

Global Innovation



James Dowey

Co-Head of
Global Innovation



Storm Uru

Co-Head of
Global Innovation



Clare Pleydell-Bouverie

Fund Manager



James O'Connor

Fund Manager

The biggest infrastructure buildout in history: accelerated computing drives a new era

- The computing infrastructure is undergoing a significant re-architecture to accommodate accelerated computing, leading to the largest infrastructure buildout in history. Over the first three years of this ten-year project, more than \$3 trillion in spending is required – an amount equivalent to the size of the United Kingdom's economy¹.
- This re-architecture includes 10,000 data centres, seven billion smartphones, and three billion PCs and laptops, not including new builds. Microsoft is standing up a new data centre every three days to keep up with demand².
- Every 15-20 years over the past 60 years, accumulations of innovations have led to significant shifts in the computing industry, resulting in markets that are ten times the size of their predecessors.
- In 2023, the market share of accelerated servers was only 10%. We expect this to rise to 30% of all servers sold by 2027, indicating that we are in the early stages of this transformative change.
- As supply constraints for accelerated servers ease, enterprises and data centre co-location companies are set to gain access to Nvidia's servers, providing another growth phase for the market.
- New data centres require significantly more electrical equipment, cooling, and power supply to support the increased demands of accelerated computing. Accelerated servers require at least ten times the power of traditional servers.
- We think an unexpected source of demand for edge devices will emerge this year as the upgrade cycle of PCs and smartphones begins, with companies competing for leadership in the next technology platform.

¹ BNP Paribas and Liontrust, 2024

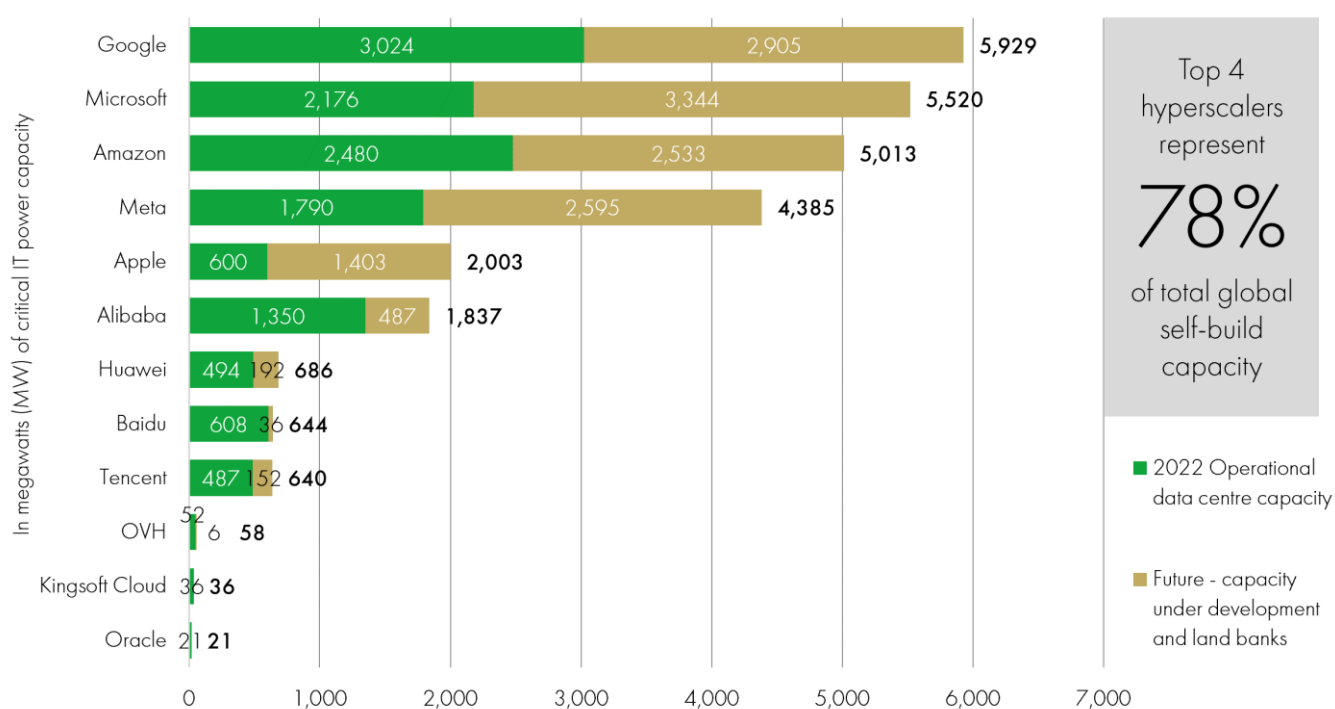
² Gartner, 2024

- New winners of this AI-driven technology cycle are emerging in various subsectors, including advanced packaging, custom application specific integrated circuits (ASICs), and computational software.

In 2022, ChatGPT marked a pivotal moment for artificial intelligence (AI), akin to the iPhone's impact on mobile technology, alerting senior executives across the world's largest companies that a major technology platform shift was underway. Artificial intelligence is now a reality, heralding a new technology cycle. And, much like the advent of electricity in the 19th century, companies today face a critical juncture: adapt or risk extinction. For new companies, this presents a once-in-a-lifetime opportunity; for established ones, it is a matter of survival. To deploy AI at scale, the global computing infrastructure must undergo a significant re-architecture to accommodate accelerated computing, leading to the largest infrastructure buildout in history.

Accelerated computing drastically reduces the cost of compute performance, enabling almost every company to deploy AI at scale. Pioneers, which already have their IT workloads in the cloud, can move quickly, while laggards must first transition to the cloud – currently, only 5% of IT workloads are cloud-based. Therefore, not only do all compute workloads need to be accelerated, but they also need to transition to the cloud. Over the first three years of this ten-year project, more than \$3 trillion in spending is required – an amount equivalent to the size of the United Kingdom's economy. This monumental investment is essential to facilitate this transformative shift in the global computing landscape.

