

# KEYNOTE INTERVIEW

## Plugging into AI



*Consumer and enterprise demand for data and generative AI points to a bullish future for US data centre operators, say Partners Group's Fentress Boyse and Morgan Jones*

AI arguably first caught the public's attention in the late 1980s, when Russian world chess champion Garry Kasparov played and lost to IBM's supercomputer Deep Blue. Since those early days, AI has continued to evolve and today demand for data is pushing the technology into uncharted territory.

Fentress Boyse and Morgan Jones, members of management in the private infrastructure Americas team at global private markets firm Partners Group, discuss the impact of AI on data centres and broader digital infrastructure.

**Q How is the current state of AI in North America translating to opportunities for digital infrastructure operators?**

**Fentress Boyse:** The broad introduction of ChatGPT in November 2022

SPONSOR  
**PARTNERS GROUP**

was an iPhone-type moment for digital infrastructure. We have been tracking the data centre sector through our thematic investing approach for some time and are now really starting to see an uptake in leasing activity from customers at our data centre assets. While it is still early days for products and monetisation, the infrastructure demand implications are dramatic.

If you think about our customers' products or services sequentially, AI applications need to be trained appropriately to first understand historical context before becoming usable for future commercial purposes. Our customers are making deployment decisions in

data centres right now in anticipation of the second phase of AI. The second inference phase is where AI can start making predictions, which requires slightly different infrastructure needs. This reflects how the inference phase has higher expected resiliency and latency requirements relative to the training phase.

Higher resiliency and latency requirements are impacting the size of leases, with higher resiliency meaning denser power requirements and therefore more demand for data centre space. Today, lease sizes have grown to record levels considering these real-time developments. Firms are targeting 300MW-600MW campus deployments, something the industry has not yet seen at this scale.

Considering deployment sizes and

the density of deployments at both the macro (campus/building) and micro (server) levels, customers are becoming more flexible in terms of location. Growing deployment sizes are also leading to challenges including supply/demand mismatches, supply chain issues, longer lead times, construction cost increases and rising financing costs.

Ultimately, increased demand towards phase two of AI will create new opportunities for nimble players with experience navigating these dynamics at scale.

**Morgan Jones:** When we get to this second phase, or inference phase, where more AI applications have been built, we expect there will be a need for compute power closer to the actual source and consumption of data at the edge of networks. We are seeing this play out through capital deployed and demand at the edge in anticipation of this phase.

We are seeing unprecedented lease sizes from the hyperscalers at both dedicated core sites as well as edge facilities. The hyperscale-sized leases of a few years ago are now being deployed at the edge and smaller data centres of less than 40MW, which are seeing an uptick in demand. It really is a landgrab right now.

### Q How will advancements in AI impact the size, location and power demands of future data centre infrastructure?

**FB:** If you translate the power requirements for AI training on a per capita basis, or look at some of the forecasts, it simply does not work to have all your locations in what are historically considered tier-one markets. The gigawatt-scale grid constraints in those locations, such as Northern Virginia and Santa Clara, are now at transmission and/or generation levels, not just distribution.

Customers will need to be somewhat more flexible, while acknowledging

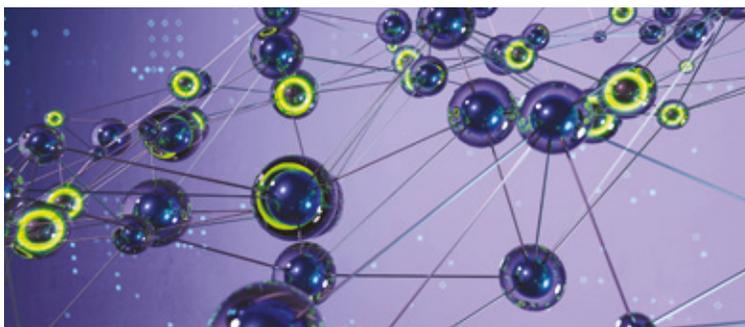
### Q How do you see demand growth being offset by technological innovation?

**Morgan Jones:** Chip technology will evolve but your biggest constraints again are power usage and cooling requirements. AI chips take three times the amount of power as traditional chips.

There is also going to be continued demand over a long period of time. If we think about wireless analogies from 2G, 3G, 4G or forward to 6G, the trend is all about more data consumption. Even if we think five to 10 years into the future, when AI is more mature, another emergent technology is likely to follow hot on the heels.

**Fentress Boyse:** If we look back over the last 20 years, the data centre sector has been an ever-evolving market. Technology companies are extremely innovative, always testing the boundaries of what is possible, and the level of competition is high. Some developments have the potential to be disruptive but there are substantial opportunities, too.

