Note: The following is a redacted version of the original report published June 25, 2024 [32 pgs].

Goldman Sachs | Global Macro Research

ISSUE 129 | June 25, 2024 | 5:10 PM EDT

TOP of MIND

GEN AI: TOO MUCH SPEND, TOO LITTLE BENEFIT?

Tech giants and beyond are set to spend over $1tn on AI capex in coming years, with so far little to show for it. So, will this large spend ever pay off? MIT's Daron Acemoglu and GS' Jim Covello are skeptical, with Acemoglu seeing only limited US economic upside from AI over the next decade and Covello arguing that the technology isn't designed to solve the complex problems that would justify the costs, which may not decline as many expect. But GS' Joseph Briggs, Kash Rangan, and Eric Sheridan remain more optimistic about AI's economic potential and its ability to ultimately generate returns beyond the current "picks and shovels" phase, even if AI's "killer application" has yet to emerge. And even if it does, we explore whether the current chips shortage (with GS' Toshiya Hari) and looming power shortage (with Cloverleaf Infrastructure's Brian Janous) will constrain AI growth. But despite these concerns and constraints, we still see room for the AI theme to run, either because AI starts to deliver on its promise, or because bubbles take a long time to burst.

"Given the focus and architecture of generative AI technology today... truly transformative changes won't happen quickly and few—if any—will likely occur within the next 10 years.

\- Daron Acemoglu

Spending is certainly high today in absolute dollar terms. But this capex cycle seems more promising than even previous capex cycles.

\- Kash Rangan

AI technology is exceptionally expensive, and to justify those costs, the technology must be able to solve complex problems, which it isn't designed to do.

\- Jim Covello

[AI] dollars spent vs. company revenues… are not materially different than those of prior investment cycles.

\- Eric Sheridan"

WHAT'S INSIDE

Allison Nathan | allison.nathan@gs.com   Jenny Grimberg | jenny.grimberg@gs.com   Ashley Rhodes | ashley.rhodes@gs.com

Macro news and views

We provide a brief snapshot on the most important economies for the global markets

US

Latest GS proprietary datapoints/major changes in views

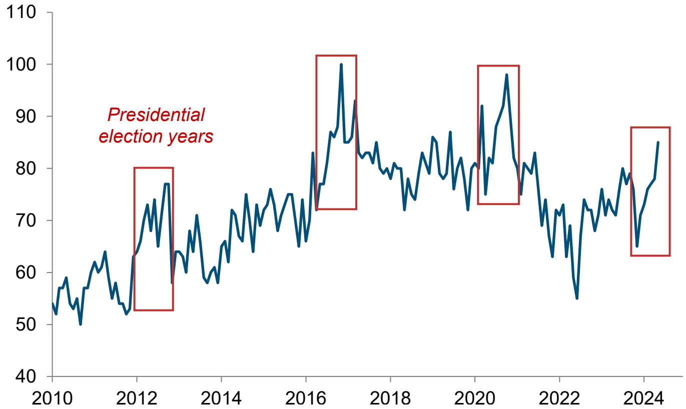
- No major changes in views.

Datapoints/trends we're focused on

- Fed policy; we expect quarterly Fed rate cuts beginning in September, for a total of two cuts this year.
- Inflation; we expect core PCE inflation to stand at 2.7% yoy by Dec 2024 before converging toward 2% next year.
- Growth; we think most of the slowdown from the 4.1% pace of real GDP growth in 2H23 is here to stay given softer real income growth, lower consumer sentiment, and election-related uncertainty that could weigh on business investment.
- Labor market, which is now fully rebalanced, likely meaning that a material softening in labor demand would hit actual jobs.

Election uncertainty: a potential growth drag

NFIB Small Business Uncertainty Index



Source: Haver Analytics, Goldman Sachs GIR.

Japan

Latest GS proprietary datapoints/major changes in views

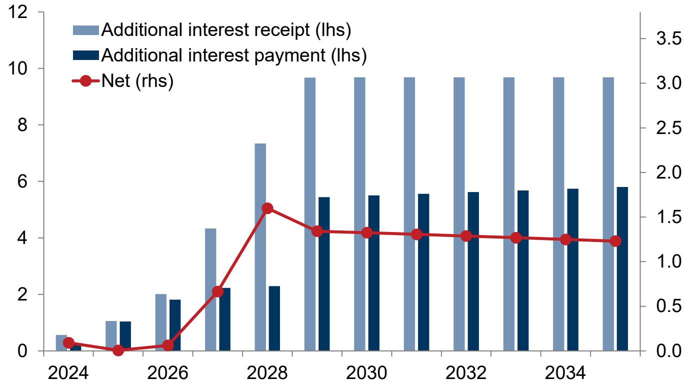
- We now expect the next BoJ rate hike in July (vs. Oct before) as the hurdle for the next hike is low given that it will likely be only 15bp and the BoJ sees the policy rate as significantly lower than the current nominal neutral rate.

Datapoints/trends we're focused on

- Japanese inflation; sequential core inflation has recently shown signs of weakness, but we expect core inflation to remain above the BoJ's target this year at 2.6% yoy.
- Japan's rising interest burden, which will likely be manageable for households and corporates given that it is occurring against a backdrop of solid activity and steady wage growth.
- Japan financial conditions, which continue to ease.

Japanese households: net interest receivers

Interest payments & receipts (lhs, ¥tn), net interest receipt (rhs, % of 2023 disposable income)



Source: Goldman Sachs GIR.

Europe

Latest GS proprietary datapoints/major changes in views

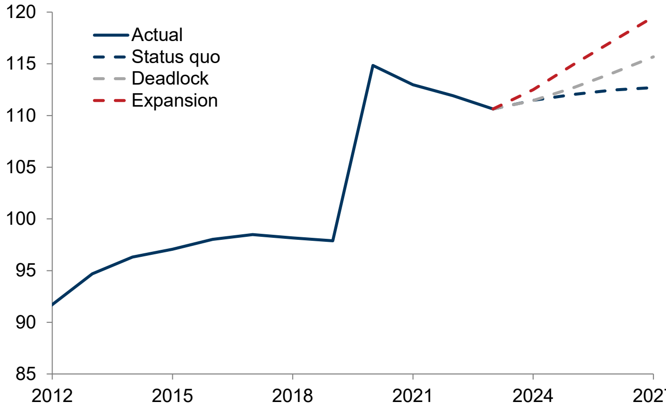
- We raised our 2024 UK GDP growth forecast to 0.9% (from 0.8%) following slightly above consensus April GDP data.

Datapoints/trends we're focused on

- ECB policy; we expect the next rate cut in Sept, though we think a pause in the easing cycle is possible if inflation and wage data surprise to the upside over the summer.
- BoE policy; we expect the BoE to embark on rate cuts in August on the back of renewed UK disinflation progress.
- French snap elections (Jun 30), which could result in a fiscal expansion that would lead the debt-to-GDP ratio to rise.
- UK general election (Jul 4), which will likely deliver relatively similar fiscal outcomes irrespective of which party wins.

French election: upside risk to debt trajectory

French government debt, % of GDP



Source: Goldman Sachs GIR.

Emerging Markets (EM)

Latest GS proprietary datapoints/major changes in views

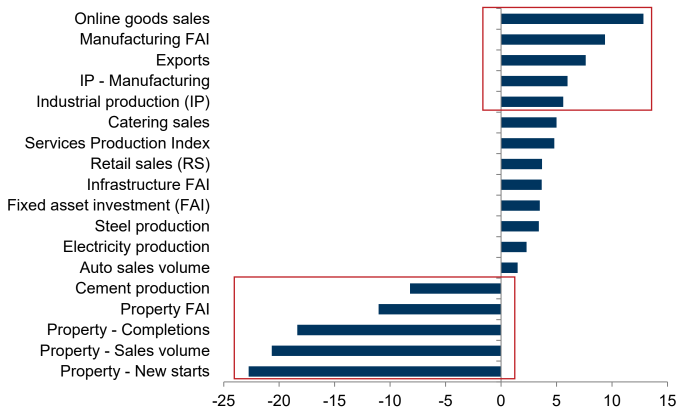
- We recently pushed back our PBOC easing forecasts by one quarter given ample near-term liquidity, and now expect a 25bp RRR cut in Q3 and a 10bp policy rate cut in Q4.

Datapoints/trends we're focused on

- China's economy, which remains bifurcated between strength in exports and manufacturing activity and weakness in housing and credit, coupled with very low inflation.
- EM easing cycle; we think the fundamental case for further EM rate cuts remains strong, though the recent unwind in EM FX carry trades following electoral surprises in Mexico, India, and South Africa could impede policy normalization.

China: a bifurcated economy

China activity indicator, % change, yoy



Source: Haver Analytics, Goldman Sachs GIR.

Gen AI: too much spend, too little benefit?

The promise of generative AI technology to transform companies, industries, and societies continues to be touted, leading tech giants, other companies, and utilities to spend an estimated ~$1tn on capex in coming years, including significant investments in data centers, chips, other AI infrastructure, and the power grid. But this spending has little to show for it so far beyond reports of efficiency gains among developers. And even the stock of the company reaping the most benefits to date—Nvidia—has sharply corrected. We ask industry and economy specialists whether this large spend will ever pay off in terms of AI benefits and returns, and explore the implications for economies, companies, and markets if it does, or if it doesn't.

We first speak with Daron Acemoglu, Institute Professor at MIT, who's skeptical. He estimates that only a quarter of AI-exposed tasks will be cost-effective to automate within the next 10 years, implying that AI will impact less than 5% of all tasks. And he doesn't take much comfort from history that shows technologies improving and becoming less costly over time, arguing that AI model advances likely won't occur nearly as quickly—or be nearly as impressive—as many believe. He also questions whether AI adoption will create new tasks and products, saying these impacts are "not a law of nature." So, he forecasts AI will increase US productivity by only 0.5% and GDP growth by only 0.9% cumulatively over the next decade.

GS Head of Global Equity Research Jim Covello goes a step further, arguing that to earn an adequate return on the ~$1tn estimated cost of developing and running AI technology, it must be able to solve complex problems, which, he says, it isn't built to do. He points out that truly life-changing inventions like the internet enabled low-cost solutions to disrupt high-cost solutions even in its infancy, unlike costly AI tech today. And he's skeptical that AI's costs will ever decline enough to make automating a large share of tasks affordable given the high starting point as well as the complexity of building critical inputs—like GPU chips—which may prevent competition. He's also doubtful that AI will boost the valuation of companies that use the tech, as any efficiency gains would likely be competed away, and the path to actually boosting revenues is unclear, in his view. And he questions whether models trained on historical data will ever be able to replicate humans' most valuable capabilities.

But GS senior global economist Joseph Briggs is more optimistic. He estimates that gen AI will ultimately automate 25% of all work tasks and raise US productivity by 9% and GDP growth by 6.1% cumulatively over the next decade. While Briggs acknowledges that automating many AI-exposed tasks isn't cost-effective *today*, he argues that the large potential for cost savings and likelihood that costs will decline over the long run—as is often, if not always, the case with new technologies—should eventually lead to more AI automation. And, unlike Acemoglu, Briggs incorporates both the potential for labor reallocation and new task creation into his productivity estimates, consistent with the strong and long historical record of technological innovation driving new opportunities.

GS US software analyst Kash Rangan and internet analyst Eric Sheridan also remain enthusiastic about generative AI's long-term transformative and returns potential even as AI's "killer application" has yet to emerge. Despite big tech's large spending on AI infrastructure, they don't see signs of irrational exuberance. Indeed, Sheridan notes that current capex spend as a share of revenues doesn't look markedly different from prior tech investment cycles, and that investors are rewarding only those companies that can tie a dollar of AI spending back to revenues. Rangan, for his part, argues that the potential for returns from this capex cycle seems more promising than even previous cycles given that incumbents with low costs of capital and massive distribution networks and customer bases are leading it. So, both Sheridan and Rangan are optimistic that the huge AI spend will eventually pay off.

But even if AI could potentially generate significant benefits for economies and returns for companies, could shortages of key inputs—namely, chips and power—keep the technology from delivering on this promise? GS US semiconductor analysts Toshiya Hari, Anmol Makkar, and David Balaban argue that chips will indeed constrain AI growth over the next few years, with demand for chips outstripping supply owing to shortages in High-Bandwidth Memory technology and Chip-on-Wafer-on-Substrate packaging—two critical chip components.

But the bigger question seems to be whether power supply can keep up. GS US and European utilities analysts Carly Davenport and Alberto Gandolfi, respectively, expect the proliferation of AI technology, and the data centers necessary to feed it, to drive an increase in power demand the likes of which hasn't been seen in a generation (which GS commodities strategist Hongcen Wei finds early evidence of in Virginia, a hotbed for US data center growth).

Brian Janous, Co-founder of Cloverleaf Infrastructure and former VP of Energy at Microsoft, believes that US utilities—which haven't experienced electricity consumption growth in nearly two decades and are contending with an already aged US power grid—aren't prepared for this coming demand surge. He and Davenport agree that the required substantial investments in power infrastructure won't happen quickly or easily given the highly regulated nature of the utilities industry and supply chain constraints, with Janous warning that a painful power crunch that could constrain AI's growth likely lies ahead.

So, what does this all mean for markets? Although Covello believes AI's fundamental story is unlikely to hold up, he cautions that the AI bubble could take a long time to burst, with the "picks and shovels" AI infrastructure providers continuing to benefit in the meantime. GS senior US equity strategist Ryan Hammond also sees more room for the AI theme to run and expects AI beneficiaries to broaden out beyond just Nvidia, and particularly to what looks set to be the next big winner: Utilities.

That said, looking at the bigger picture, GS senior multi-asset strategist Christian Mueller-Glissmann finds that only the most favorable AI scenario, in which AI significantly boosts trend growth and corporate profitability without raising inflation, would result in above-average long-term S&P 500 returns, making AI's ability to deliver on its oft-touted potential even more crucial.

Allison Nathan, Editor

Email: allison.nathan@gs.com
Tel: 212-357-7504
Goldman Sachs & Co. LLC

Interview with Daron Acemoglu

Daron Acemoglu is Institute Professor at MIT and has written several books, including *Why Nations Fail: The Origins of Power, Prosperity, and Poverty* and his latest, *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. Below, he argues that the upside to US productivity and growth from generative AI technology over the next decade—and perhaps beyond—will likely be more limited than many expect.

The views stated herein are those of the interviewee and do not necessarily reflect those of Goldman Sachs.

Allison Nathan: In a recent paper, you argued that the upside to US productivity and, consequently, GDP growth from generative AI will likely prove much more limited than many forecasters—including Goldman Sachs—expect. Specifically, you forecast a ~0.5% increase in productivity and ~1% increase in GDP in the next 10 years vs. GS economists' estimates of a ~9% increase in productivity and 6.1% increase in GDP. Why are you less optimistic on AI's potential economic impacts?

Daron Acemoglu: The forecast differences seem to revolve more around the timing of AI's economic impacts than the ultimate promise of the technology. Generative AI has the potential to fundamentally change the process of scientific discovery, research and development, innovation, new product and material testing, etc. as well as create new products and platforms. But given the focus and architecture of generative AI technology today, these truly transformative changes won't happen quickly and few—if any—will likely occur within the next 10 years. Over this horizon, AI technology will instead primarily increase the efficiency of existing production processes by automating certain tasks or by making workers who perform these tasks more productive. So, estimating the gains in productivity and growth from AI technology on a shorter horizon depends wholly on the number of production processes that the technology will impact and the degree to which this technology increases productivity or reduces costs over this timeframe.

My prior guess, even before looking at the data, was that the number of tasks that AI will impact in the short run would not be massive. Many tasks that humans currently perform, for example in the areas of transportation, manufacturing, mining, etc., are multifaceted and require real-world interaction, which AI won't be able to materially improve anytime soon. So, the largest impacts of the technology in the coming years will most likely revolve around pure mental tasks, which are non-trivial in number and size but not huge, either.

To quantify this, I began with Eloundou et al.'s comprehensive study that found that the combination of generative AI, other AI technology, and computer vision could transform slightly over 20% of value-added tasks in the production process. But that's a timeless prediction. So, I then looked at another study by Thompson et al. on a subset of these technologies—computer vision—which estimates that around a quarter of tasks that this technology can perform could be cost-effectively automated within 10 years. If only 23% of exposed tasks are cost effective to automate within the next ten years, this suggests that only 4.6% of all tasks will be impacted by AI. Combining this figure with the 27% average labor cost savings estimates from Noy and Zhang's and Brynjolfsson et al.'s studies implies that total factor productivity effects within the next decade should be no more than 0.66%—and an even lower 0.53% when adjusting for the complexity of hard-to-learn tasks. And that figure roughly translates into a 0.9% GDP impact over the decade.

Allison Nathan: Recent studies estimate cost savings from the use of AI ranging from 10% to 60%, yet you assume only around 30% cost savings. Why is that?

Daron Acemoglu: Of the three detailed studies published on AI-related costs, I chose to exclude the one with the highest cost savings—Peng et al. estimates of 56%—because the task in the study that AI technology so markedly improved was notably simple. It seems unlikely that other, more complex, tasks will be affected as much. Specifically, the study focuses on time savings incurred by utilizing AI technology—in this case, GitHub Copilot—for programmers to write simple subroutines in HTML, a task for which GitHub Copilot had been extensively trained. My sense is that such cost savings won't translate to more complex, open-ended tasks like summarizing texts, where more than one right answer exists. So, I excluded this study from my cost-savings estimate and instead averaged the savings from the other two studies.

Allison Nathan: While AI technology cannot perform many complex tasks well today—let alone in a cost-effective manner—the historical record suggests that as technologies evolve, they both improve and become less costly. Won't AI technology follow a similar pattern?

