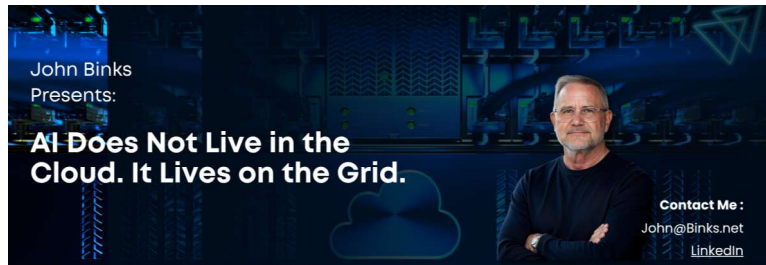


-  Edit article
-  View stats
-  View post



AI Does Not Live in the Cloud. It Lives on the Grid.



John Binks, PMP®, AWS-CCP, AMA-CPM 
Award-Winning / Best-Selling Author | Federal Growth &
Business Development Executive | Federal Systems Integrato...



July 8, 2026

For the last few years, most of the AI conversation has focused on models, chips, prompts, agents, copilots, automation, and productivity.

Those things matter.

But the next major AI constraint may be far more physical.

Electricity.

AI does not really live in the cloud. It lives in data centers, substations, transformers, transmission corridors, cooling systems, fiber routes, interconnection queues, control rooms, field trucks, permitting offices, and utility planning models that must keep the lights on as demand changes faster than traditional planning cycles can handle.

That is the practical leadership issue hiding under the AI hype. AI is not just a software story anymore. It is becoming an energy, infrastructure, security, and national competitiveness story.

The electric grid is the invisible infrastructure underpinning modern life, and it is now under strain from AI, electrification, data center growth, renewable integration, aging assets, and workforce constraints.

AI and the grid are now one strategic conversation, not two separate technology conversations.

The plain English version

The power grid was built around a simpler operating model. Large power plants generated electricity. Transmission lines moved it long distances. Distribution systems carried it to homes, businesses, bases, hospitals, factories, schools, data centers, and government facilities. Demand moved in patterns utilities could forecast with reasonable confidence.

That world is changing.

AI data centers can add enormous new electrical load in concentrated locations. Electric vehicles, building electrification, semiconductor manufacturing, industrial reshoring, advanced computing, and new cooling requirements are also adding pressure. At the same time, the supply side is becoming more diverse. Wind, solar, batteries, nuclear, natural gas, hydropower, demand response, microgrids, geothermal, and distributed energy resources are all part of the mix.

That makes the grid less like a one-way machine and more like a complex, digital, constantly shifting operating environment.

That is where AI enters the equation from both sides. AI can help utilities forecast demand, inspect assets, model grid scenarios, detect anomalies, optimize maintenance, accelerate engineering design, manage distributed resources, and support faster decision-making. But AI also creates new electricity demand through the data centers needed to train, tune, and run AI systems.

That is the paradox.

AI is both a new load on the grid and a tool for modernizing it.

Why this matters now

The demand signal is no longer theoretical.

Lawrence Berkeley National Laboratory reported that U.S. data centers consumed about 4.4 percent of total U.S. electricity in 2023. The same report estimated that data centers could consume between 6.7 percent and 12 percent of total U.S. electricity by 2028, depending on broader economic and data center growth. It also found that U.S. data center electricity use climbed from 58 terawatt-hours in 2014 to 176 terawatt-hours in 2023, with a possible increase to between 325 and 580 terawatt-hours by 2028.

The International Energy Agency projects that global data center electricity consumption could double to around 945 terawatt-hours by 2030 in its base case. The IEA also noted that data center electricity demand increased sharply in 2025 and that AI-focused facilities are growing faster than overall electricity demand.

NERC, the North American Electric Reliability Corporation, has been direct about the risk. Its 2025 Long Term Reliability Assessment stated that summer peak demand is forecast to grow by 224 gigawatts over the ten-year assessment period, a major increase over the prior outlook, with new data centers for AI and the digital economy accounting for most of the projected increase.

DOE has also moved the issue into the policy lane. Its Speed to Power initiative aims to accelerate large-scale transmission and generation projects so the grid can support AI, manufacturing, industrial growth, and broader demand for reliable and affordable energy.

