

Expert Insights

The Coming Wave of Disputes in Data Centre and AI Infrastructure

By [James Barratt](#), Vinson & Elkins

The global data centre sector is in the midst of the largest infrastructure build-out of the modern era. The Goldman Sachs Global Institute projects a cumulative AI infrastructure spend of approximately US\$7.6 trillion between 2026 and 2031, with annual spend rising from around US\$765 billion in 2026 to roughly US\$1.6 trillion in 2031 across compute, data centres and power. BloombergNEF data shows more than 23 gigawatts of IT capacity under construction globally at the end of September 2025. JLL projects that nearly 100 gigawatts of new capacity will come online between 2026 and 2030, effectively doubling the global installed base to approximately 200 gigawatts at a compound annual growth rate of around 14 per cent. By any conventional measure of industrial capital deployment, this is unprecedented. Our own *AI Power & Infrastructure Pulse Survey* of 200 senior infrastructure developers and private capital investors, conducted in March 2026, found that 89 per cent of investors expressed confidence that current capital expenditure levels are sustainable over the next one to five years. At the same time, 96 per cent of developers expect power availability to be the constraint that reshapes the geography of the build-out.

The *leitmotif* of the present market is therefore capital abundance, growth pressure, and a recognition that capacity in terms of raw molecules is scarce relative to demand from the AI build out. In a bull

market, such as the one we are witnessing at present, counterparties presented with operational problems generally renegotiate, accommodate or absorb problems. However, history suggests that in any large, project-financed sector, disputes often arise in the operational phase, generally only after the first material market dislocation — a demand correction, a step-change in input prices, a major regulatory change, a counterparty failure, or a technology shift that renders an asset less economic than the initial financing assumptions. Long-term supply contracts negotiated in a growth market are often tested.

The data centre and AI infrastructure sector is unlikely to escape that pattern. The agreements being negotiated and signed today — typically with 10- to 20-year terms, take-or-pay obligations, and detailed performance commitments — will be enforced through the next downturn, the next change in global policy, and the next generation of hardware.

Beyond construction: Looking to the operational phase

Construction-phase disputes are real and material, but they are also relatively well-charted territory. We have written recently and at length on the construction-phase issues that arise in the energy infrastructure that underpins the data centre sector, including a recent treatment

of natural gas power plant construction disputes covering notice provisions, the documentation of failures, mitigation and acceleration, temporal limitations on claims, multi-contract and multi-forum litigation strategy, first-of-a-kind technology risk, and consequential damages waivers. Colleagues across the firm have addressed broader themes in construction innovation and European data centre development during the current AI boom.

This article looks past the construction phase. The agreements that will dominate the operational phase disputes landscape — and on which the literature is significantly thinner — are the service level agreements (“SLAs”) between operators and tenants, the power purchase agreements (“PPAs”) that underwrite the energy supply, and the long-term capacity commitment and take-or-pay arrangements between hyperscalers and operators. Around those sit a layer of related agreements, including long-term service agreements (“LTSAs”) with equipment manufacturers and the joint venture commitments that increasingly govern the largest assets.

A template? Lessons from LNG and other project-financed sectors

In considering how this dispute landscape will mature, the most instructive analogue, and one that has not yet been widely

drawn, is the liquefied natural gas (“LNG”) sector. Both LNG export facilities and large data centres are capital-intensive, long-term assets whose profitability depends on long-term commitments with creditworthy counterparties. Both rely heavily on take-or-pay mechanics; indeed, market participants in the data centre sector now use the term explicitly. Applied Digital’s two AI Factory campus leases with the same US investment-grade hyperscaler — at Delta Forge 1 (announced April 2026) and Polaris Forge 3 (announced May 2026) — are publicly described in SEC filings as 15-year take-or-pay leases with a combined base-term contracted value of approximately US\$15 billion, rising to approximately US\$36 billion if all renewal options are exercised. Hut 8’s 15-year Beacon Point lease with another US investment-grade hyperscaler, announced in May 2026, is structured on what the parties describe as a “triple-net, take-or-pay” basis at approximately US\$9.8 billion over its base term. Both LNG and data centre offtake produce a recognizable family of disputes: shortfall and delivery failures, force majeure, change-in-law claims, and the vexed question of whether long-term pricing remains in line with prevailing market reality.