Daron Acemoglu: Absolutely. But I am less convinced that throwing more data and GPU capacity at AI models will achieve these improvements more quickly. Many people in the industry seem to believe in some sort of scaling law, i.e. that doubling the amount of data and compute capacity will double the capability of AI models. But I would challenge this view in several ways. What does it mean to double AI's capabilities? For open-ended tasks like customer service or understanding and summarizing text, no clear metric exists to demonstrate that the output is twice as good. Similarly, what does a doubling of data really mean, and what can it achieve? Including twice as much data from Reddit into the next version of GPT may improve its ability to predict the next word when engaging in an informal conversation, but it won't necessarily improve a customer service representative's ability to help a customer troubleshoot problems with their video service. The quality of the data also matters, and it's not clear where more high-quality data will come from and whether it will be easily and cheaply available to AI models. Lastly, the current architecture of AI technology itself may have limitations. Human cognition involves many types of cognitive processes, sensory inputs, and reasoning capabilities. Large language models (LLMs) today have proven more impressive than many people would have predicted, but a big leap of faith is still required to believe that the architecture of predicting the next word in a sentence will achieve capabilities as smart as HAL 9000 in *2001: A Space Odyssey*. It's all but certain that current AI models won't achieve anything close to such a feat within the next ten years.

Allison Nathan: So, are the risks to even your relatively conservative estimates of AI's economic impacts over the next 5-10 years skewed to the downside?

Daron Acemoglu: Both downside and upside risks exist. Technological breakthroughs are always possible, although even such breakthroughs take time to have real impact. But even my more conservative estimates of productivity gains may turn out to be too large if AI models prove less successful in improving upon more complex tasks. And while large organizations such as the tech companies leading the development of AI technology may introduce AI-driven tools quickly, smaller organizations may be slower to adopt them.

Allison Nathan: Over the longer term, what odds do you place on AI technology achieving superintelligence?

Daron Acemoglu: I question whether AI technology can achieve superintelligence over even longer horizons because, as I said, it is very difficult to imagine that an LLM will have the same cognitive capabilities as humans to pose questions, develop solutions, then test those solutions and adopt them to new circumstances. I am entirely open to the possibility that AI tools could revolutionize scientific processes on, say, a 20-30-year horizon, but with humans still in the driver's seat. So, for example, humans may be able to identify a problem that AI could help solve, then humans could test the solutions the AI models provide and make iterative changes as circumstances shift. A truly superintelligent AI model would be able to achieve all of that without human involvement, and I don't find that likely on even a thirty-year horizon, and probably beyond.

Allison Nathan: Your colleague David Autor and coauthors have shown that technological innovations tend to drive the creation of new occupations, with 60% of workers today employed in occupations that didn't exist 80 years ago. So, could the impact of AI technology over the longer term prove more significant than you expect?

Daron Acemoglu: Technological innovation has undoubtedly meaningfully impacted nearly every facet of our lives. But that impact is not a law of nature. It depends on the types of technologies that we invent and how we use them. So, again, my hope is that we use AI technology to create new tasks, products, business occupations, and competencies. In my example about how AI tools may revolutionize scientific discovery, AI models would be trained to help scientists conceive of and test new materials so that humans can then be trained to become more specialized and provide better inputs into the AI models. Such an evolution would ultimately lead to much better possibilities for human discovery. But it is by no means guaranteed.

Allison Nathan: Will some—or maybe even most—of the substantial spending on AI technology today ultimately go to waste?

Daron Acemoglu: That is an interesting question. Basic economic analysis suggests that an investment boom should occur because AI technology today is primarily used for automation, which means that algorithms and capital are substituting for human labor, which should lead to investment. This explains why my estimates for GDP increases are nearly twice as large as my estimates for productivity increases. But then reality supervenes and says that some of the spending will end up wasted because some projects will fail, and some firms will be too optimistic about the extent of the efficiency gains and cost savings they can achieve or their ability to integrate AI into their organizations. On the other hand, some of the spending will plant the seeds for the next, and more promising, phase of the technology. The devil is ultimately in the details. So, I don't have a strong prior as to how much of the current investment boom will be wasted vs. productive. But I expect both will happen.

Allison Nathan: Are other costs of AI technology not receiving enough attention?

Daron Acemoglu: Yes. GDP is not everything. Technology that has the potential to provide good information can also provide bad information and be misused for nefarious purposes. I am not overly concerned about deepfakes at this point, but they are the tip of the iceberg in terms of how bad actors could misuse generative AI. And a trillion dollars of investment in deepfakes would add a trillion dollars to GDP, but I don't think most people would be happy about that or benefit from it.

Allison Nathan: Given everything we've discussed, is the current enthusiasm around AI technology overdone?

Daron Acemoglu: Every human invention should be celebrated, and generative AI is a true human invention. But too much optimism and hype may lead to the premature use of technologies that are not yet ready for prime time. This risk seems particularly high today for using AI to advance automation. Too much automation too soon could create bottlenecks and other problems for firms that no longer have the flexibility and trouble-shooting capabilities that human capital provides.

And, as I mentioned, using technology that is so pervasive and powerful—providing information and visual or written feedback to humans in ways that we don't yet fully understand and don't at all regulate—could prove dangerous. Although I don't believe superintelligence and evil AI pose major threats, I often think about how the current risks might be perceived looking back 50 years from now. The risk that our children or grandchildren in 2074 accuse us of moving too slowly in 2024 at the expense of growth seems far lower than the risk that we end up moving too quickly and destroy institutions, democracy, and beyond in the process. So, the costs of the mistakes that we risk making are much more asymmetric on the downside. That's why it's important to resist the hype and take a somewhat cautious approach, which may include better regulatory tools, as AI technologies continue to evolve.

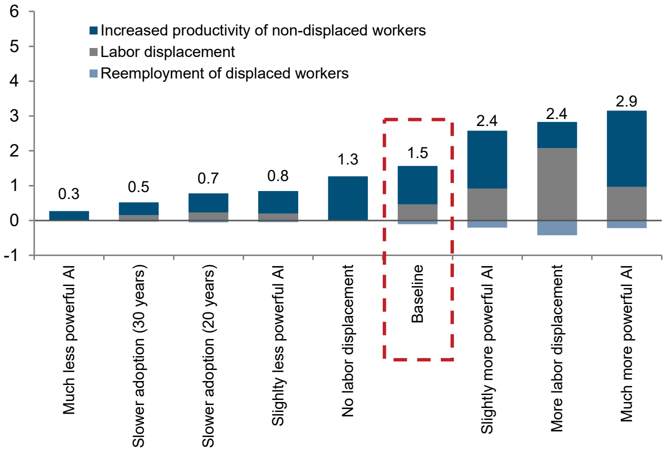
Addressing the AI growth debate

Joseph Briggs addresses the AI productivity and growth debate, arguing that generative AI will likely lead to significant economic upside

We have long argued that generative AI could lead to significant economic upside, primarily owing to its ability to automate a large share of work tasks, with our baseline estimate implying as much as 15% cumulative gross upside to US labor productivity and GDP growth[1] following widespread adoption of the technology.

A significant boost to US labor productivity from generative AI

Effect of AI adoption on annual US labor productivity growth, 10y adoption period, pp



Source: Goldman Sachs GIR.

That said, substantial debate exists around generative AI's potential macro impacts. Studies that assume generative AI will accelerate the development and adoption of robotics or that view recent generative AI advances as foreshadowing the emergence of a "superintelligence", for example, estimate even more upside to productivity and GDP than our baseline forecast. We see such outcomes as possible but premature since they generally assume AI advancements well beyond the frontier of current models.

More notably, MIT economist Daron Acemoglu sees much more limited upside to US productivity and GDP than we expect, with his baseline estimates implying that generative AI will boost US total factor productivity (TFP) by 0.53% and GDP by 0.9% over the next 10 years (see pgs. 4-5). As we take similar approaches to assessing the economic impacts of generative AI, we explore what explains the large differences in our estimates.

Breaking down the differences

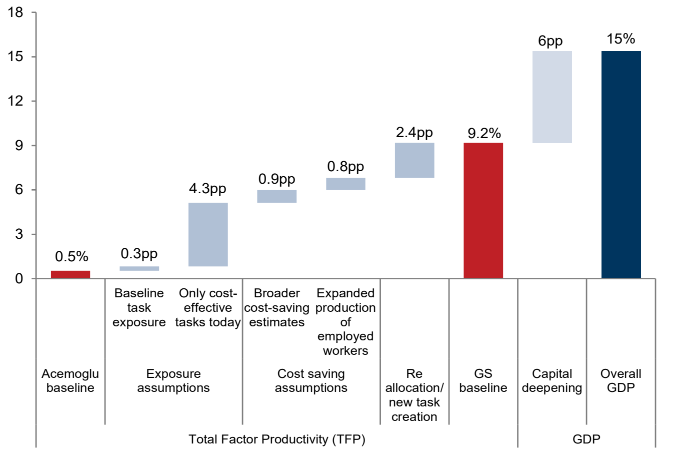
We find two main factors that explain the differences in our estimates versus those of Acemoglu. First, Acemoglu assumes that generative AI will automate only 4.6% of total work tasks, as he estimates that only 19.9% of all tasks are exposed to AI and assumes that only 23% of exposed tasks will be cost effective to automate within the next ten years. In contrast, we assume that generative AI will automate 25% of all work tasks following the technology's full adoption.

Second, Acemoglu's framework assumes that the primary driver of cost savings will be workers completing existing tasks more efficiently and ignores productivity gains from labor reallocation or the creation of new tasks. In contrast, our productivity estimates incorporate both worker reallocation—via displacement and subsequent reemployment in new occupations made possible by AI-related technological advancement—and new task creation that expands non-displaced workers' production potential.

Differences in these assumptions explain over 80% of the discrepancy between our 9.2%[2] and Acemoglu's 0.53% estimates of increases in TFP over the next decade[3]. The remaining 20% of the gap reflects differences in cost savings and marginal productivity assumptions. For instance, Acemoglu assumes 27% cost savings based on two studies that he considers the most representative of AI's real-world impact, but cost savings would rise to 36% if the full set of studies were considered. We are also more optimistic that AI will raise non-displaced workers' output, largely because we expect AI automation to create new tasks and products.

Differences in macro estimates mostly reflect differences in assumptions around tasks that can be profitably automated and the reallocation of labor to new tasks

Reconciling estimates of AI impact on GDP: Acemoglu (2024) vs. GS (2023), %



Source: Goldman Sachs GIR.

More widespread AI automation ahead

So, whose estimates regarding the share of automated tasks and new task creation—will more likely prove correct?

We are very sympathetic to Acemoglu's argument that automation of many AI-exposed tasks is not cost effective today, and may not become so even within the next ten years. AI adoption remains very modest outside of the few

[1] Our GDP estimate assumes that the capital stock evolves to match increased labor potential, which seems broadly validated by the sizable investment response aimed at facilitating the AI transition.

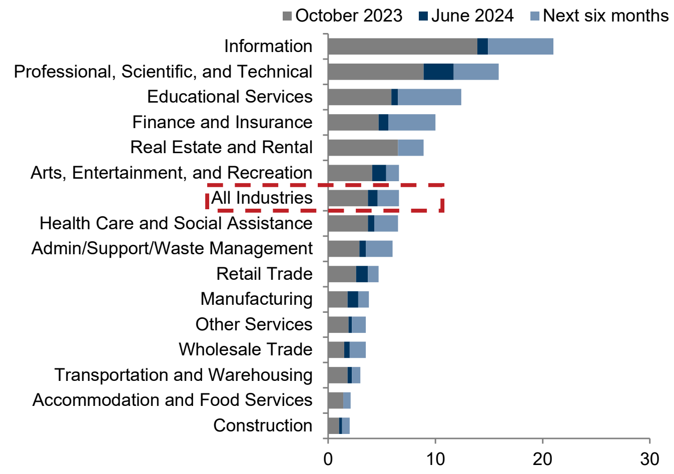
[2] This figured is calculated by multiplying the labor share of output, 62%, by our 15% estimate of the AI upside to labor productivity and growth.

[3] The quantitative contribution of different channels to the discrepancy between Acemoglu's and our estimates depends on the order that they are considered in, with differences in exposure assumptions explaining more of the gap if differences in cost savings assumptions are considered first and vice versa. To reduce this sensitivity, we consider both orderings and present the average contributions.

industries—including computing and data infrastructure, information services, and motion picture and sound production—that we estimate will benefit the most, and adoption rates are likely to remain below levels necessary to achieve large aggregate productivity gains for the next few years. This explains why we only raised our US GDP forecast by 0.4pp by the end of our forecast horizon in 2034 (with smaller increases in other countries) when we incorporated an AI boost into our global potential growth forecasts last fall. When stripping out offsetting growth impacts from the partial redirection of capex from other technologies to AI and slower productivity growth in a non-AI counterfactual, this 0.4pp annual figure translates into a 6.1% GDP uplift from AI by 2034 vs. Acemoglu's 0.9% estimate.

AI adoption remains modest on average across industries

Share of US firms using AI by sector, %



Source: Census Bureau, Goldman Sachs GIR.

That said, the full automation of AI exposed tasks that are likely to occur over a longer horizon could generate significant cost savings to the tune of several thousands of dollars per worker per year. The cost of new technologies also tends to fall rapidly over time. Given that cost-saving applications of generative AI will likely follow a similar pattern, and that the marginal cost of deployment will likely be very small once applications are developed, we expect AI adoption and automation rates to ultimately far exceed Acemoglu's 4.6% estimate.

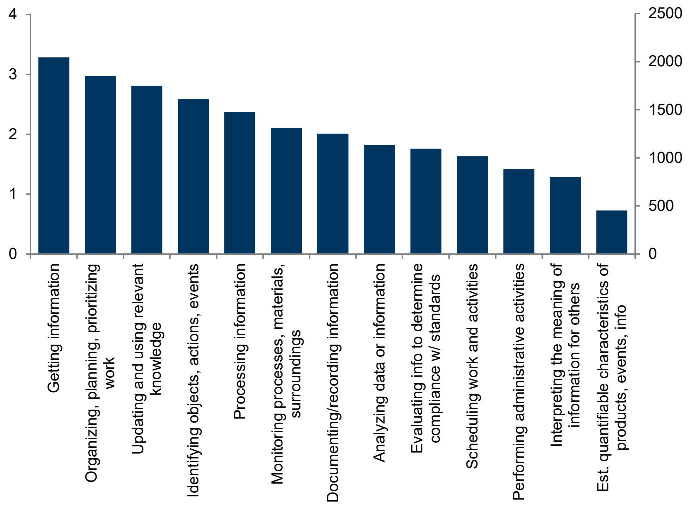
Labor reallocation and new task creation on the horizon

We also disagree with Acemoglu's decision not to incorporate productivity improvements from new tasks and products into his estimates, partly given his questioning of whether AI adoption will lead to labor reallocation and the creation of new tasks. The historical record provides strong evidence that economic growth stems mainly from technology-driven reallocation of resources and expansion of the production frontier, and we anticipate that AI will raise output both by raising demand in areas where labor has a comparative advantage and by creating new opportunities that were previously technologically or economically infeasible.

This dynamic clearly played out following the emergence of information technology—which created new occupations like webpage designers, software developers, and digital marketing professionals and indirectly drove demand for service sector workers in industries like healthcare, education, and food services—and is visible over a much longer horizon in recent work by MIT economist David Autor and coauthors. Using Census data, they find that 60% of workers today are employed in occupations that did not exist in 1940, with their estimates implying that the technology-driven creation of new occupations accounts for more than 85% of employment growth over the last 80 years.

Automation of work tasks should generate significant economic value, particularly as costs decline

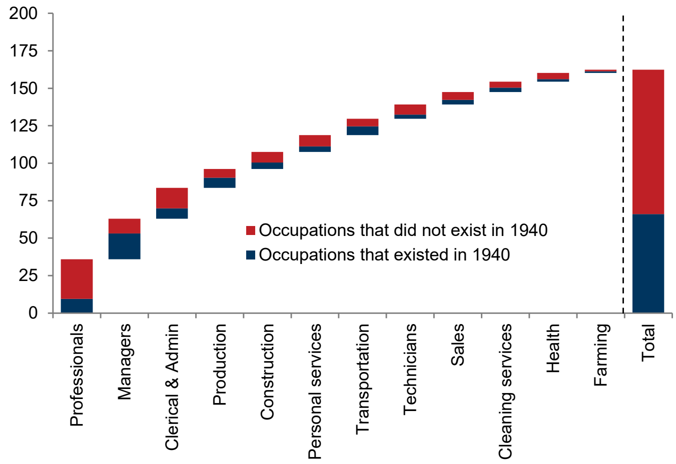
Value of automating work task categories per worker, % of time (lhs), $ (rhs)



Source: Goldman Sachs GIR.

Technological creation of new opportunities is a main driver of employment and economic growth

Employment by new and pre-existing occupations, millions



Source: Autor et al. (2022), Goldman Sachs GIR.

Accordingly, while we believe that Acemoglu's relatively pessimistic assessment of generative AI's economic potential highlights valid concerns that the macroeconomic impacts could be more backloaded than is commonly appreciated, we maintain that generative AI's large potential to drive automation, cost savings, and efficiency gains should eventually lead to significant uplifts of productivity and GDP.

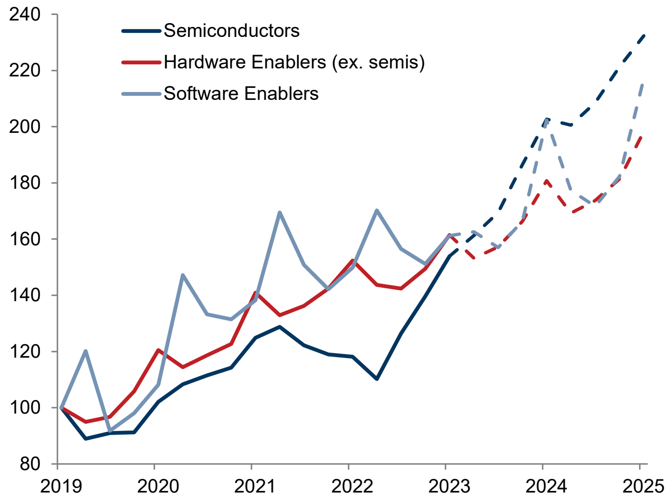
Joseph Briggs, Senior Global Economist

Email: joseph.briggs@gs.com *Goldman Sachs & Co. LLC*
Tel: 212-902-2163

The state of the AI transition...

