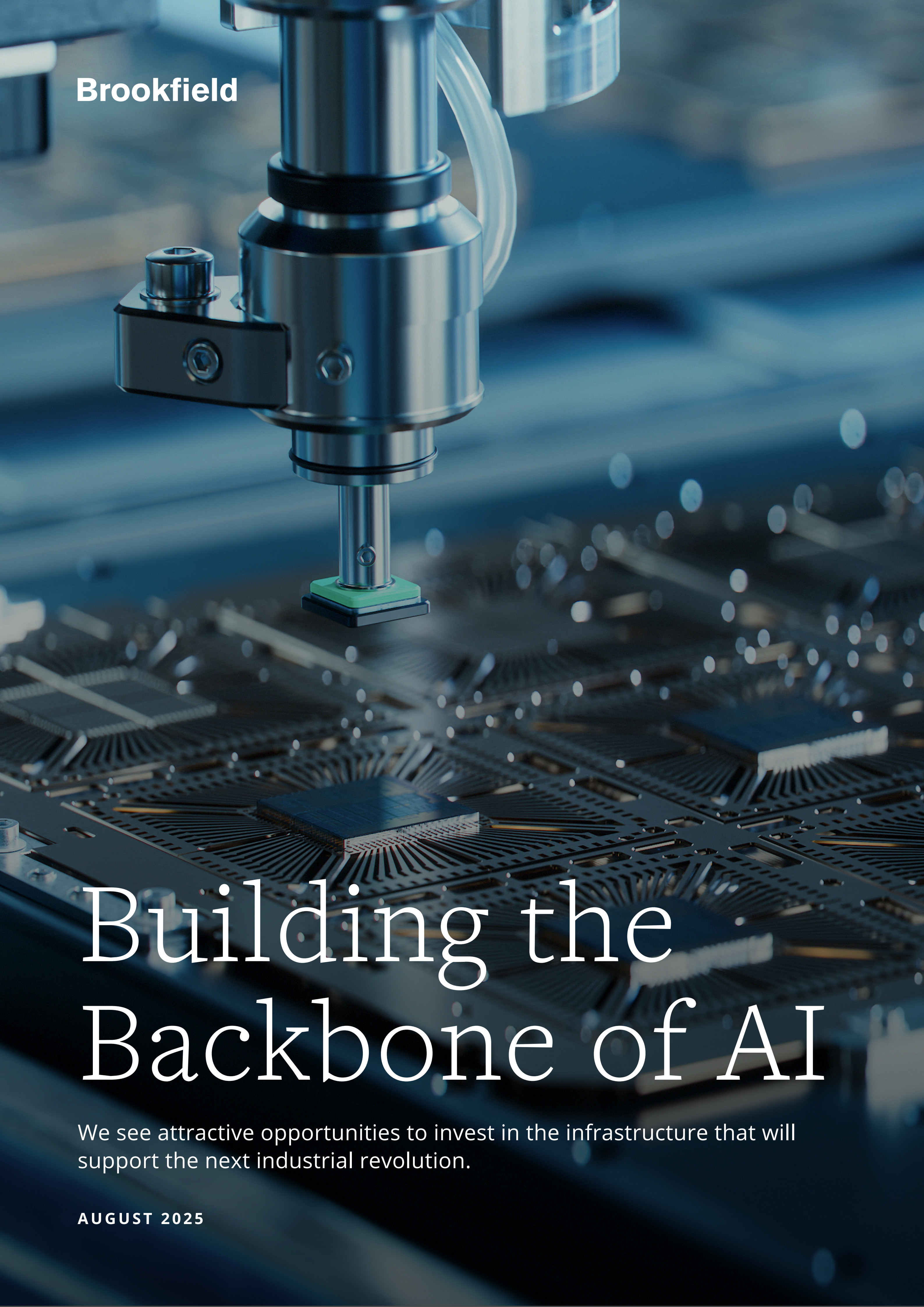


Brookfield

Building the Backbone of AI

We see attractive opportunities to invest in the infrastructure that will support the next industrial revolution.

AUGUST 2025

TABLE OF CONTENTS

Essential Insights

Throughout history, transformative technologies—from steam engines to the internet—have triggered waves of economic advancement by reshaping industries and creating new markets. Today, artificial intelligence (AI) and advances in "compute" (computational power) mark the next major inflection point, enabling machines to perform cognitive tasks and driving a new era of innovation and growth.

While chatbots have dominated public attention over the past two years, the next frontier—agentic AI and physical AI (robotics)—is poised to deliver the next wave of productivity gains.

- AI has the potential to significantly reduce the marginal costs of producing all essential resources and lead to a so-called Age of Abundance.
- AI-led automation could lead to massive growth in global gross domestic product (GDP), potentially reaching $10 trillion in annual economic productivity gains in the next decade.
- **We estimate total spending on AI-related infrastructure will exceed $7 trillion in the next 10 years.**

Despite market volatility, we expect AI growth to continue, driven by:

- **Sovereign demand for local compute.** Driven by commerce and national security priorities, governments are beginning to sponsor AI gigafactories, protect critical chip supply chains, and revise permit or grid interconnection policies to attract AI investments.
- **Cheaper compute, which will spur demand for more AI.** The **Jevons Paradox** occurs when higher efficiency leads to higher consumption. For example, as the price of electricity has fallen over the last 70 years, energy consumption has increased, leading to an approximate 5x expansion of the global electricity market.

The growth in AI is giving rise to several investable business models, each addressing a critical piece of the AI infrastructure puzzle:

- **Hyperscale compute** via AI factories can provide an attractive risk-reward proposition and a premium relative to traditional infrastructure sectors.
- **GPU as a service/compute infrastructure** can offer attractive infrastructure-like characteristics.
- **Power solutions**, through contracts and financing, are emerging as a powerful tool to accelerate power delivered "behind the meter."
- **Strategic adjacencies & capital partnerships** can unlock capital across the broader AI value chain, including dedicated fiber, circular economy networks and robotics, as well as chip manufacturing industries.

Understanding where the bottlenecks lie—and where the next leaps will come from—is essential to designing, investing in and future-proofing the physical backbone of AI.

- **Grid constraints:** Even with adequate power generation, grid connection can take up to 10 years to secure in some markets.
- **Model efficiencies:** As AI models become more efficient, fewer computing resources are needed per task.
- **Scaling laws:** Bigger models and data sets require higher compute capacity; "smarter" thinking requires more compute.
- **Training vs. inference:** We expect roughly 75% of future AI compute demand to come from inference by 2030.
- **Quantum computing:** Eventually, quantum machines will focus on certain tasks and complement AI compute.
- **Robotics:** While AI-powered robotics are still in the early innings, we expect an "S curve" of adoption within 10 years, potentially making it one of the largest global industries.
- **Technology obsolescence:** AI hubs need modular designs so that power and cooling systems can be upgraded easily. Chip technology will evolve quickly.

Expertise in developing and operating critical AI infrastructure is essential to managing the risks involved in building the backbone of AI. This is a once-in-a-generation opportunity where only strong operators with solid market understanding and access to the right operational assets will ultimately succeed in this complex, highly specialized asset class.

A Revolution in the Making

Technological progress has always hinged not merely on breakthrough ideas but also on the infrastructure that scaled them into everyday life.

The steam engine unlocked mechanized industry, but only after railways, coal supply chains and factories were built to harness its power. The telephone revolutionized communication, but only after vast networks of wires, switches and operators connected households and businesses. Electricity changed every facet of daily life, enabled by a massive buildout of power plants and transmission lines. Each of these breakthroughs catalyzed new industries, boosted productivity and redefined economic potential.

Today, artificial intelligence (AI) is poised to become the most impactful general-purpose technology in history. It could transform industries—and life as we know it—through faster drug discovery, better healthcare diagnoses, autonomous vehicles, natural disaster predictions, home robotics care and beyond. Long-term, AI has the potential to significantly reduce the marginal costs of producing all essential resources, and lead to a so-called Age of Abundance.

Yet none of these breakthroughs will be possible without the buildout of capital-intensive physical infrastructure to support the adoption of AI. This includes AI factories, power & transmission, compute infrastructure and strategic adjacencies & capital partnerships—adding up to an investment opportunity of over $7 trillion[1] over the next decade (see Figure 1).

Figure 1: **The AI Infrastructure Value Chain Represents a $7 Trillion Investment Opportunity**

AI Factories

$2T

Development of new data center capacity from land acquisition to ready-for-service

Power & Transmission

$0.5T

Baseload power and electricity transmission infrastructure to energize compute

Compute Infrastructure

$4T

GPU partnerships, as well as design and manufacturing of chips

Strategic Adjacencies & Capital Partnerships

$0.5T

Dedicated fiber connectivity, cooling solutions and semiconductor and robotics manufacturing

Source: Brookfield internal research.

