ARCAPITA

Possibilities and Pressures:

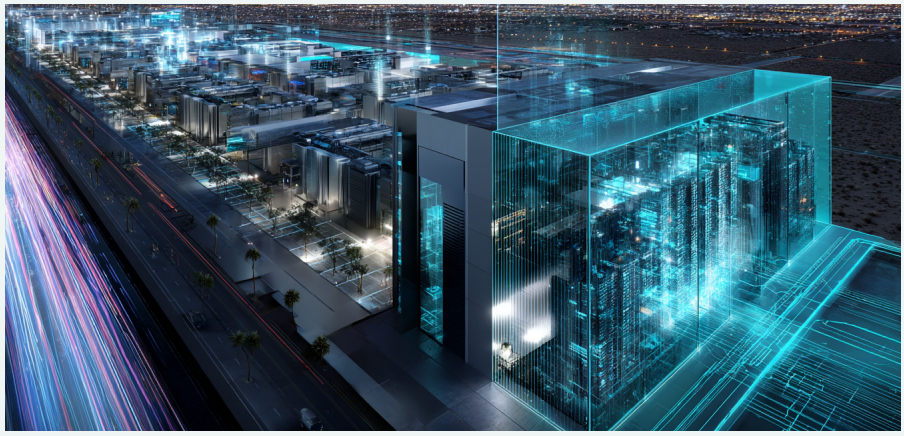
The Current State of AI Digital Infrastructure

January 2026

Overview

In all rapid technological advances, from the railways to the internet, a life-changing revolution has needed huge investment – but often with investors experiencing a leap into what may feel quite unknown. At the same time, in considering the opportunities and risks of these technological advances, investors should bear in mind Bill Gates's advice in his 1995 book "The Road Ahead", "We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next ten. Don't let yourself be lulled into inaction".

The field of artificial intelligence has advanced rapidly, creating the potential for significant change and a growing need for substantial investment.



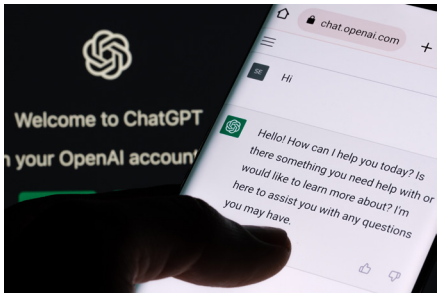
Moreover it is also reaching an inflection point at which the powerful force of demand for AI is meeting the very real challenge of constrained supply. Data centers can't be built fast enough for power-hungry hyperscalers and specialised operators even as investors are eager to provide capital for this new technology. The time horizons don't match up: we want more AI right now but the infrastructure required can take years or even decades to secure.

In this report we look at the current state of AI data centers, and bearing in mind Gates's advice about objectively assessing medium term change and avoiding inaction, explore the challenges and opportunities we see in AI digital infrastructure over the coming years.

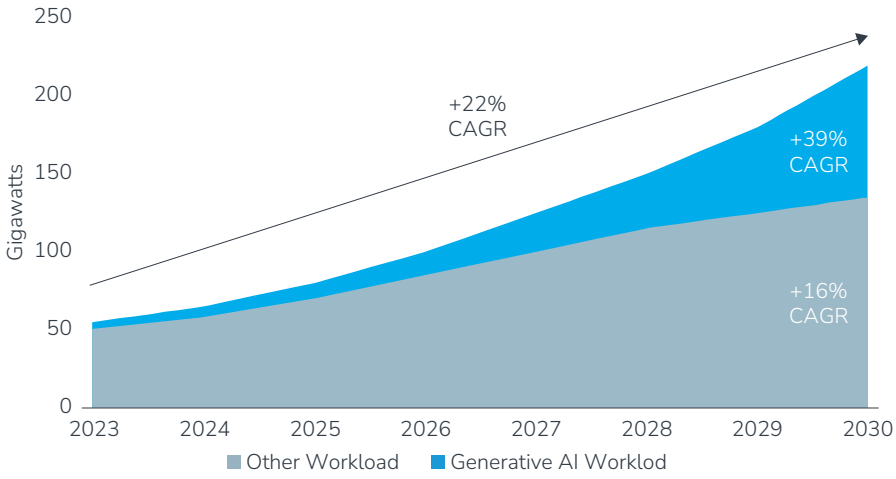
Where We Are Today

Data centers have already increased in scale and size since cloud computing became the efficient way to deliver IT resources on demand over the internet. Outsourcing hardware to secure scalable data centers has proved useful now that there is greater demand from our changing lifestyles for services such as virtual wallets, media streaming, online education, social media, virtual reality, video conferencing and so much more.

