

PROMETHEUS

RESEARCH SERIES | WHITE PAPER No. 1

INTELLIGENCE BECOMES INFRASTRUCTURE

Designing the Energy-Compute Architecture
of the Intelligence Age

David W Tietgen

Gold Proof | August 2026

EXECUTIVE BRIEF | THE THESIS

Artificial intelligence is becoming physical infrastructure. Behind every model sits an industrial system of processors, memory, networking, transformers, substations, cooling, high-capacity fiber-optic communications networks (fiber), land, skilled labor, capital and electricity.

Fiber is the communications backbone connecting an AI campus to users, data sources, cloud infrastructure and other computing campuses. Later in this paper it becomes strategically important because certain computational workloads may be moved geographically toward available energy.

The central problem is not simply that AI needs more power. AI and infrastructure operate on different clocks. Compute can scale at digital speed while generation, transmission, equipment manufacturing and construction expand at industrial speed. That mismatch turns speed-to-power into a competitive variable.

The bottleneck also moves. Solve generation and transmission may constrain the project. Solve transmission and transformers may constrain it. Then cooling, land, water, fiber, labor, capital or community acceptance can become the limiting factor.

PROMETHEUS therefore treats the AI infrastructure challenge as a coordination problem across physical systems operating at different speeds.

The Intelligence Age does not need one energy revolution. It needs an architecture capable of absorbing many revolutions.

EXECUTIVE BRIEF | THE BLANK-SHEET QUESTION

If a 1-GW AI campus were designed today from a blank sheet of paper, it should not begin as a conventional data center waiting for the grid to solve every energy problem. It should begin as an integrated energy-compute system.

The campus can combine grid access, dedicated firm generation, variable generation, batteries, longer-duration storage, advanced cooling and high-capacity communications. The exact energy portfolio changes with geography; the architecture remains common.

The key addition is computational flexibility. Some work must run now. Some can wait. Some can accelerate when energy is abundant. Some may be able to move geographically.

That creates three distinct tools: physical storage moves electricity through time; computational flexibility moves demand through time; communications can move suitable computation through space.

Consider a temporary 150-MW deficit at a 1-GW campus. One illustrative portfolio is 50 MW of battery discharge + 40 MW of flexible AI model-training workloads temporarily deferred + 30 MW of reduced cooling/electrical overhead + 30 MW of suitable work shifted to another campus = 150 MW.

Here, 'deferred' means selected noncritical AI model-training or batch-compute jobs are slowed, paused, or rescheduled - not workforce training. Real-time AI services and other critical workloads remain protected.

When information is easier to move than energy, some computation can follow the energy.

EXECUTIVE BRIEF | THE PROMETHEUS ANSWER

PROMETHEUS proposes two linked control layers. An Energy Orchestration Layer understands grid conditions, generation, storage, cooling, weather and reserves. A Compute Orchestration Layer understands workload priority, deadlines, latency, hardware, data location and mobility.

When the two layers communicate, energy becomes aware of computation and computation becomes aware of energy. The campus stops behaving like a passive electrical load and begins operating as a coordinated cyber-physical system.

The objective is not maximum megawatts, maximum GPU count or the largest campus. The objective is maximum useful intelligence per unit of energy and capital.

The full paper develops the physical constraints, energy portfolio, storage logic, water and cooling choices, power-aware geography, compute-following-energy concept, and the architecture needed to connect them.

Generate what is needed. Store what is economical. Move what can move. Delay what can wait. Protect what cannot fail.

The pages that follow are the evidence and engineering argument behind that operating rule.

MOVEMENT I - THE COLLISION

1. Two Clocks

FIGURE 1 | THE TWO CLOCKS

DIGITAL CLOCK	INDUSTRIAL CLOCK
Models: months	Transmission: years
Software: days/weeks	Generation: years
Processors: product cycles	Transformers/turbines: manufacturing lead times
Demand forecasts: rapid	Permitting/construction: slow

Artificial intelligence and infrastructure operate on different clocks.

AI models can improve in months. New processors arrive in product cycles measured in quarters and years. Software can be deployed globally almost immediately.

The infrastructure beneath those systems moves differently.

A transmission project may require years of planning, permitting, engineering and construction. New generating capacity requires equipment, fuel, interconnection and capital. Large transformers and turbines depend upon manufacturing capacity that cannot be created overnight. Nuclear projects operate on still longer schedules.

The result is not simply an electricity shortage.

It is a **speed mismatch**.

The Intelligence Age is advancing at digital speed while much of its supporting infrastructure expands at industrial speed.

That mismatch is already visible. Global data-center electricity consumption is rising rapidly, while grid bottlenecks threaten to delay substantial amounts of planned capacity.[1]

The question is therefore changing.

For years, technology companies asked:

How quickly can we obtain more compute?

Increasingly they must also ask:

How quickly can we obtain usable power for that compute?

Regulators have given the problem a revealing name:

Speed-to-power.

Large data centers and other enormous new electrical loads are forcing utilities and regulators to reconsider rules designed for a different generation of industrial customers.

This is an important transition.