> **"Just as electricity generation plants powered the last one, AI factories are driving a new industrial revolution. AI is the infrastructure to advance society, and the time to build it is now."**
>
> – Jensen Huang, Founder and CEO of NVIDIA (July 2025)

Unlike prior industrial revolutions, the AI revolution arrives with a crucial difference: It can expand both economic productivity and the effective supply of labor by augmenting and automating cognitive tasks at scale. We expect that AI-led automation could lead to massive growth in global gross domestic product (GDP), potentially reaching over $10 trillion in annual economic productivity gains in the next decade.[2]

Today, large language models (LLMs), agentic AI systems (AI agents) and advanced reasoning innovations are powering AI applications. Current AI models in production have automated targeted tasks and are increasingly generating more complex workflows via chatbots. Yet these early innovations are just scratching the surface of AI's capabilities. Continual waves of innovation and new types of AI are on the horizon, each requiring more sophisticated levels of compute.

These technological bounds are shifting how AI infrastructure is built and require a massive and more capital-intensive infrastructure stack. For example, AI compute relies on highly specialized graphics processing units (GPUs) that are much more specialized than the general central processing units (CPUs) associated with legacy cloud and software applications. These dense AI systems draw on higher levels of power, which require liquid cooling systems, while CPUs need less power and can use air-cooled systems. What's more, surging demand for AI compute and power is causing development timelines to double and interconnection queues to stretch out 3–6x longer in key geographical areas where the digital buildout will occur.[3] Today, compute infrastructure, including GPUs, interconnects and supporting hardware, accounts for up to 50% of total capital expenditure in next-generation data center deployments.

Investing in the backbone of AI requires access to capital and specialized expertise to manage the risks involved. But the long-term tailwinds are undeniable: surging demand for AI capabilities, exponential data growth and the transformative impact of intelligent systems across every industry. Despite some market concerns of overhype, we see little risk of overbuilding, as even meaningfully reduced forecasts still point to demand far outpacing current and planned infrastructure capacity. As AI capabilities become more commercial and embedded in real-world use cases, the need for scalable, high-performance infrastructure will only accelerate.

We believe this is the beginning of a once-in-a-generation opportunity to build the digital backbone of the future—one with the potential to reshape economies, drive innovation and deliver attractive risk-adjusted returns for those investors positioned to lead.

Digital rendering of AI hub to be built in Northern France. For illustrative purposes only.

Accelerating Demand for AI Infrastructure

Revolutions like these don't come easy—every industrial revolution has faced moments of exuberance followed by skepticism. AI is now at a similar crossroads.

The public markets have been sensitive to the news cycle on AI developments and the need for infrastructure buildouts. Investor sentiment around AI initially soared with the release of ChatGPT, driving a surge in valuations as markets rushed to price in its disruptive potential. But as new models pushed capabilities even further, some doubts about the extent of the need for new infrastructure began to emerge.

For example, DeepSeek made headlines in early 2025 for releasing a lower-cost AI model that achieved comparable performance to the leading AI labs. This raised concerns that, if China-backed AI is on par with the U.S., it could be developed at a fraction of the cost with less hardware, fewer data centers and less electricity than anticipated. Since then, some market participants have questioned the durability of demand for data centers and other associated assets.

However, our outlook for data center demand remains positive. In the near term, we expect to see the rise of "compound AI models," such as agentic AI, that string together multiple AI inference tasks to achieve a goal or conduct deep research. In practice, more efficient models are being used in greater numbers to tackle more complex tasks, consuming more compute overall. This ongoing research and development (R&D) will itself require more AI compute for training and experimentation. What's more, we are nowhere near the limit of artificial "intelligence." Both private companies and governments are investing billions in the race to develop the next generation of AI.

AI has become a strategic sovereign priority

Among governments, AI has become a strategic sovereign priority driven by commerce and national security concerns. Governments are beginning to sponsor AI gigafactories, protect critical chip supply chains, and revise permit or grid interconnection policies to attract AI investments. For example, European Commission President Ursula von der Leyen launched the InvestAI Initiative at the AI Action Summit in Paris in early 2025. The initiative aims to mobilize up to €200 billion in public-private investment in AI. This includes a dedicated €20 billion fund to support the construction of four "AI gigafactories" across the EU—large-scale compute hubs, each housing roughly 100,000 high-end AI chips designed for training the most advanced models.

Public-private partnerships are also emerging, whereby governments are engaging with highly strategic investors and operators to build large-scale sovereign AI infrastructure. The French government and Brookfield have announced a partnership to invest €20 billion in French AI, and Sweden and Brookfield have announced a $10 billion partnership to support the country's national AI strategy. Similar programs are emerging in North America, Europe, the Middle East and Asia, spurring robust sovereign demand for infrastructure.

AI systems stand to reshape defense, elections and politics. As a result, **governments will be at the forefront, regulating and directly investing in AI infrastructure** (much like World War II mobilizations or the Covid-19 pandemic response). In recent years, we have seen tighter export controls restricting access to cutting-edge chip technology to protect sovereign interests. Given the global shift toward deglobalization, we expect regional AI investment to accelerate.

In this environment, choosing the right partners is crucial: Companies aligned with government objectives could receive major funding and regulatory benefits, while those on the outside could face more hurdles.

Sikander Rashid (L), Global Head of AI Infrastructure at Brookfield, with Emmanuel Macron (R), President of France, in February 2025.

AI IS A STRATEGIC PRIORITY FOR WORLD LEADERS

"We have the knowledge, the talents and the decarbonized energy needed to accelerate in AI. We must not slow down, the world is accelerating. This is a battle for independence."

– French President Emmanuel Macron (February 2025)

“I think we really need to step up [AI innovation] in Europe ... the American economy, Chinese economy have been growing far faster compared to the European economies over the last 20 years.”

– Swedish Prime Minister Ulf Kristersson (February 2025)

“As our global competitors race to exploit these technologies, it is a national security imperative for the United States to achieve and maintain unquestioned and unchallenged global technological dominance. To secure our future, we must harness the full power of American innovation.”

– U.S. President Donald J. Trump (July 2025)

“It’s crucial that we’re building those capabilities on Canadian technology, to Canadian values, to protect Canadians, because we can’t rely on foreign suppliers for them. That is one of the new realities ... the natural approach here is gone. That’s one of the main priorities I would have in terms of securing Canada.”

– Canadian Prime Minister Mark Carney (February 2025)

“The AI industry needs a government that is on their side, one that won’t sit back and let opportunities slip through its fingers. And in a world of fierce competition, we cannot stand by. We must move fast and take action to win the global race.”

– U.K. Prime Minister Keir Starmer (January 2025)

“We will support capital investment for semiconductor manufacturing plants as well as public and private sector research, development and investment in such fields as artificial intelligence, quantum science and technology.”

– Former Japanese Prime Minister Fumio Kishida (January 2022)

FRANCE + BROOKFIELD

- Brookfield is accelerating its commitment to AI infrastructure in France with the confirmation of the first AI hubs under its €20 billion strategic partnership with the French government.
- The partnership was first announced at the Artificial Intelligence Action Summit in February 2025. In May 2025, President Macron announced the first planned AI hubs to be developed by Brookfield as part of this partnership and said "the €20 billion investment by Brookfield … will allow France to remain in the race alongside major AI players."
- The anchor site, E-Valley in Northern France, will be the first to launch, with construction starting in 2026. It will host at least 300 MW in the short term with a target to host up to 1 GW of capacity, representing more than €10 billion of investment and 4,000 direct and indirect jobs on this site alone.
- Along with two other sites identified in the region, it will form Europe's largest AI infrastructure cluster, with over 2 GW.

Digital rendering of Northern France AI hub. For illustrative purposes only.

SWEDEN + BROOKFIELD

- Brookfield is investing up to $10 billion to support the development of AI infrastructure in Sweden.
- This investment represents one of Brookfield's largest AI investments in Europe and extends the partnership with the Swedish government, its public authorities, academia and businesses in the region.
- The Brookfield investment will be centered on a new large AI center in Strängnäs, creating a strategic infrastructure asset to support the country's national AI strategy.
- The new site will create over 1,000 new permanent jobs and add another 2,000 jobs to support the 10-year construction process. The facility will be the first of its kind in Sweden and one of the first in Europe.

Digital rendering of Strängnäs AI center. For illustrative purposes only.