FERC has been pulled into the same conversation. The commission directed PJM, the largest U.S. grid operator, to establish clearer rules for AI-driven data centers and other large loads that are co-located with generation facilities. The stated goal was to enable service to large loads while protecting reliability and consumers in a region serving more than 67 million people across 13 states and Washington, D.C.

For business leaders, government contractors, technology teams, utility executives, and public sector planners, this is not just an energy issue. It is a strategy issue.

The ability to deploy AI at scale will increasingly depend on access to power, grid capacity, permitting speed, energy procurement, resilience planning, cybersecurity, and credible sustainability decisions.

AI infrastructure is becoming physical infrastructure.

For years, many executives treated cloud computing as a procurement choice. Pick a provider. Sign the agreement. Manage consumption. Move on.

AI makes that view too narrow.

Large-scale AI requires physical sites, electrical interconnections, cooling, backup power, fiber, trained labor, security, land, water in some cases, and utility coordination. That means AI strategy now intersects with real estate, public policy, energy markets, environmental review, community acceptance, emergency management, and resilience planning.

This is especially relevant for government contractors and federal technology leaders. National security analytics, cyber operations, autonomous systems, intelligence processing, digital engineering, scientific computing, emergency response, and mission AI all depend on reliable power.

A highly capable AI system is not very useful if the infrastructure behind it is brittle, delayed, poorly governed, or exposed to avoidable operational risk.

This changes the executive question.

The old question was, what AI capability do we want to build?

The better question is, what physical, digital, energy, security, and workforce conditions must exist for that capability to perform reliably when it matters?

The grid needs AI, but not blind automation.

I am bullish on AI for grid modernization, but not because I want a fully autonomous grid making opaque decisions no one can explain.

The better model is AI-assisted human decision-making.

Think of AI as a field assistant, an engineering accelerator, a forecasting engine, a pattern-detection system, a maintenance planner, and a scenario modeler. It can review imagery, flag damaged assets, identify vegetation risk, help prioritize crews, detect anomalies, support compliance review, and help engineers test options faster.

Examples include machine vision for pole inventory and AI assistants that help apply technical standards. That matters because grid modernization is not one giant project. It is millions of asset decisions across thousands of locations.

Moving from slow manual inspection to faster AI-supported workflows can accelerate modernization. It can also help capture institutional knowledge at a time when many senior engineers, planners, and field experts are nearing retirement.

That does not mean removing human judgment. It means giving human experts better tools, better context, better models, and better time to focus on the decisions that matter most.

In grid operations, trust is not a slogan. It is an operating requirement.

AI used in grid planning and operations must be explainable, tested, monitored, secured, and governed. The acceptable error rate for a marketing draft is not the same as that for a grid control recommendation.

Load flexibility may become as important as load growth

A common mistake is to view data centers only as massive fixed loads. Some workloads do need continuous low-latency performance, but not every AI workload has the same urgency.

Training, batch processing, testing, synthetic data generation, simulation, model evaluation, and some inference workloads may have more flexibility than traditional grid planning assumes.

That creates an important opportunity.

AI infrastructure can become more grid-aware. Workloads can be scheduled when power is cleaner, cheaper, or less constrained. Some workloads can be shifted across regions. Data centers can participate in demand response where appropriate. Batteries and backup systems can be managed more intelligently. Cooling strategies can become more predictive. Compute orchestration can consider grid stress as one of the operating signals.

This is not magic, and it is not a universal answer. Mission-critical workloads and low-latency services will still require reliable performance. But it changes the discussion.

The question is no longer only "how much power does AI need?"

The better question is, how flexible can AI demand become without harming mission performance, service quality, security, or customer trust?

That one question could separate thoughtful AI infrastructure strategies from expensive grid headaches.

Energy resilience is now part of AI resilience.

Most AI risk discussions focus on data privacy, hallucinations, bias, cyber exposure, model governance, intellectual property, and workforce impact. Those issues are real.

But AI resilience also depends on energy resilience.

What happens if a region faces extreme heat and demand spikes? What happens if a data center cluster grows faster than transmission upgrades? What happens if transformer supply chains delay expansion? What happens if local communities resist new generation or water use? What happens if backup fuel logistics are not realistic during an emergency? What happens if power quality, not just power quantity, becomes the constraint?

These are not abstract questions. They affect uptime, cost, mission assurance, customer commitments, and public trust.

