

Understanding the AI Infrastructure Buildout

A guide to the industry behind the headlines — and the companies building it

September 2026

"Artificial intelligence" has become shorthand for a stock market story, but underneath the ticker symbols is one of the largest physical construction projects in economic history. Training and running today's AI models requires purpose-built computer chips, power plants, cooling systems, and data centers on a scale that did not exist five years ago — and companies across dozens of industries are being reshaped by the effort to build it. This note walks through what is actually being built, who is building it, and how the pieces fit together, so that the next headline about a chip company's earnings or a utility's new power contract has a clear place in the bigger picture.

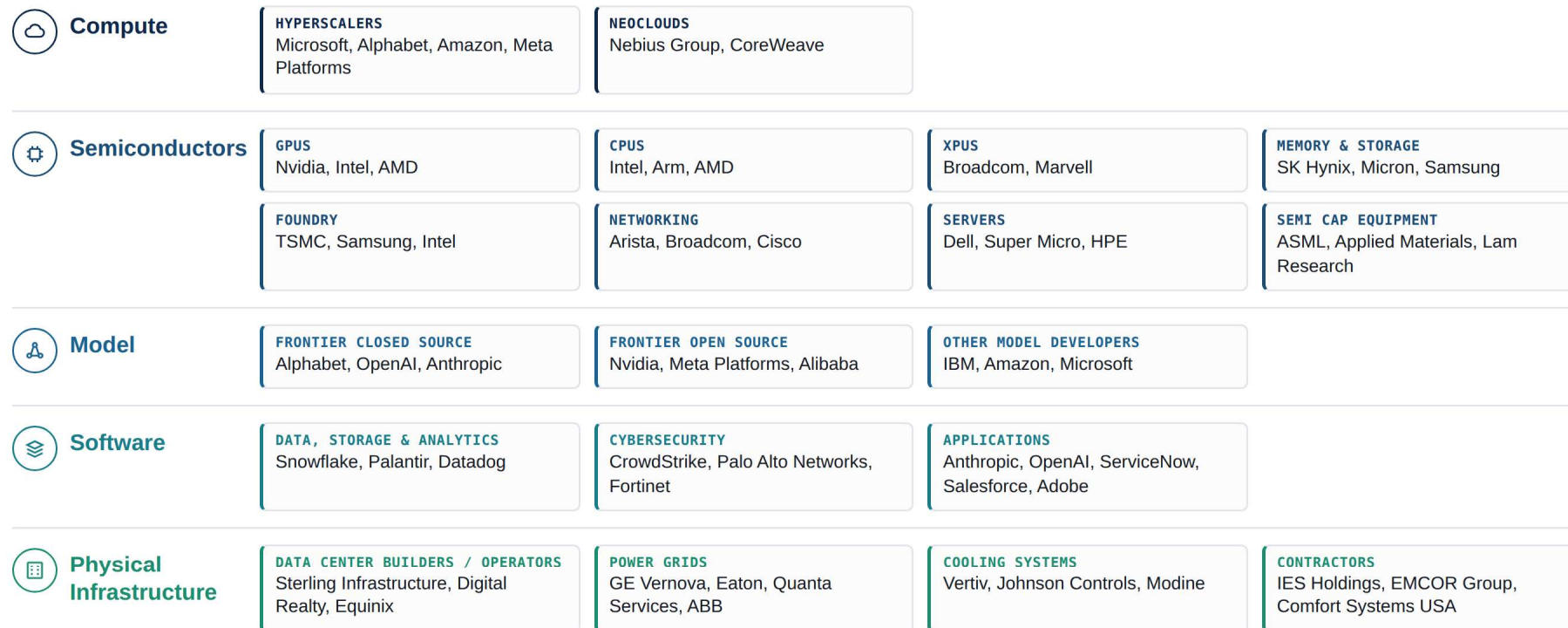
Five Layers, One Machine

It helps to think of the AI industry as a stack of five layers, each dependent on the one below it. At the top sits compute — the raw processing power that trains and runs AI models, supplied either by familiar hyperscalers (Microsoft, Amazon, Google, Meta) out of their own data centers, or by a newer class of specialist "neocloud" providers, such as CoreWeave and Nebius, that rent out AI-specific computing capacity to whoever needs it. Underneath compute sits semiconductors — the physical chips, memory, networking gear, and servers that make computation possible, dominated by designers such as Nvidia, AMD, and Broadcom, and by the foundries and equipment makers, including Taiwan Semiconductor Manufacturing Company (TSMC), ASML, Applied Materials, and Lam Research, that actually manufacture them.