The Jevons Paradox: More efficient AI will spur demand for more AI

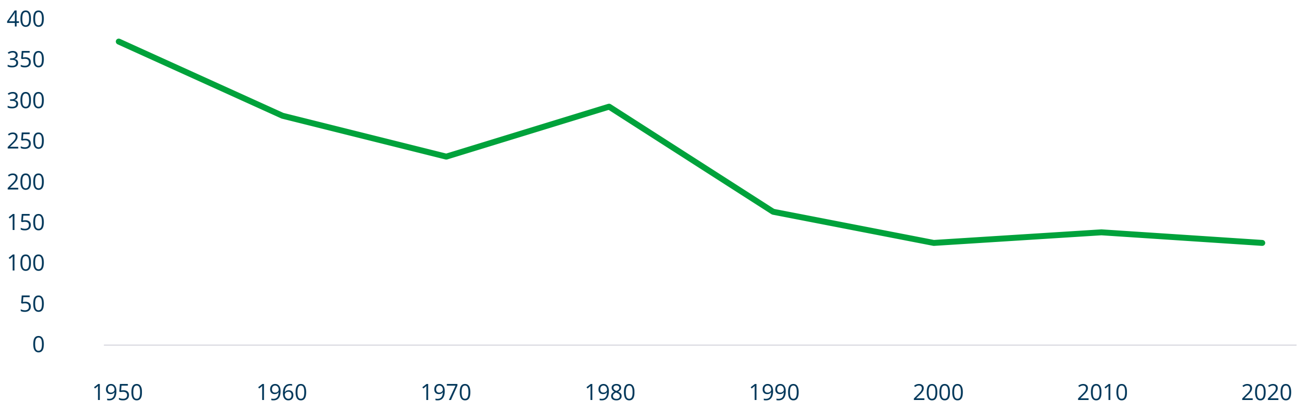
Falling unit prices, enabled by greater scale and efficiency, can spur even higher consumption—a phenomenon known as the Jevons Paradox. In 1865, economist William Stanley Jevons observed that improvements in the efficiency of coal-powered steam engines in the U.K. led to more, not less, coal consumption.

The same dynamic applies to modern technologies. Take electricity: In the last 70 years, the real (inflation-adjusted) unit price of electricity fell ~65% due to productivity improvements. At the same time, total energy consumption increased by about 15x, driven by more users and rising usage per user. The net result was a ~5x expansion of the global electricity market to ~$3 trillion (see Figure 2).[4] **We expect that AI compute will ultimately come to resemble utility products like electricity, driven by its unprecedented and sustained long-term demand.**

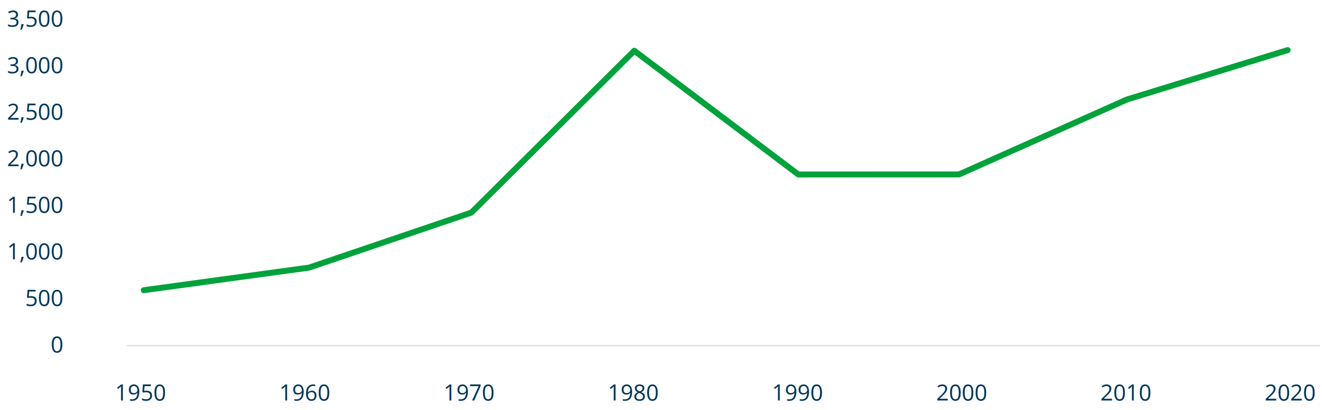
Since ChatGPT's release, the marginal cost that OpenAI charged developers (dollars per million AI tokens) fell by 99% in just 18 months. This was achieved via algorithmic improvements to increase efficiency in newer models and higher performance of GPUs (each GPU generation delivers ~2–3x more compute).[5] For example, Nvidia's Hopper (H100) architecture enabled substantial gains in training throughput over its predecessor, while the newly announced GB300 (Blackwell) is expected to push performance even further, supporting dramatically larger models at lower cost per token. Meanwhile, SEC filings and earnings releases from major hyperscalers continue to show an increase in capital expenditures over time (see Figure 3).

Figure 2: **The Jevons Paradox in Action: Electricity Prices**

U.S. Electricity Price ($/MWh, Inflation-Adjusted)

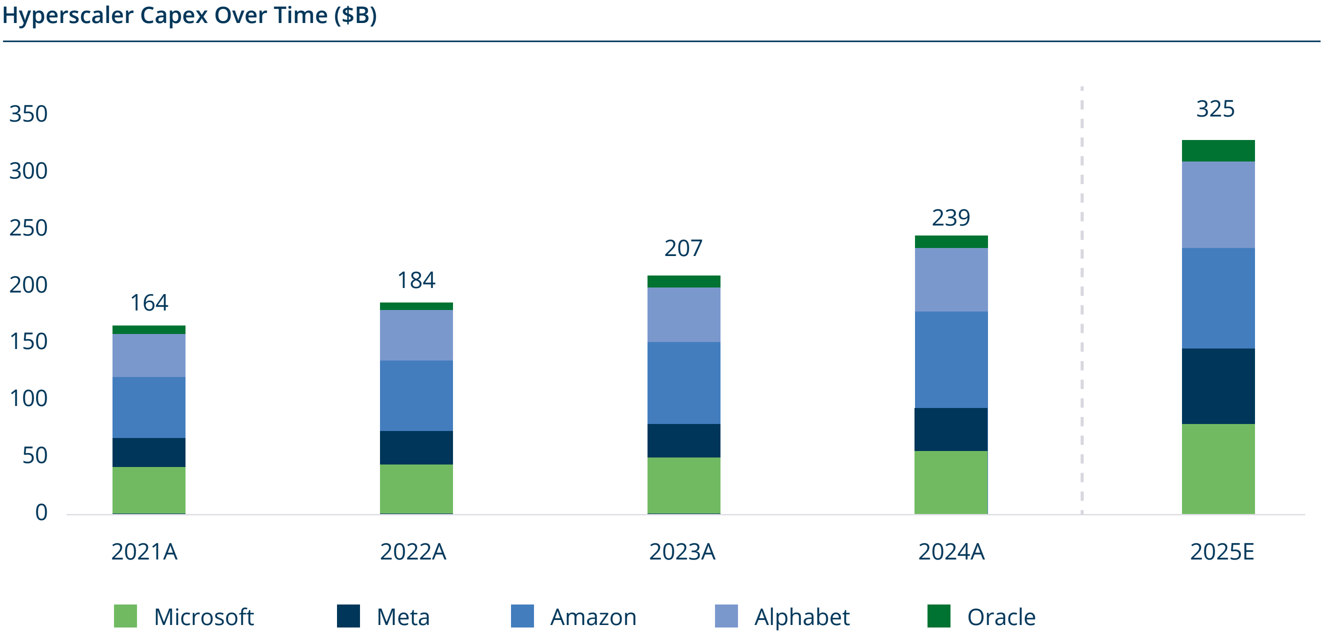


Annual Electricity Revenues ($B, Referenced on U.S. Inflation-Adjusted Prices)



Source: Energy Information Administration, U.S. Bureau of Labor Statistics.

Figure 3: **Trillions in Motion: Tech's Biggest Spend in History**



Source: Microsoft, Meta, Amazon, Alphabet and Oracle SEC filings and earnings releases, 2025. Data by calendar year.

> **"AI is becoming commercial and functional in a greater number of use cases. More efficiency in AI is going to drive more demand for AI. We see very little chance of overbuilding at this point."**
>
> – Connor Teskey, President of Brookfield Asset Management

How Infrastructure Enables AI at Scale

For AI to reach its true promise—unleashing economic growth and powering breakthroughs across healthcare, finance, science and knowledge work—it must be scaled. But that requires more than just better models; it needs robust and expansive infrastructure.