Artificial intelligence takes this to another level. Data centers provide the massive computational power, storage, and high-speed networking required for training and running complex AI models. We are in the foothills of understanding how AI can be used in practice, and what the resulting computational and power requirements may be. People are interrogating the likes of ChatGPT and Claude on topics as wide ranging as travel recommendations, relationship advice, and summarising legal documents. Companies are often immersed in experimentation as firms work out the most impactful use cases in the office and in the field. Excitable optimists have extrapolated this into expectations that it could release many workers from the clutches of mundane work, enabling humanity to leap into a more enjoyable life; doomsayers would argue that making many people redundant from their core work is a process for immiseration and civil strife, sceptics may wonder how much may actually change. The truth will be far from any extreme but it reflects the tendency of AI to veer into the realm of perception rather than reality.



Estimated Global Data Center Capacity Demand



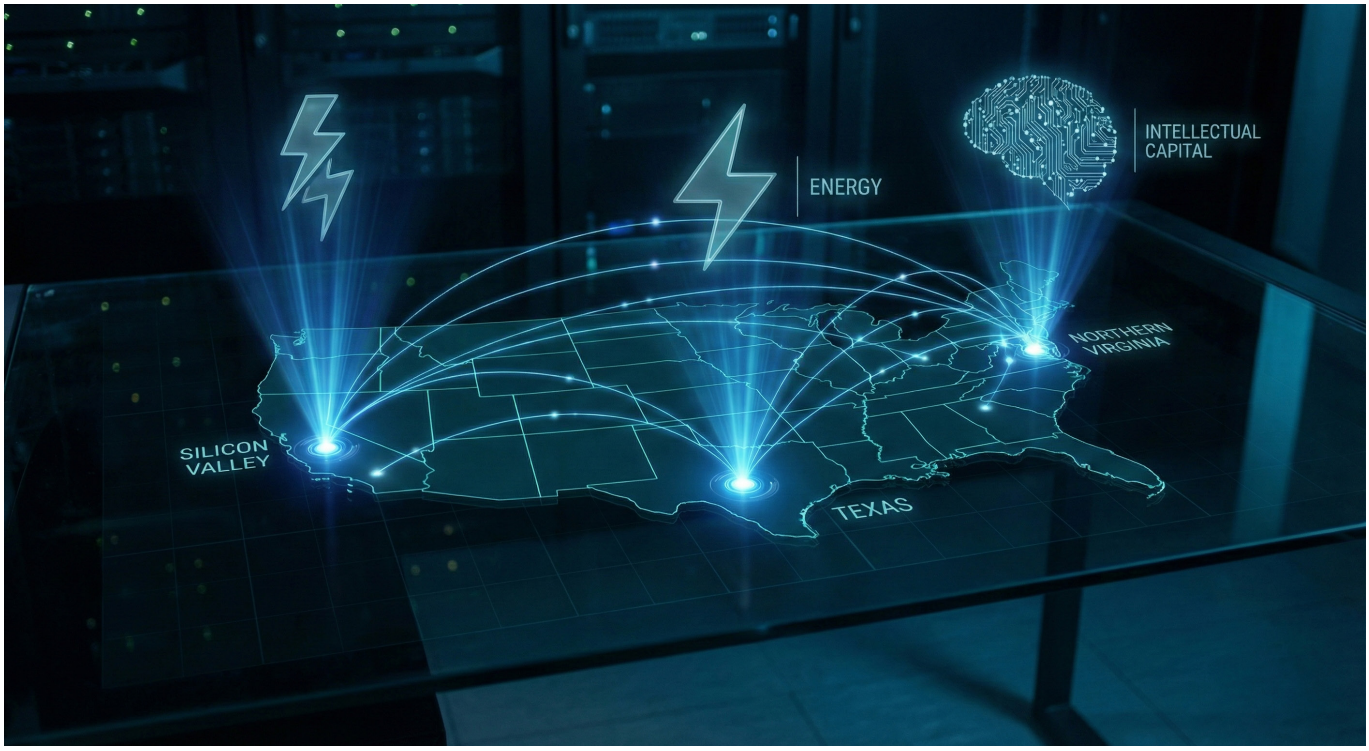
Source: McKinsey

Sticking to the facts, AI and "advanced AI" in all its forms is clearly increasing demand for data centers. According to Goldman Sachs, "The total data center development footprint now exceeds 50 million square feet – double the volume of five years ago". But there is still a shortfall. McKinsey estimates that to ensure demand meets supply by the end of this decade, "at least twice the data center capacity built since 2000 would have to be built in less than a quarter of the time".

And it is not just the amount of data center capacity, it is the cost. Artificial Intelligence requires data centers with connectivity and power inside strong and secure buildings. This does not come cheap. Goldman estimates that "The average cost to bring a 250MW AI data center online is roughly $12bn inclusive of the equipment inside".

Fortunately, there are presently deep and diversified capital markets able to, and eager to, provide this capital, and so investment has followed - but it is running to catch up. CapEx from the global hyperscalers alone equalled ~$800m per day in 2024, according to FactSet.

Beyond a strategic hyperscaler financing a project from their own balance sheet, organising the capital needed to develop new data centers is not necessarily straightforward if the initial cost is huge and the outlook, whilst sanguine, can feel uncertain given the rapid pace of change in the technology. This has led to different financing structures being developed in order to contain and spread the perceived risk. Banks might be involved in a classic construction loan which then converts to a longer term investment loan once the data center is operational; private credit may provide more flexible financing with higher loan proceeds; the CMBS market is getting more comfortable with both operational and even in-construction data centers as loan collateral. New structures are also