The LNG sector also offers a cautionary tale, and one that has continued to play out in the past eighteen months. In the early 2000s, and again in the years following the COVID-19 pandemic, the European gas sector saw waves of price-review arbitrations driven by the simple fact that the pricing provisions in long-term sale and purchase agreements had ceased to make economic sense for one of the parties. More recently, the long-running series of arbitrations brought against Venture Global LNG by Shell, BP, Edison and others — all alleging that Venture Global had wrongfully diverted contracted cargoes to the higher-priced spot market — has produced a striking pattern of divergent outcomes

from materially similar facts. Shell lost its multi-billion-dollar claim in an ICC award in August 2025, and saw a New York court decline to vacate the award in March 2026. BP won the liability phase of its claim in an ICC partial final award in October 2025, with BP reported to be seeking damages in excess of US\$3.7 billion and a quantum hearing yet to be held. Edison reached a commercial settlement in March 2026. That kind of factual, tribunal and counsel-dependent variability, even in a sector with several decades of arbitral jurisprudence behind it, should give pause to any party contemplating long-duration take-or-pay commitments in a much less developed market.

That is precisely the dynamic that we may anticipate in the data centre sector when the unit economics of the AI hardware drift materially away from the assumptions underlying today’s take-or-pay leases. The underlying mechanic is, if anything, already operative at the GPU rental layer: The hourly rental price of an NVIDIA H100 GPU fell from a peak of approximately US\$8 in 2023 to a low of around US\$1.70 in late 2025, before rebounding to approximately US\$2.35 in early 2026 as on-demand capacity sold out across all GPU types. That kind of volatility — a more than fourfold swing in less than 30 months — is the precise condition under which the assumptions underlying today’s 10- to 15-year take-or-pay leases will be tested, particularly when overlaid with continuing GPU generational change (the rapid succession from H100 to H200 to B200 and B300 to the Vera Rubin generation in the same period) that materially alters the unit economics of compute supplied under those contracts.

There are, however, four significant differences that the LNG template does not adequately capture, and each is worth remembering when negotiating resilient contracts today.

First, asset life and technological obsolescence. LNG facilities operate over 20- to 30-year horizons against relatively stable underlying technology. Data centres are exposed to rapid technological change in cooling, power density, GPU generation, and workload composition. The expected shift from training-dominant to inference-dominant AI workloads, anticipated around 2027, is itself a material structural change in what tenants will need from the underlying asset. A 15-year commitment signed today may be exposed to a risk that has no real LNG parallel.

Secondly, counterparty profile. LNG offtake counterparties have historically been national gas companies and large utilities, often with regulated or sovereign-backed demand. Data centre tenants vary considerably. The leading hyperscalers have exceptionally strong investment-grade credit, but the next tier — neocloud platforms, frontier AI laboratories and specialized GPU providers — rely on nascent business models. The dispute profile is therefore bifurcated: hyperscalers reshaping inconvenient legacy contracts are analogous to LNG price reviews, while weaker counterparties trying to escape obligations they can no longer afford are closer to the power disputes of the early 2000s.

Thirdly, asset fungibility and the depth of any secondary market. LNG cargoes are highly fungible and largely trade in a liquid spot market. Data centre capacity, by contrast, is heavily constrained by its physical location, latency, build specification and tenant fit. A defaulted lease cannot easily be replaced, which changes the strategic availability of remedies, mitigation and termination.

Finally, regulatory maturity. The LNG sector has decades of arbitral jurisprudence on force majeure, change in law and price review. The data centre sector has

very little, and the underlying regulatory environment is itself in motion. Tribunals will be writing on a comparatively blank canvas, which makes arbitrator selection, seat choice and governing law of the utmost importance.

SLA disputes

SLAs remain at the heart of the operator — tenant relationships and SLAs are among the most heavily negotiated agreements in the sector. An SLA is the contractual guarantee that a data centre provider makes to its customer, setting out the performance standards that the provider must meet — typically covering uptime, latency, throughput and processing speed. Because even brief periods of downtime can cause significant financial loss and reputational harm, SLAs are among the most heavily scrutinised agreements by the customer before entering into hosting contracts.

