cypris

APRIL 2026

Innovation Outlook

Bypassing the Grid:
On-Site Power Generation
for U.S. Data Centers

Prepared By

Cypris Team

info@cypris.ai

Table of Contents

Executive Summary

Hyperscalers aren't waiting for the grid. They’re building their own.

The consensus framing on data center power is that the grid is straining to catch up with AI demand, that interconnection queues will clear, and that behind-the-meter generation is a transitional workaround until utilities can absorb the load. The evidence indicates a different end-state. Across 46 announced projects, hyperscalers and their developers have committed to 56 GW of behind-the-meter generation capacity — approximately 30% of all planned US data center capacity. Natural gas accounts for roughly 75% of verified orders. Meta's Hyperion project alone commits to 7.5 GW of dedicated gas capacity with 240 miles of privately financed transmission, a single project representing more than a 30% increase to Louisiana's entire existing grid capacity. Twenty states have rewritten their large-load tariff frameworks in response, and FERC has issued its first formal guidance on co-location — both regulatory moves that trail the commercial shift rather than shape it.

The emerging architecture is not grid-independent. Hyperscaler patent activity — Microsoft's fuel-cell-integrated cooling systems, Google's unified PEM fuel cell and power distribution designs, Eaton's multi-layer DER control architectures — shows engineers designing for hybrid operation, where on-site generation and grid supply are co-dispatched rather than substituted for each other. When interconnection eventually catches up, the self-generated capacity does not get abandoned. It runs in hybrid mode alongside the grid, which is the architecture the IP record is already describing. The durable shift is not that hyperscalers are leaving utilities behind. It is that they are entering the next decade holding a structurally larger share of their own generation than anyone planning the grid expected.

This report assembles the commercial, financial, and IP evidence into a single ecosystem view: where demand is concentrating, where the supply and enabling layers are constrained, how the bypass is sized and priced, which technologies are absorbing which parts of the buildout, and where the most defensible investable whitespace sits. The analytical payoff of reading those signals together is that the most defensible whitespace is not at the generation layer, which is visibly competitive and already well-capitalized, but in the integration infrastructure that makes on-site power dispatchable, transferable, and operable under data center service conditions. The component blocks are public; the data-center-specific orchestration that ties them together is not.

The Setting: AI's Power Problem and How the Grid Lost

Data centers form the physical backbone of the AI economy - the facilities where compute, storage, and networking converge to train models, support real-time applications, and deliver cloud services at scale. As AI adoption accelerates, so too does the energy required to support it.

U.S. data center electricity consumption increased from 58 TWh in 2014 to 176 TWh in 2023. Current projections estimate demand could reach between 325 and 580 TWh by 2028—equivalent to as much as 12% of total U.S. electricity consumption. Capital investment is scaling in parallel. Microsoft, Amazon, Alphabet, and Meta have collectively forecast approximately $650 billion in capital expenditures in 2026, largely directed toward AI infrastructure, and hyperscaler capital spending including Oracle grew 66% year-over-year in 2024. These reflect long-term structural commitments, not cyclical expansion.

Electric grid infrastructure is not evolving at a comparable pace. The average time from interconnection request to commercial operation now exceeds four years, more than double the 2008 timeline, and broader grid upgrades typically require five to ten years. Transmission expansion operates on even longer horizons, historically designed to accommodate gradual industrial growth rather than rapid, concentrated increases in demand from a single sector.

In response, hyperscalers have increasingly pursued alternatives. Beginning in 2024 and accelerating through 2025, many initiated deployments of generation capacity at or near data center sites, provisioning power directly rather than waiting for interconnection. However, this is not a departure from utilities as long-term partners. Patent activity across major operators shows a focus on orchestration systems capable of dynamically managing power flows between on-site generation and grid supply. The objective is integration and flexibility, not substitution. In the near term though, the priority is speed.

The regulatory response is tracking this shift rather than shaping it. On December 18, 2025, FERC directed PJM Interconnection to revise its tariff to establish clearer rules for large loads co-located with generation assets — a proceeding whose significance lies less in its content than in its timing, with formal frameworks emerging after co-location strategies had already been adopted at scale. Twenty states have approved new large-load rate structures with additional proposals pending, and PJM approved an $11.8 billion transmission expansion plan in early 2026, approximately double the annual approvals in 2023 and 2024.

Individually, these measures are rational responses to emerging system pressures. Collectively, they illustrate a regulatory framework adapting to a demand profile it was not originally designed to accommodate — policy functioning as a lagging indicator of structural change. The reconfiguration of power delivery for data centers is not a prospective scenario. It is already in progress, and the remainder of this report examines how that transition is unfolding across the ecosystem, and where value is being created as a result.

The Data Center Ecosystem

The following visualization (**Figure 1)** illustrates the key functional domains and associated actors across the data center ecosystem. These domains are highly interconnected in practice and should not be interpreted as discrete segments. Reading them together is what makes the structural pressure visible: where demand is concentrating, where the supply and enabling layers are constrained, and where on-site generation is emerging as the release valve.

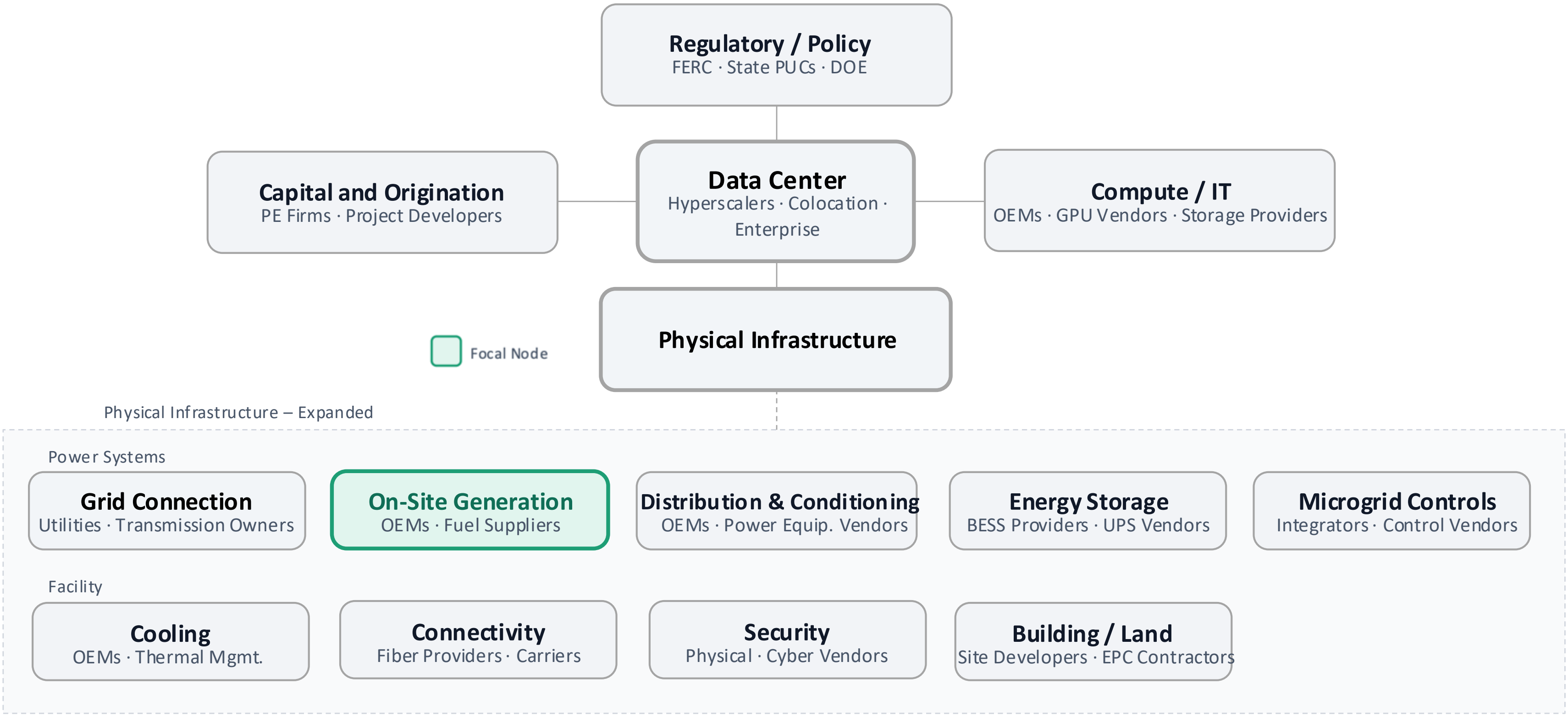


Figure 1: U.S. Data Center Ecosystem Map. On-site generation (highlighted) sits within the Physical Infrastructure layer as the focal node of this report.

Power Ecosystem: Demand, Supply, and Enabling Actors

The following sections trace how pressure is distributed across the ecosystem. Demand is concentrated and accelerating; the supply side is split between constrained incumbents and a fast-mobilizing response layer; and the enabling layer is moving reactively rather than ahead of the curve. That distribution of pressure is what is producing the migration to on-site generation.