Global self-build data centre capacity split by hyperscaler



Source: Structure Research, 2022

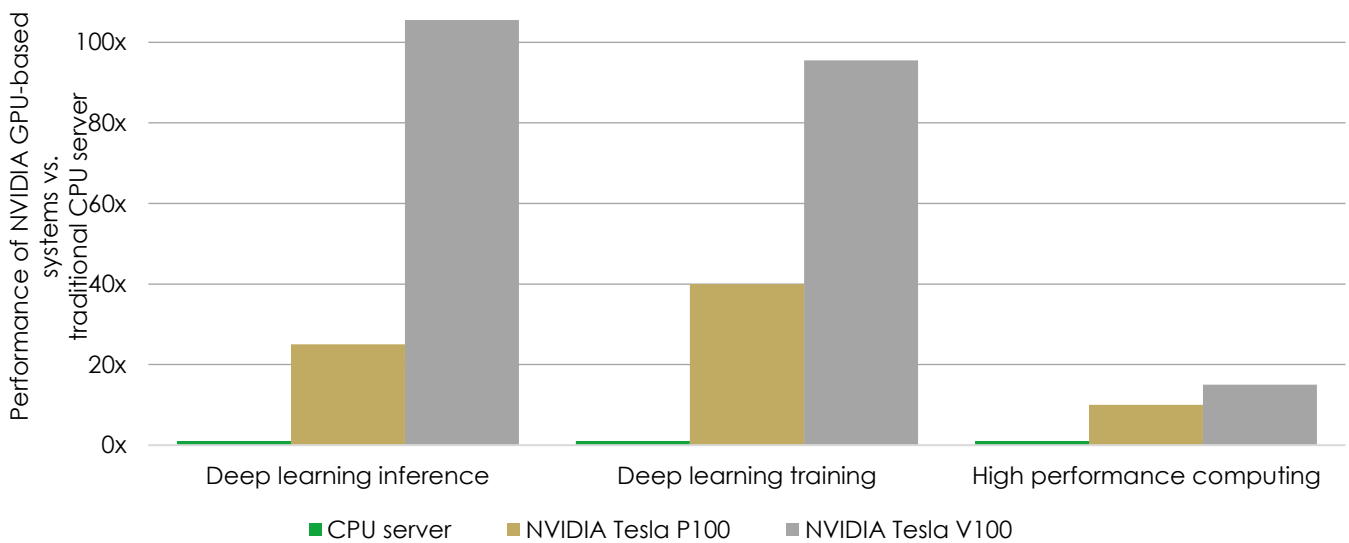
Even in its infancy, this technology is making massive leaps in drug discovery, fission energy, robotics, creativity, communication, and many other areas. Every 15-20 years over the past 60, accumulations of innovations have led to tectonic shifts in the computing industry, which history shows result in markets ten times the size of previous markets³. The shift from traditional computing using serial processing to accelerated computing using parallel processing has just begun. In order to undergo this transition, additional software is required to convert central processing unit (CPU) functions to graphics processing unit (GPU) functions for parallel execution, which explains why hyperscalers are building future data centre capacity.

³ Evercore ISI, 2024

We are in the early stages of a significant "rip out and replace" cycle within data centres, as legacy traditional computing servers can no longer compete in terms of performance and energy efficiency. Every IT workload, from high-performance computing to inference, is transitioning to accelerated computing. This shift is happening regardless of whether the workloads are infused with AI, simply because accelerated computing is significantly cheaper. It's akin to choosing between paying £10 for a McDonald's cheeseburger or £1 for the same product next door.

To illustrate the advancements in accelerated computing, below is a comparison of GPUs from the Pascal (P100, 2016) and Volta (V100, 2017) architectures, which preceded Nvidia's more recent Ampere and Hopper models. Both Volta and Pascal introduced significantly better performance for high-performance workloads, and the latest Hopper and Blackwell architectures are leaps ahead again. Indeed, Blackwell is 1,000 times faster than traditional computing from eight years ago. This transition is no longer a matter of debate.

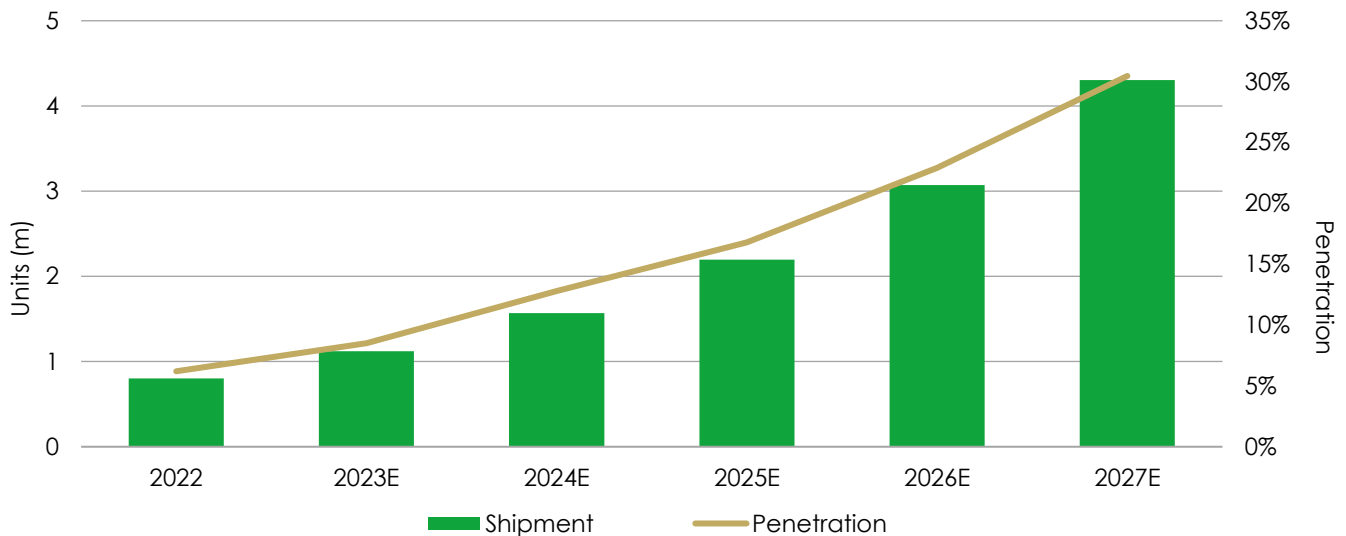
GPU-accelerated servers offer higher levels of computing performance



Source: Nvidia, 2020

As industries race to rearchitect their data centres, hyperscalers hold a first-mover advantage due to their significant buying power. This provides them with a considerable comparative edge, as they possess not only the resources to purchase accelerated computing systems but also the expertise to build increasingly complex data centres at scale and speed. However, as supply constraints in advanced packaging ease (capacity for TSMC's leading chip-on-wafer-on-substrate (CoWoS) offering is set to double in 2024), we expect the availability of GPU systems to improve. Consequently, enterprises are anticipated to become more significant purchasers in the accelerated server market. Currently, we are in the early stages of this buildout, with accelerator server shipments this year expected to achieve only about 13% market penetration. This buildout is projected to take ten years to complete, with accelerated servers anticipated to reach 30% penetration by 2027.