If you look at the early predictions, around 40GW and an investment of \$400 billion has been discussed to fully scale the AI opportunity. Those are staggering numbers. We have heard it said that AI is expected to one day be larger than the cloud in terms of power requirements, while adding another vector to cloud growth as well. This will test the eco-system.



that investments in data centres should not be just anywhere. An openness to look beyond the largest markets means turning to places in the American West, Midwest and Southeast where power has been less constrained. Salt Lake City, for example, executed a lease last year that was around 10 times the average size.

### Q How does the power demand dynamic influence the attractiveness of renewables?

**FB:** Data centre operators need to think more creatively about how renewables match with customer needs. If we cast ahead five to 10 years, management

teams will need to form a view about solutions such as small modular reactors and the associated power implications. These are large issues that the industry is grappling with because ESG has never been more important to our customers. Wind and solar are important pieces of the puzzle, but they are a baseload solution only if there is the accompanying storage. We are looking to crack the code where we can, but the answer is not always obvious.

**MJ:** We are going to see an increasing trend of large clean energy developers and operators working closely with data centre operators to offer a solution to the power problem. Today, customers

and investors are laser focused on ESG and ensuring that green energy is the dominant source powering their data centres, where it is available, with almost every major tech firm having to achieve net-zero or carbon-negative goals.

In the near- to medium-term, data centres will remain connected to the grid to ensure uptime. However, there is active dialogue and testing on how to leverage renewables and battery storage to not only provide a clean, redundant power source, but also faster time to market and less reliance on utilities with long lead times.

### **Q How do you think AI requirements will shape data centre design in the future?**

**FB:** Customer requirements are always evolving, and new use cases put tried and tested approaches at the forefront of innovation. If you think about traditional public cloud relative to AI, we are now seeing more power and rack density demand from customers for AI deployments.

AI requires more power in a smaller space and that puts pressure on cooling, or more specifically, heat rejection. You have seen more innovation in cooling, whether liquid cooling or otherwise, because there is an upper limit in terms of rack density. Larger AI-driven training clusters also require more aggregate megawatts at the outset to make the entire cluster usable, translating to larger campuses than ever before.

The other point to think about is redundancy and uptime requirements for AI versus public cloud. There are expected to be differences in uptime requirements for training and inference AI deployments. A lot of institutions have their data workloads in cloud environments where uptime is a critical focus, and we expect commercial solutions to have similar requirements for inference AI workloads. For training, there might be a bit more flexibility around redundancy and uptime.

*“Operators need to give more than they get in communities to be successful”*

**FENTRESS BOYSE**

Finally, if we look at the connectivity piece of this, there are likely more similarities between inference and public cloud in terms of latency requirements. For AI training there is potentially more flexibility, but with time, I think that those will converge.

### **Q How will AI demand impact customers that are building their own data centres compared to leasing ones owned and operated by third parties?**

**FB:** Historically, customers pursued a mix of self-built and leasing data centre capacity, with the split somewhere between half to two-thirds self-built. Recent trends suggest the partnership approach is moving towards a more even 50:50 mix or even towards two-thirds leasing, with third-party providers increasingly seen as just another supply chain option.

**MJ:** Today, hyperscale customers are trying to play catch-up. Perhaps, at first, they didn't fully appreciate the level of demand and the magnitude of computing power needed to deploy, teach and run AI applications, which is why third parties are attractive as they can deploy capacity quicker and cheaper.

A hyperscale customer's core business is selling their product to enterprise customers and ensuring a great user experience and providing high

levels of reliable service, not developing real estate. The data centre operator with high quality, available capacity in critical markets provides hyperscale customers a time to market advantage to accelerate their core business.

Equally, from a capital perspective, hyperscalers are spending so much money on equipment to enable AI services that they need to take a partnership approach. This means working with data centre operators to alleviate some of that pressure.

### **Q What are the main risks for data centre investing, and how might these be mitigated?**

**FB:** There are a few issues that need to be solved. Power availability is one important aspect. The other aspect is available pools of capital, whether that be equity or debt.

Finding power also means thinking about how local communities will be impacted. Community engagement and being ahead of potential local considerations is absolutely critical. In Mesa, Arizona, our portfolio platform EdgeCore is working to develop a programme with the local technical university to educate the workforce. Operators need to give more than they get in communities to be successful.

**MJ:** Over the last few years with covid, there have been supply chain issues creating longer lead times as well as higher costs. Operators need to develop detailed plans and strategies to try and mitigate delays and potential third-party problems to ensure capacity is developed and brought online in a timely manner. Standardised designs and ordering long-lead time items ahead of time in anticipation of future developments is a best practice for data centre operators.

Developing long-term partnerships based on trust, transparency, and reliability with your material and equipment vendors and key contractors is as important as anything for successful execution. ■