Electricity is no longer merely an operating expense of artificial intelligence.

Access to electricity is becoming a competitive variable.

2. The Gigawatt Changes the Conversation

A megawatt-scale computing facility is an electricity customer.

A gigawatt-scale AI campus begins to resemble something else.

At full load, 1 GW operating continuously represents 8.76 TWh of annual electricity consumption.

That is not a forecast for every future AI campus. Actual utilization varies, and not every announced project will operate continuously at nameplate capacity.

But the arithmetic demonstrates the scale.

At gigawatt levels, questions that once belonged primarily to utilities become technology-company questions:

Where will generation come from?

Who will finance it?

Who pays for transmission?

How much redundancy is required?

What happens during peak grid conditions?

Can the load be reduced?

Can some computation wait?

Can batteries absorb rapid load changes?

Can the campus generate part of its own electricity?

Can some workloads operate somewhere else?

Those questions reveal why the conventional data-center model is reaching its limits.

The old model treated electricity as an input.

The emerging model must treat electricity as part of the **architecture**.

3. The Grid Was Not Designed for This

The electric grid is one of the largest engineered systems ever created.

Its fundamental obligation is deceptively simple:

Supply and demand must remain balanced continuously.

AI complicates that task in two ways.

First is scale.

Individual campuses can require hundreds of megawatts, with announced projects reaching gigawatt levels.

Second is behavior.

AI workloads can produce unusually rapid changes in electrical demand. Tightly synchronized computing equipment can change power consumption much faster than many traditional industrial loads.

This creates a paradox.

The same computational systems creating a new challenge for the grid may also help solve it.

Some AI workloads are time-sensitive.

Others are not.

Some inference must occur immediately.

Some training, batch processing and other computational work can potentially be scheduled around energy availability.

That means a large AI campus need not always behave like a conventional factory operating at an inflexible load.

It can potentially become **power-aware**.

And that gives infrastructure planners a new tool.

Instead of designing the entire energy system around maximum instantaneous computational demand, part of the computational system can be designed around the availability of energy.

This distinction becomes critical later in this paper.

4. Power Is Only the First Collision

FIGURE 2 | THE MOVING BOTTLENECK

SYSTEM LAYER	WHAT CAN CONSTRAIN IT
Generation	Capacity / fuel / permitting
Transmission	Interconnection / lines
Electrical equipment	Transformers / switchgear
Cooling	Heat rejection / water / power
Compute	Processors / memory / networking
Site	Land / fiber / roads / risk
Execution	Labor / capital / community

Suppose sufficient electricity is available.

The problem is not necessarily solved.

A modern AI campus requires a chain of physical systems:

Generation

v

Transmission

v

Substations and Transformers

v

Power Electronics

v

Water and Cooling Resources

v

Processors and Networking

v

Fiber and Communications

v

Land and Site Infrastructure

v

Skilled Labor and Construction Capacity

Each layer has its own capacity limits.

A billion-dollar investment in GPUs is of little use if the building cannot energize them.

Available generation is of limited value if transmission cannot deliver it.

Transmission capacity does not solve a transformer shortage.

A completed data center cannot operate at design capacity if its cooling infrastructure cannot remove the heat.

And a technically attractive location is not necessarily a buildable location.

Land Is More Than Acreage

A gigawatt-scale AI campus requires more than an available parcel.

The site must accommodate the data-center buildings themselves, substations, transmission access, backup and dedicated generation, batteries, cooling infrastructure, fiber routes, security, roads and potentially substantial future expansion.

The surrounding geography matters as well.

Is sufficient power nearby?

Can new transmission reach the site?

Is fiber available through redundant routes?

Are water resources adequate for the proposed cooling architecture?

Can heavy equipment reach the property?

Is the site exposed to flooding, wildfire, extreme heat or other physical risks?

Can additional generation, storage or compute capacity be added later?

The cheapest land may therefore become very expensive if everything the campus requires must be brought to it.

For intelligence infrastructure, location should be valued not simply by the cost of land, but by the infrastructure the land can reach.

Water and Cooling Become Site Decisions

Water deserves particular attention because cooling architecture can influence where large AI campuses can responsibly and economically operate.

Closed-loop systems can recirculate water or another coolant repeatedly through the facility while heat is removed through heat exchangers. The full cooling supply is not replaced on every pass; losses arise mainly from evaporation, maintenance, leakage and water-quality management.

This makes water withdrawal different from water consumption. A campus may circulate a large volume internally while consuming much less fresh water. Reclaimed municipal or industrial water can further reduce reliance on potable supplies where conditions permit.

Heat recovery adds another opportunity: warm cooling water may serve district heating, industrial processes, greenhouses or other thermal loads. The better question is how much fresh water is consumed, how often it can be reused, and whether rejected heat can become a resource.

Cooling technology affects electricity consumption.

Local climate affects cooling performance.

Water availability can affect site suitability.

Processor design affects heat density.

And heat density can influence the architecture of the entire facility.

Cooling therefore becomes part of the energy-compute system itself.

And Infrastructure Still Has to Be Built

Even the ideal site remains only a plan until people construct it.