Demand Side

Demand is dominated by a cohort of hyperscale operators — principally Amazon Web Services, Microsoft Azure, Google Cloud, and Meta — whose buildouts have reshaped electricity sector planning. These operators function as both tenants and developers, self-building large campuses while also leasing from colocation providers, and their power requirements frequently exceed what local grids can deliver on the timelines their plans require. Their capacity decisions cascade directly into utility planning, transmission investment, and technology procurement across the rest of the ecosystem. The scale is commensurate: Microsoft, Google, and AWS plan to invest more than $500 billion in AI-supporting infrastructure in fiscal year 2026, and in Q3 2025 alone they leased more U.S. data center capacity — a record 7.4 GW — than in all of 2024. Colocation operators including Equinix, Digital Realty, and QTS occupy a structurally similar position on the demand side, absorbing the same interconnection delays while bearing the capital risk of delivering energized capacity on contractually committed schedules.

Supply Side — The Constraint

Electric utilities (Dominion Energy, Duke Energy, Entergy, AEP, Xcel Energy) and the RTOs and ISOs administering the wholesale markets they operate in (PJM, MISO, CAISO, ERCOT) sit at the structural bottleneck of the ecosystem. Utilities are revising long-term capital plans at significant scale and introducing new commercial terms — large-load tariffs, take-or-pay contracts, minimum demand requirements — to manage stranded infrastructure risk, but the pace of buildout remains constrained by regulatory approval timelines, supply chain pressure on transformers and switchgear, and the sheer volume of requests now entering utility planning queues simultaneously. The grid operators responsible for interconnection face the same pace mismatch: median duration from request to commercial operation has doubled from under two years for projects built in 2000-2007 to over four years for those built in 2018-2024. Transmission owners, whose assets must carry whatever new generation does get built, move on even longer cycles, with high-voltage lines typically requiring years of environmental review and right-of-way negotiation before energization. The combined result is that the layer responsible for delivering power to data centers cannot do so on the timelines data center deployment requires.

Supply Side — The Response

A parallel set of actors has mobilized to build around that constraint. OEM lead times diverge sharply across technology categories, and that divergence is itself reshaping procurement: GE Vernova's CEO expects gas turbine reservations to be sold out through 2030 by the end of 2026, while Bloom Energy has committed to delivering fuel cell systems within 90 days of order. EPC contractors and pure-play power project developers translate that equipment into operating assets, absorbing the permitting, fuel supply, and execution risk that neither OEMs nor operators are structured to manage directly — and private capital has mobilized behind them at significant scale, with Blackstone alone committing over $25 billion to digital and energy infrastructure in Pennsylvania and expecting to catalyze an additional $60 billion. Fuel supply is moving in the same direction: EQT Corporation has signed agreements to supply large-scale data center power projects in the Appalachian Basin, and integrated majors ExxonMobil and Chevron have announced dedicated gas-fired generation initiatives explicitly structured to avoid grid interconnection. Microgrid integrators and controls providers close the remaining gap between discrete assets and the automated architecture data centers require: Enchanted Rock's "Bridge-to-Grid" model allows hyperscale data centers to become operational years before permanent utility connections are available, with on-site generation transitioning to backup once the grid connection is established — a direct structural substitute for the interconnection process itself.

Enabling Layer

The regulatory response to AI-driven data center load has accelerated, but remains in a reactive posture. Federal and state regulators are now actively reshaping how large loads interconnect, how co-located generation is permitted, and how infrastructure costs are allocated.

At the federal level, the Federal Energy Regulatory Commission (FERC) governs wholesale electricity markets, interstate transmission, and generator interconnection. Its December 18, 2025 order directing PJM to revise its tariff for co-located large loads is the most consequential single action to date for behind-the-meter viability. It addresses a question commercial actors have already answered in practice: under what conditions can a data center sit adjacent to a dedicated power plant and consume its output directly. PJM is the largest and most grid-constrained RTO, and its resolution is likely to set the template other RTOs adopt. The Department of Energy's 2024 Powering AI recommendations reinforce federal attention but do not materially alter near-term deployment timelines.

State Public Utility Commissions manage the local consequences of this load growth, with actions falling into two categories. On cost allocation, twenty states have approved at least one large-load tariff, with nine more pending — Virginia's GS-5 rate class for users above 25 MW, requiring minimum demand charges to prevent residential cost-shifting, is among the most developed. On infrastructure planning, roughly $4.8 billion of PJM's $11.8 billion transmission expansion is allocated to Dominion's Virginia coverage area.

The Focal Node: On-Site and Behind-the-Meter Generation

The structural dynamics described above converge at a point of sustained tension. The gap between the urgency of AI-driven data center deployment and the multi-year timelines of traditional grid interconnection has elevated on-site and behind-the-meter generation from a contingency measure to a primary strategic option. The technologies, business models, regulatory treatments, and competitive dynamics organizing around this node are evolving rapidly and at scale, drawing in incumbents, new entrants, energy majors, and hyperscalers simultaneously.

Representative Actors Across the Data Center Power Ecosystem

Figure 2 maps the key functional categories and representative actors shaping how power is planned, delivered, and generated for U.S. data centers. It is illustrative rather than exhaustive. Category membership reflects structural role in the ecosystem, not a comprehensive directory of all participants. Reading across categories shows which actors are driving the shift toward on-site generation and which are being reshaped by it.

Figure 2: Representative Actors Across the Data Center Power Ecosystem. Color coding distinguishes demand-side actors (orange), supply-side actors (blue), and the enabling layer (gray).

Market Sizing: Quantifying the Bypass

The bypass is not a forecast. The behind-the-meter migration described previously is already being sized, priced, and contracted across the data center operator landscape — what remains unbuilt is implementation, not commitment. This section quantifies three things: the gap between what has been committed and what is currently operational, the total addressable market for on-site generation serving U.S. AI data centers, and the concentration pattern of that market across both operator segment and generation technology. The analysis is built from 2026 U.S. data center capacity, segment-weighted adoption rates, and source-tracked cost benchmarks for each generation technology.

The Contracted-vs-Installed Gap

Forward-announced behind-the-meter capacity across hyperscale and AI-driven data center projects stood at approximately 56 GW as of February 2026, representing roughly 30% of planned U.S. data center capacity. Against this, the currently installed on-site generation base across the broader U.S. data center market — built from disclosed adoption rates applied to operating and under-construction capacity — is approximately 11.4 GW (**Figure 3)**. The 44.6 GW difference is the pipeline: contracted, announced, or permitted but not yet energized.

That gap is the analytical frame. Traditional adoption-based models of on-site power deployment extrapolate from installed base forward, and those models systematically understate what is happening because they are calibrated on a pre-2024 regime in which on-site generation at hyperscale scale was uncommon. The 56 GW figure describes a different regime — one in which hyperscalers are configuring their power procurement around on-site generation as a primary strategy rather than as backup. The correct way to read these two numbers together is not as competing estimates of the same quantity but as a step-change in behavior that has already occurred and is now being executed.

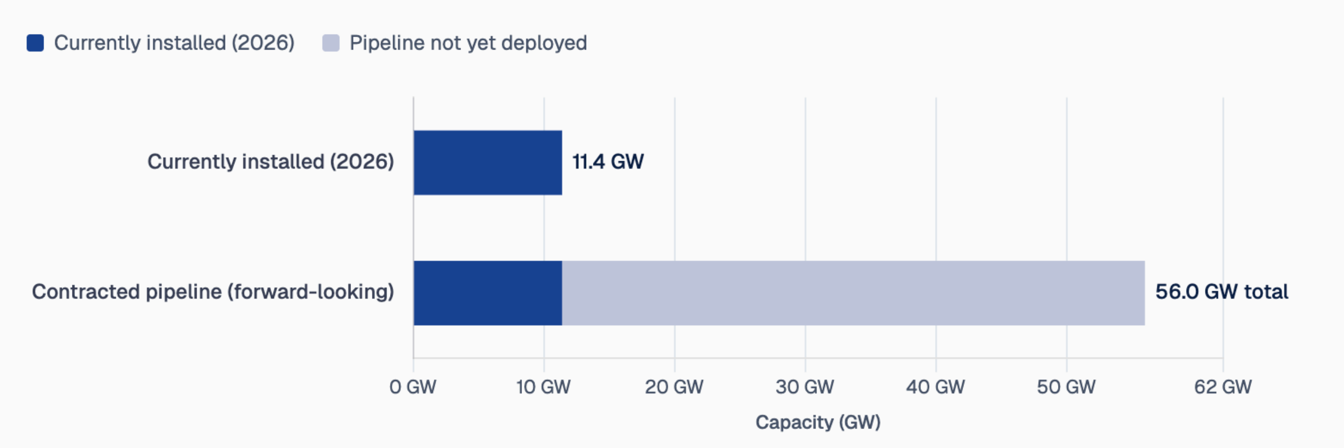


Figure 3: Contracted vs. Installed Behind-the-Meter Capacity (2026). The 44.6 GW gap between current installations and the contracted pipeline reflects a step-change in hyperscaler power strategy, not incremental growth.

Building the TAM

The U.S. data center base against which on-site generation is sized stood at approximately 98 GW in 2026, comprising roughly 46 GW operating and 52 GW under construction per the Aterio U.S. Data Centers Dashboard (April 2026). This figure is cross-referenced against Bloom Energy's 2026 citation of McKinsey data showing U.S. data center IT load at approximately 80 GW in 2025 scaling to 150 GW by 2028, which translates to approximately 100-130 GW of facility capacity after applying typical PUE adjustments. The 98 GW bottom-up count sits within this range and is used as the base for segment-level sizing.