AI investment has surged over the last several years...

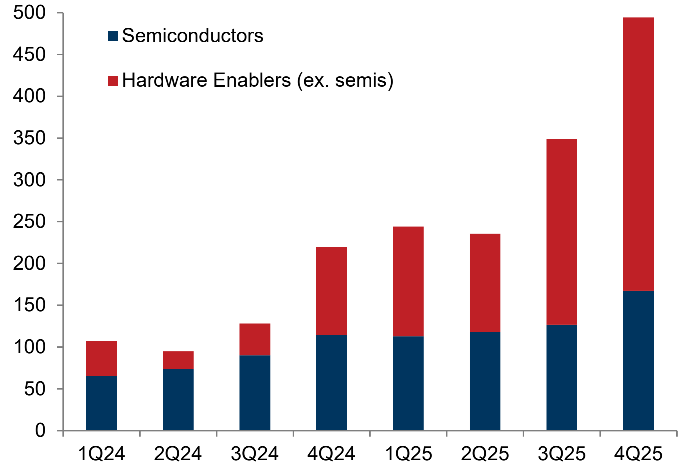
Actual and forecasted revenues by AI-exposed sector, index, 4Q19=100



Dashed lines in this chart indicate consensus revenue forecasts.
Source: FactSet, Goldman Sachs GIR.

...and the market has significantly upgraded its AI investment expectations across the AI hardware stack...

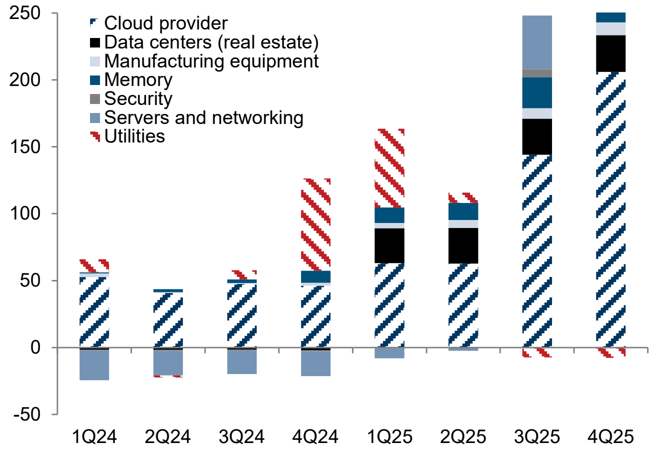
Change in consensus revenue forecasts since March 2023, $bn, annualized



Source: FactSet, Goldman Sachs GIR.

...though much less so across the broader AI space so far

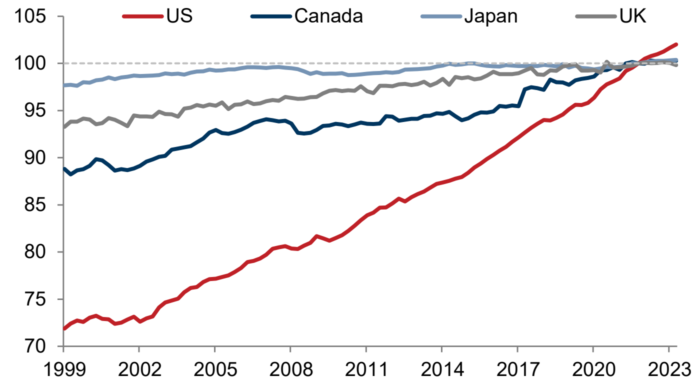
Change in consensus revenue forecasts since March 2023, $bn, annualized



Source: FactSet, Goldman Sachs GIR.

AI-related software investment isn't yet visible in the US' or other DMs' official national accounts data...

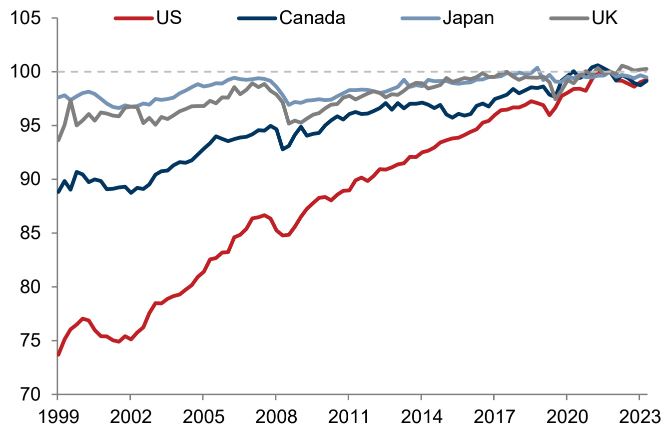
AI-related investment in software: national accounts, log index*, 3Q22=100



**Shown as log index because software investment grows by different exponential rates across countries. Steady growth in investment would appear as a line with a constant slope, while accelerating growth would appear as a line with an increasing slope.*
Source: Haver Analytics, Goldman Sachs GIR.

...nor is AI-related hardware investment, suggesting that other factors are currently playing a larger role than AI in shaping the aggregate capex outlook

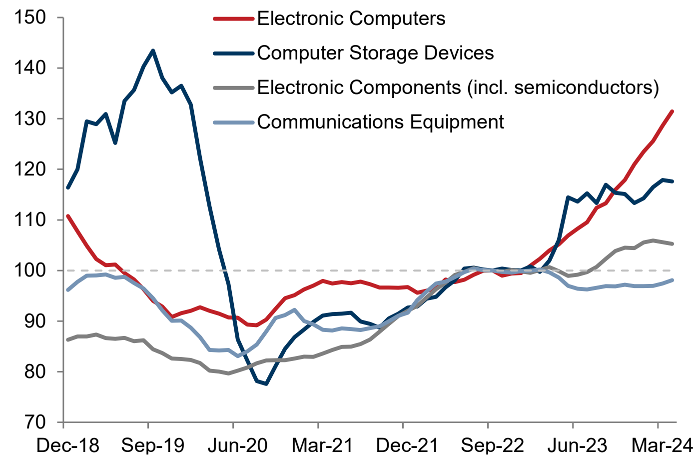
AI-related investment in hardware: national accounts, log index, 3Q22=100



Source: Haver Analytics, Goldman Sachs GIR.

However, manufacturers' shipments for some AI-related components have surged...

US nominal manufacturing sales, AI-related categories, index, Sept. 2022=100, 3m average

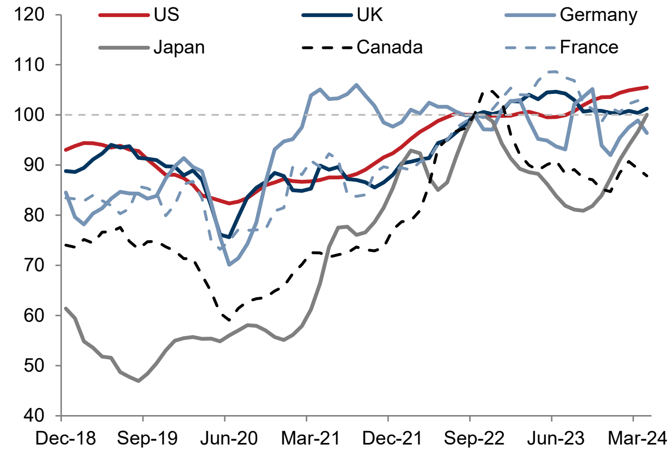


Source: Haver Analytics, Goldman Sachs GIR.

...in pics

...though this increase has not been uniform across the major developed economies, with the US leading the pack

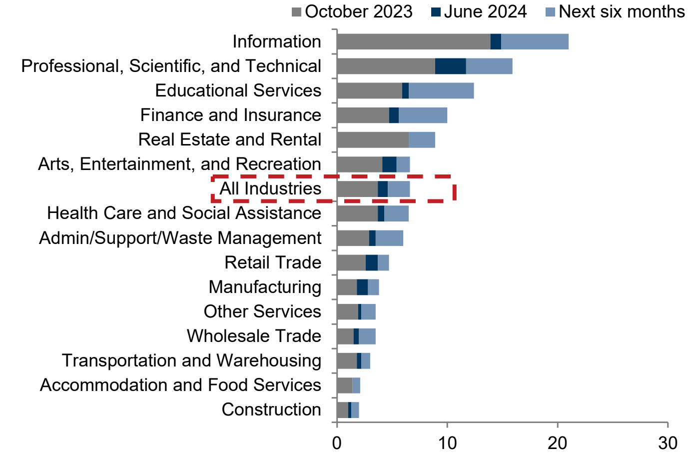
DM nominal manufacturing sales, AI-related categories, index, Sept. 2022=100, 3m average



Source: Haver Analytics, Goldman Sachs GIR.

AI adoption remains muted on average across industries, with adoption likely to pick up only modestly over the next six months...

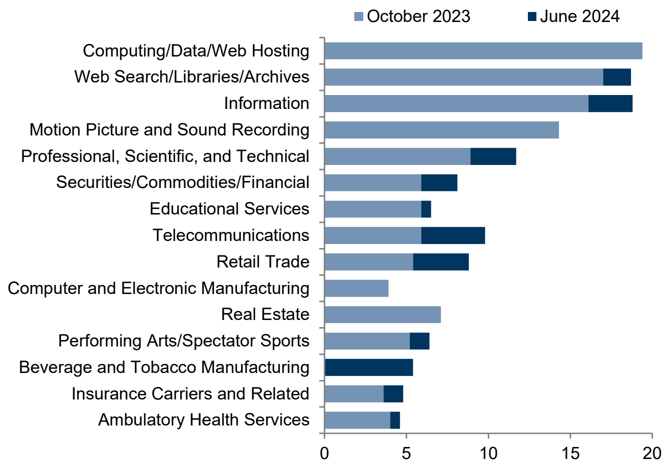
Share of US firms using AI by sector, %



Source: Census Bureau, Goldman Sachs GIR.

...though adoption rates are much higher among technology industries and other digitally-enabled fields...

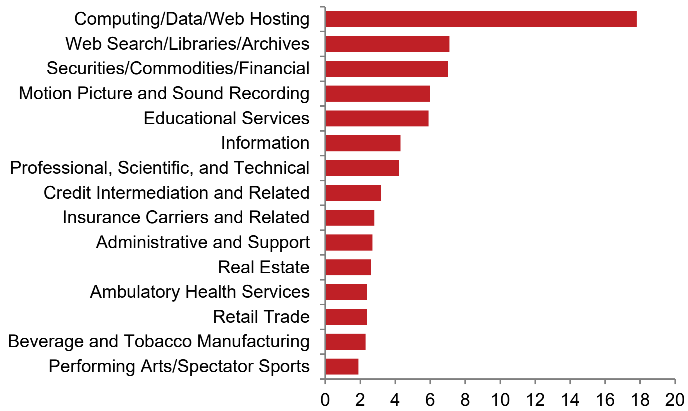
Share of US firms using AI, top 15 subsectors, %



Source: Census Bureau, Goldman Sachs GIR.

...and are expected to increase significantly across these sectors over the next six months

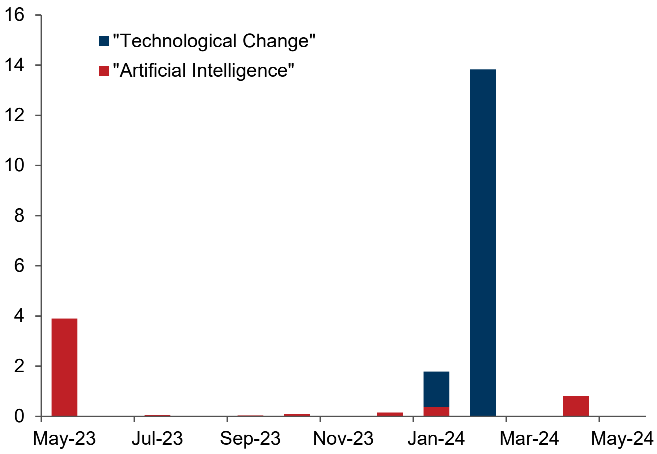
Expected change in share of firms using AI over the next six months, top 15 subsectors, pp



Source: Census Bureau, Goldman Sachs GIR.

Despite rising adoption rates, little evidence of net labor displacement from AI exists so far...

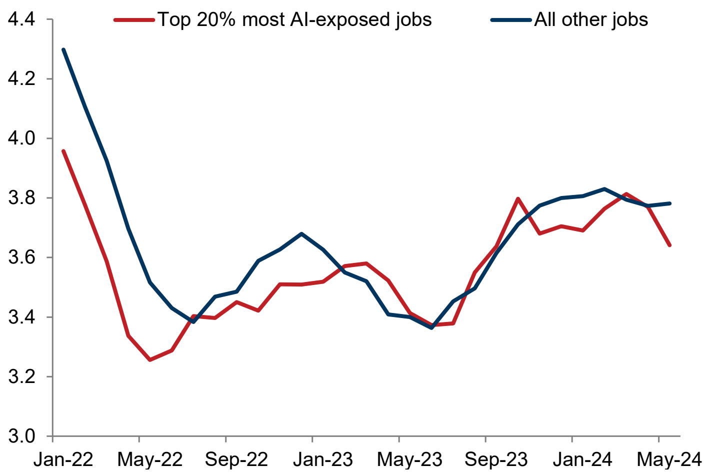
Layoffs attributed to each in corporate announcements, 000s



Source: Challenger, Gray & Christmas, Goldman Sachs GIR.

...with unemployment not looking markedly different across jobs

Unemployment rate by AI exposure, %, 3m average



Source: Census Bureau, IPUMS, Goldman Sachs GIR.

Special thanks to GS GIR global economist Devesh Kodnani for these charts, which were originally published in an April 2024 Global Economics Analyst.

Interview with Jim Covello

Jim Covello is Head of Global Equity Research at Goldman Sachs. Below, he argues that to earn an adequate return on costly AI technology, AI must solve very complex problems, which it currently isn't capable of doing, and may never be.

Allison Nathan: You haven't bought into the current generative AI enthusiasm nearly as much as many others. Why is that?

Jim Covello: My main concern is that the substantial cost to develop and run AI technology means that AI applications must solve extremely complex and important problems for enterprises to earn an appropriate return on investment (ROI). We estimate that the AI infrastructure buildout will cost over $1tn in the next several years alone, which includes spending on data centers, utilities, and applications. So, the crucial question is: What $1tn problem will AI solve? Replacing low-wage jobs with tremendously costly technology is basically the polar opposite of the prior technology transitions I've witnessed in my thirty years of closely following the tech industry.

Many people attempt to compare AI today to the early days of the internet. But even in its infancy, the internet was a low-cost technology solution that enabled e-commerce to replace costly incumbent solutions. Amazon could sell books at a lower cost than Barnes & Noble because it didn't have to maintain costly brick-and-mortar locations. Fast forward three decades, and Web 2.0 is still providing cheaper solutions that are disrupting more expensive solutions, such as Uber displacing limousine services. While the question of whether AI technology will ever deliver on the promise many people are excited about today is certainly debatable, the less debatable point is that AI technology is exceptionally expensive, and to justify those costs, the technology must be able to solve complex problems, which it isn't designed to do.

Allison Nathan: Even if AI technology is expensive today, isn't it often the case that technology costs decline dramatically as the technology evolves?

Jim Covello: The idea that technology typically starts out expensive before becoming cheaper is revisionist history. E-commerce, as we just discussed, was cheaper from day one, not ten years down the road. But even beyond that misconception, the tech world is too complacent in its assumption that AI costs will decline substantially over time. Moore's law in chips that enabled the smaller, faster, cheaper paradigm driving the history of technological innovation only proved true because competitors to Intel, like Advanced Micro Devices, forced Intel and others to reduce costs and innovate over time to remain competitive.

Today, Nvidia is the only company currently capable of producing the GPUs that power AI. Some people believe that competitors to Nvidia from within the semiconductor industry or from the hyperscalers—Google, Amazon, and Microsoft—themselves will emerge, which is possible. But that's a big leap from where we are today given that chip companies have tried and failed to dethrone Nvidia from its dominant GPU position for the last 10 years. Technology can be so difficult to replicate that no competitors are able to do so, allowing companies to maintain their monopoly and pricing power. For example, Advanced Semiconductor Materials Lithography (ASML) remains the only company in the world able to produce leading-edge lithography tools and, as a result, the cost of their machines has increased from tens of millions of dollars twenty years ago to, in some cases, hundreds of millions of dollars today. Nvidia may not follow that pattern, and the scale in dollars is different, but the market is too complacent about the certainty of cost declines.

The starting point for costs is also so high that even if costs decline, they would have to do so dramatically to make automating tasks with AI affordable. People point to the enormous cost decline in servers within a few years of their inception in the late 1990s, but the number of $64,000 Sun Microsystems servers required to power the internet technology transition in the late 1990s pales in comparison to the number of expensive chips required to power the AI transition today, even without including the replacement of the power grid and other costs necessary to support this transition that on their own are enormously expensive.

Allison Nathan: Are you just concerned about the cost of AI technology, or are you also skeptical about its ultimate transformative potential?

Jim Covello: I'm skeptical about both. Many people seem to believe that AI will be the most important technological invention of their lifetime, but I don't agree given the extent to which the internet, cell phones, and laptops have fundamentally transformed our daily lives, enabling us to do things never before possible, like make calls, compute and shop from anywhere. Currently, AI has shown the most promise in making existing processes—like coding—more efficient, although estimates of even these efficiency improvements have declined, and the cost of utilizing the technology to solve tasks is much higher than existing methods. For example, we've found that AI can update historical data in our company models more quickly than doing so manually, but at six times the cost.

More broadly, people generally substantially overestimate what the technology is capable of today. In our experience, even basic summarization tasks often yield illegible and nonsensical results. This is not a matter of just some tweaks being required here and there; despite its expensive price tag, the technology is nowhere near where it needs to be in order to be useful for even such basic tasks. And I struggle to believe that the technology will ever achieve the cognitive reasoning required to substantially augment or replace human interactions. Humans add the most value to complex tasks by identifying and understanding outliers and nuance in a way that it is difficult to imagine a model trained on historical data would ever be able to do.