Gigawatt-scale AI development requires electricians, engineers, pipefitters, welders, equipment operators, utility crews, network specialists, cooling technicians and other skilled trades - often in large numbers and often competing with other infrastructure projects for the same workers.

Manufacturing capacity and workforce capacity therefore interact.

A transformer shortage cannot necessarily be solved merely by spending more money if factories are already operating near capacity.

Likewise, an accelerated construction schedule cannot necessarily be achieved simply by adding capital if the required skilled workers are unavailable.

This creates another potential bottleneck:

The Intelligence Age may discover that some of its most important limiting resources are not digital at all. They are people capable of building the physical systems on which digital intelligence depends.

That point leads naturally into the fuller workforce discussion later in this paper.

The bottleneck therefore moves.

And that produces the first major conclusion of Movement I:

AI infrastructure is constrained not by one bottleneck, but by the slowest critical component of an interconnected system.

Solving electricity while ignoring land, water, cooling, equipment, construction capacity, communications or workforce merely moves the constraint

somewhere else.

5. Heat Becomes Architecture

For decades, cooling was largely a supporting function of computing.

AI is changing that relationship.

As power density rises, heat removal increasingly influences the design of processors, racks, buildings and electrical systems themselves.

Emerging rack-scale AI architectures illustrate the direction. Computing systems are increasingly being designed around liquid cooling rather than treating cooling as something added after the computing equipment is selected.

That is more significant than a change in cooling technology.

It demonstrates a broader pattern:

When a physical constraint becomes severe enough, the computer is redesigned around the infrastructure constraint.

That is exactly the kind of adaptation PROMETHEUS expects to see repeatedly.

Energy affects compute.

Compute affects cooling.

Cooling affects facility design.

Facility design affects site selection.

Site selection affects transmission, communications, water, labor and community impact.

Everything begins to converge.

6. Capital Meets Concrete

There is another collision underway.

AI is a digital technology requiring extraordinary quantities of physical capital.

Data centers must be financed.

Generation must be financed.

Transmission must be financed.

Semiconductor fabrication must be financed.

Equipment must be ordered years ahead.

Land must be acquired.

Construction workers must be hired.

And capital committed today may support infrastructure whose economics depend upon computing technology that will be very different by the time the project is completed.

That creates an unusual financial risk.

Infrastructure is long-lived. AI technology is not.

A power facility may operate for decades.

A data-center building may remain useful for decades.

The processors inside it may be economically superseded in a fraction of that time.

The challenge is therefore not simply raising enough money.

It is designing infrastructure that remains useful while the technology it supports changes rapidly.

That is another reason PROMETHEUS favors **technology-flexible architecture**.

Do not build an energy system that works only for today's processor.

Do not build storage around one chemistry if the electrical architecture can accommodate several.

Do not design a campus whose economics depend upon one generation technology remaining dominant for thirty years.

Build the interfaces so that components can evolve.

7. The Human Constraint

The Intelligence Age may be powered by machines.

It will still be built by people.

Electricians.

Engineers.

Welders.

Pipefitters.

Construction workers.

Nuclear operators.

Utility crews.

Network engineers.

Semiconductor technicians.

Equipment manufacturers.

Regulators.

Local officials.

And communities.

This matters because an AI campus cannot simply be downloaded into existence.

Every megawatt ultimately intersects with the physical economy.

That means workforce development is not peripheral to AI strategy.

It is AI strategy.

The same applies to community acceptance.

A project can have land, financing, processors and electricity and still fail if the community hosting it believes the costs exceed the benefits.

Those concerns can involve water consumption, electricity prices, noise, land use, environmental effects, tax structures and whether local residents actually share in the economic benefits.

PROMETHEUS therefore treats **community license** as infrastructure.

Not because every objection should stop a project.

And not because every project should proceed.

Because projects that ignore the communities around them introduce another form of infrastructure risk.

8. The Wrong Response

FIGURE 3 | THE 150-MW PROBLEM

RESPONSE	MW
Battery discharge	50
Flexible AI model-training workloads deferred	40
Reduced cooling/electrical overhead	30
Shifted computation	30
TOTAL	150

At this point the temptation is obvious.

AI needs enormous infrastructure.

Therefore:

Build everything.

More generation.

More transmission.

More batteries.

More data centers.

More pipelines.

More fiber.

Some of that will unquestionably be necessary.

But simply making every component larger is an expensive way to solve a coordination problem.

Consider a hypothetical AI campus experiencing a 150-MW temporary shortage.

The conventional response is to provide another 150 MW of supply.

But what if the campus could instead combine:

50 MW of battery discharge

40 MW of flexible AI model-training workloads temporarily deferred

30 MW of reduced cooling/electrical overhead

30 MW shifted to another campus

The same 150-MW problem has been addressed without constructing 150 MW of generation solely for a temporary condition.

Here, 'deferred' means selected noncritical AI model-training or batch-compute jobs are slowed, paused, or rescheduled - not workforce training. Real-time AI services and other critical workloads remain protected.

Before building capacity for every peak, determine whether the system itself can become more flexible.

Generation remains essential.