seen at the equity level: some hyperscalers are looking at joint ventures with private equity, and joint ventures with specialist operators are becoming increasingly common.

In particular, the US is the hub for the growth and development of this market. It houses not only the physical capital in terms of power generation and data connectivity, but also the intellectual capital emanating from its universities and Silicon Valley, and the deepest capital markets. The diverse geography of the United States also provides locations to suit any type of data center. AI models need to be trained before deployment and the "training" phase, where the data center is used to crunch through reams of data, can be more power intensive than one in the "inference" phase, where the AI model reasons out the answer. For the former, abundant energy resource is the priority, enabling chips to cool as they work at maximum power as it works through its training, while connectivity speed is less important. For the inference process, it is all about providing low latency (fast connectivity) so that answers can be delivered quickly, meaning that the data center benefits from being in an urban location, close to its end users. The US can provide locations for either stage, whether the data center would benefit from the cooling climate of the northern Midwest, the dynamic metropolis of New York or the energy abundance of Texas.

Understanding the contours of the AI and data-center surge is only the starting point. The rapid build-out of infrastructure and capability creates both significant pressures and new possibilities, and understanding these twin forces is essential to judging the medium term trajectory of the technology. This leads naturally to an examination of the challenges and opportunities that accompany this phase of expansion.

The total data center development footprint now exceeds 50 million square feet – double the volume of five years ago.

Goldman Sachs

Pressures and Possibilities

PRESSURES

1. Funding

Global data center spending commitments are forecast to hit $3 trillion over the next three years according to Morgan Stanley. JP Morgan estimates it could reach $5-7 trillion in total when including infrastructure spending on items such as chips, cooling and power systems. Even with hyperscalers covering a portion from internal funds and bond issuance, a shortfall of an estimated $1.5 trillion remains, for which private funds alone may be insufficient. For comparison, total US corporate bond issuance in 2024 was $1.5-1.8 trillion, the second busiest year on record.