Allison Nathan: But wasn't the transformative potential of the technologies you mentioned difficult to predict early on? So, why are you confident that AI won't eventually prove to be just as—or even more—transformative?

Jim Covello: The idea that the transformative potential of the internet and smartphones wasn't understood early on is false. I was a semiconductor analyst when smartphones were first introduced and sat through literally hundreds of presentations in the early 2000s about the future of the smartphone and its functionality, with much of it playing out just as the industry had expected. One example was the integration of GPS into smartphones, which wasn't yet ready for prime time but was predicted to replace the clunky GPS systems commonly found in rental cars at the time. The roadmap on what other technologies would eventually be able to do also existed at their inception. No comparable roadmap exists today. AI bulls seem to just trust that use cases will proliferate as the technology evolves. But eighteen months after the introduction of generative AI to the world, not one truly transformative—let alone cost-effective—application has been found.

Allison Nathan: Even if the benefits and the returns never justify the costs, do companies have any other choice but to pursue AI strategies given the competitive pressures?

Jim Covello: The big tech companies have no choice but to engage in the AI arms race right now given the hype around the space and FOMO, so the massive spend on the AI buildout will continue. This is not the first time a tech hype cycle has resulted in spending on technologies that don't pan out in the end; virtual reality, the metaverse, and blockchain are prime examples of technologies that saw substantial spend but have few—if any—real world applications today. And companies outside of the tech sector also face intense investor pressure to pursue AI strategies even though these strategies have yet to yield results. Some investors have accepted that it may take time for these strategies to pay off, but others aren't buying that argument. Case in point: Salesforce, where AI spend is substantial, recently suffered the biggest daily decline in its stock price since the mid-2000s after its Q2 results showed little revenue boost despite this spend.

Allison Nathan: What odds do you place on AI technology ultimately enhancing the revenues of non-tech companies? And even without revenue expansion, could cost savings still pave a path toward multiple expansion?

Jim Covello: I place low odds on AI-related revenue expansion because I don't think the technology is, or will likely be, smart enough to make employees smarter. Even one of the most plausible use cases of AI, improving search functionality, is much more likely to enable employees to find information faster than enable them to find better information. And if AI's benefits remain largely limited to efficiency improvements, that probably won't lead to multiple expansion because cost savings just get arbitraged away. If a company can use a robot to improve efficiency, so can the company's competitors. So, a company won't be able to charge more or increase margins.

Allison Nathan: What does all of this mean for AI investors over the near term, especially since the "picks and shovels" companies most exposed to the AI infrastructure buildout have already run up so far?

Jim Covello: Since the substantial spend on AI infrastructure will continue despite my skepticism, investors should remain invested in the beneficiaries of this spend, in rank order: chip manufacturers, utilities and other companies exposed to the coming buildout of the power grid to support AI technology, and the hyperscalers, which are spending substantial money themselves but will also garner incremental revenue from the AI buildout. These companies have indeed already run up substantially, but history suggests that an expensive valuation alone won't stop a company's stock price from rising further if the fundamentals that made the company expensive in the first place remain intact. I've never seen a stock decline only because it's expensive—a deterioration in fundamentals is almost always the culprit, and only then does valuation come into play.

Allison Nathan: If your skepticism ultimately proves correct, AI's fundamental story would fall apart. What would that look like?

Jim Covello: Over-building things the world doesn't have use for, or is not ready for, typically ends badly. The NASDAQ declined around 70% between the highs of the dot-com boom and the founding of Uber. The bursting of today's AI bubble may not prove as problematic as the bursting of the dot-com bubble simply because many companies spending money today are better capitalized than the companies spending money back then. But if AI technology ends up having fewer use cases and lower adoption than consensus currently expects, it's hard to imagine that won't be problematic for many companies spending on the technology today.

That said, one of the most important lessons I've learned over the past three decades is that bubbles can take a long time to burst. That's why I recommend remaining invested in AI infrastructure providers. If my skeptical view proves incorrect, these companies will continue to benefit. But even if I'm right, at least they will have generated substantial revenue from the theme that may better position them to adapt and evolve.

Allison Nathan: So, what should investors watch for signs that a burst may be approaching?

Jim Covello: How long investors will remain satisfied with the mantra that "if you build it, they will come" remains an open question. The more time that passes without significant AI applications, the more challenging the AI story will become. And my guess is that if important use cases don't start to become more apparent in the next 12-18 months, investor enthusiasm may begin to fade. But the more important area to watch is corporate profitability. Sustained corporate profitability will allow sustained experimentation with negative ROI projects. As long as corporate profits remain robust, these experiments will keep running. So, I don't expect companies to scale back spending on AI infrastructure and strategies until we enter a tougher part of the economic cycle, which we don't expect anytime soon. That said, spending on these experiments will likely be the one of the first things to go if and when corporate profitability starts to decline.

A discussion on generative AI

Kash Rangan and Eric Sheridan are Senior Equity Research Analysts at Goldman Sachs covering US software and internet, respectively. Below, they argue that while AI remains a work in progress, the large sums of money being put toward it should pay off, eventually.

Allison Nathan: When we last spoke in July 2023, you were both very enthused about the potential of generative AI. Are you just as optimistic today?

Kash Rangan: I am just as enthusiastic about generative AI's long-term potential as I was a year ago, and perhaps even more. The pace of technological change over the past 12 months has been mind-blowing, with hardly a week going by without reports of a newer, and better, AI model. The infrastructure buildout has also greatly exceeded expectations. Hyperscalers—large cloud computing companies that provide computing and storage services at scale—have spent $60-80bn in incremental capital above regular cloud capex on critical tools for building and training AI models. And rays of hope have emerged across several domains that demonstrate AI's productivity benefits. In the creative domain, generative AI has produced new design ideas in minutes that previously would've taken many hours, shortening the time it takes to bring an idea to market. In the code development domain, AI has automated low-level code writing, freeing up developers to work on more complex and productive tasks. And in the customer support domain, ServiceNow—a digital workflow software company—has reported an 80% reduction in the average time it takes to resolve a customer service problem thanks to AI technology.

That said, applications ultimately drive the success of tech cycles, and we have yet to identify AI's "killer application", akin to the Enterprise Resource Planning (ERP) software that was the killer application of the late 1990s compute cycle, the search and e-commerce applications of the 2000-10 tech cycle that achieved massive scale owing to the rise of x86 Linux open-source databases, or cloud applications, which enabled the building of low-cost compute infrastructure at massive scale during the most recent 2010-20 tech cycle. But this shouldn't come as a surprise given that every computing cycle follows a progression known as IPA—infrastructure first, platforms next, and applications last. The AI cycle is still very much in the infrastructure buildout phase, so finding the killer application will take more time, but I believe we'll get there.

Eric Sheridan: I agree that the visibility into what this infrastructure buildout will translate into in terms of AI applications and adoption rates remains relatively low. And several notable issues at the application layer—such as AI chatbots "hallucinating" or giving false answers to user prompts—have called into question the scalability of generative AI. So, the technology is still very much a work in progress. But it's impossible to sit through demonstrations of generative AI's capabilities at company events or developer conferences and not come away excited about its long-term potential.

Allison Nathan: It's well known that Nvidia has benefitted massively in the current "picks and shovels" phase of the cycle. Are firms beyond Nvidia currently monetizing the gains from generative AI technology?

Eric Sheridan: Nvidia has certainly garnered significant revenue as its graphics processing unit (GPU) chip has become the nerve center of AI systems. But the semiconductor industry more broadly has benefitted from the voracious need for chips. Cloud computing companies have also performed well owing to the enormous computing capacity required to train and run AI models, with the three large hyperscalers of Microsoft, Alphabet, and Amazon seeing an acceleration in revenue growth in the last quarter. So, capital is shifting into the AI theme, the theme and the capital are aligning against the building, and many companies exposed to semiconductors and computing workloads are monetizing these gains. So, this is not just a Nvidia story.

Allison Nathan: Are you concerned that the hundreds of billions of dollars in AI capex big tech firms are estimated to spend in coming years is a sign of irrational exuberance, and that the payoff may be low or never come?

Eric Sheridan: Those who argue that this is a phase of irrational exuberance focus on the large amounts of dollars being spent today relative to two previous large capex cycles—the late 1990s/early 2000s long-haul capacity infrastructure buildout that enabled the development of Web 1.0, or desktop computing, as well as the 2006-2012 Web 2.0 cycle involving elements of spectrum, 5G networking equipment, and smartphone adoption. But such an apples-to-apples comparison is misleading; the more relevant metric is dollars spent vs. company revenues. Cloud computing companies are currently spending over 30% of their cloud revenues on capex, with the vast majority of incremental dollar growth aimed at AI initiatives. For the overall technology industry, these levels are not materially different than those of prior investment cycles that spurred shifts in enterprise and consumer computing habits. And, unlike during the Web 1.0 cycle, investors now have their antenna up for return on capital. They're demanding visibility on how a dollar of capex spending ties back to increased revenues, and punishing companies who can't draw a dotted line between the two. We saw this with Meta a few months ago when the company's stock fell sharply after it announced plans to spend several billion dollars on AI, potentially disrupting its core business in the process, while offering little visibility into the eventual payoff. So, while I would never say I'm not concerned about the possibility of no payback, I'm not particularly worried about it today, though I could become more concerned if scaled consumer applications don't emerge over the next 6-18m.

Kash Rangan: Spending is certainly high today in absolute dollar terms. But this capex cycle seems more promising than even previous capex cycles because incumbents—rather than upstarts—are leading it, which lowers the risk that technology doesn't become mainstream. Incumbents have access to deep pools of capital, an extremely low cost of capital, and massive distribution networks and customer bases, which allows them to experiment with how the capital dollars could eventually earn a return. Leading the late-1990s investment cycle, by contrast, were companies that didn't have the financing, reputation, or knowledge to succeed, resulting in a tremendous amount of underutilized capacity. The companies spearheading the current investment cycle are also run by very capable managements, with CFOs watching expenses like a hawk, holding companies accountable for the return on investment, and standing ready to tap the brakes on spending if the returns disappoint. Of course, this could all still fail, resulting in the loss of tens of billions of dollars in capex and interest income. But the opportunity costs of pursuing these strategies despite unknown outcomes still seems small compared to the potential opportunity of building the foundation for the next big computing architecture.

Allison Nathan: Even if corporate investments in AI eventually pay off, could this take longer than expected?

Eric Sheridan: Many consumer internet companies have yet to see significant returns on their AI investments, and the timing of these returns remains uncertain because the three main channels of payout—advertising, e-commerce, and subscription fees—will depend on shifting consumer habits. In the Web 1.0 tech cycle, the Netscape web browser debuted in the mid-1990s and the market peaked in March 2000, but the return on capital only turned positive in the late 2000s/early 2010s as consumers were slow to embrace the technology. The payback period was much shorter in the Web 2.0 tech cycle that began in 2006, with most companies calling themselves mobile-first companies by 2012/13. But it's not just the timing of returns that matters—if firms continue running at current levels of annualized spend over the next several years, the *magnitude* of returns will need to be outsized to justify the costs. That said, a longer-than-expected payoff process won't kill this tech cycle. I'm loathe to use the word "bubble" because I don't believe that AI is a bubble, but most bubbles in history ended either because the cost of capital changed dramatically or end-demand deteriorated and affected companies' ability to deploy capital, not because companies retreated from investing in a technology where the payoff was taking longer than expected.

Kash Rangan: Monetization of AI technology spend for enterprise software companies, including Salesforce, SAP, and Oracle, will come from customers willing to pay a premium for AI-infused products. The consumer market is a good leading indicator for the enterprise market, so once consumers begin embracing the technology, enterprise software companies should also benefit from a revenue tailwind. However, several such companies recently issued disappointing revenue guidance, leading some to go so far as to question whether the next decade will see "hardware eat software", in stark contrast to the last decade of "software eating the world". But the recent disappointments likely owed at least in part to the high-rate environment—software is a $600-700bn industry, which makes it susceptible to high cost of capital. And, again, every computing cycle follows the IPA progression. So, while spending is currently aimed at the infrastructure, it will eventually shift to platforms and applications, which is where the software companies will come in.

Allison Nathan: Given the competitive pressures, do firms have any option other than competing in the AI arms race?

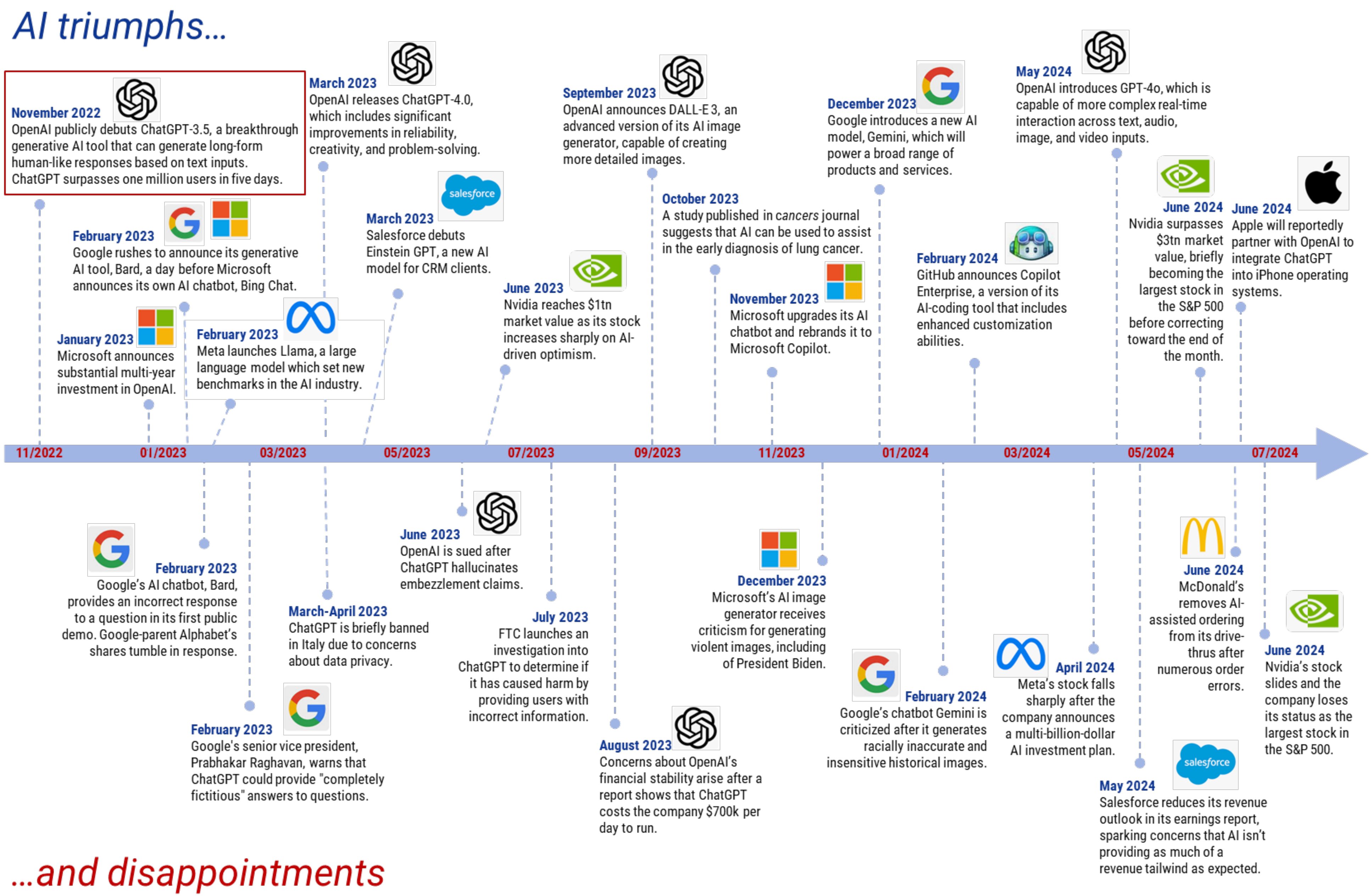
Eric Sheridan: Even if ChatGPT didn't exist, Alphabet—the poster child for this debate in my coverage universe—would probably still invest in AI. I remember sitting in the audience at Google I/O, the company's annual developer conference, in 2017 when Alphabet announced that it was now an AI-first company. But Alphabet has gone through more iteration and innovation since ChatGPT launched in 2022 than in 2017-2022, which leads me to believe that the driving force behind AI spend is as much offensive as defensive.

Allison Nathan: Some people argue that AI technology is too expensive, isn't actually fixing any real problem, and will likely never approach the cognitive abilities of humans because training on existing/historical data can only go so far. What are they missing?

Kash Rangan: AI technology is undoubtedly expensive today. And the human brain is 10,000x more effective per unit of power in performing cognitive tasks vs. generative AI. But the technology's cost equation will change, just as it always has in the past. In 1997, a Sun Microsystems server cost $64,000. Within three years, the server's capabilities could be replicated with a combination of x86 chips, Linux technology, and MySQL scalable databases for 1/50th of the cost. And the scaling of x86 chips coupled with open-source Linux, databases, and development tools led to the mainstreaming of AWS infrastructure. This, in turn, made it possible and affordable to write thousands of software applications, such as Salesforce, ServiceNow, Intuit, Adobe, Workday, etc. These applications, initially somewhat limited in scale, ultimately evolved to support a few hundred million end-users, not to mention the impressive scaling of Microsoft Azure that supported ubiquitous applications such as Office 365. Over the last decade, these applications have evolved and helped create hundreds of billions of dollars in shareholder value, providing even more evidence for the cliché that people tend to overestimate a technology's short-term effects and underestimate its long-term effects. Nobody today can say what killer applications will emerge from AI technology. But we should be open to the very real possibility that AI's cost equation will change, leading to the development of applications that we can't yet imagine.