Transmission remains essential.

Storage remains essential.

But orchestration can change how much of each is required and when.

9. The Collision Produces the Solution

We can now state the problem more precisely.

The Intelligence Age is not confronting an energy crisis alone.

It is confronting a **coordination problem across physical systems operating at different speeds.**

Compute advances rapidly.

Energy infrastructure advances more slowly.

Capital commits before all technical outcomes are known.

Cooling changes with processor density.

Communications determines where computation can occur.

Land determines where the system can physically expand.

Water and cooling resources influence where dense compute can operate efficiently.

Workforce determines how quickly infrastructure can be constructed.

Communities determine where it can be accepted.

Regulation determines how quickly many of these pieces can connect.

Treating each separately produces delay.

The alternative is integration.

That leads directly to the next movement.

MOVEMENT II - THE REINVENTION

The solution is not one enormous power plant.

It is not one miraculous battery.

It is not nuclear alone.

It is not solar alone.

It is not natural gas alone.

It is not moving every data center off-grid.

And it is certainly not waiting for an unproven technology to rescue the system.

The opportunity is to create an architecture capable of combining technologies according to what each does best.

Generate when generation is abundant.

Store when storage is economical.

Move electricity when transmission is available.

Move computation when computation is more movable than electricity.

Reduce demand when the work can wait.

Protect critical workloads when it cannot.

And continuously optimize the relationship among them.

The Intelligence Age therefore requires a different infrastructure philosophy:

Do not force every resource to solve every problem. Give each resource the problem it solves best - and orchestrate the system.

That is where **Movement II - The Reinvention** begins.

And it begins with the resource without which every other layer of machine intelligence stops:

ENERGY

ENERGY

The collision described in Movement I creates an obvious temptation:

Find more electricity.

That is necessary.

But it is not enough.

The better question is:

What energy architecture gives a 1-GW AI campus the greatest combination of reliability, speed-to-power, flexibility, resilience, and long-term adaptability?

That answer is unlikely to come from a single resource.

The emerging system will probably combine several.

10. Stop Looking for One Winner

FIGURE 4 | ENERGY RESOURCES BY FUNCTION

RESOURCE	BEST ROLE
Grid	Diversity / markets / redundancy
Firm generation	Dependable capacity
Solar & wind	Low-marginal-cost energy when available
Batteries	Fast response / hourly shifting
Long-duration storage	Extended balancing
Flexible compute	Avoided or shifted demand

The energy debate is often framed as a contest.

Nuclear versus solar.

Gas versus renewables.

Batteries versus generation.

Grid power versus behind-the-meter power.

That framing is increasingly unhelpful.

A 1-GW AI campus has different energy needs at different timescales.

It needs power in milliseconds.

It needs power through the evening.

It needs power during multi-day weather events.

It needs emergency power.

It needs firm capacity.

It needs low-cost energy when available.

It needs a path to expansion.

No single resource performs every one of those jobs equally well.

The better architecture is therefore **layered**.

Firm generation supplies dependable capacity.

**Renewables contribute low-marginal-cost energy where
resources are favorable.**

Batteries manage rapid changes and shift energy across time.

Long-duration storage addresses longer imbalances.

The grid supplies diversity, markets, and redundancy.

**Flexible computation reduces demand when that is cheaper
than increasing supply.**

The purpose is not technological compromise.

It is system optimization.

**Do not ask which energy resource wins. Ask which
problem each resource solves best.**

11. The Grid Remains Essential

The rise of dedicated generation does not make the electric grid obsolete.

Quite the opposite.

A strong grid connection gives an AI campus access to generation diversity, electricity markets, regional redundancy, reserve capacity, and the ability to import or potentially export power.

But the relationship is changing.

Historically, a large customer requested service and the grid supplied it.

Increasingly, AI campuses are large enough that their interconnection affects the system around them.

FERC's June 2026 action made that explicit. The Commission ordered all six regional grid operators under its jurisdiction to justify or reform the rules governing how data centers and other large loads connect to the transmission system.[2]

That means **speed-to-power** is no longer merely a developer complaint.

It has become a regulatory and national infrastructure question.

The implication for PROMETHEUS is straightforward:

The optimal AI campus should remain grid-connected, but it should not depend upon the grid as its only energy pathway.

That distinction is central to the design that follows.

12. Dedicated Generation Changes the Relationship

One of the clearest signs of change arrived in July 2026.

The U.S. Department of Energy's National Nuclear Security Administration selected a proposal for negotiations involving a **1-GW AI data center paired with dedicated on-site energy generation** at the Savannah River Site.[3]

That is precisely the type of convergence PROMETHEUS anticipates.

The data center and the energy project are no longer separate developments.

They become one development.

This model offers several potential advantages:

faster access to capacity,

greater control over reliability,

reduced exposure to congested transmission,

better coordination between generation and compute,

and the possibility of designing future expansion into the site from the beginning.

It also raises difficult questions.

Who carries fuel-price risk?

Who pays for stranded generation if compute demand changes?

What happens to excess power?

How should the campus interact with the surrounding grid?

What safeguards protect other customers?