Accelerated computing server shipment



Source: BNP Paribas and Liontrust, 2024

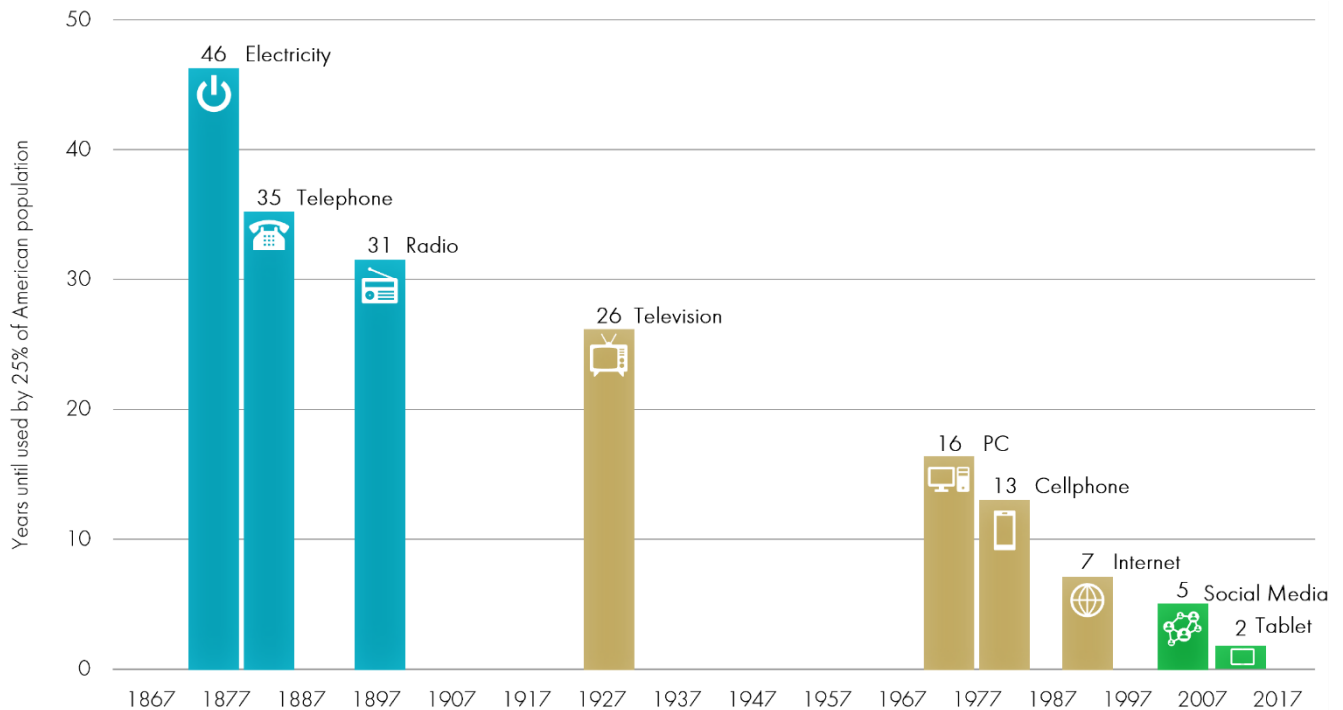
What is the current size of the market and where are we heading? According to New Street Research (NSR), data centre capital expenditure in 2023 was approximately \$300 billion, with \$125 billion coming from the top four hyperscalers. AI server (accelerated computing)-related capex was estimated at around \$90 billion, with significant growth expected. AMD's CEO, Lisa Su, predicts that the AI chip market will reach \$400 billion by 2027. When considering spending on AI networking equipment, buildings, cooling, power, and more, AI-related capex is projected to increase from \$90 billion in 2023 to \$800 billion in 2027. This represents a significant investment, but it is important to contextualise this within the broader trend of AI integration across all business functions. This projection is merely the first phase of the buildout, not the ultimate target. As AI becomes integral to every task, further investment and expansion will be inevitable.

Return on investment (ROI) is accelerating the adoption of AI, providing hyperscalers, data centre co-location providers and enterprises with the confidence to allocate significant capital to new data centre projects, which typically take three years to build. Early indications suggest a very high ROI for AI products. For instance, Microsoft Copilot has a one-day payback period for customers⁴, is margin accretive for Microsoft, and immediately doubles the size of the company's addressable market in the enterprise sector. As companies move from the experimentation stage to the trial stage, we are closely tracking the results. For full AI deployment, we expect 2027 to be a key milestone, marking the completion of the first phase of the infrastructure buildout. These developments highlight the transformative potential of AI and the substantial investments being made to support its integration across various sectors. Early examples include:

1. **Banks** – JP Morgan and Morgan Stanley (90% reduction in manual workflows)
2. **Payments** – Visa (developing the first AI-based fraud detection for real-time payments)
3. **Customer support** – Klarna (AI agent equivalent to 700 call workers, resolving calls 7x quicker with 25% more accuracy, resulting in a \$40 million annualised profit gain)
4. **Customer conversion** – L'Oreal (generative AI beauty assistant increases conversion by 60% compared to in-store advisors)
5. **Recommender systems** – Meta (50% of Instagram content is now AI-recommended)