Eric Sheridan: Again, I readily acknowledge that the return on invested capital (ROIC) visibility is currently low, and the transformative potential of AI will remain hotly debated until that becomes clearer. But AI skeptics miss three key things. One, training on existing/historical data to inform and drive analytic outcomes in the future sounds exactly like going to university—people go to learn and then improve productivity and efficiency for decades after graduation, and machines can absolutely do the same. Two, machines today can do a whole host of tasks more productively and efficiently than humans, and that will remain true for decades into the future. And three, people didn't think they needed smartphones, Uber, or Airbnb before they existed. But today it seems unthinkable that people ever resisted such technological progress. And that will almost certainly prove true for generative AI technology as well.

A short history of AI developments



Note: This does not constitute an exhaustive list of all AI-related developments.
Source: BBC, cancers, OpenAI, tech.co, Google, various news sources, compiled by Goldman Sachs GIR.

Interview with Brian Janous

Brian Janous is Co-founder of Cloverleaf Infrastructure, which develops strategies to help utilities unlock new grid capacity. Previously, he was Vice President of Energy at Microsoft. Below, he argues that US power infrastructure is not prepared for the coming surge in power demand from AI and other sources, setting up for a painful power crunch in the coming years.

The views stated herein are those of the interviewee and do not necessarily reflect those of Goldman Sachs.

Jenny Grimberg: Power demand is surging across parts of the US, and utilities and grid operators have significantly raised their estimates of US electricity demand growth over the next five years. What role are advances in AI technology playing in the US' growing hunger for electricity, and how do data centers fit into that?

Brian Janous: Cloud data centers have grown rapidly since the advent of cloud computing around 2010. However, global data center electricity consumption barely budged over the subsequent decade as these data centers cannibalized on-prem workloads, which used multiples more electricity per unit of compute than cloud data centers. So, the migration of data to the cloud resulted in a significant increase in computation with almost no rise in electricity usage. But as the cloud data center capacity of the three large hyperscalers of Microsoft, Amazon, and Google grew from a few hundred megawatts in the early 2010s to a few gigawatts by the end of the decade, power consumption began to rise. And the release of ChatGPT 3.5 in November 2022 ushered in a new layer of AI-related demand, which will likely require adding hundreds of megawatts—if not gigawatts—of data center capacity annually. So, power demand is set to continue surging over the coming years.

Jenny Grimberg: How much does electric grid capacity have to expand to meet this surge?

Brian Janous: That's the million-dollar question. Utilities are fielding hundreds of requests for huge amounts of power as everyone chases the AI wave, but only a fraction of that demand will ultimately be realized. AEP, one of the largest US electric utility companies, has reportedly received 80-90 gigawatts (GW) of load requests. Only 15 GW of that is likely real because many of the AI projects that companies are currently envisioning will never actually see the light of day. But 15 GW is still massive given that AEP currently owns/operates around 23 GW of generating capacity in the US. And even if overall grid capacity grows by only 2% annually—which seems like a reasonable forecast—utilities would still need to add well in excess of 100 GW of peak capacity to a system that currently handles around 800 GW at peak. The increase in power demand will also likely be hyperlocalized, with Northern Virginia, for example, potentially requiring a doubling of grid capacity over the next decade given the concentration of data centers in the area. So, grid capacity will need to expand substantially across the US, and likely even more in certain regions.

Jenny Grimberg: Are utility companies and the underlying power infrastructure equipped to meet the rapid surge in power demand?

Brian Janous: No. Utilities have not experienced a period of load growth in almost two decades and are not prepared for—or even capable of matching—the speed at which AI technology is developing. Only six months elapsed between the release of ChatGPT 3.5 and ChatGPT 4.0, which featured a massive improvement in capabilities. But the amount of time required to build the power infrastructure to support such improvements is measured in years. And AI technology isn't developing in a vacuum—electrification of transportation and buildings, onshoring of manufacturing driven partly by the Inflation Reduction Act and CHIPS Act, and potential development of a hydrogen economy are also increasing the demands on an already aged power grid.

Regulatory lags and interconnection and supply chain constraints are also impediments to meeting the rising power demand. The total capacity of power projects waiting to connect to the grid grew nearly 30% last year, with wait times currently ranging from 40-70 months, and lead times for critical electrical components such as transformers and switchgears have substantially increased. Until those issues can be resolved, and the grid can catch up, a significant power crunch will likely force utilities and states to pick and choose who receives power. My concern is that data centers will become an easy target because they're not perceived as major engines of job creation relative to building the next Hyundai factory, for example. This dynamic has already occurred in Dublin, where EirGrid, a state-owned power operator, enacted a de facto moratorium on new data centers by delaying their grid connection until 2028. Amsterdam also recently unveiled new rules that would impose fines on data centers that don't switch off idle servers to conserve energy. It's not out of the realm of possibility that something similar could happen in the US.

Jenny Grimberg: Didn't similar concerns arise during the era of hyperscale computing, which many worried would gobble up all the world's power, only to be proven wrong as data centers became more efficient?

Brian Janous: The experience of 2010-20 has provided a false sense of comfort. Most of the efficiency gains over that period owed to the shutdown of inefficient on-prem data centers in favor of cloud data centers. Data centers themselves did not become significantly more efficient. To put some numbers on this, the average power usage effectiveness (PUE)—a measure of data center efficiency calculated by dividing the total amount of power a facility consumes by the amount used to run the servers—of on-prem data centers was 2-3 vs. around 1.3 for cloud, but the PUE of cloud data centers themselves only declined by around 0.2 over the course of the decade. And the average PUE of data centers today is around 1.1, meaning over 90% of the power they consume goes directly to the servers vs. cooling, lighting, etc. So, only limited room exists to extract more efficiency from a data center. And even if new ways were discovered to increase the efficiency of data centers or AI chips themselves, humans' capacity to consume data is nearly insatiable. Every time we develop a more efficient chip or process, we find ways to use more of the underlying resource, not less of it, which is known as Jevons paradox. Big tech firms are currently engaging in an AI arms race to create the most powerful and capable AI model, and until we reach a level of saturation in terms of human capacity to consume data, any amount of efficiency gains will undoubtedly be gobbled up by even more demand.

Jenny Grimberg: What's required to expand the grid?

Brian Janous: Expanding the grid is no easy or quick task. The electric utility industry is highly regulated, and utility companies must go through a lengthy permitting and approval process before starting to construct new capacity. They then must contend with a supply chain that isn't prepared for every utility company to suddenly double their equipment orders, and building up the supply chain to meet the growing demand in itself will take at least months, and even years in some cases.

To help ease some of the power constraints in the meantime, utilities will need to find ways to extract more efficiencies from the current system, for example by reconductoring, or replacing, existing transmission lines to move more power over them, and investing in grid-enhancing technologies. Expanding long-duration storage to deliver electricity when and where it's most needed could also ease periods of peak capacity, as can integrating more flexibility into how, and when, energy is consumed. While there has been discussion about AI workloads being flexible, the reality is they will still need a high level of power availability. However, data centers' generation and storage assets can be leveraged for flexibility. That can help ensure that peak capacity isn't increasing at the same rate as the overall consumption of electricity. But again, the power constraint issue ultimately cannot be resolved without a significant buildout of electric grid infrastructure.

Jenny Grimberg: Big tech companies seem to have endless amounts of capital to throw at AI technology. Couldn't they just spend some of that money securing the power supply they need?

Brian Janous: If this was simply an issue of money, it would have already been solved. The big tech companies face the same constraints as the utility industry. They must go through the same regulatory processes and are subject to the same supply chain issues. Some people have suggested that these firms should just generate all their own power. But the only way to do so today is by using natural gas, which still sits on a grid that comes with its own set of constraints and requires upgrading and building new infrastructure. Nuclear power also gets tossed around as a potential solution, but building a new nuclear plant within the next decade isn't feasible. Those who believe nuclear is the answer often point to Amazon Web Services' recent purchase of a data center from Talen Energy located next to the Susquehanna nuclear power station in Pennsylvania. But nothing new was built to unlock that power—that nuclear plant has been around for over four decades. And so, if the problem we're trying to solve is how to increase power supply this decade, we have to expand the electric grid. No other solution exists.

Jenny Grimberg: So, is the US up to the task?

Brian Janous: The US has unfortunately lost the ability to build large infrastructure projects—this is a task better suited for 1930s America, not 2030s America. So, that leaves me a bit pessimistic. That said, utilities and policymakers are starting to take seriously the need to invest in America's transmission infrastructure, which isn't designed for today's energy generation mix. The transmission infrastructure was built from the coasts into the country, but today, massive wind resources are located in the center of the country and solar in the southwest. So, the transmission system ideally needs to run from the inside out.

Utilities have also begun to recognize that the significant increase in load growth that lies ahead creates a massive economic development opportunity, probably the biggest they'll ever see. The utilities that can find ways to offer more power to more customers sooner will attract that economic development and growth. And while data centers aren't perceived as major job creators, they still create significant economic activity in the form of construction jobs and large tax revenues. No utility wants to turn away customers or tell their state's governor that a new factory went elsewhere because they couldn't provide enough power. That's a significant motivator for utilities to not only invest in grid infrastructure—which, if done thoughtfully, shouldn't lead to rate increases for consumers—but also find efficiencies in the current system, which they currently lack the incentive to do because there's no money in it for them. So, on balance, I'm optimistic that America can rise to meet the challenge, though the next decade will likely prove painful as the demand for power outpaces the available supply.

Jenny Grimberg: Are the big tech companies spending hundreds of billions of dollars on AI infrastructure, including data centers, underappreciating the power constraint?

Brian Janous: Yes and no. These companies are very optimistic about what they can achieve with AI, but tech firms are starting to realize that power supply will be a significant constraint on the technology. When generative AI first exploded onto the scene, people debated what would constrain its potential—a shortage of chips or a shortage of power. That debate has now been settled, with everyone agreeing that over the medium-to-longer term the major constraint will be power. Meta's Mark Zuckerberg recently stated in an interview that energy constraints are the biggest bottleneck to building out AI data centers. Microsoft's Satya Nadella has also spoken about this. And Nvidia's Jensen Huang recently addressed the electric utility industry at the EEI conference, which would have seemed crazy even a year ago but now makes total sense. So, companies have woken up to the fact that electricity is an incredibly important commodity, and are now hyper-focused on the power constraint. But recognition of the problem is one thing—solving it is a much more difficult challenge.

Once in a generation, generation

Carly Davenport answers key questions about the coming surge in US power demand from AI technology and data centers

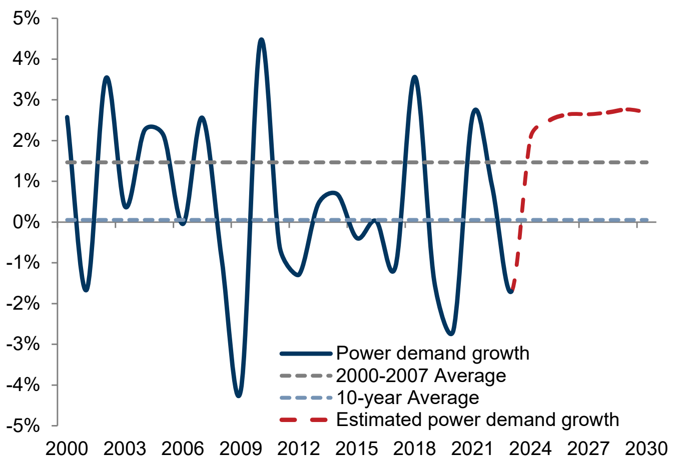
The proliferation of generative AI technology—and the data centers needed to feed it—is set to drive an increase in US power demand not seen in a generation. Here, we address key questions about the coming power demand surge, how much new generation capacity will be required to meet it, and the implications for companies and investors.

Q: How significant will the power demand growth from AI/data centers be?

A: After stagnating over the last decade, we expect US electricity demand to rise at a 2.4% compound annual growth rate (CAGR) from 2022-2030, with data centers accounting for roughly 90bp of that growth. Indeed, amid AI growth, a broader rise in data demand, and a material slowdown in power efficiency gains, data centers will likely more than double their electricity use by 2030. This implies that the share of total US power demand accounted for by data centers will increase from around 3% currently to 8% by 2030, translating into a 15% CAGR in data center power demand from 2023-2030.

After stagnating over the last decade, US power demand should grow by 2% per year on average through 2030

US power demand growth, %



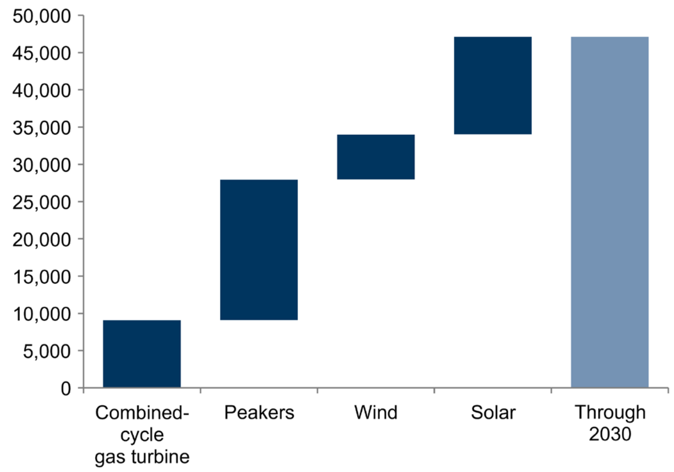
Source: EIA, Goldman Sachs GIR.

Q: How much power generation is required to support this demand growth, and where will it come from?

A: We estimate the US will require 47 gigawatts (GW) of new power generation capacity through 2030 to support the growth in data center power demand. We expect this capacity to be split 60%/40% between natural gas and renewables generation, reflecting a balance between the reliability needs of data centers and companies' green energy commitments. The data centers that power AI models must essentially run 24/7 given the nature of AI workloads, and so require a constant energy source like natural gas that can be dispatched on demand rather than renewables, which are more intermittent in nature. But we still believe renewables will play an important role given that many of the companies building data centers—especially the hyperscalers of the world—have committed to green electricity consumption and are unlikely to abandon those commitments to meet growing data center demand.

We estimate around 47 GW of incremental capacity is needed to serve data center-driven load growth in the US through 2030

Data center-driven capacity adds, megawatts (MW)



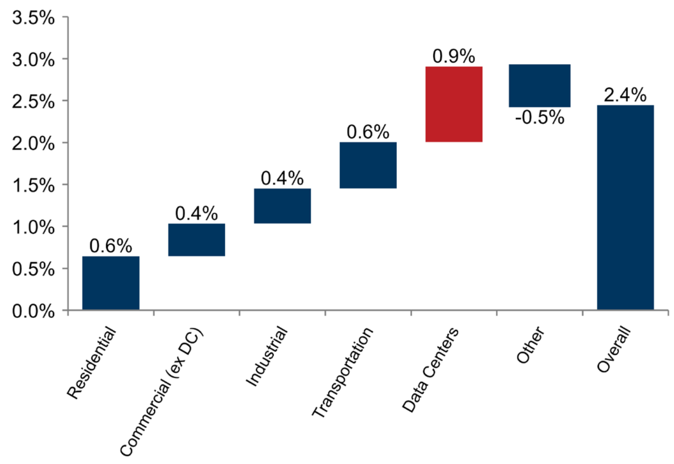
Source: Goldman Sachs GIR.

Q: Could nuclear energy also be part of the solution alongside natural gas and renewables?

A: The US historically hasn't demonstrated the best track record of building nuclear plants on time or on budget, so we don't think utilities would take on the risk of attempting to build new capacity. However, data centers could attempt to strike deals with companies that operate unregulated nuclear plants—those not regulated by a state utility commission and therefore not precluded from striking direct contracts with customers—because nuclear power solves exactly what the data centers are looking for: reliability and no carbon emissions. And some such deals have already occurred.

Data centers should contribute 90bp to our 2.4% US power demand CAGR from 2022-2030

Composition of US power demand CAGR, 2022-2030, %



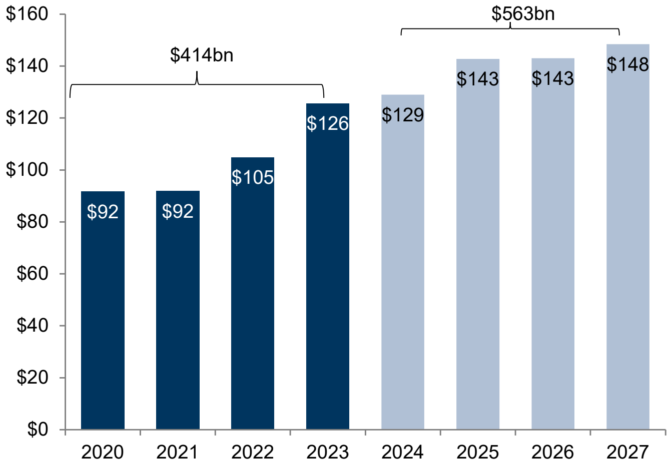
Source: EIA, Goldman Sachs GIR.

Q: How much capital investment will utilities need to make to provide the necessary capacity?

A: Approximately $50bn of investment through 2030, or roughly $7bn annually, is needed to facilitate the new power generation alone. But utilities will also need to build out the supporting infrastructure, such as the transmission wires that transport electricity over long distances and distribution cables that carry electricity to homes, so the overall investment will likely prove much higher. Between generation, transmission, and distribution needs, we expect the utility companies in our coverage universe to spend nearly 40% more from 2024-2027 relative to the prior four-year period, amounting to roughly $140bn on average annually. So, a significant increase in utility capex likely lies ahead, with this investment already materializing in regions like Northern Virginia, a hotbed for data center growth (see pg. 19).

Capex across the utility companies in our coverage should increase by roughly $140bn on average annually from 2024-2027

Capex by year, $bn



Source: SNL, Company data, Goldman Sachs GIR.

Q: What constraints—if any—could prevent the industry from delivering the required capacity?

A: The most significant constraint is the long timelines for infrastructure permitting and construction. Many power project developers start the process 5-7 years in advance to adequately plan for land acquisition, resource planning, permitting timelines, and interconnection queues—which currently range from ~40-70 months across the country—as well as any potential supply chain constraints.