Dedicated generation therefore does not eliminate complexity.

It relocates some of that complexity inside the campus.

13. Natural Gas: The Near-Term Reality

Any serious 2026 energy analysis must acknowledge natural gas.

It can be built relatively quickly.

The United States possesses extensive fuel infrastructure.

Gas turbines can provide firm power.

And data-center demand is already influencing fossil generation expectations. EIA reported in March that faster data-center load growth could increase fossil generation, particularly because power demand is rising faster than some alternative infrastructure can be constructed.

That makes gas an obvious **near-term bridge resource** for some campuses.

Its weaknesses are equally obvious.

Fuel-price exposure.

Pipeline dependence.

Emissions.

Potential methane leakage.

Community opposition.

And the risk of constructing long-lived assets around a fuel mix that may change substantially over the next several decades.

PROMETHEUS therefore does not treat gas as either villain or savior.

It treats gas according to function:

Fast, firm capacity that may solve near-term speed-to-power problems - but should be integrated into an architecture capable of evolving beyond it.

That last clause matters.

A campus built around gas alone can become locked into gas.

A campus built around a flexible electrical microgrid can use gas today and replace part of that capacity with nuclear, geothermal, storage, or another proven source later.

14. Nuclear: From Gigawatts to Modules

Nuclear power offers something particularly attractive to AI infrastructure:

firm, high-capacity-factor electricity with low operational carbon emissions.

Large conventional reactors can provide enormous quantities of reliable power.

But their capital requirements, construction schedules, regulatory complexity, and project risk make them difficult to treat as the universal answer to immediate data-center demand.

The more interesting development for PROMETHEUS is **modularity**.

Small modular reactors and microreactors raise the possibility of building firm nuclear capacity in smaller increments.

DOE describes microreactors as factory-built systems capable of serving nontraditional markets, including data centers.

And the technology has advanced materially in 2026. DOE reported in July that three distinct advanced microreactor designs achieved criticality demonstrations under federal programs within the same month.

That is still far from widespread commercial deployment.

But the direction matters.

The traditional nuclear model is:

one enormous plant -> one enormous block of power.

A future modular model could become:

small block + small block + small block -> expand as compute expands.

That aligns unusually well with AI campuses whose computational capacity may grow in stages.

The PROMETHEUS question is therefore not:

Will microreactors power every AI facility?

It is:

Can modular nuclear eventually allow firm generation to scale in increments that more closely match the growth of compute?

If the answer becomes yes, that changes campus design substantially.

15. Nuclear Microgrids

A reactor alone is not the interesting architecture.

A reactor inside a **hybrid microgrid** is.

DOE is actively studying microgrids that combine small nuclear reactors with storage and other resources for high-uptime applications.

That combination matters because nuclear and batteries solve different problems.

Nuclear supplies sustained energy.

Batteries respond almost instantly.

The grid supplies regional diversity.

Flexible compute adjusts demand.

Put them together and the system becomes more capable than any component independently.

Conceptually:

NUCLEAR

v

CAMPUS MICROGRID

<-> BATTERY STORAGE

<-> GRID

v

AI COMPUTE

Now add workload orchestration.

The campus can preserve critical loads, defer lower-priority work, island during emergencies, and potentially support the surrounding system under certain market and interconnection arrangements.

The important innovation is not merely a smaller reactor.

It is **the integration of generation, storage, grid connection, and computation into a controllable system.**

16. Geothermal: The Quiet Candidate

Advanced geothermal deserves far more attention than it usually receives in discussions about AI power.

Its appeal is straightforward.

It can provide firm electricity.

It requires relatively little surface land compared with some generation technologies.

Its output is not dependent upon sunlight or wind.

And advances in drilling technology are expanding the number of locations that may eventually become economically viable.

Traditional geothermal has been geographically constrained.

Enhanced geothermal systems and advanced drilling could loosen that constraint.

That does not make geothermal universal.

Subsurface conditions matter.

Drilling risk matters.

Capital costs matter.

Permitting matters.

But if advanced geothermal can scale commercially, it may become one of the most natural partners for round-the-clock computing.

Why?

Because its operating profile resembles the demand profile AI campuses often want:

continuous electricity without continuous fuel delivery.

For PROMETHEUS, geothermal belongs in the category:

SCALABLE NEXT - potentially strategic where geology and economics align.

17. Solar and Wind: Abundance Requires Coordination

Solar and wind will remain important because they can add large quantities of generation relatively quickly in favorable regions.

EIA expects U.S. developers to add a record **86 GW of utility-scale generation and storage capacity in 2026**, with solar representing 43.4 GW of planned additions.[4]

That scale cannot be dismissed.

But nameplate capacity is not the same thing as firm AI capacity.

Solar disappears at night.

Wind changes with weather.

A campus requiring highly reliable electricity cannot simply treat intermittent generation as if every installed megawatt is continuously available.

The solution is coordination.

Renewables + batteries.

Renewables + firm generation.

Renewables + transmission.

Renewables + flexible compute.

If abundant solar appears at noon, selected computation can accelerate.

If wind surges overnight, batteries can charge and training workloads can expand.