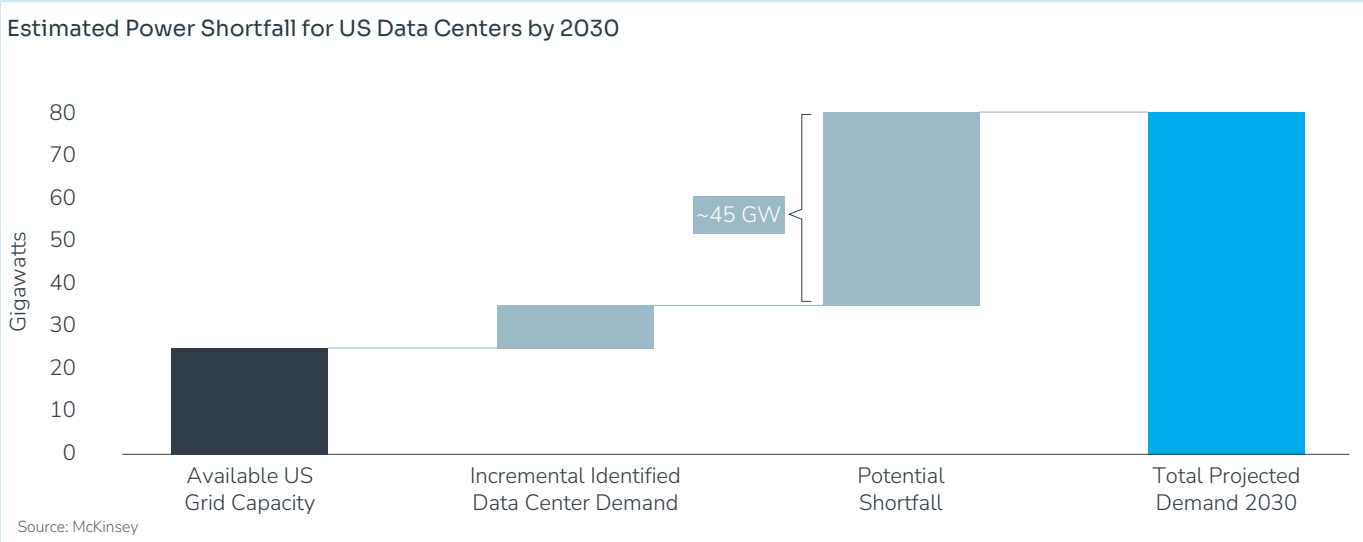
AI chips by their nature are at the cutting edge of technology and may become obsolete for the most demanding applications within as little as three years. Such relatively rapid obsolescence compares to 4-5 years for traditional server equipment and 15-20 years for the data center building's internal electrical and mechanical systems.

The large spending required to get the data center off the ground before revenues roll in leaves a narrower path to profitability in the short term. Given the huge amount of upfront investment required for a relatively new technology that is susceptible to evolving technical obsolescence and depreciation, some investors have required alternative financing approaches to become comfortable with deploying capital.

Private credit, joint ventures, securitisation and government subsidies are all being used to plug the gap.

PRESSURES

2. Power



Data centers use a lot of power and their needs are growing According to McKinsey, "Ten years ago, a 30-megawatt (MW) center was considered large. Today, a 200-MW facility is considered normal".

According to the International Energy Agency, electricity consumption from data centers is estimated to account for 1.5% of global energy consumption. Whilst that sounds a small proportion, the growth has been rapid, at 12% per year for the last five years.

The speed and scale has created development bottlenecks. There is an inherent timing mismatch between the time for a data center to be completed and its insertion within the broader energy ecosystem. Connection to the grid has a far longer lead time than the two or three years that it takes to build the data center, particularly for urban areas with existing priorities to develop more residential real estate and where grid capacity has less headroom.

These bottlenecks can be exacerbated by the high power components within AI data centers:

- Training large models requires thousands of GPUs and specialist chips to run at high capacity for months.
- Reliable and continuous uninterrupted power is paramount. Power failures can be extremely costly. This requires built-in redundancy and a guaranteed UPS (uninterruptable power supply) which must in turn be capable of supporting high power density.
- With the chips working overtime, the components become very hot. Cooling systems are integral to the process and account for around 40% of a data center's total power consumption.

Outside of the physical building which houses the data center, its operability is prey to wider power constraints such as congestion on the grid, long interconnection queues and the general peaks and troughs of energy demand.