Affordability is also an important constraint for utilities, which is a highly regulated industry. Regulators are focused on ensuring that electricity bills remain affordable for residential customers and that the capital investment necessary to meet data center growth isn't borne by the residential customer. This ultimately puts a cap on the rates utility companies can charge and still get their project approved.

Q: If utilities don't have much leeway to raise prices, where will the funding for the capacity investment come from?

A: Utilities don't generate a significant amount of free cash flow, so they will need to add debt capacity or issue equity to facilitate this massive investment. But part of the investment could also come from the data center customers themselves. If a utility needs to build infrastructure that will only support a data center, several ways exist for the utility to structures its contracts and rates to ensure that the capital is sourced only from that customer and the cost isn't socialized across the broader customer base.

Q: What are the main risks to your demand/investment forecasts?

A: AI/data center-led demand could be lower than we expect if advancements in energy-efficient hardware materialize or data center customers overestimate their near-term power needs. However, the increasing proliferation of AI technology and demand for data, as well as a slowdown in efficiency gains, should ultimately drive stronger power demand from data centers. The ongoing electrification of transportation, buildings, and oil & gas operations, as well as increased manufacturing activity in the US due to reshoring, strengthens our confidence that power demand growth will rise to levels not seen since the turn of the century. This growth, together with the energy transition, need to address aging existing infrastructure, and increased climate risk, should ultimately support a significant rise in grid investments from utilities as well as infrastructure contractors and industrials making products that support the buildout.

Q: What companies will benefit the most from the coming surge in power demand?

A: We see two broad categories of beneficiaries: demand growth beneficiaries and supply chain/infrastructure beneficiaries. **Demand growth beneficiaries** include companies levered to power needs/prices, including unregulated power producers, gas companies, energy storage players, and those that provide power solutions to data centers. This category also includes companies involved in building power generation capacity to help meet the growing load, including regulated utilities, merchant power producers, renewables companies, and generation kit suppliers. **Supply chain/infrastructure** beneficiaries include companies positioned to invest in infrastructure or equipment to help facilitate the buildout of power infrastructure and support grid reliability. Although some stocks have moved higher on the potential for increased power demand, many of the downstream exposed names have continued to trade on more cyclical factors—like elevated interest rates—versus the secular tailwinds that we believe will contribute to future earnings growth.

Carly Davenport, Senior US Utilities Equity Research Analyst

Email: *carly.davenport@gs.com* *Goldman Sachs & Co. LLC*
Tel: *212-357-1914*

Virginia is for data centers

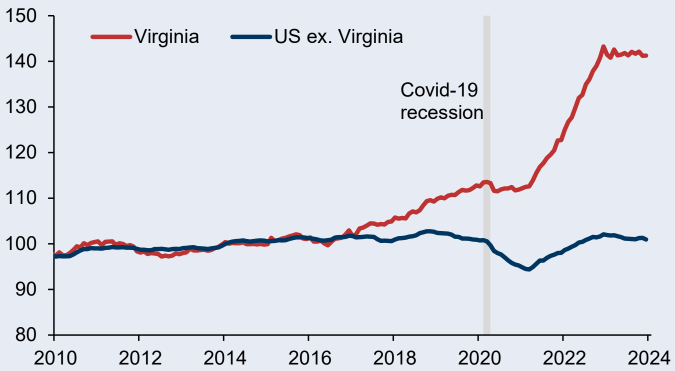
Following two decades of stagnation, investors are increasingly focused on the potential boost to US power demand from AI and data centers. While the expected increase in AI-driven power demand is in its early days (see pgs. 17-18), evidence from Virginia—the likely epicenter of this demand growth—provides a glimpse of the coming US power demand surge.

Evidence from the data center capital…

Virginia is a useful starting point to assess the boost to overall power demand in the US given its concentration of data centers. Data centers have grown rapidly in Virginia since late 2016 despite a brief pause during the pandemic, with Northern Virginia home to the most data centers in the US. Alongside this explosion of data centers, commercial power consumption in Virginia rose 37% from 2016 to 2023, while remaining flat in most other states. And within Virginia, commercial power consumption growth has also outstripped non-commercial power consumption, with both residential and industrial power consumption decreasing over 2016-2023 by 3% and 4%, respectively.

Virginia commercial power consumption growth has outpaced other states since late 2016…

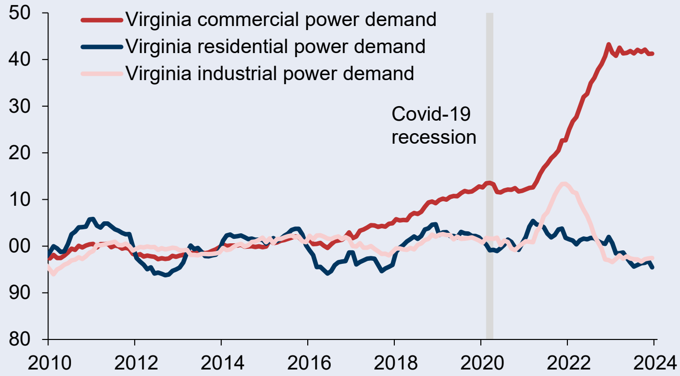
Commercial power consumption, 12m moving avg, indexed, 2010-2016 avg=100



Source: Haver Analytics, EIA, Goldman Sachs GIR.

…and has also outpaced other power consumption sectors

Power consumption across power sectors, 12m moving avg, indexed, 2010-2016 avg=100



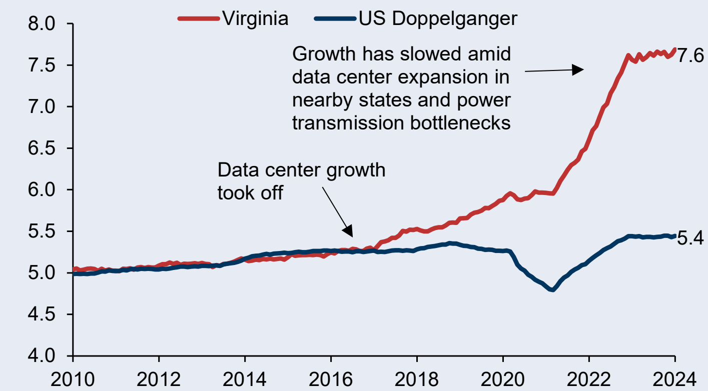
Source: Haver Analytics, EIA, Goldman Sachs GIR.

…points to a boost in power consumption…

We use a statistical "doppelganger" technique to estimate how much data centers have contributed to the observed rise in Virginia power consumption data. The doppelganger method uses the historical relationship (2010-2016) between commercial power consumption in Virginia and the US more broadly to estimate what Virginia commercial power consumption would have looked like without data centers. Taking the difference between actual Virginia power demand and doppelganger demand, we find that data centers boosted Virginia power consumption by 2.2 gigawatts (GW) in 2023, accounting for 15% of the total power consumption in the state that year, compared to virtually 0% in 2016 and roughly 3% in 2019.

Doppelganger demand for Virginia commercial power consumption points to lower consumption levels…

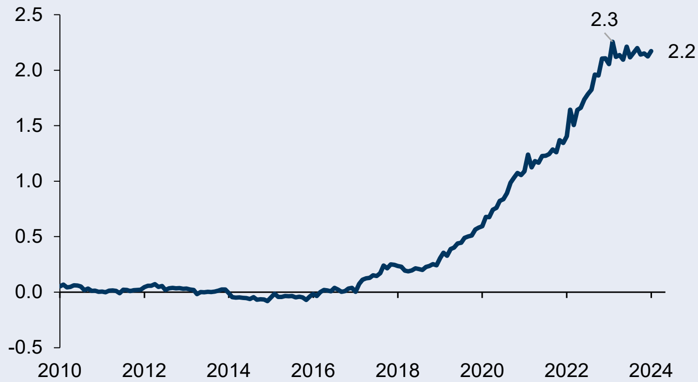
Commercial power consumption, gigawatts



Source: Haver Analytics, EIA, Goldman Sachs GIR.

…implying data center power consumption of 2.2 GW in 2023

Implied data centers power consumption of Virginia, gigawatts



Source: Haver Analytics, EIA , Goldman Sachs GIR.

…although only a modest one so far

While the evidence suggests that AI and data centers are boosting US power demand, the overall magnitude of the boost remains modest compared to both the current level of total US power demand as well as the level of data center power demand expected later this decade. We estimate the 2.2 GW of Virginia data center power demand in 2023 makes up only 0.5% of the 470 GW of total US power demand and 7% of the roughly 30 GW increase in overall data center demand our equity analysts expect by 2030. But the magnitude of the recent increase in data center power demand in Virginia provides a glimpse of the large boost in US power demand likely ahead.

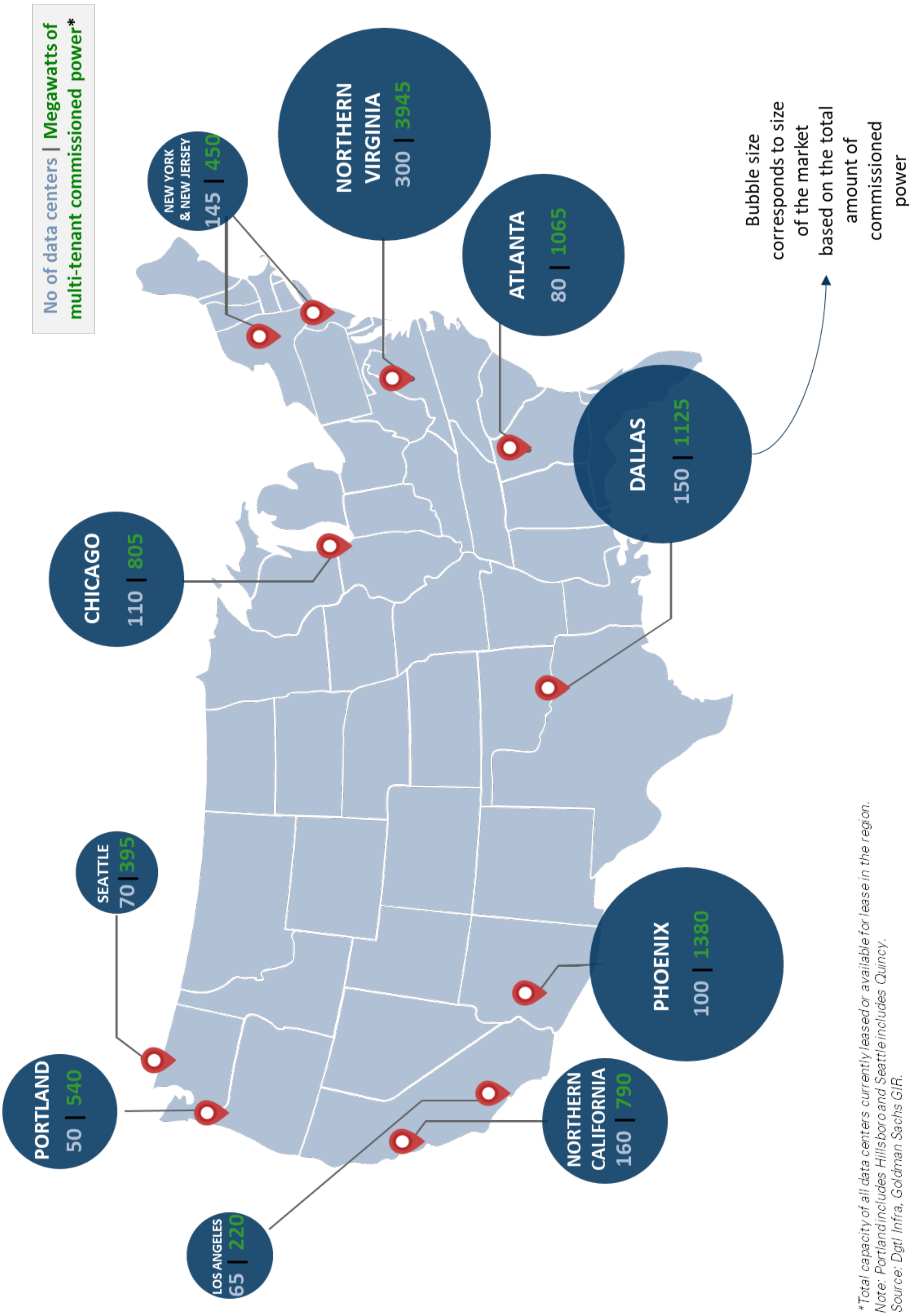
Hongcen Wei, Commodities Strategist

Email: *hongcen.wei@gs.com* *Goldman Sachs & Co. LLC*
Tel: *212-934-4691*

US data centers, mapped out

Northern Virginia is the US data center capital, though Phoenix, Dallas, and Atlanta are also hotbeds for data center growth

Top 10 US data center markets



*Total capacity of all data centers currently leased or available for lease in the region.

Note: Portland includes Hillsboro and Seattle includes Quincy.

Source: Dgtl Infra, Goldman Sachs GIR.

AI: powering up Europe

Alberto Gandolfi argues that the expansion of AI data centers will boost European power demand over the next decade, which should benefit "Electrification Compounders"

Over the past fifteen years, a series of exogenous shocks—the Global Financial Crisis, Covid pandemic, 2022 energy crisis, a slower-than-expected electrification process, and the ongoing de-industrialization of the continent's economy—have hit power demand in Europe. As a result, electricity consumption has declined by around 10% from its 2008 peak. However, this negative trend may be on the verge of reversing. We estimate that the rapid expansion of data centers amid the increasing proliferation of generative AI technology and gradual acceleration of the electrification process could boost Europe's power demand by around 40-50% over the coming decade.

AI data centers: a new driver of power demand

Traditional data centers have rapidly expanded to meet higher demand from retail customers (owing to the increased popularity of cloud storage, social media, and movie streaming), the service industry (on increased computational and storage needs), and the large tech companies such as Google, Amazon, Meta, and Microsoft. However, data centers currently account for only just over 1% of power demand globally. Our conservative base case scenario assumes that the expansion of traditional data centers could boost European power demand by around 10-15% over the coming ten years.

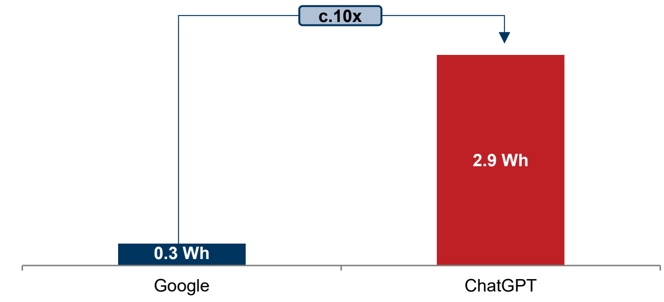
Studies show that AI data centers can consume up to around 10x more energy than traditional data centers, particularly during their training phase. We estimate that AI data centers and electrification could boost European power demand by +40% over the coming decade based on our Tech analysts' expectations for global AI shipments, conservative assumptions on energy efficiency, and a declining market share in Europe vis-à-vis the US. However, a bull case for AI data centers—which assumes a slightly higher data center market share of 25% for Europe and no efficiency gains on future server deliveries—could see cumulative electricity consumption growth of around 50% over the next decade. But even in our base case, the incremental power consumption we expect from AI and traditional data centers in Europe over the next decade would be equivalent to the current consumption of the Netherlands, Portugal, and Greece, combined.

A highly regional impact

We expect this power demand to be highly concentrated in two areas. First, countries with cheap, abundant baseload power (i.e, those that enjoy a higher proportion of wind, solar, hydro, and nuclear in their energy mix): the Nordics, Spain, and France. Second, countries with a strong financial services presence and those acting as big tech hubs as well as those willing to offer incentives to attract data centers and to support a faster adoption of electrification technologies: Germany, the UK, and Ireland. Assuming these two groups of countries—which currently account for nearly three-quarters of Europe's total power consumption—absorb 85-95% of the total incremental power demand from data centers, electricity consumption in these regions could rise by 10-15% over the next decade.

ChatGPT queries are 10x as power intensive as Google searches

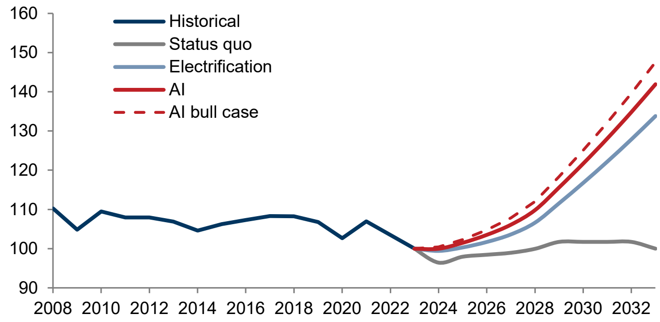
Power consumption per query/search, Watt-hour (Wh)



Source: Google, SemiAnalysis, compiled by Goldman Sachs GIR.

AI data centers and electrification could boost European power by over 40% in the coming decade

EU-27 power demand scenario analysis, index, 2023=100



Source: EMBER, Goldman Sachs GIR.

Meeting higher demand: grids and renewables to the rescue

We estimate that the rapid expansion of data centers in these areas, together with the REPowerEU Plan—which is set to kickstart a major electrification process in Europe—will lead European power demand to grow by around 40-50% over the next ten years. Investments in power grids and renewables will likely prove pivotal in meeting this substantial rise in demand. On the power grid front, we expect a secular capex supercycle ahead with European investments in power grids accelerating by 80-100%, depending on the region. And on the renewables front, we expect Europe to add nearly 800 gigawatts (GW) of wind and solar over the coming 10-15 years, nearly tripling the amount currently installed in the region.

Investment implications: look to electrification compounders

For European utilities, an industry with elevated operational and financial gearing, the coming inflection in power demand should have significant positive implications for revenues and, in turn, profits. This revenue boost will also likely trigger secular organic growth in power grids and renewables—the key infrastructure enabling the proliferation of data centers and the electrification process. And we see "Electrification Compounders"—utilities that mostly grow profits from power grids and renewables—as the main beneficiaries of the trend toward rising power demand given their highly attractive risk/reward profiles.