Disputes are likely to arise in four recurring areas.

The first is the identification of the responsible party when performance levels fall below the contractual standard. A modern AI-ready data centre is a network of highly interdependent systems supplied and installed by different contractors and OEMs — uninterruptible power supplies, switchgear, chillers, liquid cooling loops, network infrastructure and, increasingly, the GPU servers themselves. When the service drops below the agreed standard, allocating legal responsibility between operator, subcontractor, and OEM is rarely straightforward.

The second is evidence and proof. Disputes will frequently turn on whose monitoring data is authoritative, whether an outage falls within a contractual exclusion (most

commonly scheduled maintenance), and whether the tenant took reasonable steps to mitigate the impact of the outage. Tiered remedy structures, under which service credits depend on the precise classification of the outage, make these evidentiary disputes commercially significant.

The third is the adequacy of contractual remedies. Service credits are often capped at a percentage of monthly fees, an amount tenants will frequently argue is wholly inadequate when set against the downstream commercial harm of an outage. The AI workload context is impactful. A frontier-model training run interrupted by an outage may cost a tenant weeks of compute-time rather than minutes of latency, and the regulatory and reputational fallout from a major customer-facing failure may be substantial. The recoverability of those downstream losses will be constrained by consequential or indirect loss exclusion clauses, the proper interpretation and scope of which will themselves be litigated.

The fourth is an increasing mismatch between traditional SLA architecture and the demands of AI workloads. Many SLAs in current use were designed for general-purpose hyperscale cloud workloads, not for the power-density, thermal and throughput consistency requirements of frontier-model training, which can require an order of magnitude more power density per rack than traditional cloud workloads. As workloads evolve faster than contracts, the underlying assumptions about what constitutes “service” may themselves be a source of dispute.

PPA disputes

Data centres are now one of the largest single drivers of incremental electricity demand globally. The International Energy

Agency estimates that data centres were responsible for around 1.5 per cent of world electricity consumption in 2024, equivalent to roughly 415 terawatt-hours per year, with consumption projected to more than double to approximately 945 terawatt-hours by 2030 under current predictions about AI adoption; in advanced economies, data centres are projected to account for more than 20 per cent of all electricity demand growth in that period. That demand is typically secured through long-term PPAs, structured either between the power generator and the operator (which in turn supplies energy to tenants as part of the lease) or directly between the generator and an anchor tenant such as a hyperscaler.

It seems to us that three categories of PPA dispute are likely to dominate the operational phase.

The first is generation profile risk: The mismatch between when renewable generation is available and when the data centre needs it. This is especially a concern where a PPA is linked to renewable (and intermittent) energy sources. For example, a solar PPA generates nothing at night; a hydroelectric PPA tracks seasonal hydrology; a wind PPA is exposed to multi-year variability. Where the contract does not adequately allocate the financial consequences of this mismatch, disputes will follow.

The second category is curtailment and outage risk. When a generation facility goes offline or is curtailed by the grid operator, the data centre will typically have to procure replacement power at short notice and at a premium. The question of who pays for that replacement power, and whether the supplier is excused under the force majeure or curtailment provisions of the PPA, is a familiar source of dispute in

mature power markets and is likely to be intensely litigated in this one.

The third, and increasingly important, is performance risk on nascent generation technology. As the sector turns to small modular reactors, behind-the-meter natural gas generation, nuclear restarts, and other novel power solutions to escape grid constraints — as illustrated by Microsoft's 2024 20-year, 835 MW power purchase agreement with Constellation Energy for the planned 2028 restart of Three Mile Island Unit 1, by corporate PPAs for early-stage small modular reactor capacity, and by multi-gigawatt behind-the-meter natural gas commitments by AI-focused developers — PPAs are now being signed for technology that has yet to deliver to specification at scale. Disputes over performance shortfall, schedule delay and warranty allocation in this context are likely to look more like construction disputes than power-market disputes, blurring the operational and construction lines.