⁴ Gartner, 2024

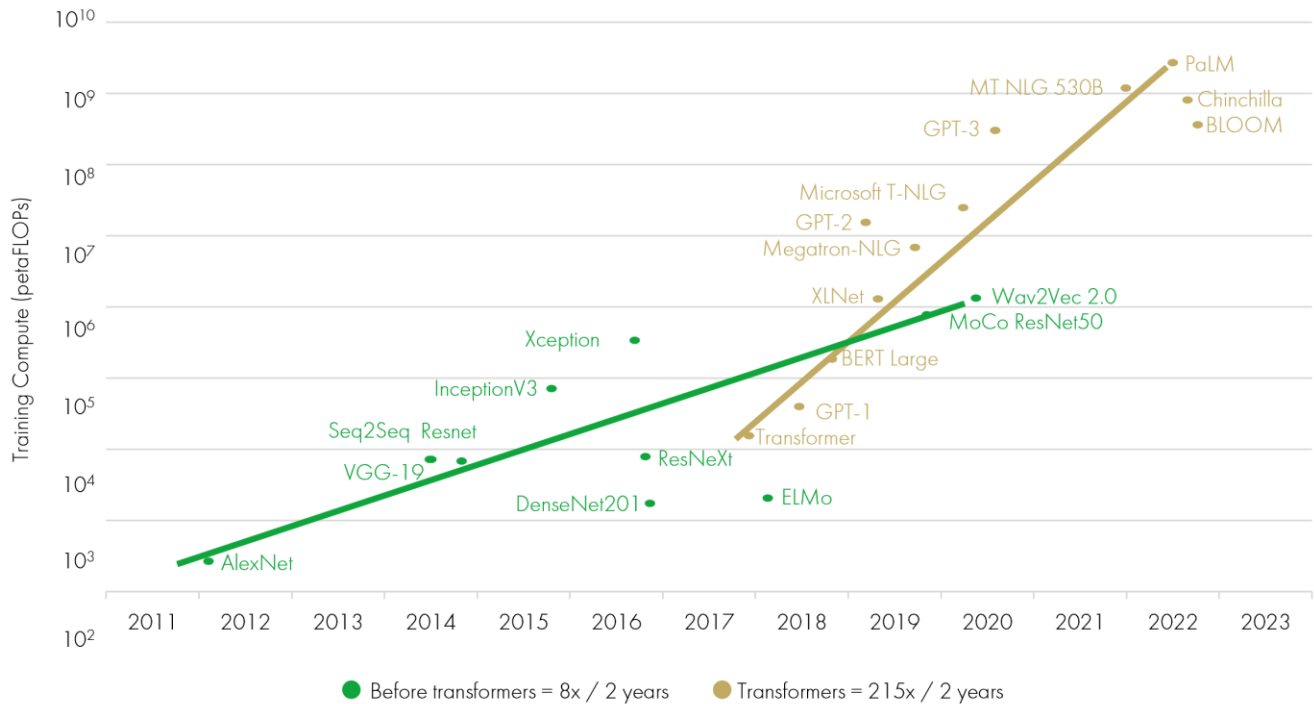
Major new product adoption speeds are accelerating



Source: Liontrust, 2024

Two factors could potentially slow down this buildout. Firstly, there is the possibility that large language models (LLMs) may require less compute to train as the industry races towards Artificial General Intelligence (AGI). However, this scenario seems unlikely at this stage. We are still in the early days of developing and training LLMs, and every indication from industry pioneers suggests that significant increases in training compute will be needed for subsequent versions. The exponential trend in transformer training compute demand, which has increased 215-fold in two years, shows no signs of abating. Additionally, Sam Altman of OpenAI recently suggested that models will require continuous training rather than a simple "complete and deploy" approach, reinforcing the need for ongoing, substantial compute resources.

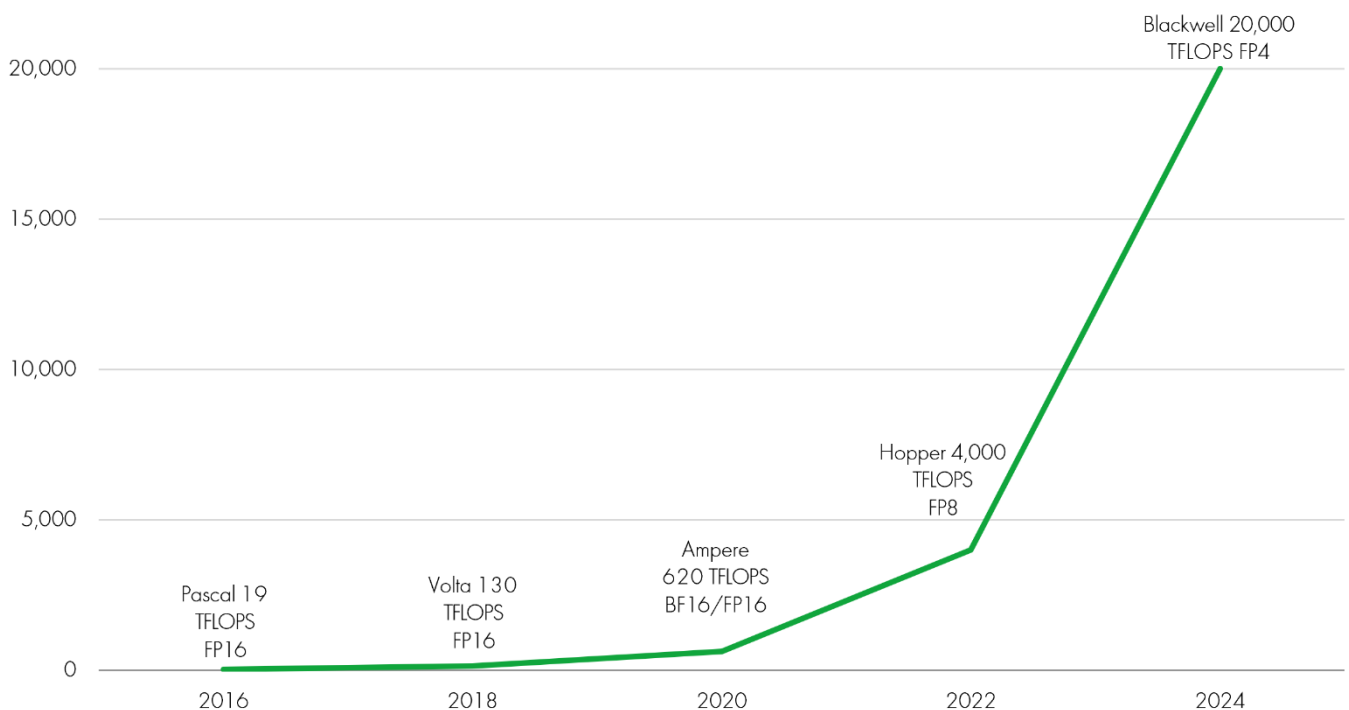
Data centres are becoming AI factories: data as input, intelligence as output



Source: Nvidia, 2024

Secondly, the rate of progress achieved by Nvidia with each new computer architecture shift – from Hopper to Blackwell and beyond – indicates that we are still in the early stages of this technology's development and potential. Nvidia is making increasing strides with each new architecture, suggesting that the pace of advancement in AI hardware is accelerating rather than slowing down. These factors highlight the robust growth and potential challenges in the AI buildout, emphasising the importance of sustained investment and innovation in AI infrastructure.