Alberto Gandolfi, Head of European Utilities Equity Research

Email: alberto.gandolfi@gs.com *Goldman Sachs and Co. LLC*
Tel: 39-02-8022-0157

AI's chip constraints

Our US semiconductor team led by Toshiya Hari expects supply constraints in the semiconductor industry to remain a limiting factor on AI growth over the next few years

As the popularity of generative AI technology continues to grow, the demand for AI chips—including everything from Nvidia's GPUs to custom chips designed by large cloud computing companies—has skyrocketed, leading to questions around whether the semiconductor industry can keep up. We expect industry supply, rather than demand, to dictate AI chip shipments through 2H24 and into early 2025 given constraints on two key fronts: High-Bandwidth Memory (HBM) technology and Chip-on-Wafer-on-Substrate (CoWoS) packaging.

An undersupplied HBM market

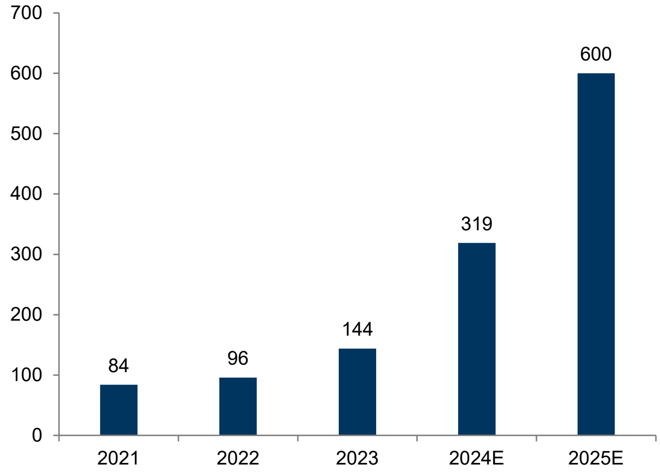
AI applications use two types of dynamic random-access memory (DRAM): HBM and DDR SDRAM. HBM is a revolutionary memory technology that stacks multiple DRAM dies—small blocks of semiconducting material on which integrated circuits are fabricated—on top of a base logic die, thereby enabling higher levels of performance through more bandwidth when interfacing with a GPU or AI chips more broadly. We expect the HBM market to grow at a ~100% compound annual growth rate (CAGR) over the next few years, from $2.3bn in 2023 to $30.2bn in 2026, as the three incumbent suppliers of DRAM (Samsung, SK Hynix, and Micron) allocate an increasing proportion of their total bit supply to meet the exponential demand growth.

Despite this ramp-up, HBM demand will likely outstrip supply over this period owing to growing HBM content requirements and major suppliers' supply discipline. We therefore forecast HBM undersupply of 3%/2%/1% in 2024/2025/2026. Indeed, as Nvidia and AMD recently indicated, updated data center GPU product roadmaps suggest that the amount of HBM required per chip will grow on a sustained basis. And lower manufacturing yield rates in HBM than in traditional DRAM given the increased complexity of the stacking process constrains suppliers' ability to increase capacity.

Packaging, bottlenecked

The other key supply bottleneck is a specific form of advanced packaging known as CoWoS, a 2.5-dimensional wafer-level multi-chip packaging technology that incorporates multiple dies side-by-side on a silicon interposer to achieve better interconnect density and performance for high-performance computing (HPC) applications. This advanced packaging capacity has been in short supply since the emergence of ChatGPT in late 2022. Although TSMC and a few other CoWoS suppliers are in the midst of expanding capacity, this undersupply is clearly serving as a gating factor to meeting AI chip demand, with the likes of Nvidia and AMD in the merchant space as well as ASIC providers continuing to highlight the tightness in CoWoS capacity. While predicting the point at which supply will catch up to demand is difficult, our Technology analysts expect TSMC's CoWoS capacity to more than double this year and nearly double again in 2025 to cater to this continued strength in AI chip demand. Accordingly, we expect chip supply to eventually catch up with robust demand, though the next few years will likely prove painful amid the constraints in critical components.

TSMC's CoWoS capacity should grow >2x in 2024 and ~2x again in 2025, easing the packaging bottleneck
TSMC's annual CoWoS capacity (k wafer per year)



Source: Goldman Sachs GIR.

Toshiya Hari, Senior US Semiconductor & Semiconductor Equipment Equity Research Analyst
Email: toshiya.hari@gs.com Goldman Sachs & Co. LLC
Tel: 646-446-1759

Anmol Makkar, US Semiconductor & Semiconductor Equipment Equity Research Analyst
Email: anmol.makkar@gs.com Goldman Sachs India SPL
Tel: 212-934-5814

David Balaban, US Semiconductor & Semiconductor Equipment Equity Research Analyst
Email: david.balaban@gs.com Goldman Sachs & Co. LLC
Tel: 212-357-7022

Full steam ahead for AI beneficiaries

Ryan Hammond sees room for the AI trade to run within US equities despite recent gains

The AI theme continues to drive the US equity market to new all-time highs, with many AI beneficiaries experiencing large rallies and Nvidia briefly becoming the largest stock in the S&P 500 following a 709% return since the start of 2023. But while AI optimism has pushed valuations for many of the large tech stocks to elevated levels, the valuation of the largest 10 TMT stocks at 31x pales in comparison to the peak of the Tech Bubble (52x) and late 2021 (43x). More broadly, we believe the AI theme has room to run, with scope for its beneficiaries to broaden as investors look to the next phase of the AI trade, and think this will benefit Utilities in particular.

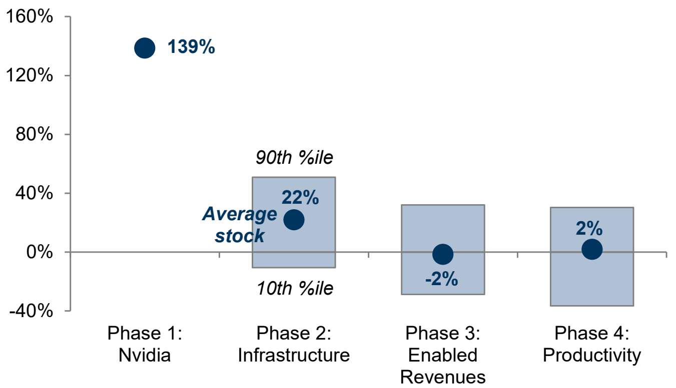
The next phases of the AI trade

We outline four phases of the AI trade. "Phase 1", which kicked off in early 2023, focuses on Nvidia, the clearest near-term AI beneficiary. "Phase 2" focuses on AI infrastructure, including semiconductor firms more broadly, cloud providers, data center REITs, hardware and equipment companies, security software stocks, and utilities companies. "Phase 3" focuses on companies with business models that can easily incorporate AI into their product offerings to boost revenues, primarily software and IT services. "Phase 4" includes companies with the biggest potential earnings boost from widespread AI adoption and productivity gains.

With Phase 1 now well underway—Nvidia has returned 139% year-to-date, accounting for 28% of the S&P 500's 15% YTD return—investors have increasingly turned their attention to Phase 2 beneficiaries. 1Q24 earnings season confirmed that many of the largest mega-cap technology companies plan to spend billions of dollars on AI-related capex investments, benefitting Phase 2 companies involved in the AI infrastructure stack. Indeed, the average Phase 2 stock has returned 22% YTD, compared with -2% for Phase 3 and 2% for Phase 4.

Phase 1 is well underway, with Phase 2 companies now reflecting more signs of AI optimism

Distribution of stock returns YTD by AI phase, %



Source: FactSet, Goldman Sachs GIR.

Utility for utilities

Utilities have emerged as a popular AI trade within Phase 2. Indeed, Utilities returned 16% between March and May, making them the best-performing sector in the S&P 500. The 3-month return ranked in the 98th percentile since 2002, surpassed only by rallies in 2003 and 2020. And relative to the equal-weight S&P 500, Utilities outperformed by 14pp, ranking in the 97th historical percentile.

Despite this strong performance, the sector remains attractive given that it offers investors two benefits: AI exposure and defensiveness. First, our equity analysts believe that a generational combination of AI demand, ex-AI demand, and a deceleration in the pace of energy efficiency gains will lead US power demand to grow at a 2.4% CAGR from 2022-2030 (see pgs. 17-18) after nearly no growth over the last decade. This increase in power demand should lead to additional capex spend among Utilities companies, which, given that many are regulated entities with capped returns, is a requirement to capture the incremental demand. Second, from a macro perspective, Utilities offer a defensive tilt in investor portfolios that should benefit as economic growth slows, as it has since 2H23, with many of the factors contributing to this slowdown here to stay, according to our economists.

While the sector's valuation is high relative to its long-term history—the sector's P/E of 16.8x ranks in the 77th percentile since 1995—it remains in line with the 10-year average, well below peak valuations of 21x in 2020 and 2022, and ranks in the 58th percentile since 1995 relative to the equal-weight S&P 500. And after adjusting valuations for the improvement in long-term EPS growth expectations that the sector has experienced, Utilities' PEG (P/E to long term growth (LTG)) ratio is 2x, well below the historical average of 3x.

The Utilities sector PEG ratio is well below its historical average

Utilities sector PEG (P/E to LTG)

Source: FactSet, Goldman Sachs GIR.

Full steam ahead, but beware of risks

While we expect Utilities to continue benefitting from the rise of AI technology, higher interest rates represent a risk to the sector, which is considered a bond proxy and typically underperforms when bond yields rise. However, our rates strategists do not expect higher rates ahead, and Fed Chair Powell also recently pushed back against renewed rate hikes.

More broadly, we expect continued investment in AI will drive outperformance of the companies exposed to that investment (Phase 2). However, signs that economic growth is slowing more sharply than expected, AI is not generating a sufficient return on investment, or outright earnings misses from the AI leaders could lead investors to reduce the valuations of perceived AI beneficiaries.

Ryan Hammond, Senior US Equity Strategist

Email: *ryan.hammond@gs.com* *Goldman Sachs & Co. LLC*
Tel: *212-902-5625*

AI optimism and long-term equity returns

Christian Mueller-Glissmann finds that a quite favorable AI scenario may be required for above-average long-term equity returns

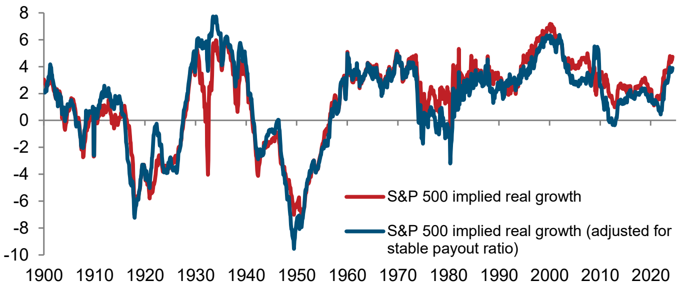
AI optimism has substantially boosted equity valuations as well as long-term market-implied S&P 500 growth expectations—similar to the experience of the 1920, 1950/60s, and late 1990s technology-led productivity booms—while US 10y yields are above their long-run median and much higher compared to the post-Global Financial Crisis (GFC) cycle. As a result, US equity risk premia (ERP) estimates, which reflect the prospective excess returns investors can expect to earn on stocks compared to risk-free assets such as bonds, are nearing some of the lowest levels since the GFC. Although the productivity pick-up that AI promises could benefit equities via higher profit growth, we find that stocks often anticipate higher productivity growth before it materializes, raising the risk of overpaying. And using our new long-term return forecasting framework, we find that a very favorable AI scenario may be required for the S&P 500 to deliver above-average returns in the coming decade.

High valuations, high implied growth

The substantial recent outperformance of the tech sector owing to AI optimism has pushed S&P 500 Shiller P/Es to exceptionally high levels—in the 97th percentile since 1900. At the same time, the market is pricing long-term real growth near 4.7%—well above its average of 2.7% since 1950 but still quite a bit below the Tech bubble peak[1].

S&P 500-implied L-T growth is elevated, but not at bubble levels

L-T real growth = ERP + 10y yield – Dividend yield – 10y breakeven inflation (assuming an ERP of 4%)



Source: Robert Shiller, Bloomberg, Goldman Sachs GIR.

Profitability, not just growth, matters

However, elevated equity valuations not only reflect economic growth optimism but also higher US corporate profitability. We find that trends in inflation and corporate profitability have materially impacted S&P 500 valuations since WWII. Equity valuations were much lower during the 1970s, but higher amid low and anchored inflation since the 1990s. And a higher Return on Equity (ROE) since the 1990s, aided by a growing weight of the highly profitable US tech sector, further explains the higher average S&P 500 Shiller P/Es of the last three decades compared with the 1960s. The high profitability and compounding of returns due to high reinvestment of large-cap US tech companies has boosted shareholder returns and supported higher valuations.

The emergence of generative AI and related potential task automation could continue this pattern by boosting productivity growth, and, in turn, corporate profitability. But this will also depend on the distribution of the technology's benefits between consumers, corporates, and governments. The US corporate profit share of GDP is already near its highest levels since WWII, with the labor share near all-time lows, resulting in very high corporate profitability. And several potential headwinds may make it more difficult for the corporate sector to outperform the economy going forward, including labor scarcity, supply chain diversification and re-shoring, more antitrust and other regulation, higher interest rates, risks from decarbonization/climate change, and potentially higher taxes.

Lower ERP in all but the most optimistic AI/macro scenarios

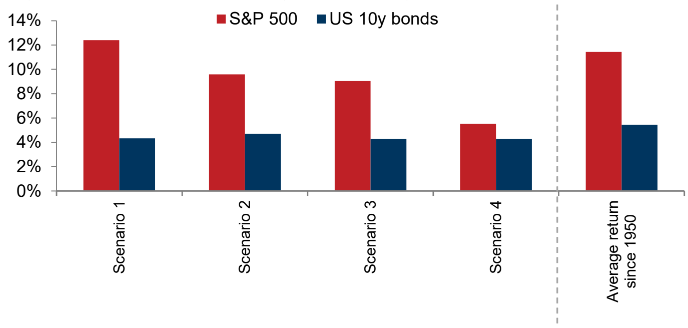
To assess how the current elevated equity valuations, together with changing macro conditions, may impact equity and bond returns over the longer term, we built a cross-asset long-term return forecasting framework that incorporates starting valuations and expectations of trend GDP growth, inflation, and profitability, to control for different structural cycle scenarios.

We focus on four scenarios with different impacts of AI on trend growth, inflation, and profitability (ROE): (1) Large trend GDP growth boost to 4% with 2% inflation similar to the post-1990s average and the S&P 500 ROE back to peak levels (21%), (2) Small trend GDP growth boost to 3%, 2% inflation, and ROE at current levels, (3) Small trend GDP growth boost to 3% but higher inflation (3%) and ROE at current levels (4) Small trend GDP growth boost to 3%, 2% inflation, and ROE declines to its post-1990s average of 15%.

Outside of the most bullish AI scenario that includes a material improvement to the structural growth/inflation mix and peak US corporate profitability, we forecast that S&P 500 returns would be below their post-1950 average. AI's impact on corporate profitability will matter critically—reverting to the post-1990s average ROE would materially weigh on equity valuations and returns in the coming decade. In such a scenario, equities would struggle to outperform bonds, pointing to very little reward for taking on equity risk and, in turn, low expected ERP.

A very favorable AI scenario may be required for above-average S&P 500 returns in the coming decade

10y nominal S&P 500 and US 10y bond return per annum across scenarios



Source: Goldman Sachs GIR.

Christian Mueller-Glissmann, Sr. Multi-Asset Strategist

Email: christian.mueller-glissmann@gs.com Goldman Sachs International
Tel: 44-20-7774-1714

[1] Based on a single-stage dividend discount model. Assuming the same payout ratio since 1900 (as many US tech stocks do not pay dividends) results in slightly lower implied real growth of 3.9%.