On-site generation adoption is not uniform across operator segments. The analysis applies segment-specific adoption rates derived from the Aterio data sample (n=200, April 2026), which shows 22.7% BTM adoption at AI-flagged facilities versus 5.1% at non-AI or unclassified facilities. Blended against segment-level AI workload intensity — 45% for hyperscale, 20% for colocation, 12% for enterprise, 7% for edge — this yields forward-looking on-site adoption rates of 13.1% (hyperscale), 8.6% (colocation), 7.2% (enterprise), and 6.3% (edge). The blended average of 11.6% is consistent with Aterio's 9% current disclosed adoption (likely undercounted due to disclosure lag) and with Bloom Energy's stated 10-15% near-term projection.

Applying these adoption rates to segment capacity and multiplying through segment-blended installed costs ($/MW, derived from EIA AEO 2026 generation cost assumptions weighted by segment-specific technology mix) yields a total installed TAM of approximately **$35.8 billion** for on-site power generation serving U.S. data centers. Applying a 65% equipment share of total installed cost — separating hardware from EPC and soft costs — yields an **equipment-only TAM of approximately $23.3 billion.**

Market Concentration

The TAM is heavily concentrated in the hyperscale segment. Hyperscalers account for approximately $27.3 billion of the $35.8 billion installed market, ~76%, driven by the combination of the largest segment share of total U.S. capacity (70%) and the highest on-site adoption rate (13.1%). Colocation operators represent the second-largest share at approximately $6.5 billion (18%), reflecting a combination of more moderate adoption rates and the structural position colocation providers occupy as intermediaries serving hyperscaler tenants with grid-independent power requirements. Enterprise ($1.9 billion, 5%) and edge ($0.2 billion, <1%) are commercially marginal at the current buildout stage.

	Total MW	% On-Site	Install Base (Millions USD)
Hyperscale	68,600	13.1	$27,279
Colocation	19,600	8.6	$6,461
Enterprise	8,820	7.2	$1,921
Edge	980	6.3	$166
			$35,828

The concentration pattern has two implications for how this market should be read. First, on-site generation is effectively a hyperscaler-driven phenomenon with secondary colocation exposure; enterprise and edge segments are not where the commercial opportunity compounds. Second, the segment concentration mirrors the concentration of grid interconnection pressure — the same operators facing the longest interconnection queues and the largest individual load requests are the ones deploying on-site generation at the highest rates. This is a confirmation signal that adoption is constraint-driven rather than preference-driven.

The technology mix across the TAM shows gas turbines capturing approximately $22.8 billion (63.7% of installed TAM), fuel cells $6.4 billion (18.0%), solar-plus-storage $3.6 billion (10.1%), and reciprocating engines $2.9 billion (8.2%). This distribution converges with an independent measurement from a third-party dataset, which found that natural gas equipment accounts for approximately 75% of verified behind-the-meter generation orders by capacity. The two numbers are sliced differently — one is bottom-up by dollar-weighted TAM, the other is a ground-truth count of contracted GW — and their convergence from different directions provides strong evidence that gas is the structural majority of the current buildout rather than an artifact of measurement.

The reason gas dominates is mechanical, not preferential. Historically, behind-the-meter deployment in data centers was dominated by diesel backup and limited fuel cell installations, with the grid serving as primary supply. Primary natural gas generation at hyperscale scale was uncommon before 2024. The regime change visible in the 2025-2026 data — gas as dominant primary generation at hyperscale — reflects what is deployable in the 2026-2030 window under current OEM lead times, not a long-run preferred mix. Hyperscaler net-zero commitments remain publicly in force (Google and Meta targeting 2030, Microsoft carbon-negative by 2030, Amazon by 2040), implying that the current gas-heavy mix will require offsets, carbon capture retrofits, or replacement with nuclear and other low-carbon sources in the post-2030 window. Gas dominates not because it is cheapest, but because it is the only technology that can be deployed at scale within the timeline the buildout requires.

The Deployment Window

The deployment-timeline segmentation makes the constraint visible (**Figure 4**). The 2026-2027 near-term window is served by fuel cells, reciprocating engines, mobile gas gensets, aeroderivative gas turbines, and repurposed aircraft or cruise ship turbines — technologies with lead times under 24 months, in some cases under 12. The 2028-2030 mid-term window is where large-frame gas turbines (GE 7HA, Siemens SGT, Mitsubishi M501), new-build combined-cycle plants, and modular gas plants come online, reflecting the 3-4 year lead times characteristic of that equipment class when prioritized (5-7 years for non-priority deliveries). The 2030+ long-term

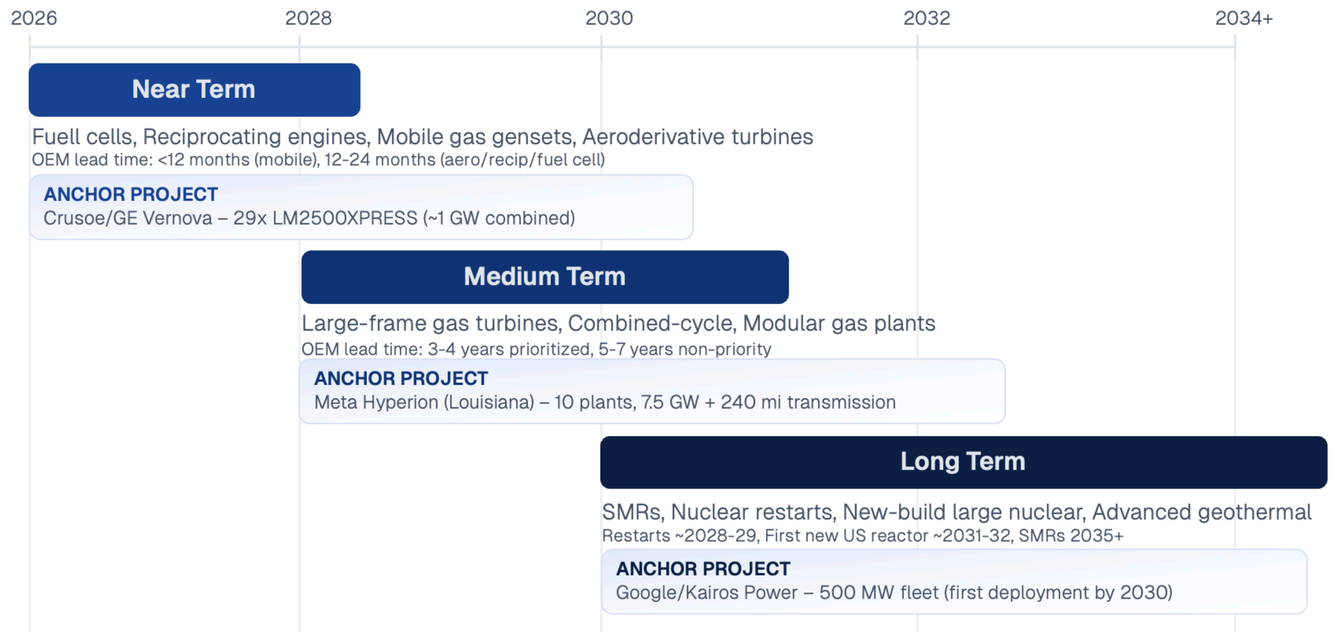


Figure 4: On-Site Generation Deployment Timeline by Technology. OEM lead times determine which technologies can serve each window, with anchor projects illustrating the scale of current commitments.

window is where small modular reactors, restarts of retired nuclear, and new-build large nuclear become commercially available.

Read together with the technology mix in the TAM, the timeline segmentation is what makes the gas-dominance finding hold true. Fuel cells are deploying first because their lead times are shortest, capturing the time-to-energization segment of the market. Gas turbines capture the largest dollar share because they are the only technology capable of delivering gigawatt-scale firm capacity in the window the hyperscaler buildout occupies. SMRs are absent from the near- and mid-term market not because commitments are missing but because their commercial availability begins after the buildout window closes. The shape of the technology mix is a function of the deployment window, not of relative preference.

The supply side is already constrained in a way that validates this timeline reading. Gas turbine deliveries for 2026 and 2027 are largely sold out, 2028 is filling, and new utility-scale gas-fired plants are not expected to come online until approximately 2032 per recent OEM commentary. This means the 56 GW of contracted BTM capacity will not deploy uniformly across the window; it will deploy as fast as supply permits, with the binding constraint on near-term installation being manufacturing throughput rather than demand.

Technology Trends

The technologies now being deployed for behind-the-meter data center power are not being selected on a single axis. Operators are optimizing across three variables that trade off against each other: time-to-energization, scalable capacity, and long-term emissions profile. No single technology currently dominates all three, which is why the market is forming as a stack rather than converging on a winner.

The pattern that has emerged is legible: fuel cells have captured the early lead on time-to-energization and are now deploying as primary generation at data center scale; gas turbines and reciprocating engines are absorbing the bulk of planned capacity because they are the only technology that can deliver firm power in gigawatt-scale blocks on a near-term timeline; and small modular reactors are being procured as long-dated options for the post-2030 window, not as near-term supply. Two cross-cutting enablers — waste heat recovery and power conversion architecture — determine whether any of this generation translates into usable data center capacity at the rack.