Capacity-commitment and take-or-pay disputes

A growing share of the long-term contracted revenue in the sector is now structured as take-or-pay capacity commitments. Two broad models are starting to predominate: physical colocation leases under which the tenant commits to take and pay for a defined quantity of MW of critical load over a long term, and GPU-as-a-service or compute reservation agreements under which the tenant commits to take and pay for a defined quantity of compute, generally over shorter terms. Both raise distinct, but overlapping, dispute risks.

The scale and concentration of these commitments is now extraordinary. On

the colocation side, Applied Digital's 15-year take-or-pay leases at Delta Forge 1 and Polaris Forge 3 together commit a single US investment-grade hyperscaler to approximately US\$15 billion of base-term rent (up to approximately US\$36 billion if all options are exercised), and Hut 8's 15-year US\$9.8 billion "triple-net, take-or-pay" lease at Beacon Point commits a different hyperscaler to a similar long-term obligation. On the compute reservation side, CoreWeave's expanded agreement with Meta (now approximately US\$21 billion through 2032), which is in addition to a prior US\$14.2 billion commitment announced in September 2025 and brings Meta's aggregate contracted commitment to CoreWeave to approximately US\$35 billion, its cumulative agreements with OpenAI (approximately US\$22.4 billion) and a series of similar contracts with other neoclouds illustrate the equivalent lock-in dynamics in the GPU-as-a-service market. At platform level, the US\$40 billion acquisition of Aligned Data Centers by a consortium including BlackRock's Global Infrastructure Partners, MGX, and the AI Infrastructure Partnership in late 2025, give a sense of the underlying scale.

The most familiar dispute trigger is the customer's inability or unwillingness to absorb the contracted capacity. This could be due to a number of causes: processor shortages, grid connection delays, model design changes, business model failures, or a shift in the customer's priorities. The customer will then typically test whether force majeure relieves it of the take-or-pay obligation; the operator will typically argue that it does not. The outcome is heavily fact-, market- and language-dependent.

Layered on top is the obsolescence problem. Compute prices are likely to fall materially over the life of a 15-year commitment, and the contracted price

may come to look distinctly out-of-market. There is no settled market practice on price review or break mechanics in this context, a sharp contrast with LNG, where the price review architecture is mature. As contracts age, we may expect disputes over whether implied terms, change-of-circumstances doctrines or contractual hardship clauses provide any relief. The Goldman Sachs Global Institute's recent *Tracking Trillions* analysis makes the same point: The economic useful life of AI silicon is the single most influential variable in determining the scale of the build-out, and small shifts in assumed replacement cadence move cumulative spend by hundreds of billions of dollars. As the report puts it, "those shifts can arrive abruptly." Contracts that price compute on assumptions about silicon useful life that prove materially wrong are precisely the contracts most likely to generate the disputes we describe.

Take-or-pay exposure is also typically embedded in the operator's project finance covenants. A dispute over whether a take-or-pay obligation has been triggered or breached can therefore simultaneously create a default under the operator's financing, exposing additional creditor stakeholders to the dispute and significantly raising the stakes for early resolution.

Regulatory change as a risk

Another major area of dispute is likely to arise from regulatory changes. Regulatory change cuts across SLA, PPA and capacity disputes alike. Governments worldwide are introducing rules that constrain how data centres access the grid, source their energy, locate their workloads and account for their emissions. Where such a change makes contractual performance more expensive or impossible, the affected party

will usually seek relief through a change-in-law clause; whether such relief is in fact available will turn on the language of the underlying contract.

Recent examples illustrate the breadth of the issue.

In the United Kingdom, Ofgem's TM04+ *First Ready and Needed, First Connected* reforms, approved in April 2025, replaced the previous "first come, first served" grid connection model and require large energy users, including data centres, to demonstrate strategic alignment with UK energy planning, meet strict readiness criteria, and validate planning progress before securing a grid connection.

In the United States, the strain on regional grids has driven a cascade of state and federal policy responses. Texas SB 6, signed in June 2025, together with the associated ERGOT large-load interconnection rules effective from December 2025, imposes a new mandatory interconnection process, mandatory curtailment obligations and substantial financial security requirements on data centre loads of 75 MW or more; in 2025, ERCOT's large-load interconnection queue is reported to have grown by approximately 300 per cent over the 2024 year-end total, to in excess of 230 gigawatts. State legislatures from Virginia to Maine, Oklahoma to Vermont have introduced data centre moratorium bills. At the federal level, FERC's ongoing review of co-located generation arrangements has now opened into review of tariff treatment of co-located load, with hearings underway through 2026. The Talen-AWS-Constellation matter is the most prominent example to date of an operational phase dispute already producing live regulatory adjudication.