- In an attempt to mitigate any delays, data center owners and operators are increasingly looking at ensuring proximity to strong and stable power supplies. Going "behind the meter" and acting as their own power supplier has led operators to build microgrids on-site.
- Beyond the use of Power Purchase Agreements (PPAs), which enables the operators to lock in the purchase of electricity at a fixed or formula-based price, operators are increasingly considering building their own small nuclear modular reactors (SMRs).
- Such decisions are not without controversy. At the xAI Project Colossus facility in Memphis, Tennessee, local residents have expressed concerns about air and water pollution.
- Communities have also become concerned that the demands of data centers on the local grid will freeze them out. As McKinsey notes, "Ireland has stopped issuing new grid connections to data centers in the Dublin area until 2028. Ireland's transmission system operator estimates that data centers will account for 28% of the country's power use by 2031".

PRESSURES

3. Division of Responsibilities

The rapid pace of technological change may require frequent hardware upgrades and potentially changes to the systems powering or cooling the servers, meaning the data center should ideally be scalable and adaptable to change. With operators often being separate from the owners – or indeed, multiple operators being housed underneath one landlord – there is the potential for inefficiency to build up, weighing on the capacity of the data center.

One potential solution is in landlord-occupier Service Level Agreements that set minimum performance thresholds for the owners to deliver in areas such as power and temperature control. That doesn't guarantee they will be able to meet such thresholds in all circumstances but it does at least attempt to align owner interests with the operator.

Such agreements would tend to make the landlord responsible for the building's structural integrity such as its power, cooling and backup systems, while the tenant is responsible for their own IT equipment, racks, servers and cabling. However not all landlords are looking for this level of operational involvement in a data center, and not all occupiers are necessarily willing to allow it, for operational or security reasons.

None of this can prevent wider power shortages, for example under an extreme weather event where power is intermittent or diverted elsewhere. The many steps in the interconnection chain between power producer and data center consumer creates a vulnerability that is difficult to mitigate.

4. Security

A data center houses such advanced technology and requires such a constant power source that it is uniquely vulnerable to external threats, such as:

- Extreme weather or climate events which disrupt power supplies
- Intellectual property theft or commercial sabotage
- Hostile activity such as hacking or phishing attacks

Protection from these threats is expensive. It can include:

- Perimeter defence
- Access control
- 24/7 onsite security guards
- Internal security zones
- Environmental controls
- Network segmentation and cybersecurity
- A Zero trust architecture which requires continuous verification and least privilege access.

Comprehensive physical and cybersecurity policies incorporate access control, surveillance and data breach notification procedures. This may also require coordination between the occupier and the owner.

PRESSURES

5. Geopolitics

With the potential to deliver life-changing technological advances for an economy, governments have become ever more fascinated by the global AI arms race.

Technology companies have returned the embrace from politicians, knowing that government subsidies and financing can be a vital part of the solution for them to achieve the scale required to end up as victors.

However in the wake of the pandemic, many governments have become encumbered by increased debt, reducing their fiscal room for manoeuvre. Whilst they may fully anticipate that AI can boost the potential growth rate of the economy, budget deficits may prevent them from offering resources.

This is where countries are cooperating for extra help. The recent visit of the Crown Prince and Prime Minister of Saudi Arabia to meet President Trump in the White House provided an opportunity for both countries to establish a deeper trading relationship in the area of technology. Nvidia's chips will be used in Saudi Arabia's new 500 MW data center and Elon Musk's xAI has promised to be its first customer. OpenAI is planning the 1GW Stargate UAE factory in collaboration with Nvidia, SoftBank and the Emirati AI firm G42, with an initial 200 MW facility opening by the end of 2026.

Nvidia CEO Jensen Huang has described how each nation needs to use its culture, knowledge, language and history to create what he calls "sovereign AI". With the US placing export controls of Nvidia's most advanced chips to those nations it considers competitors, such as China, it is clear that AI is on the front line of global politics.

POSSIBILITIES

1. Funding

As mentioned, there is an estimated shortfall of ~$1.5 billion in data center investment needs over the next few years, which provides an enticing opportunity for private capital.