If renewable output falls, firm generation and the grid carry more of the burden.

The objective is not to force renewable power into a workload pattern it cannot always match.

It is to make the workload more capable of taking advantage of renewable abundance when it appears.

18. Batteries: The Electrical Shock Absorber

Battery storage is expanding rapidly enough that it must be treated as major infrastructure rather than experimental technology.

EIA expects developers to add **24 GW of utility-scale battery storage in the United States during 2026.**^[4]

But batteries should not be asked to perform one job.

They perform several.

At the shortest timescale, batteries can stabilize power and absorb rapid AI load changes.

Over hours, they can shift electricity from periods of abundance to periods of scarcity.

They can bridge failures.

They can reduce peaks.

They can help a campus transition into islanded operation.

And eventually, different battery chemistries may address different durations.

This suggests a layered storage architecture:

milliseconds to minutes - power quality, UPS, load smoothing

hours - lithium/LFP and similar systems

longer-duration - flow, sodium, iron-based and other systems

multi-day - emerging long-duration storage technologies where economically viable

The important point is not which chemistry wins.

It is that **storage duration should match the problem.**

Using a four-hour battery to solve a four-day problem is poor system design.

19. Storage Is Not the Only Storage

FIGURE 5 | THREE WAYS TO MOVE THE PROBLEM

TOOL	WHAT MOVES
------	------------

Physical storage	Electricity through time
Flexible computation	Demand through time
High-capacity communications	Suitable computation through space

This is where the PROMETHEUS architecture becomes more interesting.

A battery stores electricity.

But a flexible computational workload can sometimes provide a related system benefit by **delaying the need for electricity**.

These are not identical.

A delayed computation is not a stored electron.

But from the perspective of a temporary grid constraint, both can reduce immediate power demand.

That leads to a useful distinction:

Physical storage moves electricity through time.

Computational flexibility moves demand through time.

And high-capacity communications may allow certain workloads to move through geography.

The future energy system can use all three.

Move electricity through time.

Move computation through time.

Move computation through space.

That gives the AI campus more options than a conventional industrial load has historically possessed.

20. The Missing Resource: Avoided Demand

Energy planning traditionally focuses on supply.

How much generation must be built?

How much storage?

How much transmission?

But a flexible AI campus introduces another resource:

power that does not need to be consumed at that moment.

Suppose a campus expects a temporary 100-MW constraint.

One solution is to build or buy another 100 MW.

Another is to combine:

battery discharge,

temporary reduction of lower-priority training,

more efficient cooling,

rescheduling,

and workload movement.

That does not eliminate the need for additional infrastructure over the long term.

But it may reduce the amount of infrastructure built solely for rare peaks.

This suggests a hierarchy:

First eliminate waste. Then shift what can move. Store what is economical. Build what remains necessary.

That is a much more disciplined infrastructure philosophy.

21. Cooling Creates Virtual Capacity

Cooling deserves to be treated as part of energy supply because every megawatt consumed by cooling is a megawatt unavailable to computation.

Improving cooling efficiency therefore has an effect similar to creating additional usable capacity inside the site.

If a campus reduces its non-compute electrical overhead, the grid does not necessarily deliver more electricity.

But more of the electricity already delivered becomes useful compute.

That is why advanced liquid cooling matters.

The processor, rack, cooling loop, building, and electrical system begin to function as a single engineered unit.

In practical terms:

Efficiency is generation that does not need to be built.

Not literally.

But economically and operationally, the comparison is useful.

A 5% improvement on a 1-GW campus represents **50 MW** of reduced demand.

At this scale, percentages become infrastructure.

22. Fuel Cells and Distributed Generation

Fuel cells deserve a place in the architecture because they offer modular, on-site generation and can potentially be deployed faster than some large central-station projects.

Their usefulness will depend heavily upon fuel source, economics, emissions profile, maintenance, and scale.

But their architectural attraction is clear:

They can place generation **near the compute load**.

That reduces dependence on long-distance transmission for at least part of the site's power requirement.

The broader point extends beyond fuel cells.

AI infrastructure is likely to encourage a revival of **distributed generation** at industrial scale.

Not residential rooftop generation.

Hundreds of megawatts of generation located at or near enormous computing loads.

That changes how utilities, regulators, pipelines, grids, and data-center developers interact.

23. Hydropower: The Geographic Advantage

Hydropower offers another lesson.

Not every region begins with the same energy endowment.

Areas with abundant hydroelectric generation can offer unusually attractive combinations of firm or dispatchable electricity, existing transmission, and relatively low operational emissions.

The Pacific Northwest is an obvious example.

But hydropower cannot simply be replicated everywhere.

Geography determines availability.

That reinforces one of PROMETHEUS's central design principles:

The future AI energy system should be geographically optimized rather than nationally standardized.

A campus near hydro should exploit hydro.

A campus in a geothermal-rich region should evaluate geothermal.

A solar-rich region should use solar aggressively.

A region with nuclear capacity should value it differently.

The architecture remains common.

The energy portfolio changes with location.

24. Power-Aware Geography

This leads directly to site selection.