On top of that hardware sits the model layer — the large language models and related systems built by labs such as OpenAI, Anthropic, Google, and Meta, which is probably what most people picture when they hear the word "AI." Above the models sits software: the tools that turn a raw model into something a business can use, spanning data infrastructure (Snowflake, Palantir, Datadog), security (CrowdStrike, Palo Alto Networks, Fortinet), and applications (Salesforce, ServiceNow, Adobe, and the labs' own products). And underneath everything, often overlooked, sits physical infrastructure — the buildings, power grids, cooling systems, and construction crews that make it possible to put all of that hardware somewhere and keep it running.

The exhibit on the following page maps this stack layer by layer, with a sample of the public companies operating in each.

AI Layers Map



For informational purposes only; not a recommendation to buy or sell any security. Company inclusion is illustrative of an industry segment and does not indicate a current holding or future profitability.

Illustrative mapping of major public companies operating in each layer of the AI stack. Not an exhaustive list, and not a recommendation to buy or sell any security.

The Scale of What Is Being Built

Goldman Sachs Research estimates that global AI-related investment will exceed \$1 trillion in 2026 alone, with roughly \$581 billion of that in the United States^[1] — a figure that captures not just the well-known hyperscaler capital budgets, which several industry trackers now put north of \$700 billion for the largest cloud companies alone^[5,6], but also private AI companies, chipmakers, and international spending, particularly out of Asia. Goldman's researchers note that AI capital spending is on pace to rise from roughly 1.8% of U.S. GDP in 2026 to 2.8% by 2028^[1], a trajectory they compare to the peak years of prior transformative technology buildouts, such as electrification and the early internet.

A large and growing share of that spending is not going to chips at all — it is going to power. The International Energy Agency projects that global data center electricity consumption will more than double by 2030, and that data centers will account for more than a fifth of all electricity demand growth worldwide over that period; in the United States, it is expected to be closer to half.^[2,3] Gartner, for its part, expects data center electricity demand to jump 26% in 2026 alone.^[4] That is one reason names with no obvious connection to software — utilities, power-equipment makers, even an oilfield-services company that recently acquired a data-center cooling specialist — have become part of the AI conversation.

How the Pieces Connect

The relationships between these companies run in both directions rather than forming a simple hierarchy. A model lab such as OpenAI or Anthropic needs enormous amounts of compute to train and serve its models; it obtains that compute by renting it from a hyperscaler's cloud, contracting directly with a neocloud, or, increasingly, taking an equity stake in — or signing a multi-year capacity commitment with — the very infrastructure providers it depends on. Those compute providers, in turn, place correspondingly enormous orders for chips: largely GPUs from Nvidia, but a growing share of custom silicon ("XPU") designed in-house by the hyperscalers themselves with help from chip designers such as Broadcom and Marvell, a way of reducing dependence on any single supplier and lowering the cost per unit of compute over time.

Those chips, whoever designs them, are almost all physically manufactured by a small number of foundries, led overwhelmingly by TSMC, using equipment that in turn comes from an even smaller number of specialist toolmakers such as ASML, Applied Materials, and Lam Research. Once chips exist, they still need memory to work with — a market currently so tight that AI demand is pushing up memory prices across the entire technology industry, benefiting suppliers such as Micron, SK Hynix, and Samsung while squeezing everyone else who needs a memory chip for an unrelated product.

From there, chips are assembled into servers and networking equipment by companies such as Dell, Super Micro, and Hewlett Packard Enterprise, wired together by networking specialists such as Arista and Cisco, and finally installed in physical data centers built and operated by infrastructure companies such as Digital Realty, Equinix, and Sterling Infrastructure. Those buildings need power delivered by utilities and power-

equipment makers (GE Vernova, Eaton, Quanta Services, ABB), kept cool by thermal-management specialists (Vertiv, Johnson Controls, Modine), and physically constructed and wired by electrical and mechanical contractors (EMCOR, Comfort Systems USA, IES Holdings). Once all of that exists and a model is running, the software layer is what actually turns it into a product a business will pay for — which is why so much of the visible “AI” software boom, in tools from Salesforce, ServiceNow, Microsoft, and Adobe, is really a story about monetizing everything built underneath it.

Dynamics Worth Watching

A few dynamics are worth understanding because they shape how durable — and how risky — this buildout might be. One is the increasingly circular nature of the financing behind it: chipmakers investing directly in the cloud companies and model labs that buy their chips, cloud providers signing multi-year, multibillion-dollar compute contracts with the same labs they have invested in, and model labs committing to spend that in some cases exceeds their current revenue by a wide margin. None of this is necessarily unsound — vertical investment between suppliers and customers is common in capital-intensive industries — but it does mean a slowdown in AI demand anywhere in the chain could echo through several companies’ earnings at once, rather than being contained to a single name.

