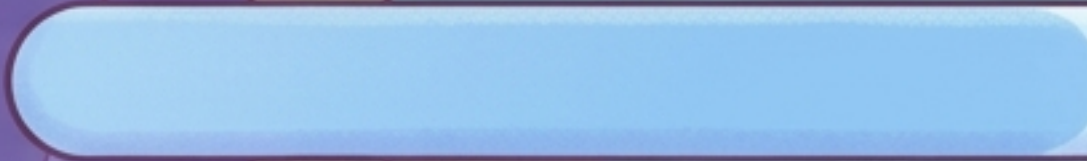


AI Moves Fast. The Grid Moves Slow.



AI Release Cycle (Months)



Grid Connection (Years)

A model team works toward its next release. A utility has to expand a substation. Those schedules are very far apart.



Developers cannot get data centers connected on the schedule their business actually needs.

Building Power Means Joining a Different Queue.

Gas turbines, permits, and the redundancy needed for reliable operation create their own constraints.



(Capacity)



(Consumption)



**An onsite power plant doesn't automatically solve the timing problem.
Moving to a different supply chain just shifts the delay.**

Managing the Peak: The Curtailment Button

AI Power Draw

Grid Stress Level

In a field demonstration, a 256-GPU cluster reduced power by 25% for three hours while meeting service commitments.

Wa!

We can pause or throttle deferrable AI workloads during difficult grid hours to relieve temporary stress.

A Patient User Does Not Make Idle GPUs Cheap.

Depreciation continues while the machines are idle. Waiting for cheap electricity reduces the amount of work the equipment delivers.

Job Submitted!

Move Region

Trade-off: High latency & data movement costs.

Curtail at Peak

Trade-off: Requires recovery overhead & spare capacity.

Wait for Cheap Power

Trade-off: Lost output & higher fixed cost per task.

Depreciation never sleeps. Saving on energy costs by waiting increases the fixed cost of every completed task.

Who Pays for the Empty Room?

Suppose models become more efficient or demand falls short. A 20-year power commitment does not disappear because the job queue is empty.



Infrastructure outlasts AI hardware cycles. Without guaranteed long-term demand, the facility owner carries immense financial risk.