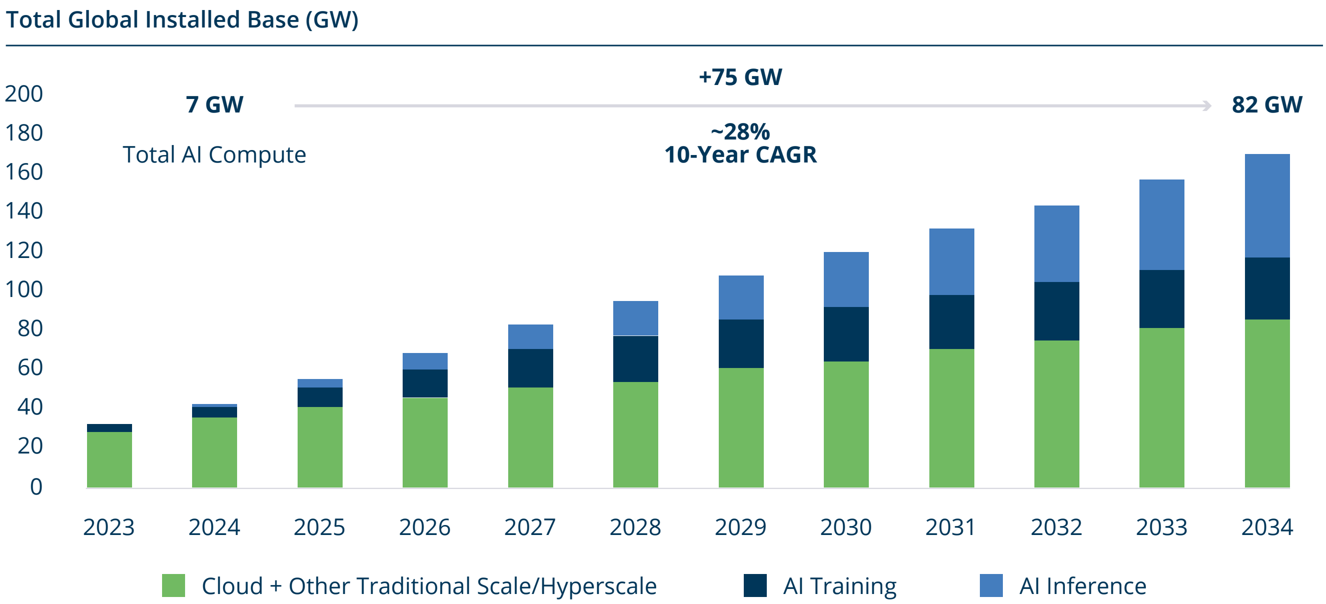
The AI value chain has many components, but we focus on the critical infrastructure layer—that is, the physical assets and services that enable AI at scale. These include AI factories, power & transmission, compute infrastructure and strategic adjacencies & capital partnerships.

AI factories

Unlike the cloud data centers that are predominant today, AI factories are modern digital hubs that incorporate advanced cooling (liquid or immersion) and specialized networking (InfiniBand or ethernet) to cluster thousands of chips. GenAI's rapid adoption is accelerating data center demand at a scale never seen before (see Figure 4).

By the end of 2025, we expect that AI factories will expand to ~15 GW of power capacity online from only ~7 GW at the end of 2024.[6] Over the next 10 years, we expect them to add an additional ~75 GW. This would bring total AI data center capacity to roughly 82 GW by 2034—more than a tenfold increase in a decade (see Figure 5).[7]

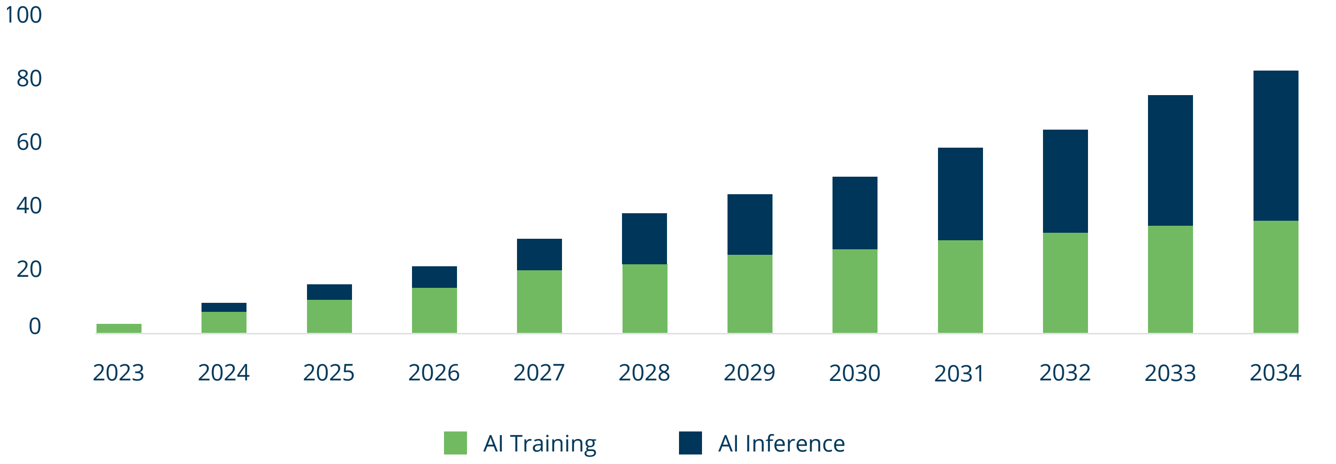
Figure 4: **Data Center Demand for Cloud Versus AI Training and Inference**



Source: Brookfield internal research.

Figure 5: **AI Data Center Capacity to Increase 10x by 2034**

Global AI Factories Installed Base (GW)



Source: Brookfield internal research.

As a result, capital requirements to build AI hubs and acquire compute hardware are skyrocketing. The user base for intense compute is also expanding beyond just AI labs to include enterprise R&D groups and governments, with many choosing to rent AI compute rather than own it.

Power & transmission

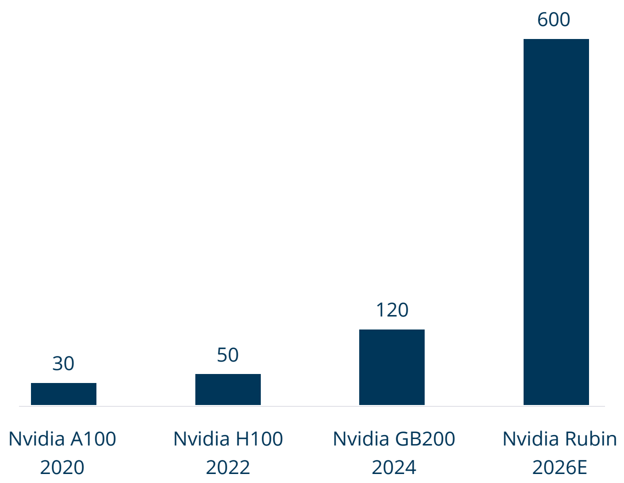
Supplying reliable power has become a top priority, as AI workloads consume far more electricity than traditional IT—driven by the intense compute demands of large-scale models. The power density of AI chips is ~10x higher than that of typical servers and is expected to rise another 5–10x in the coming years. A high-density AI rack can draw over 120 kW per rack, versus 10–15 kW for a standard data center rack. At the same time, new chips have become dramatically more efficient in their energy use (see Figure 6).

Yet, even as efficiency improves, aggregate power must continue to surge as workloads scale exponentially and demand spreads across industries. **This means that power and cooling infrastructure has become just as critical as the chips themselves in enabling AI to scale.**

Figure 6: **Powering Advanced AI Chips**

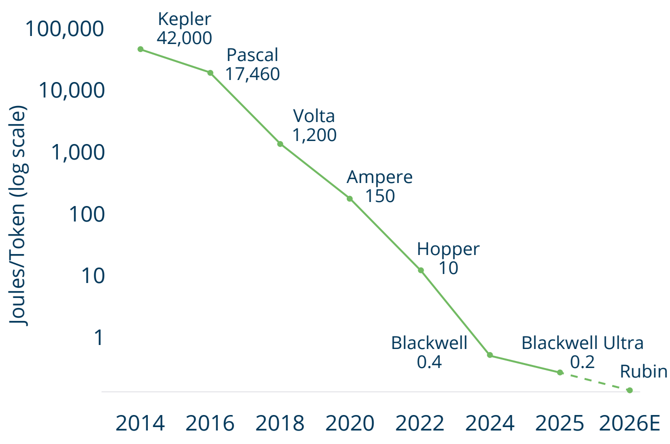
Advanced AI Chips Have Growing Power Needs

Power Density (kW Per Rack)



LLM Inference Continues to Get More Energy-Efficient

GPT-MoE-1.8T



Source: Nvidia.