The binding constraint across all three pathways is no longer utility capacity. It is OEM capacity, fuel supply, and, increasingly, the electrical architecture that moves power from the generator to the server.

Gas Turbines and Gas Engines: The Scale Layer

Gas-based generation is the largest component of the current behind-the-meter buildout, and the pace at which it is being deployed is the single clearest evidence that the bypass is operating as a coordinated industrial response rather than a collection of opportunistic projects.

Meta's Hyperion project in Richland Parish, Louisiana commits to 10 gas-fired power plants delivering 7.5 GW of dedicated capacity through Entergy, more than tripling the initial three-plant plan approved in August 2025 and representing over a 30% increase to Louisiana's entire grid capacity. Meta is funding the full cost of service, including 240 miles of new transmission lines — effectively financing a parallel power system at its own expense. In Ohio, the Power Siting Board approved a 200 MW behind-the-meter gas plant (the "Socrates South" project) to power a Meta-affiliated data center campus in New Albany, combining turbines from Solar Turbines and Siemens Energy with reciprocating engines from Caterpillar in a fully islanded configuration with zero grid connection.

Zoomed out, a February 2026 analysis identified 46 data centers with 56 GW of planned behind-the-meter capacity — roughly 30% of planned U.S. data center capacity — with natural gas equipment accounting for approximately 75% of verified generation orders. That concentration is a deliberate response to the one variable gas solves uniquely well at this moment: scalable firm capacity deliverable in 2026-2028 rather than 2030+.

The supply-side signal confirms the demand pattern. NextEra Energy's CEO has stated that gas turbine prices have tripled in the last 24 months and that new gas-fired generation facilities cannot come online until 2032. GE Vernova separately reported 50 GW of gas turbines under contract or slot reservation as of April 2025, with 2026 and 2027 largely sold out, 2028 filling up, and commercial activity accelerating for 2029 and 2030 deliveries. The OEM backlog is itself the thesis: the industry collectively decided to bypass rather than wait for the grid, and that decision has now propagated up the supply chain to the point where turbine manufacturing capacity is the binding constraint on the near-term buildout.

Deployment innovation is moving toward pre-packaged, rapidly deployable modules. The GE Vernova LM2500XPRESS delivers 35 MW per unit with 95% factory assembly into simplified modules, with each dual-fuel package capable of starting independently of the power grid in five minutes — Crusoe's combined 29-unit order (nearly 1 GW combined) demonstrates the scalability from single-unit to multi-unit behind-the-meter deployments. Rolls-Royce Power Systems' new 20-cylinder mtu Series 4000 L64 gas engine, available from 2026 for the 60 Hz North American market, delivers its full 2.8 MW output in 45 seconds — a significant ramp-time reduction over prior models, purpose-built for continuous primary power in data center applications. The June 2025 Siemens Energy-Eaton partnership announced a modular power plant concept generating 500 MW from SGT-800 gas turbines with built-in redundancy, battery storage, and Eaton's electrical distribution integrated from medium-voltage switchgear through to the chip — a turnkey architecture designed to enable simultaneous construction of data centers and their on-site power generation.

Waste heat recovery is being integrated into these gas deployments as a margin-improvement mechanism rather than as a standalone trend. A 2025 technical analysis from Rolls-Royce Power Systems found that utilizing waste heat for trigeneration holds a cost-cutting potential of up to 22% for engine-based self-generation configurations compared to power-only deployments. The Fortum-Microsoft project in Finland, where Microsoft data centers will supply approximately 40% of district heating for the Espoo and Kirkkonummi region through 350 MW of recovered thermal power, demonstrates the scale at which thermal integration can operate when the policy and infrastructure environment supports it.

Fuel Cells: The Speed Layer

Fuel cells have captured the behind-the-meter market segment where time-to-energization is the binding variable. Their value proposition is narrower than gas — they are not yet deploying at gigawatt scale in single sites — but on the specific axis of speed, they are the commercially dominant option.

Bloom Energy's supply agreement with AEP for up to 1 GW of fuel cell products (signed November 2024, with an initial 100 MW order) is the largest commercial fuel cell procurement globally to date, with systems designed to be co-located at customer sites behind the meter. Bloom's collaboration with Oracle, announced July 2025, commits to delivering on-site fuel cell power for an entire data center within 90 days — a deployment speed no other generation technology currently matches. Bloom's February 2025 expansion with Equinix surpassed 100 MW across 19 data centers in six states, with approximately 75 MW already operational, making it the most operationally mature fuel cell deployment in the data center sector. Scaling from a 1 MW pilot in 2015 to this footprint took ten years; the pace of new commitments announced since 2024 suggests the next equivalent buildout will take a fraction of that time.

The category is also broadening beyond Bloom. FuelCell Energy, Diversified Energy, and TESIAC announced a strategic partnership in March 2025 to supply up to 360 MW of electricity to three data center locations in Virginia, West Virginia, and Kentucky using natural gas and captured coal mine methane in an off-grid configuration targeting operational power within two years. FuelCell Energy separately introduced in March 2026 a standardized 12.5 MW packaged power block bundling ten proven 1.25 MW modules with shared balance-of-plant infrastructure, alongside plans to expand manufacturing capacity from approximately 100 MW to 350 MW at its Torrington, Connecticut facility — directly targeting the behind-the-meter data center market with a utility-grade product.

Underlying technology advances are extending the commercial envelope. Bloom's solid oxide platform has achieved approximately 60% electrical efficiency on 100% hydrogen with 90% combined heat and power efficiency, preserving fuel flexibility from natural gas to hydrogen as longer-term fuel infrastructure develops.

The patent signal reinforces that fuel cells are becoming a component of integrated data center architecture rather than a plug-in generator. A Microsoft patent (US-12402278-B2) describes a datacenter system in which liquefied gas cools cryogenic compute resources and the resulting vaporized gas is reused as fuel for fuel cells that generate electrical power back to the datacenter load. A Google patent (US-12469861-B2) proposes an integrated power architecture where PEM fuel cells, hydrogen storage, batteries, cooling, and power distribution are designed as a single unified system. Hyperscalers are not simply procuring fuel cells; they are patenting fuel-cell-integrated data center architectures, which suggests the category will remain a strategic layer of the stack rather than being displaced as gas capacity catches up.

Small Modular Reactors: The Hedge

Small modular reactor commitments are real, large, and advancing on the regulatory side. They are also long-dated. The commercial pattern that has emerged is not one of SMRs solving the near-term bottleneck but of hyperscalers buying long-duration options while simultaneously building out gas for the 2026-2030 window.

The scale of commitment is significant. In October 2024, Google committed to a 500 MW fleet of Kairos Power advanced reactors with first deployment by 2030. In December 2024, Oklo and data center developer Switch signed a non-binding 12 GW master power agreement for deployments through 2044. Amazon announced three simultaneous nuclear agreements in the same month, including an investment in X-energy and a partnership with Energy Northwest for up to 960 MW of X-energy SMRs in Washington state. In January 2026, Meta announced deals with three nuclear companies totaling over 6 GW, combining 20-year agreements for 2.1 GW from existing Vistra reactors with 1.2 GW from Oklo's Aurora SMRs (targeting 2030) and funding for TerraPower reactors (targeting 2032).

The regulatory pathway is also moving. On March 4, 2026, the NRC issued the first construction permit for a commercial-scale advanced reactor to TerraPower's Natrium project in Kemmerer, Wyoming — a 345 MW sodium-cooled fast reactor with a molten salt energy storage system that can boost output to 500 MW during peak demand. On February 13, 2026, the NRC issued a 40-year Special Nuclear Material License to TRISO-X, establishing the first new NRC-licensed fuel fabrication facilities in over 50 years and addressing the domestic HALEU fuel supply bottleneck that has constrained advanced reactor deployment timelines. Tennessee Valley Authority submitted the first U.S. construction permit application for GE Vernova Hitachi's BWRX-300 SMR in May 2025, leveraging a multi-utility standard design collaboration that creates a repeatable licensing framework. Oklo's September 2025 Principal Design Criteria topical report was accepted by the NRC for accelerated review, which, once approved, can compress future Aurora license applications.

Read together, these developments signal that advanced nuclear is a genuine pathway, not a speculative one. What they do not signal is that advanced nuclear will meet data center demand in the 2026-2030 window. First-of-a-kind construction risk, HALEU supply ramp, and the fact that none of the named developers has yet operated a commercial-scale advanced reactor mean the announced timelines are targets, not guarantees. The commercial structure of the deals reflects this: they are announced as gigawatt-scale agreements that deploy incrementally over 10-20 years, with optionality for the hyperscaler and execution risk that sits with the developer. The correct way to read hyperscaler nuclear announcements is as strategic positioning for the post-2030 compute layer, while gas and fuel cells handle everything before that.

Power Conversion and Distribution: The Scaling Constraint

Another trend is the one that determines whether any of the above generation translates into usable capacity at the rack. As power densities climb toward megawatt-scale racks, traditional AC-based distribution architectures are reaching physical limits, and the electrical architecture required to move power from a 500 MW on-site plant into a 1 MW rack is itself becoming a scaling constraint.