The grid is not just behind the wall outlet anymore. For AI, the grid is part of the operating architecture.

Regulation and market design are now part of AI strategy

Power availability is becoming a competitive factor in AI. It is also becoming a consumer protection and reliability issue.

Large loads raise hard questions. Who pays for grid upgrades? Which projects get priority? How should interconnection queues work? How should co-located generation be treated? How should communities be protected from unfair cost shifts? How much flexibility should large loads provide? What transparency should data center developers offer to utilities and regulators?

In plain English, AI growth is forcing a rethink of who pays, who benefits, who gets priority, how risk is allocated, and how fast infrastructure should move.

Technology companies cannot solve that alone. Utilities cannot solve it alone. Regulators cannot solve it alone. State and local governments cannot solve it alone. Customers cannot solve it alone.

This will require better coordination among hyperscalers, utilities, grid operators, regulators, data center developers, public agencies, defense and intelligence customers, energy providers, local communities, and the contractors who build and secure the infrastructure.

That is why leaders should treat AI energy strategy as a board-level and mission-level issue, not a facilities footnote.

What leaders should do next?

1. Treat power as a core AI planning variable.

Do not build an AI road map that ignores energy availability, facility constraints, interconnection timelines, cooling, backup power, grid stress, and resilience requirements. AI planning should include power planning from the beginning.

2. Ask better questions of cloud and data center partners.

Where is the compute located? How is power sourced? What are the resilience assumptions? What happens during regional grid stress? What workload flexibility exists? What is the cooling approach? What transparency is available around energy use and emissions? What are the outage and failover assumptions?

3. Build cross-functional AI infrastructure teams.

AI strategy now needs input from technology, facilities, security, finance, legal, procurement, operations, sustainability, resilience, and risk management. If energy is not in the room, the AI plan is incomplete.

4. Use AI to improve infrastructure work, not just office work.

Some of the most valuable AI use cases may be in inspection, maintenance planning, engineering design support, asset management, outage prediction, cyber monitoring, vegetation management, construction planning, and operational analytics.

5. Plan for workforce augmentation.

The grid workforce challenge is real. AI can help capture institutional knowledge, support junior engineers, accelerate field workflows, reduce repetitive analysis, and preserve expert judgment. The goal should be to make experts more valuable, not invisible.

6. Grid-design-aware compute.

Segment workloads by urgency, flexibility, geography, data sensitivity, and mission criticality. Some workloads may be able to move in time or location. Others cannot. Leaders need that map before power constraints become urgent.

7. Strengthen cyber and operational governance.

More sensors, models, APIs, automation, and external data feeds can improve visibility, but they also expand the environment that must be secured. AI used near operational technology needs strict access controls, logging, model validation, incident response planning, and human accountability.

8. Watch the local impact.

National electricity statistics can hide regional stress. A data center cluster in one county can create planning, water, permitting, rate, land use, and reliability concerns that do not show up in national averages. Community trust will matter.

Final takeaway

AI is often described as a form of digital transformation. That is only half right.

The next chapter of AI is also physical transformation. It depends on wires, substations, generation, cooling, construction crews, field engineers, regulators, utility planners, cybersecurity teams, and communities.

The companies and agencies that understand this first will have an advantage. They will design AI strategies that are not only impressive in the lab, but workable in the real world.

And that means leaders who want to win with AI need to understand the grid.

If you enjoyed this article, be sure to share and repost!

Remember, the bots are getting smarter, but you're still the boss! ©

Copyright © 2026 John Binks / Bots & Bosses LLC, All Rights Reserved

#ArtificialIntelligence #AI #PowerGrid #GridModernization #DataCenters
#Energy #EnergyTransition #TechnologyStrategy #BusinessTransformation
#GovCon #Infrastructure #Cybersecurity #AllLeadership
#DigitalInfrastructure #EnergyResilience #BotsAndBosses

Comments



Like



Comment



Share

Add a comment...



No comments, yet.

Be the first to comment.

[Start the conversation](#)



John Binks, PMP®, AWS-CCP, AMA-CPM

Award-Winning / Best-Selling Author | Federal Growth & Business Development Executive |
Federal Systems Integrator (FSI) | AI & Emerging Technology | Public Speaker / Podcaster |
Former FEMA Technology Executive