As a result, the energy and utility sectors are facing unprecedented pressure—and opportunity. Industry experts estimate that power generation and transmission buildout to support AI-driven demand could represent over $0.5 trillion in capital investment over the next decade. Strategic coordination between AI infrastructure developers, utilities and regulators will be essential to unlocking this next wave of digital growth.

Compute infrastructure

This category includes GPU partnerships as well as the design and manufacturing of chips, including the high-performance Nvidia GPUs that power AI training and inference.

In the push toward ever-larger AI supercomputers, state-of-the-art designs like Nvidia's Blackwell product now connect 72 GPUs with 36 CPUs in a single AI system. In 2024 alone, Microsoft reportedly acquired nearly 500,000 Nvidia GPUs and Meta over 200,000.[8] Meanwhile, supply constraints are intensifying the global race to secure next-gen semiconductor capacity.

Given the expansive economic gains tied to GenAI, GPUs have become one of the largest and fastest-growing tech markets. We project the installed base of GPUs will grow about 7x, from ~7 million in 2024 to 45 million by 2034—representing over $4 trillion in cumulative GPU hardware sales over that period (see Figure 7).[9] Over the last decade, GPU performance improved 1,000x,[10] effectively outpacing Moore's Law. The total available compute for training AI models has increased more in the past 10 years than in the previous 40.[11]

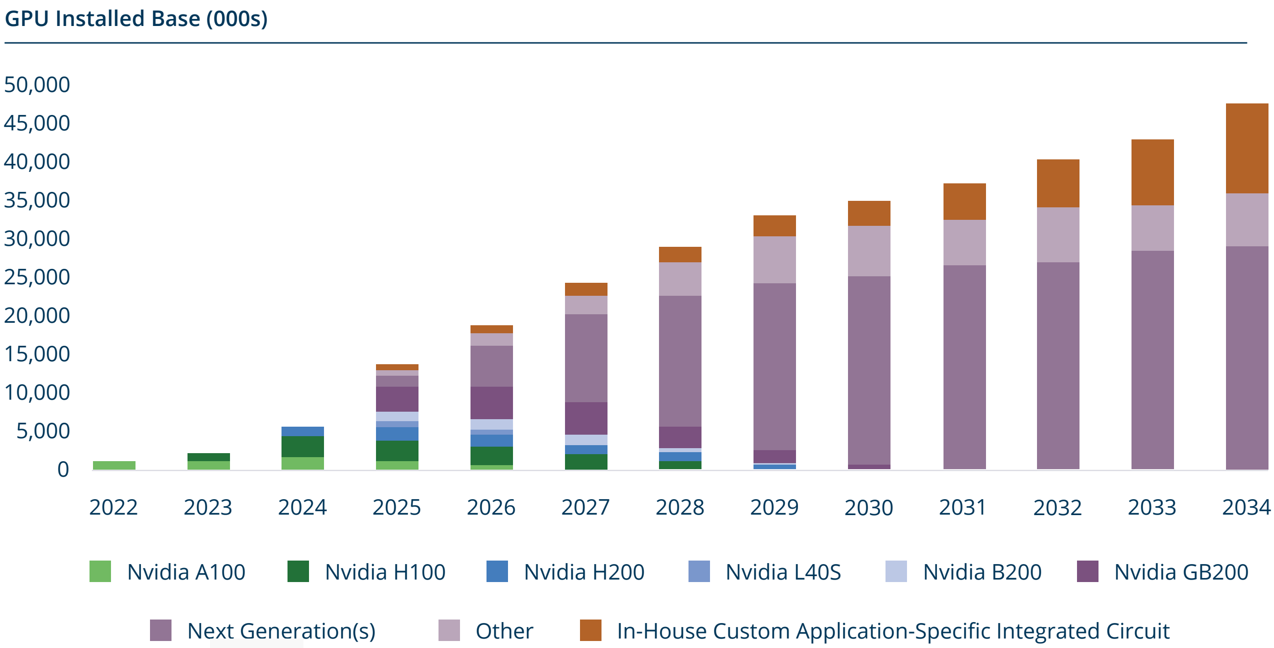
Experts suggest that reaching artificial general intelligence will require repeating the last five years of compute gains. Based on current chip roadmaps and massive data center buildouts, we believe the industry is on track to achieve that level of compute this decade.

Strategic adjacencies & capital partnerships

As AI workloads grow in complexity and scale, they drive demand for supporting infrastructure, including dedicated fiber connectivity, liquid cooling and circular economy networks. These adjacent sectors are critical enablers of sustained AI growth and represent compelling opportunities for infrastructure-oriented capital.

Additionally, the onshoring of supply chains in Western markets—accelerated by shifting geopolitical views and industrial policy—has created a new wave of investment requirements in semiconductor fabrication, robotics manufacturing and model training hubs. Projects like Intel's and TSMC's U.S. fabrication facilities and emerging robotics production hubs in North America and Europe highlight the physical infrastructure being built to secure AI competitiveness. Capital partnerships driving projects like these not only support the AI value chain but also offer resilient, long-duration investment opportunities supported by governments and large enterprises and aligned with national priorities and technological sovereignty.

Figure 7: **Installed Base of GPUs to Grow 7x by 2034**



Source: Brookfield internal research.

Business Models That Are Investable Today

AI's potential is limitless—but compute and energy are the gatekeepers. Building the AI foundation will require an unprecedented amount of resources and redefine infrastructure investing as we know it.

Already, the cost to train leading AI models has jumped ~10x in just a few years, following steep scaling laws.[12] Meanwhile, the data center industry needs to more than triple capacity by 2030 to keep up with surging AI workloads.[13] Frontier AI labs, for example, have announced multibillion-dollar expansion plans.

This growth is giving rise to several investable business models, each addressing a critical piece of the AI infrastructure puzzle.

Data centers & AI factories

AI-purposed hyperscale data centers built with custom power and cooling represent one of the fastest-growing infrastructure segments. Brookfield's existing data center platforms in North America (Compass), Europe (Data4) and Asia-Pacific (DCI) have been early beneficiaries of this trend, leveraging our longstanding relationships and deep knowledge of digital infrastructure needs. By expanding campuses and tailoring them for AI (e.g., higher rack densities and onsite power), we aim to capture outsized demand from hyperscalers and AI labs.

As a result of the large capital requirement and complexity of building next-generation infrastructure, the unit economics in this segment are attractive and present a premium relative to the traditional infrastructure verticals.

Power solutions

At the upper end of AI campus scale (projects of 1 GW or more), electricity supply becomes a mission-critical concern—and the energy solutions needed are "all of the above." Some large developments are already deploying onsite natural gas turbines to meet demand where renewables or grid upgrades can't ramp up fast enough. Beyond traditional power generation, advanced nuclear is moving from concept to reality as Google signed the first-ever small modular reactor (SMR) power agreement to secure 24x7 carbon-free energy for data centers.[14]

Modular onsite power solutions enable data center developers to accelerate their go-to-market timelines. These scalable systems provide flexibility to scale up capacity in line with future demand growth. Developers are increasingly looking for new forms of financing to fund the buildout of this behind-the-meter power infrastructure. We are seeing opportunities to provide capital partnerships by structuring financeable contracts with attractive infrastructure characteristics such as strong cash yields, long-term contracts and downside protections with investment-grade offtakers.

Pairing behind-the-meter generation with energy storage solutions can also help AI factories manage demand changes and integrate intermittent renewables more effectively. These battery storage systems can provide critical power resilience by supporting backup power and grid stability during peak GPU loads or outages.

The AP300™ SMR. Westinghouse is a leading global provider of mission-critical technologies, products and services to the nuclear power industry—and a Brookfield portfolio company.

GPU as a service/compute infrastructure

Traditional cloud providers today earn attractive returns by leasing out high-end GPU capacity. Amazon Web Services (AWS), for example, has achieved ~25–30% unlevered returns from its AI compute offerings.[15] The hyperscalers package raw compute with machine learning tools and support, but specialized "neo-cloud" players like CoreWeave have also entered this market.