The nature of the transactions needing this capital is also often appealing to private capital:

- While the hyperscalers may not have the funds to develop some data centers themselves, they still want to occupy them. This translates into long term lease commitments from the largest companies in the S&P500, and often with the occupiers taking on the operational aspects of the asset
- It is not just the hyperscalers – there are also the neo-clouds (the likes of CoreWeave, Lambda, Core42 and Nebius) that offer GPU-as-a-service, and model providers including OpenAI and Anthropic are reportedly examining direct deals as they race to expand their footprint beyond hyperscaler availability. This supports occupier competition and rental growth, particularly for core assets in sought after locations
- In particular, securing pre-leases from hyperscalers, neo-clouds or corporate users derisks development or value-add projects, and can unlock significant appetite from the equity and financing markets
- The sheer physical scale, financial cost, and complexity of modern data-center developments limits the numbers of investors able to invest in the sector. At the same time, some investors that are able may do the opposite of Bill Gates's advice, by underestimating the medium-term changes and defaulting to inaction. This imbalance can provide for outsized returns for those investments where observant capital that is able to invest can appreciate an overlooked risk/reward balance
- Debt markets are increasingly supportive of data center investments, with CMBS and private credit markets able to both appreciate their operational and cashflow nuances, and deploy substantial loan amounts over an investment period

Furthermore, the sheer scale of the current and growing capex commitments in AI, as well as the power and security challenges already evident, mean that it is quite likely that not all of the projected data center projects will actually happen. Even if they do, they cannot all be situated in the most sought after urban areas, given the power, security and geo-political challenges set out previously. So for those investors able to capture current opportunities in these primary US urban areas, the present circumstances can offer not just investor-friendly lease commitments and structures, and attractive returns, but also potentially enhanced market dynamics going forward if the competing supply pipeline is restricted.

POSSIBILITIES

2. The Risk/Return Equation

For investors, data centers provide a useful portfolio diversifier with an AI exposure at a time when concerns have been raised over the valuations placed on AI-driven equities. By investing in a data center, an investor is protected from potential drops in AI equity valuations, with AI-derived cashflows that have real estate collateral, compared to private credit exposure to the AI companies themselves. By combining the upside potential of AI with the cashflow and inflation protection of a real asset, data centers can hit a sweet spot in the risk/return calculation.

Additionally, the present environment allows for structuring to tilt the asset-level risk/return equation in an investor's favour:

- For single occupier data centers, the ability to transfer operational, security and technological aspects to the occupier, creating a simplified income stream for an investor
- For multiple occupier data centers, a growing ecosystem of specialist data center operators enables an investor to benefit from active management with the operational, security and technological aspects being handled by the operator
- In some cases, there is the opportunity to create a joint venture with a hyperscaler or other major occupier, sharing capital requirements and aligning interests. Equally, there may be opportunities for geo-political buy-in, that can provide advantages in terms of power connections, security, and financial support

3. Power

With power being so integral to the success of data centers, any owner or operator that can secure long term energy supply will be at an advantage. If utility companies seek private capital co-investment to finance new energy supply (which would appear quite possible given the scale of the required investment, the limits on utility companies' balance sheets, and the appetite of private capital for such investments), then high quality owner or tenant corporate credit may prove advantageous in unlocking such opportunities.

Fortunately we are also at a key development point in energy infrastructure.

- Renewables have been growing at pace, whether solar, wind, geothermal, or hydroelectric, as the climate agenda gained ground throughout the world.
- The Russian invasion of Ukraine went on to deliver a national security dimension to the importance of countries securing their own power supplies.
- President Trump was elected on a mandate to "Drill, Baby, Drill". Even America's Three Mile Island energy plant, the site of the worst nuclear accident in US history, is preparing to reopen after Microsoft signed a twenty year deal to purchase power from the Pennsylvania plant.

This means the political will is now in alignment with the economic demand for increased energy supplies, creating a rare opportunity for hyperscalers and other data center occupiers to pursue their own path to growth whilst doing the same for the electorate.

However, this opportunity is not indefinite and there remain limits to new energy supplies and to the time for them to come online. Nuclear fission plants, both at traditional scale and in the form of small modular reactors, are notoriously slow and expensive to construct, while the timeframe for new nuclear fusion facilities is unclear (though here again, a reminder of Gates's advice not to underestimate the medium term change from this technological leap is well warranted). Renewable power has immense potential, but in the US has less political support and at the scale and stability of supply required by data centers, has further hurdles before becoming a straightforward solution.