NVIDIA published a technical blog in January 2026 establishing 800 VDC as the power distribution architecture for next-generation AI factories, naming Eaton, Schneider Electric, and Vertiv as data center power system partners alongside silicon providers including Infineon, STMicroelectronics, and Texas Instruments. The initiative targets 1 MW IT racks by 2027 using NVIDIA Kyber rack-scale systems and addresses the physical scaling limits of legacy 54 VDC distribution, which at 1 MW per rack could require up to 200 kg of copper busbar. A March 2026 IEEE Spectrum analysis quantified the performance gains of transitioning from 415V AC to 800 VDC at scale: 85% more power transmitted through the same conductor size, 45% reduction in copper requirements, approximately 5% improvement in end-to-end efficiency, and 30% lower total cost of ownership for gigawatt-scale facilities.

The commercial response is visible in both funding and product announcements. Amperesand closed an $80 million Series A in November 2025 to commercialize its medium-voltage solid-state transformer platform, with 30 MW of commercial systems scheduled for delivery to hyperscale customers in 2026. SolarEdge and Infineon announced a November 2025 collaboration to jointly design a modular 2-5 MW solid-state transformer building block combining DC-coupled topology with silicon carbide switching, targeting over 99% efficiency for direct medium-voltage to 800-1500V DC conversion. At NVIDIA GTC 2026, Delta Electronics showcased an integrated microgrid solution combining a newly developed solid-state transformer with SOFC on-site generation and all-in-one energy storage, and Eaton debuted the Beam Rubin DSX platform, a pre-engineered end-to-end power and cooling architecture integrated with the NVIDIA Vera Rubin DSX reference design that scales from megawatts to hundreds of megawatts.

The integration infrastructure that makes this electrical architecture deployable is examined alongside other whitespace opportunities in the next section. For the technology trend picture, the relevant point is that the bypass cannot scale past 2027 without it. The $650 billion in committed hyperscaler capex is running on an electrical architecture that is itself being redesigned in parallel.

Why the Mix Looks the Way It Does

The trends above are not converging on a single solution because they are each solving a different piece of the same problem. Gas delivers scale, fuel cells deliver speed, SMRs deliver long-dated optionality, waste heat recovery extracts efficiency margin from whatever generation is deployed, and power conversion determines whether the output reaches the rack. The reason gas has captured the largest share of the current buildout is not that it is the preferred technology — the procurement patterns show hyperscalers deploying all five simultaneously — but that it is the only one that could deliver gigawatt-scale firm capacity in the 2026-2028 window when the buildout had to happen. The shape of the stack is a function of what each technology could deliver on the timeline the industry was working against, not of which technology is best.

Competitive Positioning

The competitive structure of the on-site power market for data centers is legible in the patent data before it is legible in press releases. Reading the IP landscape first — where activity is concentrated, which domains are consolidating around a small set of assignees, and where the innovation base remains diffuse — provides the structural map against which commercial positioning can be read.

What the Patent Landscape Shows

Recent innovation across the patent landscape indicates that on-site data center power is not being pursued as a single hardware solution but as a stack of generation, storage, conversion, and control technologies designed to bypass the latency and capacity limits of utility interconnection (**Figure 5**). The stack includes natural gas turbines and gas-to-power systems, fuel cells and hydrogen-based systems, UPS architectures, and hybrid microgrid assets, with control and orchestration systems coordinating these resources around the data center load.

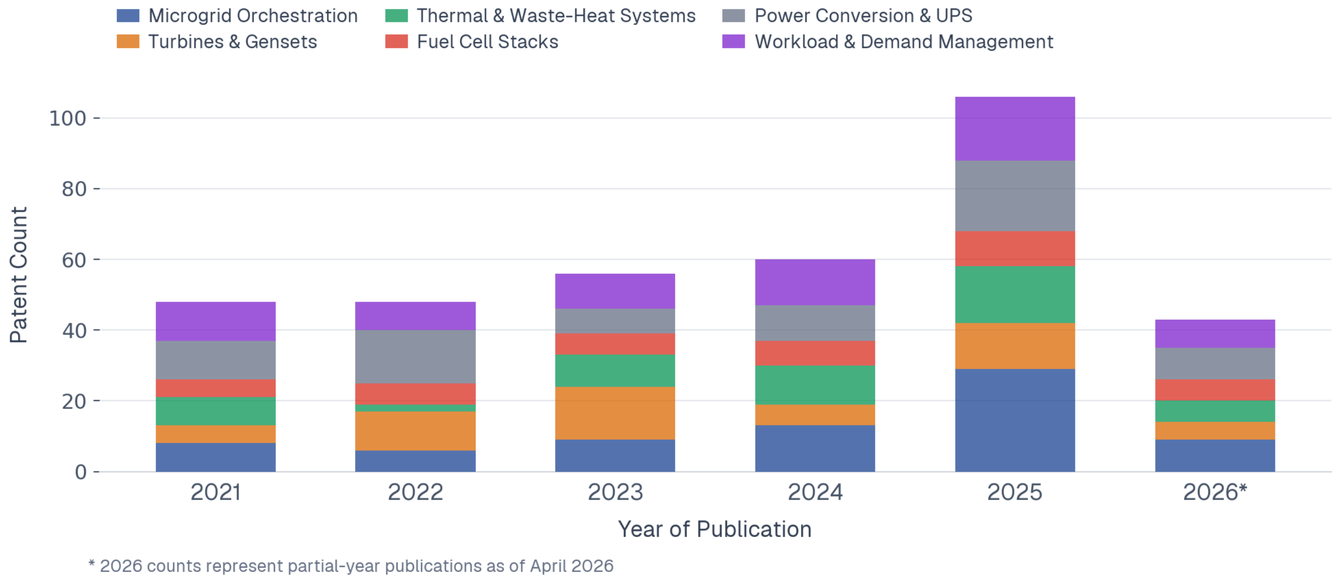
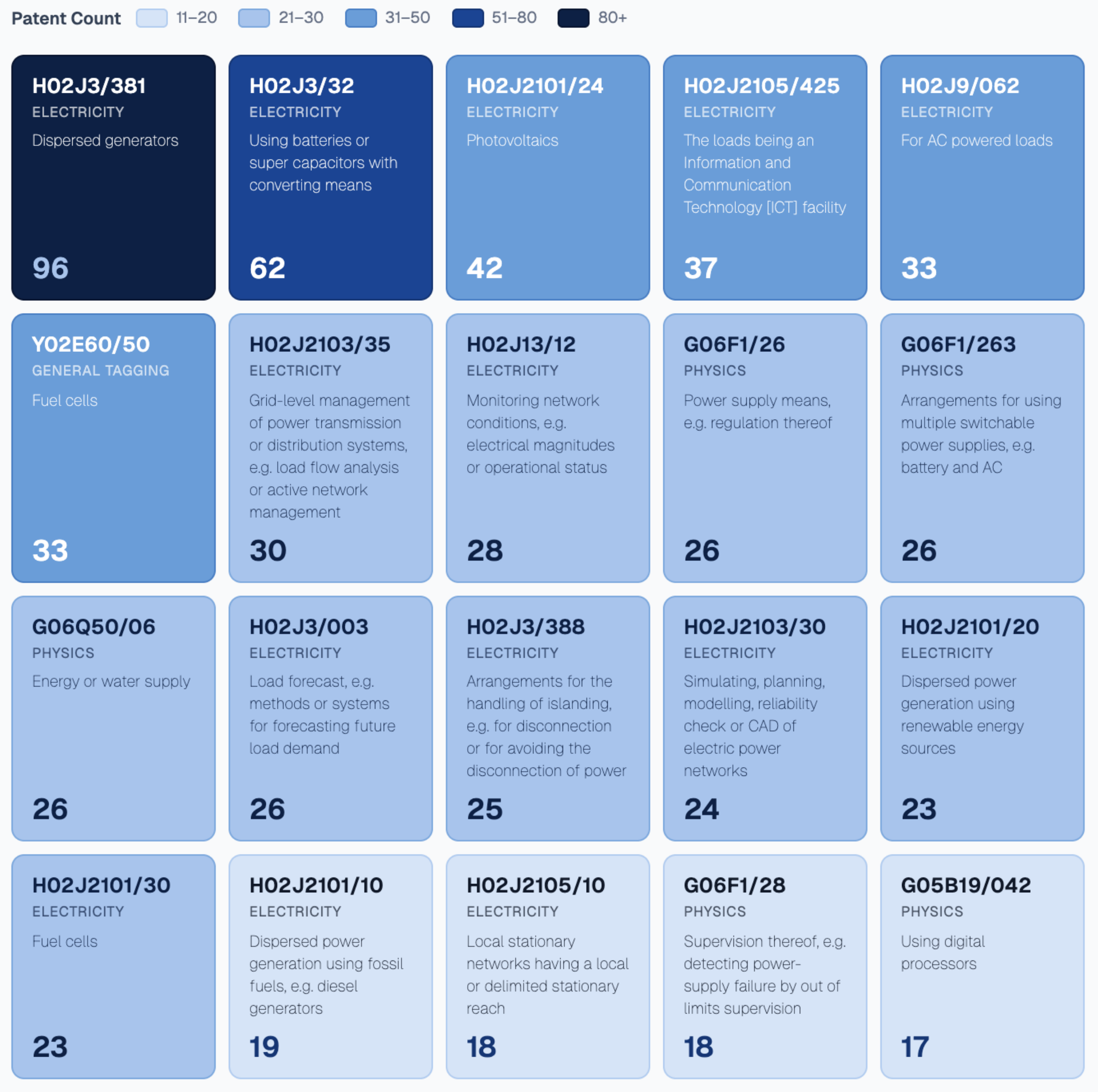


Figure 5: Annual patent publication counts indicate an increase in innovation activity across data center on-site power and operability technologies through 2025, with particularly strong contributions from microgrid orchestration and power conversion and UPS systems.