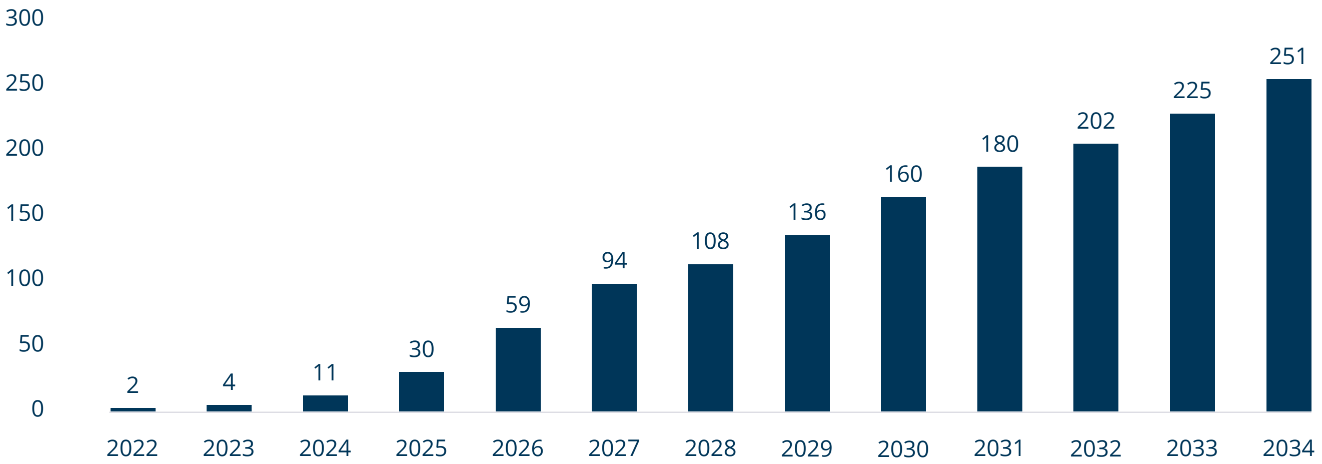
We believe our approach to GPU financing offers an attractive infrastructure-like risk profile by securing four- and five-year take-or-pay contracts with creditworthy counterparties. These contracts are often paired with long-term underlying data center leases, which increases the likelihood of additional upside from renewals following the initial term.

We project GPU as a service to grow from about $30 billion in 2025 to over $250 billion by 2034[16] as companies large and small seek flexible, on-demand access to AI horsepower without the capex commitment (see Figure 8).

> **"Energy is the bottleneck today—the answer is behind-the-meter solutions. Over time, whether it's solar, wind or nuclear, data centers will move to where the energy is."**
>
> – Sikander Rashid, Global Head of AI Infrastructure at Brookfield

Figure 8: **GPU as a Service to Top $251 Billion by 2034**

Global Demand ($B)



Source: Brookfield internal research.

Planning for Tomorrow

The emergence of increasingly capable models, soaring demand for real-time inference, and the physical limits of current compute capacity are reshaping the strategic landscape for infrastructure providers and investors alike.

Understanding where the bottlenecks lie—and where the next leaps will come from—is essential to designing, investing in and future-proofing the physical backbone of AI.

Grid constraints

In regions where the grid cannot keep up with AI's appetite, onsite generation or dedicated renewable capacity will be pivotal to keep growth on track. That's because it takes much longer to clear an interconnection queue than it does to build a new facility (see Figure 9). In the U.S., interconnection queues average six years, pushing 10 years in markets like California.[17]

Approximately 2.7 GW of new off-grid data center projects have been announced this year. Both specialist developers and the cloud giants are exploring these self-powered designs, with individual campus proposals as large as 1 GW.[18]

Model efficiencies

AI models are becoming dramatically more efficient in their use of compute. For instance, Meta's latest Llama 4 model employs a mixture-of-experts architecture, activating only a subset of its 400 billion parameters during inference, which significantly lowers computational costs and latency.[19] Other model architectures like DeepSeek continue this trend, delivering output comparable to GPT-4 at a ~75% lower inference cost just months after GPT-4's debut. DeepSeek achieved this via algorithmic breakthroughs—such as using lower-precision math, predicting multiple words per step and employing reinforcement learning to self-improve.[20]

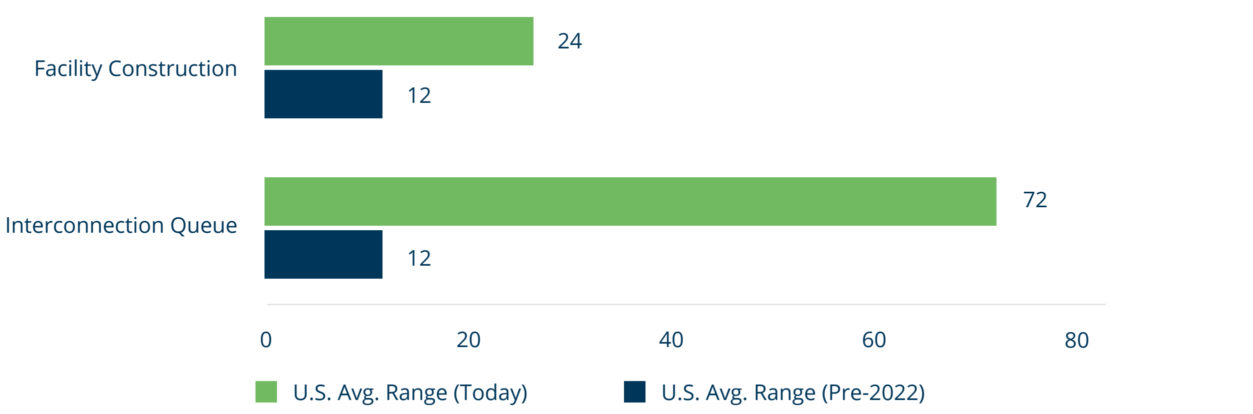
These efficiency gains mean fewer computing resources are needed per task. However, as observed through the Jevons Paradox, they should also increase overall demand.

Scaling laws

Frontier AI labs continue to observe that bigger models and data sets yield better results—a phenomenon captured in AI called "scaling laws." Realizing these gains requires massive compute capacity.

Figure 9: **Why Power Is a Bottleneck to Developing Hyperscale Data Centers**

Max Time to Secure Power and Build Hyperscale Facility (Months)



Source: Brookfield internal research.

The latest-generation foundation model training demands tens of thousands (rapidly expanding to hundreds of thousands) of GPUs wired together with ultra-low latency, as even microsecond delays can bottleneck learning. On the application side, more complex AI queries (more "thinking tokens" per answer) similarly drive the need for larger GPU pools located close to end users to maintain sub-second response times.

Scaling laws shift the constraint from algorithms to infrastructure. The ability to build dense, co-located compute clusters is now a primary requirement for progress—physical capacity, not algorithmic innovation, is becoming the binding constraint.

Training vs. inference

While training new AI models has consumed the majority of compute resources to date, inference (i.e., running models in production) is poised to become the dominant workload (see Figure 10). Once models are deployed globally, they will serve millions of queries and applications, requiring enormous compute.

Our forecasts suggest that roughly 75% of future AI compute demand will come from inference by 2030.[21] The rise of complex AI agents, which chain dozens of model calls to accomplish a single goal, will multiply inference needs even further. This shift means that data center designs will increasingly be optimized for high-volume inference traffic—not just concentrated large training jobs.

Quantum computing

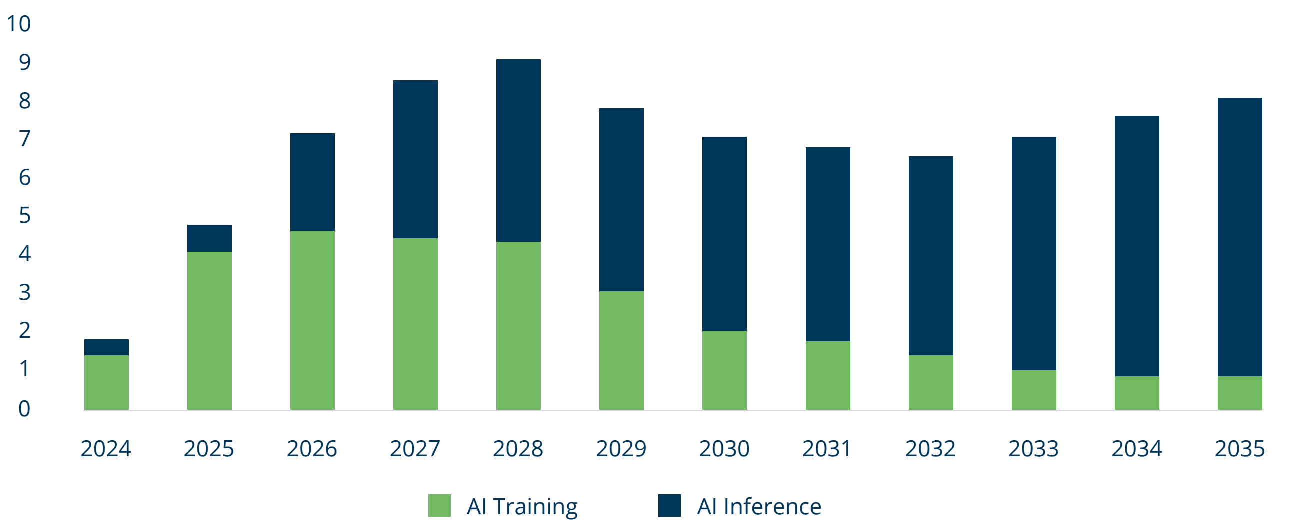
Quantum computing has the potential to digest complex problems much faster than classical computers and holds long-term promise to enhance AI. However, it remains at least five years out from making an impact on mainstream AI workloads. Early prototype systems (e.g., Google's Willow chip) have shown progress in error reduction, and some co-location providers (Equinix, Telefónica) have even started hosting quantum computers alongside classical servers.