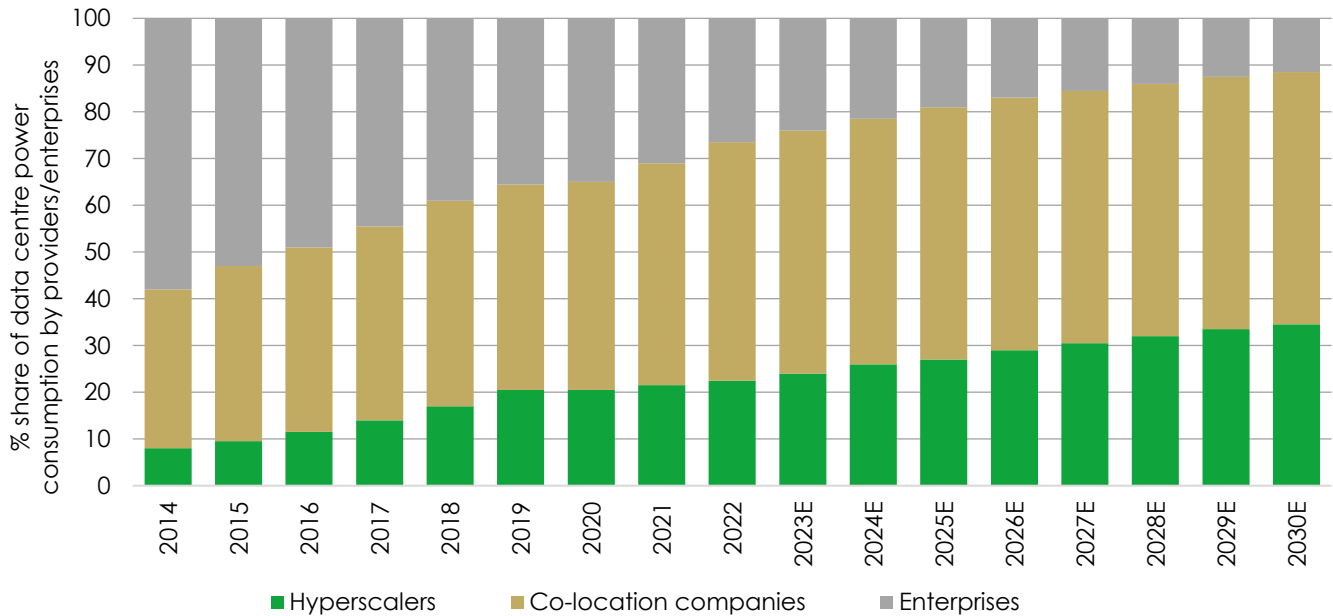
1000x AI compute in 8 years



Source: Nvidia, 2024

But who is funding this infrastructure buildout? Hyperscalers are first in line, but they represent only about 25% of the overall data centre market. Next in line are co-location companies and enterprises like JP Morgan, Visa, and SoftBank Corp. However, we should not underestimate the financial resources that the hyperscalers possess. For instance, Microsoft has little debt and \$80 billion in cash. In the last 12 months, despite a significant ramp-up in compute capex, they have spent less than 30% of their operating cash flow on cloud capex and still generated \$70 billion in free cash flow (FCF). This financial strength enables hyperscalers to continue investing heavily in the infrastructure required for AI development and deployment.

Hyperscalers & co-location companies are an increasing proportion of data centres