Within the generation layer, two pathways dominate the IP record. **Natural gas and turbine-based on-site generation** appears across patents describing mobile electric power generation trailer systems, gas turbine power generation devices, clean-power gas turbine combustion systems, and natural-gas-to-power systems that feed distributed computing units. These systems convert local fuel availability into immediate electrical capacity without waiting for grid upgrades or transmission buildout.

Fuel cells for on-site generation — particularly solid-oxide fuel cell architectures — constitute the second concentrated pathway, with patents emphasizing modular fuel-cell cabinets, fuel-cell backplanes, distributed fuel-cell power supply cabinets, and integrated designs that co-locate generation with compute and cooling. The practical value in both cases is the same: compact, dispatchable on-site sources that can be scaled in modules and paired with storage to smooth output. Energy storage and backup supply, while not the focus of this analysis, appear throughout the landscape because storage is frequently embedded in on-site power systems via backup power, transition management, and the control layers of on-site energy distribution (**Figure 6**).

These assets are often integral to making on-site generation operationally usable — absorbing transients, bridging startup delays, and maintaining uptime during source changes.



Patent Count 11–20 21–30 31–50 51–80 80+

H02J3/381 ELECTRICITY Dispersed generators **96**	**H02J3/32** ELECTRICITY Using batteries or super capacitors with converting means **62**	**H02J2101/24** ELECTRICITY Photovoltaics **42**	**H02J2105/425** ELECTRICITY The loads being an Information and Communication Technology [ICT] facility **37**	**H02J9/062** ELECTRICITY For AC powered loads **33**
Y02E60/50 GENERAL TAGGING Fuel cells **33**	**H02J2103/35** ELECTRICITY Grid-level management of power transmission or distribution systems, e.g. load flow analysis or active network management **30**	**H02J13/12** ELECTRICITY Monitoring network conditions, e.g. electrical magnitudes or operational status **28**	**G06F1/26** PHYSICS Power supply means, e.g. regulation thereof **26**	**G06F1/263** PHYSICS Arrangements for using multiple switchable power supplies, e.g. battery and AC **26**
G06Q50/06 PHYSICS Energy or water supply **26**	**H02J3/003** ELECTRICITY Load forecast, e.g. methods or systems for forecasting future load demand **26**	**H02J3/388** ELECTRICITY Arrangements for the handling of islanding, e.g. for disconnection or for avoiding the disconnection of power **25**	**H02J2103/30** ELECTRICITY Simulating, planning, modelling, reliability check or CAD of electric power networks **24**	**H02J2101/20** ELECTRICITY Dispersed power generation using renewable energy sources **23**
H02J2101/30 ELECTRICITY Fuel cells **23**	**H02J2101/10** ELECTRICITY Dispersed power generation using fossil fuels, e.g. diesel generators **19**	**H02J2105/10** ELECTRICITY Local stationary networks having a local or delimited stationary reach **18**	**G06F1/28** PHYSICS Supervision thereof, e.g. detecting power-supply failure by out of limits supervision **18**	**G05B19/042** PHYSICS Using digital processors **17**

Figure 6: The top twenty Cooperative Patent Classification (CPC) categories in the patent set are concentrated in dispersed generators, power conversion, photovoltaics, information and communication technology (ICT) facility load management, and AC-powered load architectures. The distribution also shows notable activity in fuel cells, network monitoring, islanding arrangements, and load forecasting.

The more important finding — and the one that shapes the competitive picture — is that the enabling layer around the generators is where the innovation base is broadest and most active. Microgrid controllers, predictive power management, islanding logic, load balancing, and bidirectional energy routing appear repeatedly across the patent record. These systems decide when to draw from the grid, when to shift to on-site sources, how to charge storage, and how to preserve stable operation during outages or grid disturbances. They matter because grid interconnection bottlenecks are not only physical; they are operational. Even a permitted interconnect may be unable to support rapid load growth or resilience requirements, and the orchestration layer is what allows operators to manage that constraint.

Power conditioning and conversion hardware — rectifiers, inverters, DC buses, converters, power electronic transformers, and conditioned backup outputs for data centers and computer loads — forms a parallel enabling theme. On-site generation must match highly sensitive compute loads with stable voltage, frequency, and redundancy characteristics, not just raw energy production, and the patent record reflects the engineering work required to make that match. Thermal and systems integration is a third recurring theme: fuel cell and turbine patents for data centers repeatedly couple power generation with heat recovery, cooling, ventilation, and containerized or modular packaging. Several patents illustrate how thermal integration is being embedded directly into power and cooling architecture.

A Southeast University patent (US-11889663-B1) describes an immersion dual-cycle liquid cooling system for data centers that uses both single-phase and two-phase heat-rejection loops, with cooling water routed through a waste heat recovery device after absorbing IT equipment heat — a design explicitly aimed at reducing auxiliary energy consumption and reusing server waste heat. A Mitsubishi Heavy Industries patent (US-12258902-B2) presents a gas turbine cogeneration system in which exhaust heat is recovered via a heat recovery steam generator and directed to external uses, with fuel blending adjusted to maintain electrical output as recovered steam is diverted to downstream loads. These filings are consistent with the broader observation that thermal integration is appearing in the patent record as a system-architecture problem rather than as a standalone recovery mechanism.

Taken together, the patent landscape suggests that the industry is treating grid constraints and interconnection delays as a system design problem rather than a utility problem alone. On-site generation reduces reliance on scarce grid capacity, but its usefulness depends on a broader architecture including local fuel or renewable inputs, modular generation, storage for buffering, controllers for dispatch, and power conditioning hardware that makes the output accessible. The clearest strategic pattern is that these inventions make on-site power deployable, not just possible. Mobile gas turbine systems provide immediate localized capacity; fuel cell systems provide scalable low-emission dispatchable generation; and microgrid control innovations provide the logic that lets these assets function as a resilient power plant rather than as isolated devices.

Key Innovations

Multi-unit fuel cell system with microgrid

Applicant	Publication No.	Published	Filed
HyAxiom, Inc.	US-20250385524-A1	December 18, 2025	June 12, 2024

Summary & Relevance: Scalable multi-unit fuel cell power block with an energy management system capable of islanding from the utility grid to supply microgrid loads continuously. Supports behind-the-meter on-site generation for data centers independent of utility interconnection constraints.
Classification: (H02) Generation; Conversion or Distribution of Electric Power

Fuel cell system architecture for artificial intelligence data centers

Applicant	Publication No.	Published	Filed
Bloom Energy	US-20260074522-A1	March 12, 2026	July 24, 2025

Summary & Relevance: Solid oxide fuel cell architecture purpose-built for artificial intelligence data centers, with hybrid energy buffering to smooth highly variable loads.
Classification: (G01) Measuring; Testing | (H02) Generation; Conversion or Distribution of Electric Power

Multi-layer architecture for control of distributed energy resources

Applicant	Publication No.	Published	Filed
Eaton	US-12046907-B2	July 23, 2024	August 27, 2021

Summary & Relevance: Hierarchical DER control architecture pairs cloud-based forecasting with on-site dispatch controllers to optimize on-site generation across microgrid configurations.
Classification: (H02) Generation; Conversion or Distribution of Electric Power | (G06) Computing; Calculating or Counting | (H04) Electric Communication Technique | (G05) Controlling; Regulating

System and method for intelligent power converter control of fuel cells and other auxiliary power sources

Applicant	Publication No.	Published	Filed
Vertiv	US-12424870-B2	September 23, 2025	November 9, 2022

Summary & Relevance: Intelligent power conversion architecture that centrally manages fuel cell, battery, and grid sources (e.g., managing startup, load sharing, and transient conditions) enabling fuel cells to serve as primary behind-the-meter power for data centers independent of grid availability.
Classification: (H01) Electric Elements | (H02) Generation; Conversion or Distribution of Electric Power

System for sharing loads across generator sets, generator set for providing power, and method of synchronizing generator sets

Applicant	Publication No.	Published	Filed
Cummins	US-12519319-B2	January 6, 2026	December 13, 2023

Summary & Relevance: Safety interlock system for synchronized gensets that automatically disconnects all units from load upon control panel access, providing critical maintenance safety infrastructure for multi-unit on-site generator deployments.
Classification: (H02) Generation; Conversion or Distribution of Electric Power

How the Market Is Organized Around Those Signals

The competitive landscape mirrors the layered stack the patent data reveals: generation, conversion, distribution, and conditioning. At the generation layer, recent momentum has concentrated in vendors that can deliver dispatchable capacity on timelines materially shorter than utility interconnection. Bloom Energy has become one of the most visible fuel-cell suppliers through large hyperscaler-aligned deployments that reinforce fuel cells as a modular prime-power option where speed to power is critical. GE Vernova, Wärtsilä, and other gas-based suppliers occupy a parallel position in the combustion segment, where the value proposition is fast deployment of large blocks of firm capacity. Recent deals illustrate how gas turbines and reciprocating engines are being used as dedicated on-site power for new campuses facing grid delays.