Eventually, quantum machines will require infrastructure support similar to AI: reliable energy, specialized cooling (often cryogenic), and isolation from vibrations and electromagnetic noise. This could mean that future data centers will need retrofits like cryogenic cooling loops and magnetic shielding to accommodate quantum hardware. For now, practical AI infrastructure planning remains GPU- and tensor processing unit (TPU)-centric.

Brookfield's approach is to closely monitor advancements in quantum computing and be ready to integrate quantum nodes into our facilities once the technology matures and customer demand emerges.

Figure 10: **Approximately 75% of Future AI Compute Demand to Come From Inference by 2030**

Total Global Absorption (GW)



Source: Brookfield internal research.

Robotics

We are entering a new era where AI models and robotics are merging into physical infrastructure systems—demanding scalable, real-world environments and infrastructure to operate.

There are two types of robotics: purpose-built and general-purpose. In the words of Jensen Huang, CEO of Nvidia, general-purpose robotics in a human form factor (humanoids) represent the most important opportunity because the world is already built for humans.

Over the last two years, incredible advances in multi-modal AI, synthetic data generation and dexterity have led to rapid progress in humanoids. We expect production of these humanoid robots to reach the millions within the decade, leading to a transformation of the workforce and its impact on GDP.

While this is still in the early innings, we expect to see an "S curve" of adoption within 10-15 years, creating a large capital formation cycle.[22]

Technology obsolescence

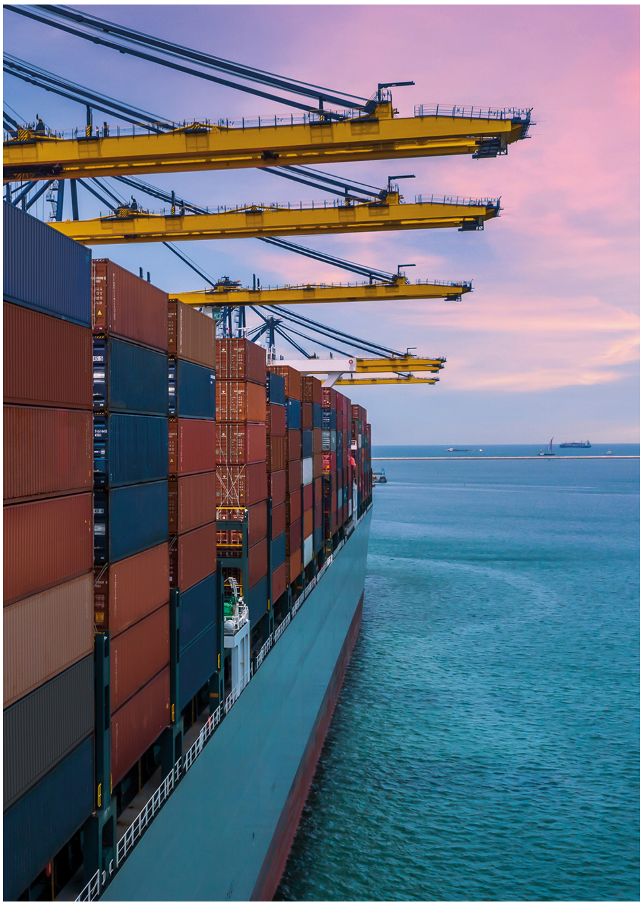
The rapid pace of innovation in AI hardware presents a refresh risk for infrastructure. Cloud hyperscalers are introducing proprietary accelerators (e.g., Google TPUv7 and AWS Trainium), and major GPU/application-specific integrated circuit (ASIC) generations now refresh every 12–18 months. Likewise, cooling and electrical standards are in flux, with rack power densities expected to exceed 100 kW.[23] On average, we see that a new build with liquid cooling can add ~$1 million per MW of capex relative to traditional lower-power-density data centers.[24]

To avoid obsolescence, AI hubs need modular designs so that power and cooling systems can be upgraded quickly as new chips and form factors roll out. Flexibility is key: For example, facilities can be designed with space and piping for future immersion cooling.

Brookfield mitigates obsolescence risk by building adaptability into our projects and maintaining deep partnerships with equipment providers to ensure we can refresh our sites in stride with technology evolutions. On the GPU deployment side, we look for long-term take-or-pay agreements with high-quality counterparties to earn strong contracted returns over the asset's peak useful life.

Mitigating Enterprise Risks in AI Adoption

Enterprises face critical decisions when adopting AI. They must choose between closed-source or open-source AI models, and decide whether to buy or build core infrastructure, develop in-house AI talent or hire external partners to deploy the model. They also need to consider any new privacy and customer-experience implications of AI-powered products.



Triton, a Brookfield portfolio company, is the world's largest lessor of intermodal freight containers.

We see many large corporations pursuing a hybrid infrastructure strategy. In practice, this means leasing short-term GPU capacity from cloud providers for experimentation and surge needs, while building or co-locating private AI clusters for sensitive workloads. Enterprises can de-risk their expansions by partnering with experienced operators and tapping into their expertise in site selection, power procurement, cooling design and operations.

We anticipate that firms in finance, healthcare, manufacturing and other sensitive sectors will increasingly invest in dedicated infrastructure to develop proprietary models, safeguard data/IP and meet regulatory requirements.

As AI infrastructure investment scales up, it's important for stakeholders to navigate several key challenges:

- **Technology flux:** Infrastructure must remain adaptable as new models and hardware continually evolve. Market participants should be prepared to upgrade or reconfigure data centers as chip generations advance and "System 2" AI architectures evolve.

- **Market competition:** High demand for AI resources can drive up valuations, requiring disciplined entry strategies and deep expertise in structurally mitigating against downside risk.

- **Supply chain bottlenecks:** GPU lead times, specialized cooling components and advanced memory remain scarce. Deep supply chain relationships developed over decades of experience are critical to operating effectively.

- **Regulatory environment:** Governments may impose regulatory mandates that shift project economics. For example, the EU's Artificial Intelligence Act will impose new requirements on "high-risk" AI systems, which could include data localization or transparency mandates that favor local data center investments. Maintaining close relationships with the leading nations provides genuine differentiation.

Brookfield's global platform approach and long-term experience owning both digital and renewable assets offer a strategic edge in providing integrated solutions (by building AI-purposed data centers powered by onsite clean energy, for example). Our scale, risk awareness and operational expertise support long-term partnerships with hyperscalers, governments and AI developers alike—converting the AI value chain into durable risk-adjusted returns.

From AI Ambition to Scale

AI's evolution from chatbots to artificial general intelligence to potentially superintelligence relies on physical infrastructure that is both capital-intensive and technologically complex, with early movers controlling the next era of technology.

Reflecting on our extensive research in this space, our core beliefs are:

- **AI is a transformative general-purpose technology** that has far-reaching implications. In the medium term, AI has the potential to add more than $10 trillion of total economic value to the global economy through automation. Long-term, AI can lead to an Age of Abundance led by superhuman intelligence and mass robotic automation.
- **Market growth will continue**, driven by scaling laws, declining unit costs driven by model efficiencies, and sovereign demand for nationalized compute.
- **Total AI demand will exceed 100 GW of power in the next 10 years**, based on our adoption forecasts. In our view, this will lead to more than $7 trillion of required capital across core markets.

Like the steam engine, electricity and the telephone before it, AI will transform the global economy, but only if the necessary infrastructure is built to support it at scale. In each past revolution, it was not the breakthrough alone, but the systems that enabled its mass adoption, that changed the course of history. The trajectory of AI may not be linear, but it's unmistakably upward: more data, more compute and more power.

"The productivity advances coming out of AI models in advanced robotics and services over the next 20 years will be unprecedented ... We're in the midst of an enormous investment era."

– Bruce Flatt, CEO of Brookfield

As AI transitions from the headlines to deeply integrated real-world applications, the task of building and operating the requisite infrastructure at scale stands as both a tremendous challenge and the defining investment opportunity of our time. Brookfield's long history in both digital and energy infrastructure positions us exceptionally well to lead this buildout of the physical backbone of AI. Just as we have done for decades in building the railways, power grids and communication networks that enabled the breakthroughs of past industrial revolutions, we are building the systems that will power the AI age.