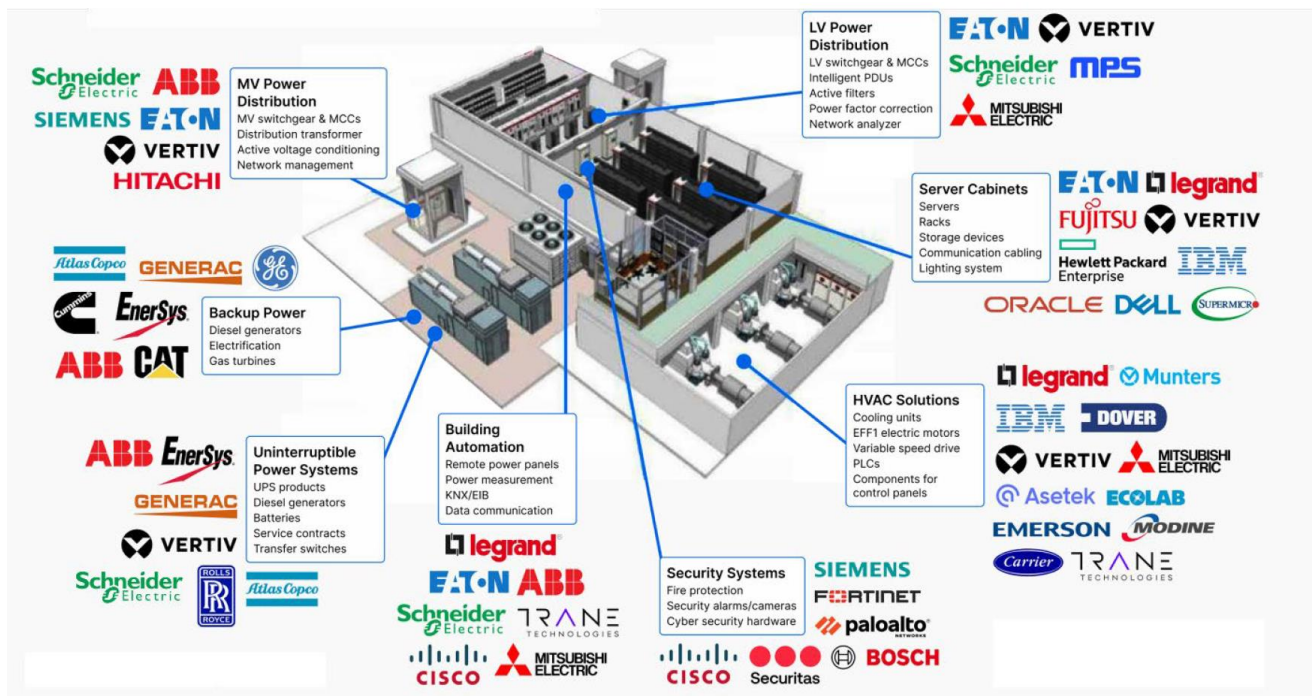


Source: McKinsey, 2023

What is often overlooked is that 50% of the capital expenditure of a data centre is not related to GPUs/CPUs, networking, or switches, but rather to land, buildings and numerous other components. Excluding the servers, the most significant difference in infrastructure required for accelerated computing compared to traditional computing is the substantial increase in power demand. The power consumption of accelerated servers compared to traditional servers jumps from around 1kW to 10kW for the H100, without even accounting for another leap with Blackwell. This necessitates rearchitecting data centres to cope with the increased power demands, creating opportunities for many new winners in the market. Industrial companies like Schneider, Vertiv, and Eaton are well-positioned to benefit from these changes. Below is a general layout of the data centre with suppliers of different components. Alongside product innovation, the key bottleneck in this buildout is the power supply.

Schneider competes within LV/MV (low voltage/medium voltage) distribution, uninterruptible power supply (UPS), software/automation and cooling

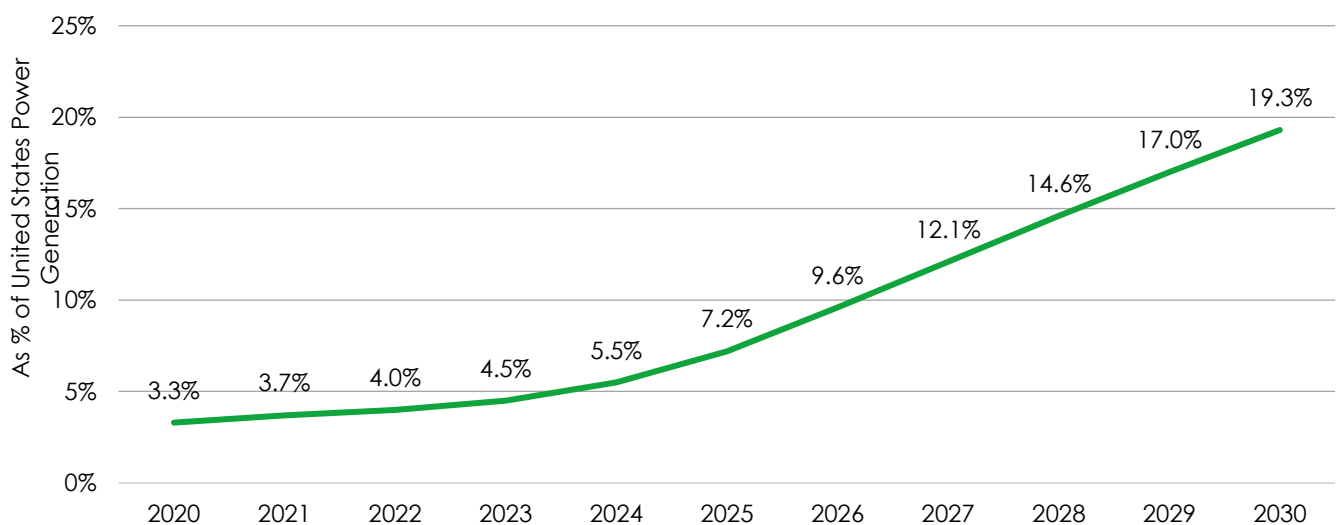
Summary of data centre power, cooling and security equipment and a selection of key providers



Source: Quartr, 2024. All use of company logos, images or trademarks in this presentation are for reference purposes only.

The tenfold increase in power consumption for accelerated servers combined with the construction of new data centres will significantly increase the power load on the electrical grid. Dylan Patel from SemiAnalysis has modelled the anticipated power load required for US data centre buildouts, showing that US power demand is set to start growing again after 15 years of stagnation. The power usage of data centres as a percentage of total US power generation is expected to rise significantly. However, with net-zero commitments in place, the ability to build low-cost gas turbines to capitalise on the US's affordable gas supply is limited.