Emerging companies are extending the field in different directions. VoltaGrid is pushing rapidly deployable natural-gas microgrids at campus scale, while HyAxiom represents an earlier-stage fuel-cell entrant explicitly targeting data centers. Farther out on the time horizon, Kairos Power, X-energy, and NuScale form a set of strategic nuclear partners collaborating with hyperscalers such as Google and Amazon — though as established earlier, advanced nuclear remains a long-dated hedge rather than a near-term answer to the current bottleneck.

The enabling layer has become increasingly central because on-site generation only functions as a practical substitute for delayed grid service when controls, electrical architecture, and delivery models allow it to operate as an integrated system. Eaton, Vertiv, Delta, and Schneider Electric compete in this layer by combining generation, switchgear, protection, UPS, storage, and microgrid controls into systems that can island, black-start, load-follow, and synchronize with the grid. Higher-voltage DC is becoming part of the same competitive space since it can reduce conversion complexity between the utility or on-site plant, distributed resources such as batteries, and the rack. NVIDIA continues to promote 800 VDC for high-density AI environments, while Vertiv, Schneider Electric, and Delta position around the conversion, protection, and distribution hardware needed to implement that shift. Adjacent power-electronics suppliers such as SolarEdge and Infineon are also engaged as technologies like solid-state transformers could further ease integration. Hyperscalers and large developers including Google, Microsoft, Meta, and Amazon continue to shape demand across the stack through power procurement, clean-energy targets, and advanced AI infrastructure investment.

The patent clustering analysis reveals an additional pattern about competitive concentration (**Figure 7**). The cluster dominated by fuel cell innovations is one of the most spatially isolated, with concentrated positions held by a few named players including Bloom Energy and HyAxiom — a combination of recent filings and cluster isolation that suggests fuel cells are an emerging area a small set of players are actively trying to defend. In contrast, the overlapping central region where several clusters meet is more spatially diffuse and populated with a wide cast of assignees including hyperscalers and power equipment incumbents. This central region is where much of the integration innovation exists, including hybrid systems combining turbines, hydrogen, and workload-aware control. The competition for fuel cell technologies appears to be consolidating, while the central integration region is where competition remains open.

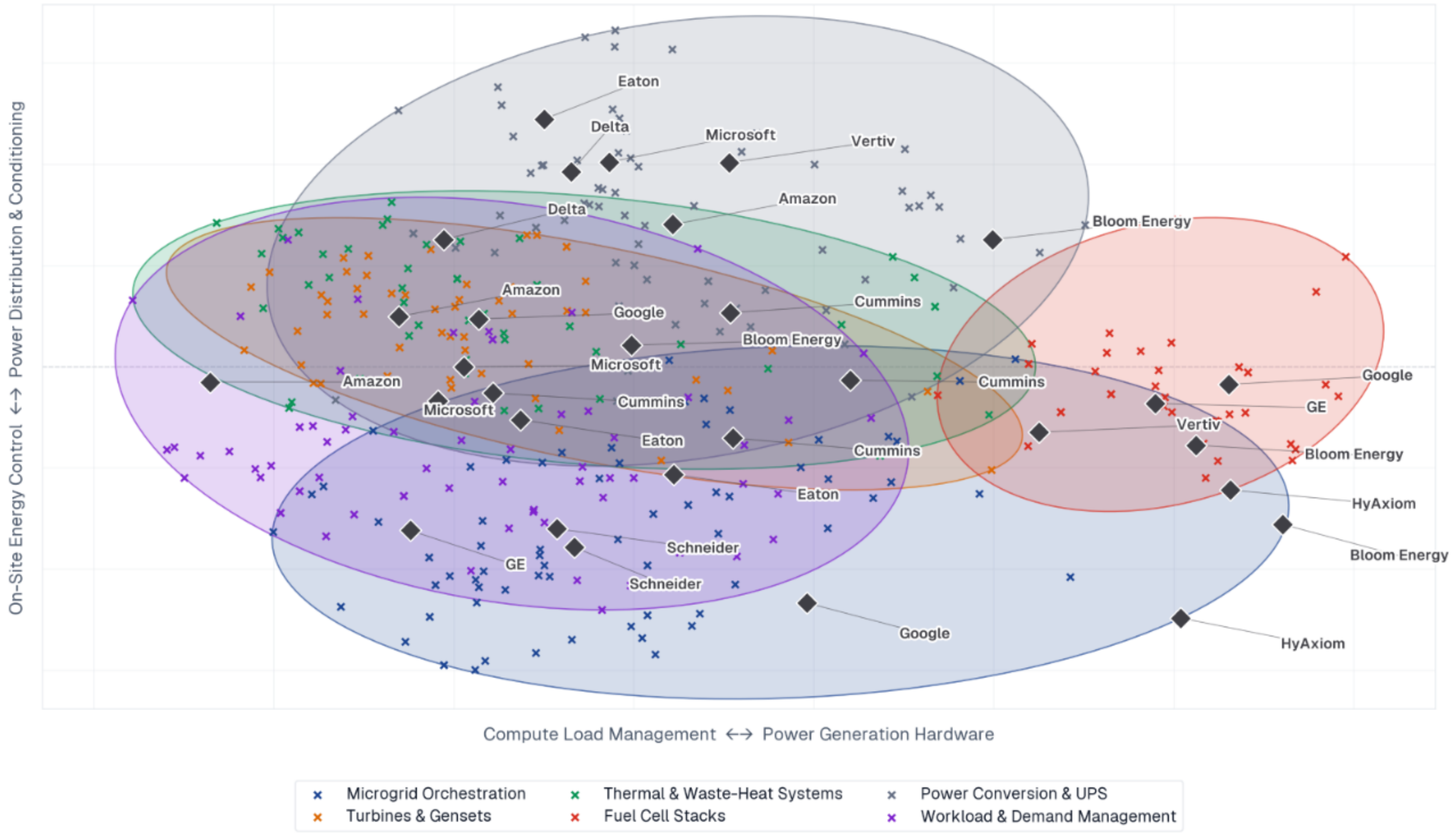


Figure 7: Scatter plot mapping patent activity across six innovation domains in the on-site power ecosystem for data centers. Each patent is represented by an 'x' marker positioned along axes that span Compute Load Management to Power-Generation Hardware and On-Site Energy Control to Power Distribution & Conditioning. Labeled company markers (diamonds) indicate representative portfolio positions of ecosystem leaders (centroids of a company's patents within each cluster). Shaded ellipses illustrate the broader clustering and spread of activity within each domain.

Ecosystem Leaders

This section highlights notable players across the on-site power stack for data centers, spanning generation providers, electrical and microgrid integrators, and power-architecture vendors. These company profiles illustrate how the on-site power market is evolving from standalone backup equipment toward primary integrated systems that combine dispatchable generation, controls, and power conversion. Further, strategic partnerships are becoming central to making on-site power more deployable in response to grid interconnection delays and rising AI load density.

Bloom Energy is one of the commercially established on-site power suppliers in this ecosystem, built around its solid-oxide fuel-cell platform for distributed generation. In July 2025, Bloom announced that it would deploy fuel cells at select Oracle Cloud Infrastructure data centers in the United States. In April 2026, they announced the expansion of the Oracle partnership to as much as 2.8 GW to support AI infrastructure build-out. Bloom Energy also expanded its longstanding Equinix relationship in February 2025 to more than 100 MW across 19 data centers.

HyAxiom is focused on on-site fuel-cell power for data centers and other critical infrastructure. They are an emerging supplier moving from demonstration-scale data-center references (Ilsan Data Center) toward larger commercial stationary-power deployments. Their current products include the PureCell® Model 400 platform and a fuel-cell portfolio designed for hydrogen, natural gas, or dual-fuel operation. In April 2025, the company announced the Charter Oak Combined Heat and Power project, partnering with Scale Microgrids, NuPower, and C.E. Floyd on a multi-megawatt, multi-story fuel-cell installation.

GE Vernova is a leading combustion-based supplier for data centers through its LM2500 family of aeroderivative gas-turbines. In July 2025, GE Vernova and Crusoe announced a 29-unit turbine deal for AI data centers, with the combined order expected to provide nearly 1 GW of electricity. In this ecosystem, they provide large, fast-deployable blocks of firm generation for campuses that need dedicated capacity before utility interconnection can be delivered.

Cummins supplies power generation equipment to the data centers through a broad portfolio that includes diesel and natural-gas generator sets, battery energy storage, transfer switches, hydrogen technologies, and the PowerCommand® microgrid controller for integrated on-site power systems. In March 2025, Cummins announced the QSK60-G29 generator set for data-center applications. Further, their Power Integration Center work is centered around configuring and validating microgrid systems that combine generator sets, PV, battery storage, fuel cells, and transfer switches for data centers. Cummins' role in the ecosystem is to provide a mature bridge between conventional generation and integrated microgrid architecture.