So once again, this creates opportunities for those investors able to capitalise on current projects where power supply is available, as future developments may increasingly face constraints on securing the necessary power capacity.

POSSIBILITIES

4. Geography

With data centers requiring proximity both to largescale power providers and also to their end users, the United States is well positioned to exploit the growth opportunity of the sector.

- This has led to extremely low vacancy rates for US data centers. As McKinsey noted, “In Northern Virginia, dubbed the data capital of the world because of the high number of data centers concentrated there, the vacancy rate was less than 1% in 2024”.

The US can boast a plethora of options for a variety of data center requirements, whether it requires a remote location with a cooler climate, or needs to be jammed into the heart of a city.

The broad geography of the country also reduces environmental risks such as earthquakes or other natural disasters.

The skill, depth and flexibility of the US labour market also provides human capital to support all aspects of data center operation and maintenance.

Latency is becoming ever more important as more end users interact with data centers. With so much projected user demand being US based, this drives hyperscalers and operators to seek US based data centers, particularly near major cities, and limits the potential for off-shoring this infrastructure.

However, while the US can provide for all of these data center location options, not all may appeal to private capital investors. For example, an investor may prefer to gain exposure to urban data centers providing ‘inference’ services (that is, the running of AI), if they view these as being more resilient or more in demand over the longer term. And particularly for primary urban areas in the US, there are only so many sites that fit the required profiles for power, security, fibre connectivity and demand from most occupiers.

Once again, this points to investors being able now to gain exposure to these locations and projects in the most sought after locations, with the potential for the quality of future competing projects to be a step below, thereby enhancing the longer term appeal of current projects.

In Conclusion

Much like the Industrial Revolution, the rise of railroads, and the dot-com boom, the surge in demand for artificial intelligence and the associated need for massive infrastructure investments represents more than a simple evolution of existing technologies. It is a potential paradigm shift that will likely reshape the economy, society, and perhaps the very fabric of daily life. Each of these historical periods shared a familiar pattern: initial excitement and rapid expansion, followed by periods of doubt, disruption, and economic correction, and ultimately, the emergence of a new equilibrium that forever changed how people lived and worked. AI and its infrastructural backbone are probably no different.

The analogy to the railroads of 19th-century America is particularly instructive. At the time, building the national rail network required enormous upfront capital, massive coordination, and long lead times, all factors that initially generated scepticism and financial risk. Investors who understood the transformative potential though ultimately reaped disproportionate rewards, as the railroads opened markets, accelerated trade, and catalysed industrialization. Similarly, data centers also demand substantial investment in land, construction, connectivity, and power, while some investors remain sceptical or cautious, and tend towards inaction, leaving open opportunities for those who are prepared to capitalise on the current inflection point in AI's growth.

From an investor's perspective, AI data centers provide exposure to the growth and maturation of the Artificial Intelligence sector, with the defensive characteristics of consistent cash flows derived from tangible assets, while also offering development projects that can deliver higher returns in a largely investor-favorable environment.