US AI and non-AI data centre power usage as % of total US power generation



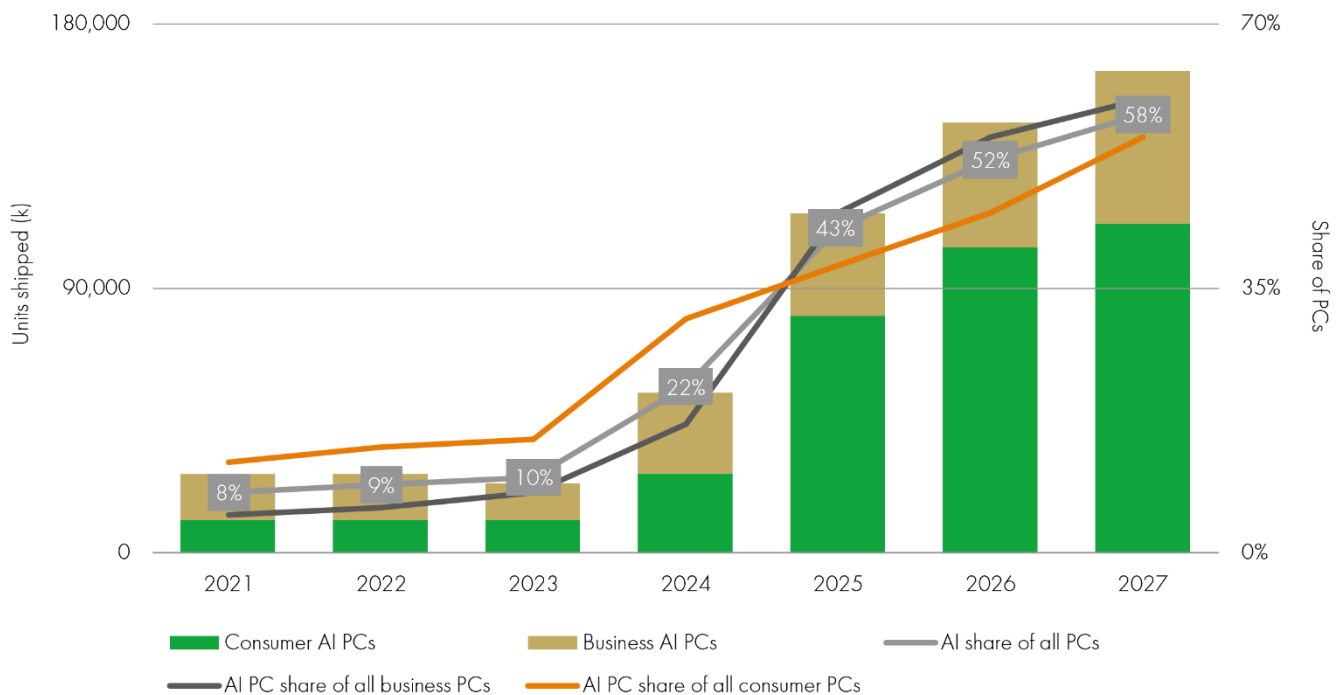
Source: SemiAnalysis, 2024

Consequently, the data centre industry will turn to reliable renewable energy sources. The scale of this compute infrastructure build necessitates a corresponding expansion in power capacity, and the first major deal has recently been announced by Microsoft. Microsoft has agreed to support an estimated \$10 billion in renewable

electricity projects to be developed by Brookfield Asset Management. This commitment aims to bring 10.5 gigawatts of generating capacity online, enough to power approximately 1.8 million homes. Brookfield stated that this capacity is about eight times larger than the previous largest corporate renewable purchase agreement. This presents a tremendous opportunity for companies such as Constellation Energy, Brookfield Renewables, and NextEra, which have substantial renewable energy generation bases and significant plans to expand capacity.

As the infrastructure buildout begins, the upgrade cycle for PCs and smartphones will also commence. Often overlooked, the demand for faster compute at the edge will catalyse a significant replacement cycle, with a short lead time. Upgrading to Windows 11, which requires more memory to run, will be the first domino to fall, the adoption of co-pilots and edge inferencing will be the second.

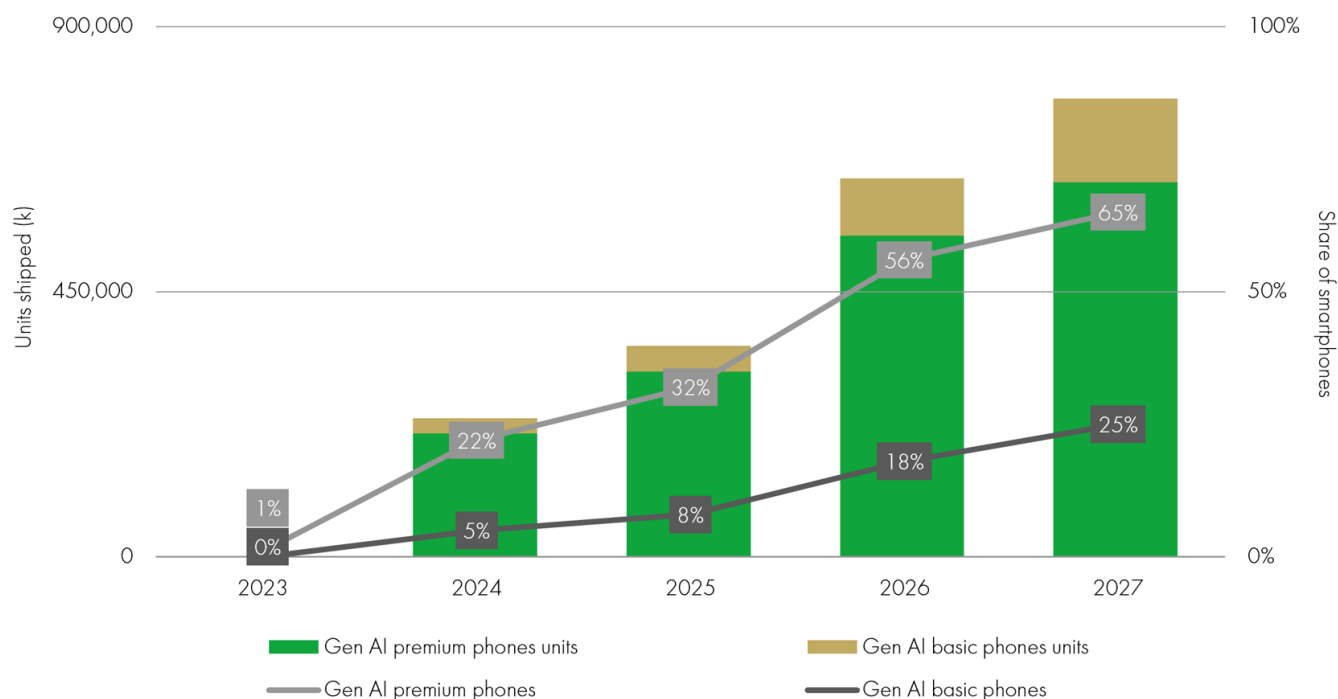
AI PC unit shipments worldwide: 2021-2027



Source: Gartner, 2024

In the smartphone market, software upgrades to Apple's iOS 18 and Android 15 will impose higher energy demands and have limited utility without a dedicated neural processing unit (NPU). Gartner charts illustrate the approaching upgrade cycle, with new software updates and AI applications compelling users with older smartphones to upgrade. We believe this upgrade cycle is likely to be faster and more substantial than Gartner's predictions due to the rapid progress of key AI applications. The convergence of these factors should drive a significant and accelerated upgrade cycle, impacting both the PC and smartphone markets.