For decades, data centers clustered around several familiar advantages:

fiber,
customers,
tax policy,
land,
and network interconnection.

Power availability is now moving rapidly toward the top of that list.

The optimal AI geography may increasingly be determined by:

available generation,
transmission headroom,
fuel infrastructure,
renewable resources,
geothermal potential,
water conditions,
climate,
construction workforce,
fiber connectivity,
and community acceptance.

This creates a powerful inversion.

Instead of asking:

Where do we want the data center and how do we bring power there?

we increasingly ask:

Where can energy, communications, land, and community support a long-lived intelligence campus most effectively?

That is a different site-selection philosophy.

25. Compute Follows Energy

Now we arrive at one of the most important ideas in the paper.

Electricity is difficult and expensive to move across long distances when transmission is constrained.

Information can sometimes move far more easily.

Therefore, for suitable workloads, the better choice may not always be to move electricity toward computation.

It may be to move computation toward electricity.

Not every workload can do this.

Latency-sensitive inference may need to remain close to users.

Data-sovereignty requirements may restrict geography.

Large datasets can be expensive to move.

Security matters.

Hardware availability matters.

But some training and batch workloads may be substantially more flexible.

That leads to a continental architecture.

Texas experiences an electricity constraint.

Utah has abundant available geothermal capacity.

The Pacific Northwest has favorable hydro conditions.

Another region has excess overnight wind.

Suitable computational work migrates accordingly.

Now fiber becomes part of the energy system.

Communications allows certain forms of intelligence to move toward available power.

That is a new infrastructure relationship.

26. Space Extends the Geography

The same logic eventually extends beyond Earth.

If suitable computation can move toward terrestrial energy abundance, some future workloads may eventually move toward abundant energy in space.

Solar-powered orbital compute remains a frontier concept.

The economics, thermal management, radiation environment, launch costs, communications, servicing, and hardware replacement cycles remain major challenges.

It should not be presented as a near-term solution to terrestrial data-center power constraints.

But its architectural significance is profound.

The traditional system says:

generate energy -> transmit electricity -> power compute.

Orbital computing asks:

**generate energy where it is abundant -> perform compute there
-> transmit information.**

That is the extreme form of the PROMETHEUS principle:

When moving information is easier than moving energy, move the computation.

It belongs in our frontier category precisely because it expands the possible geography of intelligence infrastructure.

27. The Energy Portfolio Becomes Programmable

FIGURE 6 | THE PROMETHEUS 1-GW CAMPUS

**GRID + FIRM + VARIABLE GENERATION -> CAMPUS MICROGRID <->
STORAGE -> ENERGY ORCHESTRATION <-> COMPUTE ORCHESTRATION ->
AI COMPUTE + COOLING <-> HIGH-CAPACITY COMMUNICATIONS / OTHER
CAMPUSES**

We can now assemble the pieces. A future AI campus may simultaneously have access to grid power, dedicated firm generation, renewables, battery storage, long-duration storage, distributed generation, flexible compute, advanced cooling, and connections to other campuses.

Historically, these would have been managed as separate systems. PROMETHEUS proposes that they become a coordinated system.

An Energy Orchestration Layer continuously evaluates electricity prices, grid conditions, weather, generation availability, battery state of charge, cooling conditions and reserves.

A Compute Orchestration Layer evaluates compute demand, latency, deadlines, hardware requirements, data location and workload priority - determining what must run now, what can wait, what can accelerate, and what can move.

The two layers interact continuously. Energy becomes aware of computation. Computation becomes aware of energy.

That is the reinvention.

28. The PROMETHEUS Energy Rule

Movement II can now reduce the entire energy argument to one principle:

Generate what is needed. Store what is economical. Move what can move. Delay what can wait. Protect what cannot fail.

That is a far more resilient architecture than designing a 1-GW computing campus around one fuel, one grid connection, or one prediction about the future.

It also gives us a technology-neutral way to evaluate every new energy development. When a better battery appears, ask where it fits. When a better reactor appears, ask where it fits. When geothermal improves, ask where it fits. When compute becomes more efficient, ask how much infrastructure it avoids. When orbital compute becomes credible, ask which workloads justify it.

The architecture does not need to predict the future correctly. It needs to remain useful when the future arrives.

The PROMETHEUS View

The Intelligence Age does not require one energy revolution. It requires an energy architecture capable of absorbing many revolutions. Solar will improve. Storage will improve. Nuclear may become more modular. Geothermal may become more geographically flexible. Gas may serve as a bridge. Cooling and processors will improve. Some technologies will disappoint. Others we have barely considered may emerge.

The correct response is not to select a single winner in advance. It is to design intelligence infrastructure so the winning technologies can enter when they prove themselves.

It also gives us a technology-neutral way to evaluate every new energy development.

When a better battery appears, ask where it fits.

When a better reactor appears, ask where it fits.

When geothermal improves, ask where it fits.

When compute becomes more efficient, ask how much infrastructure it avoids.

When orbital compute becomes credible, ask which workloads justify it.

The architecture does not need to predict the future correctly.

