d-market data center
nds in the age of AI

Varun Sablok
Managing Director, North America at Igneo Infrastructure Partners

What is AI?

After more than a decade of academic and industry development, Artificial Intelligence has gone mainstream over the past two years, following the launch of OpenAI in November 2022. Today, the most prevalent AI systems are based on Large Language Models (LLMs). These are complex models that need to be trained with copious amounts of data, operating as a "neural network" that is inspired by how the human brain functions. Building and training these models is power-hungry work!

There are several market participants pursuing the largest, most developed LLMs today, with the ultimate goal to build an intelligence framework that can seamlessly apply past experience and reasoning to future situations. There are innumerable potential applications for AI in our day-to-day activities, providing the opportunity to make each of us more productive with our time. Just think of all the time we spend on our screens today researching topics of interest, writing reports, and sharing or presenting this knowledge to our colleagues; used effectively, AI can significantly streamline this effort and enhance workplace efficiency.

Current landscape of Data Centers

In 2022, we estimate there were approximately 13 GW of third-party owned, colocation data centers in the United States. During the last 2 years, approximately 10GW of new data center capacity has either come online or has commenced construction, of which, 1-2GW are colocation facilities. The vast majority of these facilities are (a) classified as hyperscale facilities with greater than 50MW of power consumption (and in some cases up to 1GW), and (b) increasingly located in more rural areas relative to older established data centers. This transition to much larger data centers away from cities is being driven by a focus on facilities that can be used to train LLMs most efficiently.

Current large-scale data center demand is being driven by the ever-increasing compute needs of the hyperscale customers: Microsoft, Meta, Google and Amazon. They have driven the construction of facilities wherever utility power can be sourced. In 2024, these customers collectively spent over $200b in capex projects driven by investments in AI and cloud computing. They are projected to collectively spend another $250-300b in capex in 2025, requiring another 5-10GW of new data center capacity. We are seeing a new "data center alley" being developed in Ohio where several gigawatts of data centers are currently under construction.

> "In 2024, Microsoft, Meta, Google and Amazon collectively spent over $200b in capex projects driven by investments in AI and cloud computing."

Add to this the recent announcement by Stargate: a JV between OpenAI, Softbank and Oracle to jointly spend $500b in AI development projects, starting with $100b in the near term. The vast majority of new data center developments happening today are in the hyperscale sector. It is fair to assume these facilities will continue to be developed wherever hundreds of MW of power can be sourced and land is available, frequently further away from population centers. A typical 100MW greenfield development will cost $1-1.5b to develop and a portfolio of multiple hyperscale facilities can quickly add up to several billions of dollars in equity.

It is costly and time consuming to upgrade electric transmission networks to bring power from rural areas to population centers. However, it is significantly cheaper and quicker to lay high-count fiber routes to transmit data. This cost disparity underpins why most of today's new data center developments occur in rural areas, connecting to population centers and other data centers through new fiber routes. Large scale facilities in areas with relatively low power costs and available land have become a priority versus the previous focus on low-latency and the network benefits of co-locating data centers near demand.

A primary concern for investors and financiers is the long-term use case of these mega facilities in rural locations. Once these large AI models are trained, will they continue requiring significant amounts of power to re-train in the future? Or will these rural facilities eventually be replaced by facilities closer to population centers to facilitate low latency use cases?

Don't forget the cloud!

Until 2021, most data center discussions were centered around enterprise workload migration to the cloud. Although most of today's headlines focus on AI applications, the bulk of current IT infrastructure is still being installed to support enterprise workloads. We project that this trend will continue for a long time to support the continued growth of enterprise workloads. In the future, we expect cloud infrastructure to also be used for deployment of AI models.

The under-served customers

The exponential growth in demand, combined with the shift to hyperscale facilities, is driving up the cost of data center developments while crowding out smaller customers. Current vacancy rates in most large US cities are now consistently below 10% and it is challenging to source more than a few megawatts in a single facility in these markets. As enterprises look to develop their own AI models using proprietary data it is becoming increasingly difficult for them to locate 10-30MW of IT workload in a single facility in a major metropolitan area.

Furthermore, once an AI LLM has been "trained", it needs to be deployed for use, referred to as "Inference". Ideally, this inference phase requires computation capacity to be located

Expected shift from Training to Inference is a big opportunity for infrastructure providers
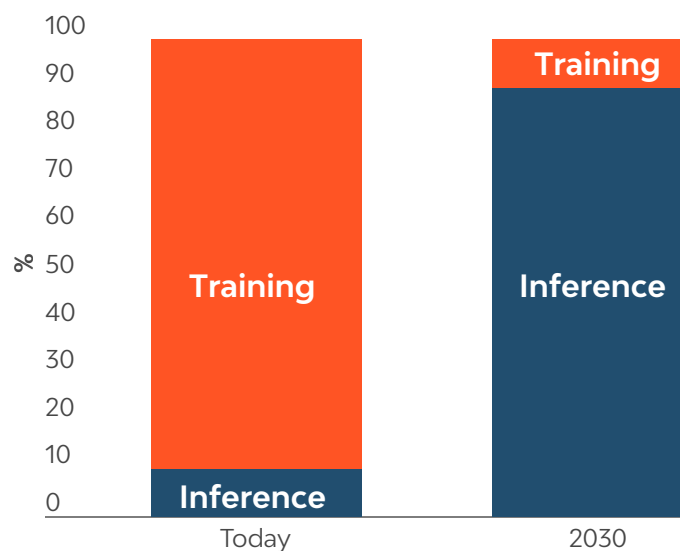
close to end users to mitigate latency – the delay between when a user queries an AI model and when they receive a response. We expect future demand for inference needs to outstrip the current demand for developing and training AI models. As a result, data centers located close to major metropolitan areas will become invaluable.

Source: US Signal, Jan 2025

The opportunity for mid-market infrastructure investors

We see two primary investment opportunities for mid-market infrastructure investors:

1. There is an increasing need from enterprise customers for high power compute that doesn't require a 100MW built-to-suit facility. These customers will typically co-locate their demand with other users to share infrastructure and want to locate their IT needs close to their operations, in major metropolitan areas;

2. We also see upcoming data center demand for AI inference applications which will require compute capacity to be close to end users to provide lower latency and faster response times.

As a result, we believe mid-size 5-30MW data centers located in core urban markets are interesting investments for mid-market infrastructure investors today. These assets are expected to play a critical role servicing computational needs in the future, even as the ultimate use case of data centers changes over time.

Locating new data centers close to population centers will require innovative strategies such as developing smaller data centers that are interconnected by dedicated dark fiber connections. In addition, permitting challenges for new facilities are likely to force the industry to reassess existing sites for potential refurbishment. As a result, investors need experience not only with technology management, but also with power development, facility design, customer acquisition and (depending on the situation) asset redevelopment to successfully navigate the evolving data center investment environment in the coming decade.

Looking ahead, we see continued demand for both new and refurbished data center assets in dense urban areas underpinned by both expanding cloud demand and fast-growing AI applications. We believe mid-market infrastructure investors who have direct experience in data center investments are well positioned to participate in future greenfield and brownfield data center growth in core markets. These assets are well positioned to become core infrastructure assets over time, supporting the increasing demand for latency sensitive workloads.

osophy
astructure
ble investment approach

eases

Contact us
Important Information
Complaints Procedure
Supplier Code of Conduct
Privacy notice
Whistleblower
Cookies policy
Cookie Preference Manager

igneo Infra Part

An affiliate of First Sentier Grou