Summary of our key forecasts

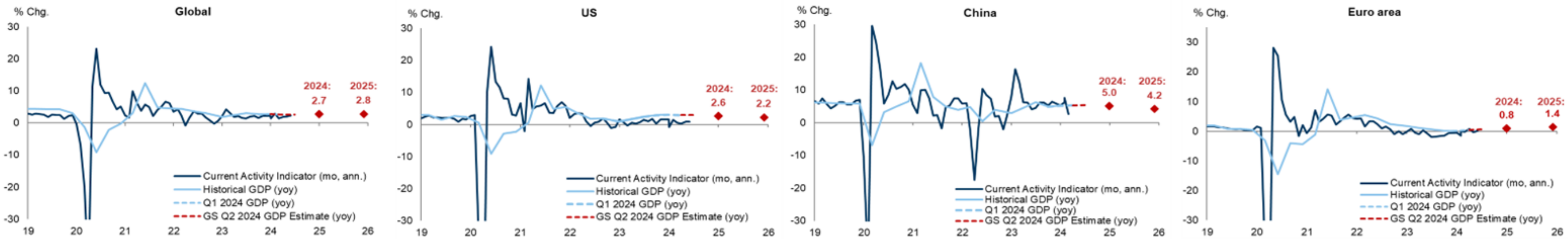
GS GIR: Macro at a glance

Watching

- **Globally,** we expect real GDP growth of 2.7% yoy in 2024, reflecting tailwinds from real household income growth, a gradual recovery in manufacturing activity, and broadening rate cuts. We expect global core inflation to fall below 3% by end-2024 and converge towards 2% by end-2025 as core goods inflation continues to decline, shelter inflation falls further, and both services inflation and wage growth continue to slow in response to the improved supply-demand balance across the global economy.
- **In the US,** we expect a moderate growth pickup in 2H24 as financial conditions ease for real GDP growth of 2.2% in 2024 on a Q4/Q4 basis, although we think most of the growth slowdown from the 4.1% pace in 2H23 is here to stay as real income growth has softened, consumer sentiment has fallen, and early signs of an increase in election-related uncertainty that could weigh on business investment have emerged. We expect core PCE inflation to stand at 2.7% yoy by December 2024 before converging toward 2% next year, reflecting further rebalancing in the auto and housing rental markets. We expect the unemployment rate to end 2024 at 4.0% and remain there for the next few years.
- **We believe the Fed** will remain on hold at the current fed funds rate range of 5.25-5.5% until the first 25bp cut in September, after which we expect rate cuts to proceed at a quarterly pace with the next 25bp cut in December until the terminal rate range reaches 3.25-3.5%.
- **In the Euro area,** we expect real GDP growth to increase to 0.8% yoy in 2024, reflecting a pickup in real disposable income and a fading credit drag amid ECB rate cuts, which should more than offset the continued fiscal drag. We expect core inflation to slow further to 2.6% yoy by December 2024, reflecting continued declines in services inflation, normalizing wage growth, and further scope for energy-related effects to fade.
- **We expect the ECB** to deliver a 25bp cut in September, after which we expect cuts to continue at a quarterly pace until the policy rate reaches 2.25% in 4Q25, although the risks are skewed toward a semiannual pace of rate cuts.
- **In China**, we expect real GDP growth of 5.0% yoy in 2024 as growth headwinds such as a prolonged property downturn and the lack of confidence among households and private businesses are offset by strong exports and continued policy easing. We expect inflation to remain low in 2024, with continued PPI deflation and moderate CPI reflation amid the ongoing food price deflation, property downturn, and manufacturing overcapacity. Over the longer term, we remain cautious on China's growth outlook given deteriorating demographics, property and local government deleveraging, and global supply chain de-risking.
- **WATCH GEOPOLITICAL RISKS AND US AND OTHER ELECTIONS.** Geopolitical tensions remain elevated as the situation in the Middle East remains highly uncertain, the Russia-Ukraine war drags on, and US-China relations continue to be fraught, which could have material market implications. In the US, the all-but-certain Trump-Biden rematch in the November general election could have important macro and market implications, especially if it brings the possibility of fresh unfunded fiscal expansion or a further rise in tariffs. And economic and market implications of recent election-related surprises elsewhere—including in India, Mexico, and South Africa as well as the announcement of snap parliamentary elections in France—are also worth watching.

Goldman Sachs Global Investment Research.

Growth



Source: Haver Analytics and Goldman Sachs Global Investment Research.
Note: GS CAI is a measure of current growth. For more information on the methodology of the CAI please see "Improving Our Within-Month CAI Forecasts," Global Economics Comment, Mar. 06, 2023.

Forecasts

Economics

GDP growth (%)	2024 GS (Q4/Q4)	2024 Cons. (Q4/Q4)	2024 GS (CY)	2024 Cons. (CY)	2025 GS (CY)	2025 Cons. (CY)
Global	2.7	--	2.7	2.6	2.8	2.6
US	2.2	1.6	2.6	2.3	2.2	1.8
China	4.7	4.7	5.0	4.9	4.2	4.5
Euro area	1.4	1.3	0.8	0.7	1.4	1.4

Policy rates (%)	2024 GS	2024 Mkt.	2025 GS	2025 Mkt.
US	4.88	4.86	3.88	3.87
Euro area	3.25	3.22	2.25	2.50
China	1.70	1.34	1.70	--
Japan	0.13	0.27	0.63	0.54

Markets

Interest rates 10Yr (%)	Last	E2024	E2025
US	4.25	4.25	4.10
Germany	2.42	2.25	2.00
Japan	0.99	1.25	1.80
UK	4.02	3.75	3.75

FX	Last	3m	12m
EUR/$	1.07	1.05	1.08
GBP/$	1.27	1.24	1.28
$/JPY	160	155	150
$/CNY	7.24	7.30	7.20

Commodities	Last	3m	12m
Crude Oil, Brent ($/bbl)	86	86	81
Nat Gas, NYMEX ($/mmBtu)	2.81	2.70	4.00
Nat Gas, TTF (EUR/MWh)	34.28	30	32
Copper ($/mt)	9,513	10,500	13,000
Gold ($/troy oz)	2,329	2,600	2,700

Credit (bp)		Last	1H24	2H24
USD	IG	93	89	90
	HY	310	297	291
EUR	IG	134	121	120
	HY	348	341	336

Equities

S&P 500	E2024 GS	E2024 Cons.	E2025 GS	E2025 Cons.
Price	5,600	--	--	--
EPS	$241	$243	$256	$278
Growth	8%	9%	6%	14%

Returns (%)	12m	YTD	E2024 P/E
S&P 500	5	14.2	22.7x
MXAPJ	9	6.8	14.6x
Topix	6	15.8	16x
STOXX 600	4	8.3	14.1x

Consumer	2024 CPI (%, yoy)	2024 Unemp. Rate	2025 CPI (%, yoy)	2025 Unemp. Rate
US	3.1	4.0	2.5	4.0
Euro area	2.5	6.7	2.2	6.7
China	0.4	--	1.5	--

Wage Tracker 2024 (%)	Q1	Q2	Q3	Q4
US	4.1	4.0	--	--
Euro area	--	--	--	--
China	--	--	--	--

Source: Bloomberg, Goldman Sachs Global Investment Research. For important disclosures, see the Disclosure Appendix or go to www.gs.com/research/hedge.html.

Market pricing as of June 24, 2024

Glossary of GS proprietary indices

Current Activity Indicator (CAI)

GS CAIs measure the growth signal in a broad range of weekly and monthly indicators, offering an alternative to Gross Domestic Product (GDP). GDP is an imperfect guide to current activity: In most countries, it is only available quarterly and is released with a substantial delay, and its initial estimates are often heavily revised. GDP also ignores important measures of real activity, such as employment and the purchasing managers' indexes (PMIs). All of these problems reduce the effectiveness of GDP for investment and policy decisions. Our CAIs aim to address GDP's shortcomings and provide a timelier read on the pace of growth.

For more, see our CAI page and Global Economics Analyst: Trackin' All Over the World – Our New Global CAI, 25 February 2017.

Dynamic Equilibrium Exchange Rates (DEER)

The GSDEER framework establishes an equilibrium (or "fair") value of the real exchange rate based on relative productivity and terms-of-trade differentials.

For more, see our GSDEER page, Global Economics Paper No. 227: Finding Fair Value in EM FX, 26 January 2016, and Global Markets Analyst: A Look at Valuation Across G10 FX, 29 June 2017.

Financial Conditions Index (FCI)

GS FCIs gauge the "looseness" or "tightness" of financial conditions across the world's major economies, incorporating variables that directly affect spending on domestically produced goods and services. FCIs can provide valuable information about the economic growth outlook and the direct and indirect effects of monetary policy on real economic activity.

FCIs for the G10 economies are calculated as a weighted average of a policy rate, a long-term risk-free bond yield, a corporate credit spread, an equity price variable, and a trade-weighted exchange rate; the Euro area FCI also includes a sovereign credit spread. The weights mirror the effects of the financial variables on real GDP growth in our models over a one-year horizon. FCIs for emerging markets are calculated as a weighted average of a short-term interest rate, a long-term swap rate, a CDS spread, an equity price variable, a trade-weighted exchange rate, and—in economies with large foreign-currency-denominated debt stocks—a debt-weighted exchange rate index.

For more, see our FCI page, Global Economics Analyst: Our New G10 Financial Conditions Indices, 20 April 2017, and Global Economics Analyst: Tracking EM Financial Conditions – Our New FCIs, 6 October 2017.

Goldman Sachs Analyst Index (GSAI)

The US GSAI is based on a monthly survey of GS equity analysts to obtain their assessments of business conditions in the industries they follow. The results provide timely "bottom-up" information about US economic activity to supplement and cross-check our analysis of "top-down" data. Based on analysts' responses, we create a diffusion index for economic activity comparable to the ISM's indexes for activity in the manufacturing and nonmanufacturing sectors.

Macro-Data Assessment Platform (MAP)

GS MAP scores facilitate rapid interpretation of new data releases for economic indicators worldwide. MAP summarizes the importance of a specific data release (i.e., its historical correlation with GDP) and the degree of surprise relative to the consensus forecast. The sign on the degree of surprise characterizes underperformance with a negative number and outperformance with a positive number. Each of these two components is ranked on a scale from 0 to 5, with the MAP score being the product of the two, i.e., from -25 to +25. For example, a MAP score of +20 (5;+4) would indicate that the data has a very high correlation to GDP (5) and that it came out well above consensus expectations (+4), for a total MAP value of +20.

Top of Mind archive: click to access

Source of photos: www.gettyimages.com, www.istockphoto.com, www.shutterstock.com, US Department of State/Wikimedia Commons/Public Domain.

Disclosure Appendix

Reg AC

We, Allison Nathan, Jenny Grimberg, Ashley Rhodes, Joseph Briggs, Christian Mueller-Glissmann, Ryan Hammond, and Hongcen Wei hereby certify that all of the views expressed in this report accurately reflect our personal views, which have not been influenced by considerations of the firm's business or client relationships.

We, Jim Covello, Alberto Gandolfi, Toshiya Hari, Kash Rangan, Eric Sheridan, Carly Davenport, Gili Naftalovich, Anmol Makkar, and David Balaban, hereby certify that all of the views expressed in this report accurately reflect our personal views about the subject company or companies and its or their securities. We also certify that no part of our compensation was, is or will be, directly or indirectly, related to the specific recommendations or views expressed in this report.

Unless otherwise stated, the individuals listed on the cover page of this report are analysts in Goldman Sachs' Global Investment Research division.

Logo Disclosure

Please note: Third party brands used in this report are the property of their respective owners, and are used here for informational purposes only. The use of such brands should not be viewed as an endorsement, affiliation, or sponsorship by or for Goldman Sachs or any of its products/services.

Disclosures

Regulatory disclosures

Disclosures required by United States laws and regulations

See company-specific regulatory disclosures above for any of the following disclosures required as to companies referred to in this report: manager or co-manager in a pending transaction; 1% or other ownership; compensation for certain services; types of client relationships; managed/co-managed public offerings in prior periods; directorships; for equity securities, market making and/or specialist role. Goldman Sachs trades or may trade as a principal in debt securities (or in related derivatives) of issuers discussed in this report.

The following are additional required disclosures: **Ownership and material conflicts of interest:** Goldman Sachs policy prohibits its analysts, professionals reporting to analysts and members of their households from owning securities of any company in the analyst's area of coverage. **Analyst compensation:** Analysts are paid in part based on the profitability of Goldman Sachs, which includes investment banking revenues. **Analyst as officer or director:** Goldman Sachs policy generally prohibits its analysts, persons reporting to analysts or members of their households from serving as an officer, director, or advisor of any company in the analyst's area of coverage. **Non-U.S. Analysts:** Non-U.S. analysts may not be associated persons of Goldman Sachs & Co. LLC and therefore may not be subject to FINRA Rule 2241 or FINRA Rule 2242 restrictions on communications with subject company, public appearances and trading securities held by the analysts.

Additional disclosures required under the laws and regulations of jurisdictions other than the United States

The following disclosures are those required by the jurisdiction indicated, except to the extent already made above pursuant to United States laws and regulations. **Australia:** Goldman Sachs Australia Pty Ltd and its affiliates are not authorised deposit-taking institutions (as that term is defined in the Banking Act 1959 (Cth)) in Australia and do not provide banking services, nor carry on a banking business, in Australia. This research, and any access to it, is intended only for "wholesale clients" within the meaning of the Australian Corporations Act, unless otherwise agreed by Goldman Sachs. In producing research reports, members of Global Investment Research of Goldman Sachs Australia may attend site visits and other meetings hosted by the companies and other entities which are the subject of its research reports. In some instances the costs of such site visits or meetings may be met in part or in whole by the issuers concerned if Goldman Sachs Australia considers it is appropriate and reasonable in the specific circumstances relating to the site visit or meeting. To the extent that the contents of this document contains any financial product advice, it is general advice only and has been prepared by Goldman Sachs without taking into account a client's objectives, financial situation or needs. A client should, before acting on any such advice, consider the appropriateness of the advice having regard to the client's own objectives, financial situation and needs. A copy of certain Goldman Sachs Australia and New Zealand disclosure of interests and a copy of Goldman Sachs' Australian Sell-Side Research Independence Policy Statement are available at: https://www.goldmansachs.com/disclosures/australia-new-zealand/index.html. **Brazil:** Disclosure information in relation to CVM Resolution n. 20 is available at https://www.gs.com/worldwide/brazil/area/gir/index.html. Where applicable, the Brazil-registered analyst primarily responsible for the content of this research report, as defined in Article 20 of CVM Resolution n. 20, is the first author named at the beginning of this report, unless indicated otherwise at the end of the text. **Canada:** This information is being provided to you for information purposes only and is not, and under no circumstances should be construed as, an advertisement, offering or solicitation by Goldman Sachs & Co. LLC for purchasers of securities in Canada to trade in any Canadian security. Goldman Sachs & Co. LLC is not registered as a dealer in any jurisdiction in Canada under applicable Canadian securities laws and generally is not permitted to trade in Canadian securities and may be prohibited from selling certain securities and products in certain jurisdictions in Canada. If you wish to trade in any Canadian securities or other products in Canada please contact Goldman Sachs Canada Inc., an affiliate of The Goldman Sachs Group Inc., or another registered Canadian dealer. **Hong Kong:** Further

European Economic Area: GSI, authorised by the PRA and regulated by the FCA and the PRA, disseminates research in the following jurisdictions within the European Economic Area: the Grand Duchy of Luxembourg, Italy, the Kingdom of Belgium, the Kingdom of Denmark, the Kingdom of Norway, the Republic of Finland and the Republic of Ireland; GSI - Succursale de Paris (Paris branch) which is authorised by the French Autorité de contrôle prudentiel et de resolution ("ACPR") and regulated by the Autorité de contrôle prudentiel et de resolution and the Autorité des marches financiers ("AMF") disseminates research in France; GSI - Sucursal en España (Madrid branch) authorized in Spain by the Comisión Nacional del Mercado de Valores disseminates research in the Kingdom of Spain; GSI - Sweden Bankfilial (Stockholm branch) is authorized by the SFSA as a "third country branch" in accordance with Chapter 4, Section 4 of the Swedish Securities and Market Act (Sw. lag (2007:528) om värdepappersmarknaden) disseminates research in the Kingdom of Sweden; Goldman Sachs Bank Europe SE ("GSBE") is a credit institution incorporated in Germany and, within the Single Supervisory Mechanism, subject to direct prudential supervision by the European Central Bank and in other respects supervised by German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin) and Deutsche Bundesbank and disseminates research in the Federal Republic of Germany and those jurisdictions within the European Economic Area where GSI is not authorised to disseminate research and additionally, GSBE, Copenhagen Branch filial af GSBE, Tyskland, supervised by the Danish Financial Authority disseminates research in the Kingdom of Denmark; GSBE - Sucursal en España (Madrid branch) subject (to a limited extent) to local supervision by the Bank of Spain disseminates research in the Kingdom of Spain; GSBE - Succursale Italia (Milan branch) to the relevant applicable extent, subject to local supervision by the Bank of Italy (Banca d'Italia) and the Italian Companies and Exchange Commission (Commissione Nazionale per le Società e la Borsa "Consob") disseminates research in Italy; GSBE - Succursale de Paris (Paris branch), supervised by the AMF and by the ACPR disseminates research in France; and GSBE - Sweden Bankfilial (Stockholm branch), to a limited extent, subject to local supervision by the Swedish Financial Supervisory Authority (Finansinpektionen) disseminates research in the Kingdom of Sweden.

General disclosures

This research is for our clients only. Other than disclosures relating to Goldman Sachs, this research is based on current public information that we consider reliable, but we do not represent it is accurate or complete, and it should not be relied on as such. The information, opinions, estimates, and forecasts contained herein are as of the date hereof and are subject to change without prior notification. We seek to update our research as appropriate, but various regulations may prevent us from doing so. Other than certain industry reports published on a periodic basis, the large majority of reports are published at irregular intervals as appropriate in the analyst's judgment.

Goldman Sachs conducts a global full-service, integrated investment banking, investment management, and brokerage business. We have investment banking and other business relationships with a substantial percentage of the companies covered by Global Investment Research. Goldman Sachs & Co. LLC, the United States broker dealer, is a member of SIPC (https://www.sipc.org).

Our salespeople, traders, and other professionals may provide oral or written market commentary or trading strategies to our clients and principal trading desks that reflect opinions that are contrary to the opinions expressed in this research. Our asset management area, principal trading desks and investing businesses may make investment decisions that are inconsistent with the recommendations or views expressed in this research.

We and our affiliates, officers, directors, and employees will from time to time have long or short positions in, act as principal in, and buy or sell, the securities or derivatives, if any, referred to in this research, unless otherwise prohibited by regulation or Goldman Sachs policy.

The views attributed to third party presenters at Goldman Sachs arranged conferences, including individuals from other parts of Goldman Sachs, do not necessarily reflect those of Global Investment Research and are not an official view of Goldman Sachs.

Any third party referenced herein, including any salespeople, traders and other professionals or members of their household, may have positions in the products mentioned that are inconsistent with the views expressed by analysts named in this report.

This research is focused on investment themes across markets, industries, and sectors. It does not attempt to distinguish between the prospects or performance of, or provide analysis of, individual companies within any industry or sector we describe.

Any trading recommendation in this research relating to an equity or credit security or securities within an industry or sector is reflective of the investment theme being discussed and is not a recommendation of any such security in isolation.

This research is not an offer to sell or the solicitation of an offer to buy any security in any jurisdiction where such an offer or solicitation would be illegal. It does not constitute a personal recommendation or take into account the particular investment objectives, financial situations, or needs of individual clients. Clients should consider whether any advice or recommendation in this research is suitable for their particular circumstances and, if appropriate, seek professional advice, including tax advice. The price and value of investments referred to in this research and the income from them may fluctuate. Past performance is not a guide to future performance, future returns are not guaranteed, and a loss of original capital may occur. Fluctuations in exchange rates could have adverse effects on the value or price of, or income derived from, certain investments.