A second dynamic is the shift toward custom silicon: hyperscalers designing their own AI chips rather than relying solely on Nvidia, a trend that has been a significant tailwind for chip designers such as Broadcom and Marvell and a genuine, if still early, competitive question for Nvidia’s dominance. A third is that the binding constraint on this buildout is shifting from chip supply toward power and physical construction — permitting a new substation or securing a grid interconnection now often takes longer than manufacturing the chips that will eventually sit inside the building, which is why utilities and construction-adjacent names have become somewhat unlikely AI beneficiaries. And a fourth is simple concentration: a remarkably small number of companies — essentially one dominant foundry, one dominant GPU designer, and a handful of hyperscalers — sit at structural chokepoints in this entire value chain, so their individual decisions, earnings, and capital budgets carry outsized weight for the whole ecosystem.

What This Means for a Portfolio

The most useful mental model here is probably not “AI stock” as a single category, but five layers with meaningfully different risk and return characteristics. The physical infrastructure and semiconductor-equipment layers tend to be capital-intensive and cyclical, with returns tied closely to the pace of construction and chip-cycle timing. The chip-design layer has historically offered high margins alongside concentrated competitive risk. The model layer is the most speculative — capital-intensive, often pre-profit, and dependent on continued access to financing. The software and applications layer, sitting furthest from the physical buildout, tends to be more asset-light with the potential for durable margins once a product finds its footing, but it is also the layer most exposed to the risk that the underlying models become commoditized.

Thinking in terms of layers, rather than chasing whichever name is in the headlines that week, is generally a more useful way to gauge where a given holding sits in the cycle — and what has to keep happening in the rest of the stack for that holding to do well. As always, we are happy to walk through how any of this connects to your specific portfolio and goals; please reach out anytime with questions.

References

Superscript numbers in the text above correspond to the sources below.

- [1] Goldman Sachs Research, "Global AI Investment Is Forecast to Exceed \$1 Trillion in 2026," 2026. <https://www.goldmansachs.com/insights/articles/global-investment-is-forecast-to-exceed-1-trillion-in-2026>
 - [2] International Energy Agency, "Energy Demand from AI," Energy and AI, 2026. <https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai>
 - [3] S&P Global Commodity Insights, "Global Data Center Power Demand to Double by 2030 on AI Surge: IEA," 2026. <https://www.spglobal.com/energy/en/news-research/latest-news/electric-power/041025-global-data-center-power-demand-to-double-by-2030-on-ai-surge-iea>
 - [4] Gartner, "Gartner Says Data Center Electricity Consumption to Grow 26% in 2026," press release, 2026. <https://www.gartner.com/en/newsroom/press-releases/2026-06-10-gartner-says-data-center-electricity-demand-to-grow-26-percent-in-2026>
 - [5] CNBC, "Tech AI Spending Approaches \$700 Billion in 2026, Cash Taking Big Hit," 2026. <https://www.cnbc.com/2026/02/06/google-microsoft-meta-amazon-ai-cash.html>
 - [6] TMT Finance, "2026 Hyperscaler Capex Tops US\$700bn — Analysis," 2026. <https://www.tmtfinance.com/intel/2026-hyperscaler-capex-tops-us700bn-analysis>
-

Important Information

This material is for informational and educational purposes only. It does not constitute investment advice, and nothing in it should be interpreted as a recommendation to buy, sell, or hold any security. Company names are included as illustrative examples of a segment of the AI industry, should not be assumed to be current portfolio holdings, and their inclusion is not indicative of — nor a guarantee of — future performance or profitability. Figures cited are drawn from third-party sources (Goldman Sachs Research, the International Energy Agency, Gartner, and industry trackers) believed to be reliable as of the date of this note but have not been independently verified and are subject to change. Past performance is not indicative of future results. Please consult your advisor before making any investment decision.

The opinions voiced in this material are for general information only and are not intended to provide specific advice or recommendations for any individual. All performance referenced is historical and is no guarantee of future results. The economic forecasts set forth in this material may not develop as predicted and there can be no guarantee that strategies promoted will be successful. Stock investing includes risks, including fluctuating prices and loss of principal. Any company names noted herein are for educational purposes only and not an indication of trading intent or a solicitation of their products or services. LPL Financial doesn't provide research on individual equities. Because of their narrow focus, investments concentrated in certain sectors or industries will be subject to greater volatility and specific risks compared with investing more broadly across many sectors, industries, and companies.