In the European Union, the AI Act's obligations on general-purpose AI providers to disclose model training

energy consumption, the Energy Efficiency Directive 2023/1791 Article 12 reporting requirements for data centres above 500 kW of IT load, and the European Commission's forthcoming Cloud and AI Development Act (on which consultation finished in July 2025) together represent a layered regulatory framework that did not exist when many existing data centre contracts were signed.

In Asia, Singapore's Green Data Centre Roadmap and the second Data Centre Call for Application launched in 2025 require successful applicants to meet stringent power usage effectiveness targets (PUE of 1.25 at full load) and to source 50 per cent of energy from green sources — among the most demanding regulatory architecture in any major market and likely to drive material reshaping of contracts negotiated under the previous regime.

Where regulatory change is sufficiently severe and discriminatory, it may also support investor-state claims under applicable bilateral investment treaties, particularly where a foreign-invested data centre is denied a permit, loses grid access, or is subjected to a tax or data-localization requirement that singles it out for adverse treatment. That category of dispute is presently rare in this sector but is well-established in adjacent energy and infrastructure sectors and is likely to feature here in time.

The dispute resolution forum

As the build-out becomes increasingly global, the sector is likely to follow the well-trodden template, where international arbitration is the preferred mechanism for resolving disputes. This is typically under the rules of leading institutions such as the ICC, LCIA, SIAC, HKIAC or under the auspices of UNCITRAL, with parties typically selecting their arbitrations to be seated in London, Singapore, Paris, Hong Kong or

New York, for example. Parties typically value the cross-border enforceability of awards under the New York Convention in over 170 jurisdictions, the confidentiality of submissions and awards (a material consideration in this sector) and the technical expertise of party-appointed arbitrators in disputes turning on uptime measurement, power availability and commissioning. Careful attention to arbitrator selection, the seat, the governing law and the scope of the arbitration agreement (including its interaction with parallel regulatory proceedings) will, in our view, be among the most important drafting choices made in the current cycle.

Other operational phase disputes

Two further categories deserve at least brief mention.

The first is the LTSA, under which equipment manufacturers provide maintenance, parts and performance support for chillers, switchgear, GPU servers and other critical equipment. The dispute archetypes — performance shortfall, latent defects, the interaction of warranty and LTSA periods, and the scope of recovery for plant downtime — are well-developed in the power generation context and will translate directly into the data centre sector.

The second is joint venture and co-investor disputes. With an increasing concentration of value in very large platforms held through complex structures — as the US\$40 billion acquisition of Aligned Data Centers by a consortium including BlackRock's Global Infrastructure Partners, MGX and the AI Infrastructure Partnership in late 2025 illustrates — deadlock, exit, valuation, transfer-restriction and pre-emption disputes between sponsors are likely to become a regular feature of the landscape. These are not new dispute types, but the scale, cross-border nature and capital

structure of the assets involved will give them an unusually high profile.

Conclusion

The growth and scale of investment in AI and physical infrastructure, and the corresponding energy demands of the data centre and AI infrastructure sector, have few parallels in modern infrastructure history. The scale of capital being deployed, the speed of build, and the strength of the demand signal all point

to continued growth, and no one in the market is currently expecting or inviting disputes. However, as the market matures and inevitably encounters its first material dislocation — whether driven by demand, technology, regulation or counterparty distress — a new generation of operational phase disputes will emerge. The experience of LNG offers a useful yardstick for thinking about how that disputes curve will develop, while reminding us that the data centre cycle is likely to move faster and be less orderly than its predecessors. Parties

negotiating contracts in the current market should take particular care that their SLAs, PPAs, capacity commitments and supporting documentation are designed to be enforceable not just in the conditions of today, but in those that develop as and when the market matures.

This information is provided by Vinson & Elkins LLP for educational and informational purposes only and is not intended, nor should it be construed, as legal advice.