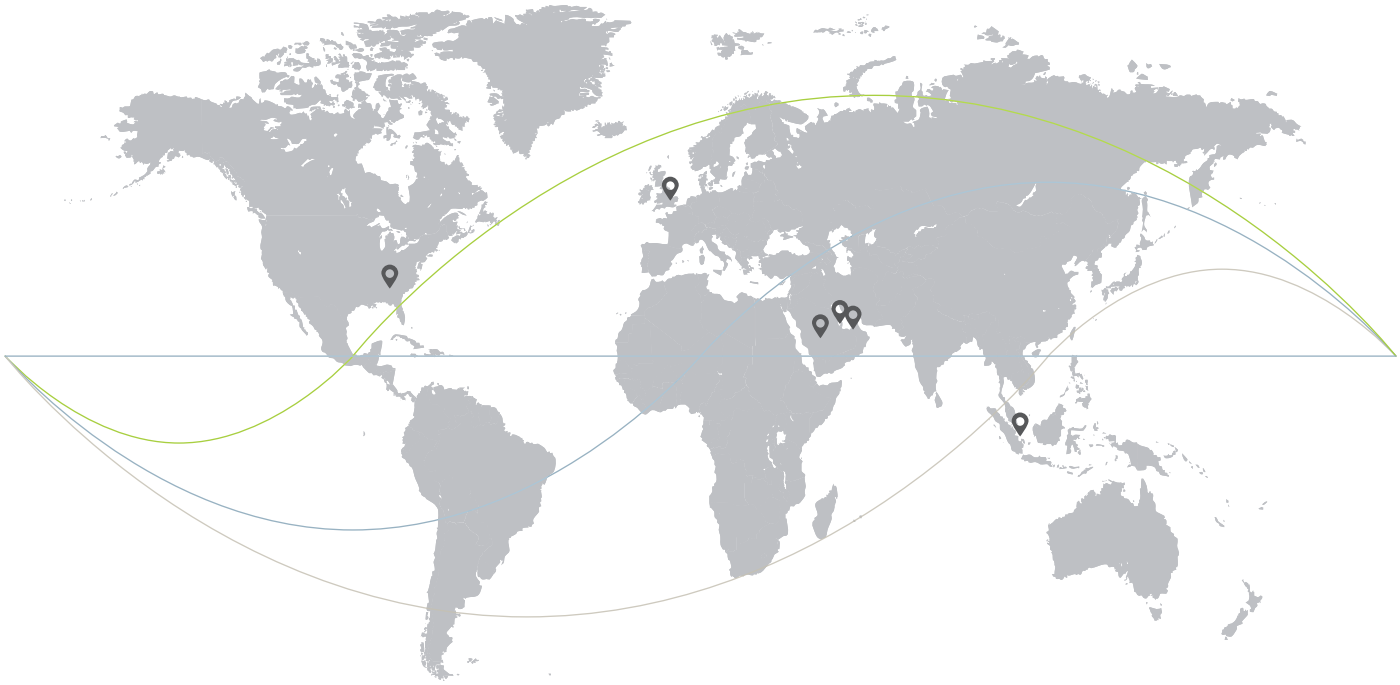
The potential challenges are undoubtedly significant, including power supply constraints, security considerations, and geopolitical pressures, but these also offer opportunities for those with foresight, understanding, and the ability to access the most compelling investments.

ARCAPITA

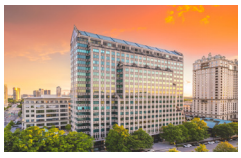
Overview

Arcapita is a premier asset manager offering diverse investment opportunities, focusing on private equity and real estate. At the center of one of the fastest growing wealth markets in the world, Arcapita's management has been serving an exclusive group of investors in the GCC region over the past two decades. With offices in Bahrain, US, UK, Saudi Arabia, UAE, and Singapore, Arcapita's management team has completed transactions worth a total value of approximately $30 billion and possesses a footprint to invest on a global scale. Arcapita focuses on defensive and counter-cyclical sectors supported by long-term macroeconomic and demographic trends.

With two decades of experience, Arcapita's management has built a global investment platform to access the opportunities that exist in our core markets of the US, Europe, Middle East and Asia.



United States



Arcapita Investment Management US Inc.

One Buckhead Plaza
3060 Peachtree Street NW, Suite 1650,
Atlanta, GA 30305 – United States

Tel: +1 404 920 9000

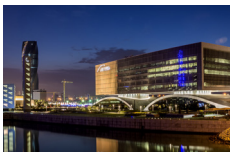
United Kingdom

Arcapita Investment Advisors UK Limited

16 Berkeley Street, Floor 5,
London W1J 8DZ
United Kingdom

Tel: +44 207 824 5600

Bahrain



Arcapita Investment Management B.S.C. (c)

Arcapita Building, P.O. Box 1357,
Manama,
Kingdom of Bahrain

Tel: +973 1721 8333

Saudi Arabia

Arcapita Capital Company

Kingdom Centre, Floor 14,
Riyadh,
Kingdom of Saudi Arabia

Tel: +966 114667610

Singapore

Arcapita Investment Management Singapore Pte. Ltd.

AIA Tower, 1 Robinson Road, #17-00
Singapore 048542
Republic of Singapore

Tel: +65 6513 0395

United Arab Emirates

Arcapita Investment Management Limited

Signature Centre, Al Maqam Tower, Office 1119,
Floor 11, ADGM Square, Al Maryah Island
Abu Dhabi, UAE

Tel: +971 816 4400