Eaton provides the integration layer to data centers rather than the prime mover itself, with a commercial portfolio spanning microgrid controls, battery energy storage, UPS systems, switchgear, and broader data-center power infrastructure. In June 2025, Eaton and Siemens Energy announced a fast-track approach to building data centers with integrated on-site power, combining standardized modular construction with grid-independent energy supply. They have also tied this strategy to AI factory build-outs through its NVIDIA-linked Beam Rubin DSX platform announced in March 2026.

Vertiv provides critical digital infrastructure and continuity systems for data-center operations. Their commercial portfolio includes UPS, battery energy storage, and integrated microgrid solutions. Recent partnership activity includes a strategic technology agreement with Ballard Power Systems focused on PEM fuel-cell backup-power applications, a case study with EdgeConneX where the Vertiv™ EXL S1 UPS was validated for AI variable loads, and advancing converged physical infrastructure designs for NVIDIA Vera Rubin DSX AI factories. Vertiv's Delaware Customer Experience Center microgrid also demonstrates how fuel cells and advanced controls can be integrated into commercial data-center power architectures.

Delta is positioned in the power management and controls layer for data centers, with commercial offerings in energy storage, microgrid control, hydrogen fuel cells, and higher-voltage DC power systems. In 2025, they introduced an AI data-center microgrid solution built around an energy storage system and microgrid controller that integrates renewables, generator sets, and hydrogen fuel cells. The architecture is designed to handle 100% step-load changes at millisecond response times while maintaining voltage regulation. Separately, Delta is partnering with NVIDIA to link these control capabilities to 800 VDC distribution architectures in next-gen AI factories.

Opportunity Mapping & Whitespace

The clearest opportunity zones in this ecosystem are not at the generation layer but in the **enabling infrastructure** that allows on-site assets to function as a practical substitute for delayed or constrained grid service. Market signals continue to point towards sustained demand for that layer. The IEA expects electricity demand from data centers to keep rising with AI adoption, while grid congestion and long interconnection queues are already putting a meaningful share of planned projects at risk of delay. In parallel, suppliers are commercializing on-site power offerings specifically for data-center deployment. Together, these signals indicate that the market need has advanced beyond backup generation alone and now centers on deployable, controllable, utility-alternative site power.

Against that demand backdrop, the broadest on-site generation categories appear relatively well occupied from an IP standpoint. Fuel-cell power for data centers is no longer a thin field, and hydrogen logistics is not a true blank space at the system level. For example, published patent filings already claim optimization models for storing liquid hydrogen to power data-center fuel cells using fuel-consumption rates, vendor refueling rates, response times, storage-area limits, and logistical refueling constraints.[1] Likewise, generic microgrid control already appears to be a mature product and patenting domain, with major vendors offering pre-engineered control centers or master controllers that integrate generators, storage, utility feeds, and other distributed assets. These facts reduce the defensibility of broad whitespace claims around concepts such as “fuel-cell-powered data centers,” “hydrogen-fueled backup or prime power,” or “microgrid control for data centers” unless the claim scope is especially narrow.

Some **opportunity likely remains in the operability and integration layer** that is specific to data center electrical architecture. Prior art exists on reserve power transfer switches for data centers, and the wider technical literature already treats black start, restoration, and grid-forming inverter behavior as active research and engineering topics.[2, 3, 4] However, the open area is likely narrower and more specific in coordinated control and protection across mixed on-site assets (e.g., generators, fuel cells, switchgear, UPS-supported loads, and grid-forming resources) with restoration and operating logic tailored to data center uptime and staged service conditions rather than generic microgrid objectives. In other words, the component building blocks are already present in the public record, but much of that record appears framed either at the generic microgrid level or around isolated transfer, backup, or local control functions, potentially leaving openings in data-center-specific orchestration and operability.

Another **potential opportunity zone is in the gas-side integration infrastructure**. There is some activity in this area, but the landscape seems much thinner compared to other innovation areas.[5] This suggests more room around gas conditioning, pressure management, gas-to-power conversion at the site edge, and coordinated controls spanning gas supply constraints and electrical dispatch. These sparse areas matter because on-site generation is increasingly being

---

[1] Dhruv GUPTA et al., Hydrogen Fueling and Storage Optimization Model, United States Patent US20240288122A1, filed February 24, 2023, and issued August 29, 2024.

[2] Paul Andrew Churnock et al., Reserve power system transfer switches for data center, United States Patent US10978904B2, filed January 15, 2018, and issued April 13, 2021.

[3] Alston D. Costa et al., System and method for automated clean energy blackstart for backup auxiliary power, microgrid customer loads, and utility grid, United States Patent US20240405566A1, filed June 3, 2024, and issued December 5, 2024.

[4] Abhishek Banerjee et al., Minimum-resource, multiple-microgrid black start driven by grid forming inverters, United States Patent US20260081430A1, filed May 31, 2023, and issued March 19, 2026.

[5] Brent Breon and Christopher Halvorson, Natural gas letdown generator system and method, United States Patent US11761705B2, filed December 2, 2022, and issued September 19, 2023.

positioned as a response to power availability constraints, and recent supplier activity shows the industry moving toward converged infrastructure models that reduce deployment risk and compress time to energization. In practical terms, the most defensible whitespace is likely not another generation prime mover, but integration infrastructure that makes on-site power dispatchable, transferable, and operable under data center service conditions. This becomes especially important where electric interconnection is delayed and gas or multi-asset site infrastructure becomes the critical enabler.

Methodology & Appendix

This research was conducted using a structured, multi-stage approach to identify, validate, and synthesize key market, technology, and innovation signals shaping on-site and behind-the-meter power generation for data centers.

The analysis leveraged the Cypris platform in combination with secondary sources, including company disclosures, industry reports, regulatory filings, patent data, academic literature, and market news. An iterative keyword refinement process was applied throughout to capture evolving terminology across data center infrastructure, grid interconnection, distributed and on-site generation, microgrids, and related enabling technologies. Boolean search strategies were used to systematically identify relevant datasets, ensuring comprehensive coverage while filtering out adjacent but non-relevant domains. For our foundational query, we used Cypris' Boolean searching functionality with the following search terms:

- Boolean Query: ("data center" OR "data centre" OR "comput* facility" OR hyperscale* OR colocation OR supercomput*) AND ("on-site generation" OR "behind-the-meter" OR "distributed generation" OR microgrid OR islanding OR "grid-independent" OR "fuel cell" OR "gas turbine" OR "combined heat and power" OR "reciprocating engine" OR genset OR generator OR "small modular reactor" OR microturbine) AND ("power supply" OR "electricity generation" OR "power generation" OR cogeneration) NOT (vehicle OR EV)

The research process began with landscape-level analysis to establish the overall market context, including the impact of grid constraints, interconnection delays, and accelerating demand from AI-driven data center expansion. From this foundation, the ecosystem was mapped across three functional layers—demand, supply, and enabling actors—reflecting how power is planned, delivered, and managed. On-site and behind-the-meter generation was designated as the focal node due to its central role in addressing deployment timelines and mitigating grid dependency.

Following ecosystem mapping, the research focused on discrete technology trends. Each trend was analyzed across two primary dimensions:

- **Commercial activity**, including funding, partnerships, project announcements, and deployments, to assess market traction and adoption; and
- **Technical advancement**, including improvements in performance, efficiency, scalability, and system integration.

Where relevant, analysis also examined deployment considerations such as modularization strategies, permitting and licensing pathways, and construction requirements, with particular attention to how these factors influence speed to deployment in data center environments.

Patent analysis and targeted review of research publications were incorporated to identify emerging technical directions, system architectures, and longer-term innovation signals. These sources were used selectively, prioritizing relevance and analytical value over completeness, and focusing on innovations directly related to on-site power generation and its integration into data center operations. Adjacent areas such as energy storage, cooling, and downstream power distribution were included only where they materially contributed to enabling on-site generation.

The final stage involved synthesizing insights across all sources, integrating market signals, technical developments, and innovation activity into a cohesive view of the ecosystem. This methodology emphasizes signal over volume, prioritizing credibility, relevance, and direct applicability to understanding how on-site power generation is evolving as a practical response to data center power constraints.

Clustering Analysis Workflow

Patent abstracts were clustered using a text analysis and K-means clustering workflow. The abstract text dataset was cleaned including normalizing whitespace and removing standalone numbers and common measurement expressions. These steps help to reduce potential noise from formatting differences and routine measurement reporting. Then, each abstract was converted into numeric form using Term Frequency–Inverse Document Frequency (TF-IDF) based on individual words and short phrases. Common English words, routine patent drafting terms, general domain-specific terms (e.g., data, center, power, energy, system) were removed, and extremely rare or overly common terms were excluded to focus the analysis on technically meaningful language. The size of the resulting feature set was limited to keep the model stable and interpretable. The TF-IDF matrix was reduced to several latent components using truncated singular value decomposition (SVD) to capture the main patterns of term usage across the dataset. These component vectors were normalized and used as inputs to K-means clustering.

The number of clusters was selected using an inertia-based elbow analysis. Briefly, K-means models were fit across a range of cluster counts, and the within-cluster sum of squares was calculated for each solution. The final number of clusters was chosen at the point of maximum curvature in the inertia curve, where additional clusters yielded diminishing reductions in within-cluster variance. Clusters were interpreted by identifying the most influential terms within each cluster based on average TF-IDF weights. For visualization, the pair of latent components that best separated the clusters in two dimensions were selected and used to generate scatter plots to visualize the resulting patterns.