GenAI smartphone unit shipments worldwide 2021-2027



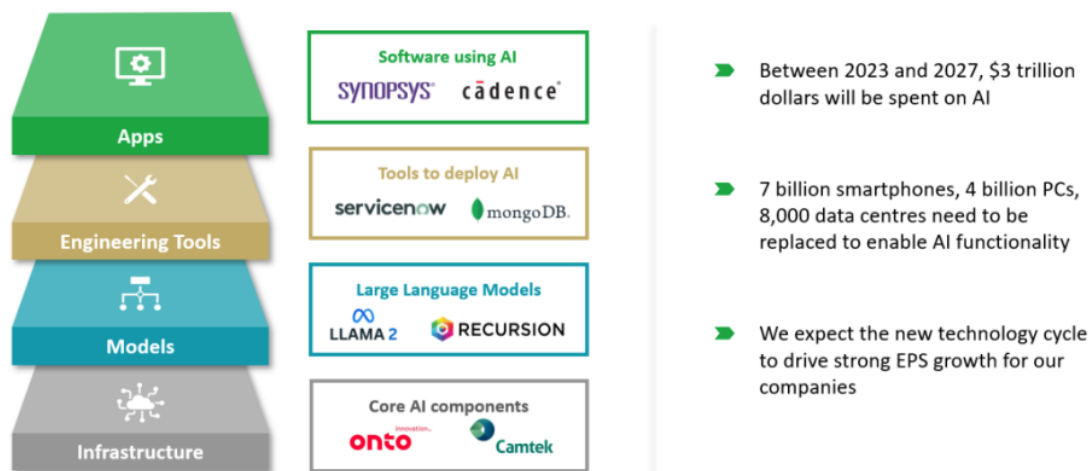
Source: Gartner, 2024

New companies are set to emerge as winners of the AI-driven technology cycle, because an entirely new technology stack is required, opening up the playing field. At the start of a major technological breakthrough, the value is captured by the infrastructure layer before moving up the stack. A unique opportunity lies in small to medium companies plugged into new strong structural growth drivers, expected to compound growth over 30% per annum over the next five years. Advanced packaging companies like Onto Innovation, Camtek, and Besi are well-positioned to become critical suppliers as companies build out capacity to meet Nvidia's demands.

Additionally, as a portion of AI workloads transition to custom AI chips, we expect the custom application specific integrated circuit (ASIC) market to experience tremendous growth from a small base. Key players like Broadcom, Marvell, and AI chip companies are well-positioned to enable companies to move workloads from Nvidia AI servers, lowering the total cost of ownership and driving differentiation. This market is expected to compound at over 30% per annum over the next five years.

AI is also enabling computational design software to make great strides in driving semiconductor performance and energy efficiency, simultaneously lowering the technological expertise required and cost of design. This democratisation – powered by electronic system designers Cadence and Synopsys – means that computational software can now be used in a growing number of industries, from drug discovery to shipbuilding to designing new factories. Tools used by pioneering semiconductor companies will be used across the economy within the next decade, unlocked by cheaper compute power and AI. This presents exciting demand from key industries including healthcare, energy, design, and engineering.

New technology stack for AI



Source: Gartner 2023 and Liontrust 2024. All use of company logos, images or trademarks are for reference purposes only.

We are in the first five minutes of the biggest infrastructure buildout in history, leading to the most innovative decade ever. While large technology companies are grabbing the headlines, a new cohort of companies is set to thrive in this AI-driven technology cycle. The structural drivers are clear, driven by the collapse in computing costs enabling major technological breakthroughs. Just like electricity in the 19th century, we expect this to seed many new industries, with Generative AI being just the first. It's time to build, and there are great opportunities for the most innovative companies in the world.

Key Risks

Past performance does not predict future returns. You may get back less than you originally invested. We recommend any fund is held long term (minimum period of 5 years). We recommend that you hold funds as part of a diversified portfolio of investments.

The Funds managed by the Global Innovation Team:

May hold overseas investments that may carry a higher currency risk. They are valued by reference to their local currency which may move up or down when compared to the currency of a Fund.

May have a concentrated portfolio, i.e. hold a limited number of investments. If one of these investments falls in value this can have a greater impact on a Fund's value than if it held a larger number of investments.

May encounter liquidity constraints from time to time. The spread between the price you buy and sell shares will reflect the less liquid nature of the underlying holdings.

Outside of normal conditions, may hold higher levels of cash which may be deposited with several credit counterparties (e.g. international banks). A credit risk arises should one or more of these counterparties be unable to return the deposited cash.

May be exposed to Counterparty Risk: any derivative contract, including FX hedging, may be at risk if the counterparty fails.

Do not guarantee a level of income.

The issue of units/shares in Liontrust Funds may be subject to an initial charge, which will have an impact on the realisable value of the investment, particularly in the short term. Investments should always be considered as long term.

The risks detailed above are reflective of the full range of Funds managed by the Global Innovation Team and not all of the risks listed are applicable to each individual Fund. For the risks associated with an individual Fund, please refer to its Key Investor Information Document (KIID)/PRIIP KID.

Disclaimer

This document is issued by Liontrust Fund Partners LLP (2 Savoy Court, London WC2R 0EZ), authorised and regulated in the UK by the Financial Conduct Authority (FRN 518165) to undertake regulated investment business.

It should not be construed as advice for investment in any product or security mentioned, an offer to buy or sell units/shares of Funds mentioned, or a solicitation to purchase securities in any company or investment product. Examples of stocks are provided for general information only to demonstrate our investment philosophy. The investment being promoted is for units in a fund, not directly in the underlying assets.

The document contains information and analysis that is believed to be accurate at the time of publication but is subject to change without notice. Whilst care has been taken in compiling the content of this document, no representation or warranty is given, whether express or implied, by Liontrust as to its accuracy or completeness, including for external sources (which may have been used) which have not been verified.

This is a marketing communication. Before making an investment, you should read the relevant Prospectus and the Key Investor Information Document (KIID) and/or PRIIP/KID, which provide full product details including investment charges and risks. These documents can be obtained, free of charge, from www.liontrust.co.uk or direct from Liontrust. If you are not a professional investor please consult a regulated financial adviser regarding the suitability of such an investment for you and your personal circumstances. All use of company logos, images or trademarks in this document are for reference purposes only.