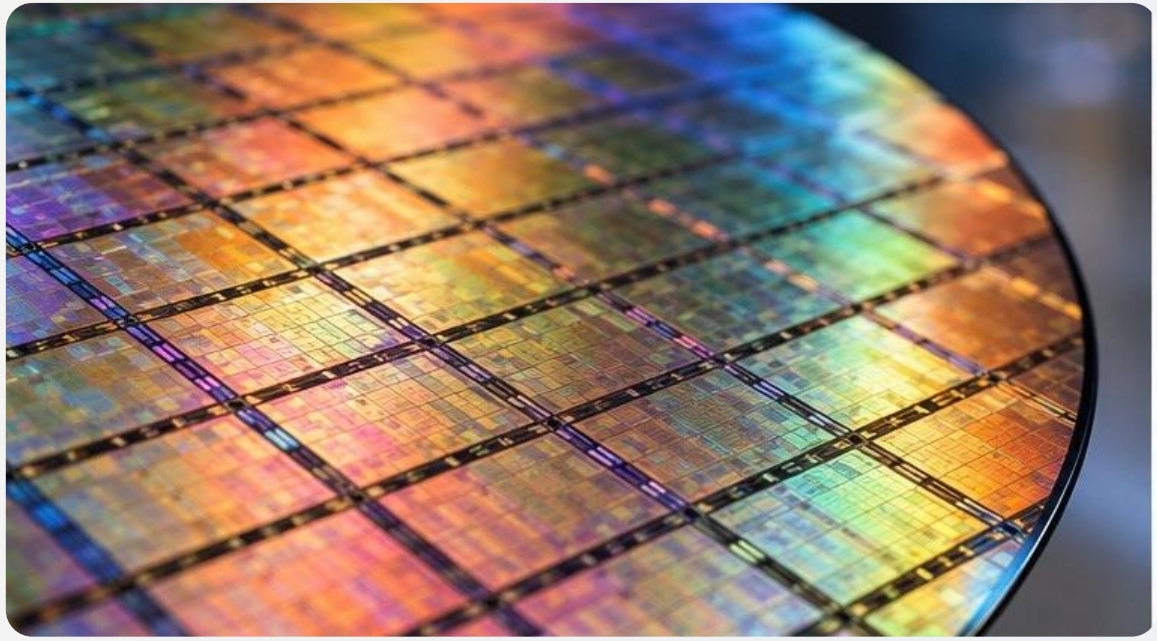
Compute and Hardware
GATED — fabrication concentrated, export-controlled

Cloud and Infrastructure
BUILDABLE — host data centers on home soil

Energy
OWN IT — sovereign, buildable, powers every layer


ENGINEERING COMPLEXITY
SEMICONDUCTOR FOUNDRY
ADVANCED PROCESS: 3nm NODE
NANOLITHOGRAPHY
PHOTOMASK
ETCHING
SEMICONDUCTOR FOUNDRY
GLOBAL SUPPLY CHAIN
COSTS & INVESTMENT
R&D ($BILLIONS)
FAB CONSTRUCTION ($BILLIONS)
CAPITAL
TRADE
COSTS & INVESTMENT
MARKET VOLATILITY
SUPPLY CRUNCH
DEMAND SURGE
TARIFFS
S.KORES
TARIFFS
SUBSIDIES
ECONOMIC COMPLEXITY
POLITICAL COMPLEXITY
CHIP ACTS
COMPETITION
INTELLECTUAL PROPERTY
NATIONAL SECURITY
DATA SOVEREIGNTY
CHIP ACTS
EXPORT CONTROLS
TRADE SANCTIONS
GEOPOLITICS & REGULATION
LOGISTICS
GLOBAL CESAND
DATA CENTER
GLOBAL DATA DEMAND
CLOUD
AI TRAINING
DATA CENTER

AI Infrastructure





Engineering, Geopolitics, Economics, and the Environment

Semiconductor supply chains · Data centers · Energy, climate, water and minerals

Nate Persily | Stanford Law School | October 2026