It needs to remain useful when the future arrives.

The PROMETHEUS View

The Intelligence Age does not require one energy revolution.

It requires an **energy architecture capable of absorbing many revolutions.**

Solar will improve.

Storage will improve.

Nuclear may become more modular.

Geothermal may become more geographically flexible.

Gas may serve as a bridge.

Fuel cells may find specific niches.

Cooling will improve.

Processors will become more efficient.

Flexible workloads will become more sophisticated.

Some technologies will disappoint.

Others we have barely considered may emerge.

The correct response is therefore not to select a single winner in advance.

It is to design the intelligence infrastructure so the winning technologies can enter when they prove themselves.

GOLD SYNTHESIS | THE MACHINE

BENEATH THE MACHINE

Artificial intelligence is not weightless. Behind every model is a computer; behind the computer is an electrical system; around it are cooling, communications, buildings and land; behind all of them are manufacturing, capital, workers and communities.

The Intelligence Age therefore cannot be built by the technology industry alone. It will require energy companies, utilities, semiconductor manufacturers, telecommunications providers, construction firms, equipment makers, capital markets, governments, universities, skilled trades and communities.

PROMETHEUS does not attempt to predict exactly which technology wins. It proposes an architecture capable of incorporating the winners. The campus becomes a long-lived platform designed around the assumption that short-lived technologies will repeatedly change.

The conventional data center asks: **Where will we find enough electricity?**
The PROMETHEUS campus asks: **How do we continuously find the most effective combination of energy, storage, computation, communications, geography and time?**

That is the transition from artificial intelligence as software, to artificial intelligence as industry, to artificial intelligence as infrastructure.

**MAXIMUM USEFUL INTELLIGENCE PER UNIT OF
ENERGY AND CAPITAL.**

NOTES & SOURCES | AUGUST 2026

GOLD AUDIT

[1] International Energy Agency, *Energy and AI - Executive Summary*. IEA projects global data-center electricity consumption to more than double to about 945 TWh by 2030 and identifies grid bottlenecks, power-aware siting, and more flexible operation of servers/on-site resources as material issues.

[2] Federal Energy Regulatory Commission, *FERC Launches Aggressive Targeted Action to Speed Large Load Integration*, June 18, 2026. FERC issued show-cause orders to all six regional grid operators under its jurisdiction concerning rules for connecting data centers and other large loads and explicitly used the phrase speed-to-power.

[3] U.S. Department of Energy, National Nuclear Security Administration, *NNSA Selects Amentum for AI Data Center and Energy Project at Savannah River Site*, July 20, 2026. The proposed project pairs a 1-GW AI data center with approximately 2 GW of on-site generation, described as natural gas bridging to nuclear; selection was for negotiations and is not a final lease award.

[4] U.S. Energy Information Administration, 2026 capacity-additions outlook. EIA expects approximately 86 GW of new U.S. utility-scale capacity in 2026, including about 43.4 GW solar and 24 GW battery storage.

[5] Williams et al., *Power-Flexible AI Data Centers: A New Paradigm for Grid-Responsive Compute*, June 2026 preprint. A real-world 130-kW GPU-cluster deployment demonstrated rapid load reduction, sustained curtailment, and geographically distributed workload shifting while preserving priority service levels. This is emerging research, not yet a universal operating standard.

[6] Liang, *Inference as Flexibility: Ramp Management for Transmission-Connected AI Data Centres*, June 2026 preprint. A proof-of-concept simulation evaluates coordinated inference flexibility and battery storage for ramp management. It is cited as emerging research and not as established commercial practice.

[7] Chaudhary, Ben-Idris and Mitra, *Large-Load Demand Flexibility as Virtual Storage*, July 2026 preprint. The paper develops a mathematical virtual-storage equivalence for flexible large loads. PROMETHEUS uses the broader conceptual distinction that delayed computation is not literally stored electricity, even when both can reduce immediate grid demand.

SOURCE DISCIPLINE

Confirmed fact: sourced to primary government, international-agency, technical documentation, or clearly identified research.

PROMETHEUS analysis: organizing concepts such as The Two Clocks, Moving Bottleneck, power-aware geography, Useful Intelligence per Unit of Energy and Capital, and the integrated 1-GW campus are presented as analytical frameworks.

Forward-looking architecture: modular nuclear scaling, geographically mobile workloads, orbital compute, and the Infrastructure Operating System are scenarios/proposals whose feasibility depends on technology, economics, regulation and deployment evidence.

PROMETHEUS RESEARCH SERIES

WHITE PAPER No. 1

INTELLIGENCE BECOMES INFRASTRUCTURE

Designing the Energy-Compute Architecture of the Intelligence Age

Next: WHITE PAPER No. 2

POWERING INTELLIGENCE

The New Energy Architecture for AI

WP-02 will deepen the energy portfolio introduced here: nuclear and modular reactors, natural gas and bridge generation, advanced geothermal, renewables, batteries and long-duration storage, distributed generation, grid architecture, emerging energy technologies, and the economics of firm power for the Intelligence Age.

David W Tietgen

PROMETHEUS Research Series | August 2026