Neoen, a leading global renewables developer and a Brookfield portfolio company, provides crucial baseload power to its power generation facilities through battery storage.

Appendix: What Is AI?

Artificial intelligence is a general-purpose technology designed to mimic human cognitive and decision-making abilities.

At its heart is software code that learns by reading massive amounts of data and finding patterns. The model then processes those patterns to make predictions based on its learning. These two steps are more generally known as training and inference. Although the term AI was developed over 50 years ago, rapid advancements in both science and access to large compute clusters have accelerated AI innovation, leading to the viral ChatGPT moment in 2022.

Broadly speaking, the capabilities of AI in practice are often described in three categories of increasing sophistication: artificial narrow intelligence, artificial general intelligence and artificial superintelligence (see Figure 11).

Artificial narrow intelligence (ANI)

Artificial narrow intelligence consists of algorithms that excel at specific tasks (e.g., targeted ads and voice recognition) but lack broad adaptability. Today's production AI models fall in this category because they are advanced in a single domain but cannot reliably transfer skills elsewhere. Even cutting-edge LLMs such as OpenAI's GPT-4, Google's Gemini 2.5, xAI's Grok 3, Anthropic's Claude, Meta's Llama or DeepSeek's R1 are still specialists: predicting text, summarizing documents or generating code, but unable to autonomously tackle unfamiliar problems without new training or human guidance. In practice, ANI powers use cases like customer support chatbots, content recommendation engines and advanced driver-assistance (lane-keeping) systems in cars.

Artificial general intelligence (AGI)

Artificial general intelligence refers to a future system with human-level versatility, reasoning, planning and learning across tasks it has never seen. Many researchers argue that reaching true AGI will require the same scale of effective compute increase we have seen in the last half-decade. Given that the compute used to train recent models grew 4-5x each year for the last decade, we have reason to believe that AGI is achievable within this decade if current scaling laws persist.[25]

Early techniques such as "scratchpad" reasoning already reveal sparks of general problem-solving in today's AIs, hinting that the gap from current GPT-4-level performance to human-level competence may close sooner than expected. For example, future AGI could enable autonomous research assistants or multi-skilled digital workers that adapt to new problems on the fly, without explicit reprogramming. Despite these leaps in AI, human-level intelligence doesn't appear to be the end goal but merely a springboard to something more extreme: superintelligence.

Artificial superintelligence (ASI)

Beyond AGI lies artificial superintelligence, a hypothetical phase where AI systems far surpass human cognition in every field, from science and strategy to creativity and emotional intelligence. Once AI can design even better AI, an intelligence explosion feedback loop could compress decades of R&D into months. The result would be a productivity shock: hundreds of millions of tireless virtual "researchers" iterating continuously, likely reshaping every industry. In an ASI scenario, a self-improving AI could deploy millions of cooperating agents to tackle climate change, medical research and real-time global strategic planning simultaneously, at a speed no human organization could match.

Figure 11: **AI Capabilities by Category**

INTELLIGENCE LEVEL	CORE CAPABILITY	EXAMPLES	USE CASES
Artificial narrow intelligence (ANI)	Single-domain proficiency	ChatGPT/GPT-4 class LLMs	Customer support bots, content generation, product recommendations, lane-keeping vehicles
Artificial general intelligence (AGI)	Human-level, cross-domain reasoning	Future LLM systems capable of self-directed problem-solving	Autonomous research assistants, multi-skill digital workers, adaptive robotics
Artificial superintelligence (ASI)	Beyond-human cognition; recursive self-improvement	Hypothetical self-improving LLM running millions of virtual agents	Rapid scientific discovery, global macro-optimization, real-time strategy at planetary scale

Glossary

AI Factories: These large digital hubs house high-performance computing power, specialized hardware such as GPUs, enormous storage capacity and cooling systems that all work together to train and deploy AI models.

Age of Abundance: This vision of the future—fueled by technology breakthroughs like AI—provides low-cost, readily available access to goods, services and opportunities, making it easier for people to achieve their goals and live fulfilling lives.

Agentic AI: These AI systems can plan, strategize and execute tasks independently, often adapting to changing conditions.

Behind the Meter: These power solutions are installed on the consumer's side of the electricity meter, allowing them to generate, store or manage their electricity independently of the grid.

Circular Economy: This is a model of production and consumption that involves sharing, leasing, reusing, repairing, refurbishing and recycling existing materials and products as long as possible. It extends the life cycle of products and reduces waste to a minimum.

Compute: This is shorthand for computational power and refers to the processing capacity of computer systems, especially in performing complex calculations and running software applications.

Frontier AI Labs: These organizations work at the cutting edge of AI, particularly in developing general-purpose or highly capable models that push the boundaries of what AI can do.

Gigawatts (GW): This is a unit of power equal to 1 billion watts—enough to power a midsize city like San Francisco.

Jevons Paradox: This theory, named after economist William Stanley Jevons, suggests that the increasing use of a resource (e.g., electricity) can, paradoxically, lead to even more consumption because efficiency gains often lower the resource's cost and increase its attractiveness.

Moore's Law: This conclusion, made by Intel co-founder Gordon Moore, states that the number of transistors on a chip will double approximately every two years with a minimal increase in cost.

Quantum Computing: This technology uses the principles of quantum mechanics, where subatomic particles can exist in multiple states simultaneously, to perform calculations far faster than traditional computers.

S Curve: All the technological revolutions of the past few decades tend to follow a similar behavior, referred to as the S curve. Technology starts out expensive, bulky and not widely adopted; improvement is slow as the fundamental concepts are being figured out. A period of rapid innovation and massive adoption follows, up to a slowdown in meaningful improvement and fewer new customers. [26]

Scaling Laws: These concepts describe how the performance of an AI model improves as the training data, model parameters and computational power increase.

Scratchpad Reasoning: This technique instructs AI models to show their work and decision-making process before providing a final answer.

System 2 Architecture: This AI technology takes a slow and deliberate approach that is required for planning and deduction tasks, as opposed to System 1 architecture that relies on quick decisions for simpler tasks such as pattern recognition.

Thinking Token: These are the fundamental units of text that AI models use to process and understand language.

Endnotes

[1] Estimated capital requirement over the next 15 years based on Brookfield internal research.

[2] IDC, "The Business Opportunity of AI," November 2023.

[3] Brookfield internal research.

[4] Energy Information Administration, U.S. Bureau of Labor Statistics.

[5] Artificial Analysis, "OpenAI: Models Intelligence, Performance & Price," April 2025.

[6] Brookfield internal research.

[7] Brookfield internal research.

[8] Financial Times, "Microsoft acquires twice as many Nvidia AI chips as tech rivals," December 17, 2024.

[9] Brookfield internal research.

[10] Brookfield internal research.

[11] Brookfield internal research.

[12] Brookfield internal research.

[13] Brookfield internal research.

[14] Google, "New nuclear clean energy agreement with Kairos Power," October 14, 2024.

[15] Brookfield estimates.

[16] Brookfield internal research.

[17] ENVERUS, "Unveiling ISO Dynamics and Market Trends for 2025."

[18] Brookfield estimates.

[19] Meta, "The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation," April 5, 2025.

[20] Artificial Analysis, "DeepSeek: Models Intelligence, Performance & Price," May 2025.

[21] Brookfield internal research.

[22] Brookfield internal research.

[23] Data Center Dynamics, "The path to power," March 14, 2025.

[24] Brookfield internal research.

[25] Epoch AI, "Training Compute of Frontier AI Models Grows by 4-5x per Year," May 28, 2024.

[26] Medium, "Technology's Favorite Curve: The S-Curve (and Why It Matters)," November 2, 2022.

Disclosures

About Brookfield

Brookfield Asset Management is a leading global alternative asset manager, headquartered in New York, with over $1 trillion of assets under management. Brookfield invests client capital for the long term with a focus on real assets and essential service businesses that form the backbone of the global economy. Brookfield offers a range of alternative investment products to investors around the world—including public and private pension plans, endowments and foundations, sovereign wealth funds, financial institutions, insurance companies and private wealth investors.

With more than $200 billion in assets, Brookfield is one of the world's largest investors, owners and operators of critical infrastructure assets. Our portfolio, grounded in 120 years of investment experience, provides diversified exposure to scarce, high-quality businesses with significant barriers to entry. With a global team of approximately 61,000 operating employees and 220 professionals, we own and operate assets across data, transport, utilities and midstream sectors with a focus on cash flow stability and resilience.

Learn more about brookfield.com/our-businesses/infrastructure

Brookfield