Decoding the Agentic Economy

The Coming Inflection in AI Usage and Margins

Agentic AI has burst onto the evolving AI scene with significant promise, capturing the imagination of industry practitioners and investors alike. At its fullest, Agentic AI can handle a wide range of tasks currently done by humans in a fully autonomous way. On the other hand, Agentic AI could be misdirected and counterproductive, consuming vast resources with little return. In this report, we outline some of the likely use cases we see for Agentic AI across the enterprise and consumer sphere – and quantify potential upside to business outcomes, along with the investment levels required.

Moving past concepts to numbers: We see Agentic AI driving a dramatic increase in token consumption of 24X or 120 quadrillion tokens per month by 2030. We think enterprise agents will be the largest token multiplier, lifting token consumption 55X by 2040. We believe consumer agents will broaden usage away from episodic chats to utility beyond traditional search, driving 12X token consumption by 2030.

Why now is the time – an economic inflection point for hyperscalers and model makers: Critically, we outline why we see an inflection in token economics ahead for hyperscalers and model providers, enabled by continued token cost declines (powered by leading semi companies driving 60%-70% lower annualized cost per token) and stabilizing token prices (which have moved from ~40% annual declines to flat or increasing pricing). We believe these improving economics are likely to improve margins and the overall economic model for hyperscalers and model providers - with a positive gross margin inflection likely over the next 3-12 months - thus making infrastructure spending more sustainable for the entire ecosystem.

Where we are differentiated: We have taken a bottom-up approach to arrive at our forecasts, building "real world" implementable agents in pseudo-code to estimate token consumption and overall costs. Similarly, on the compute cost side, we have used chip performance, benchmark, and pricing data to arrive at our estimates of all-in token costs and estimate margin inflection points for the industry.

Investment views: In Semiconductors, we prefer Broadcom (leader in custom silicon), **Nvidia** (high-performance merchant solutions leader), and **AMD** (enterprise CPU leader & emerging GPU player). **In Internet, we prefer Alphabet** (cloud & consumer computing utility), **Amazon** (leader in cloud computing, eCommerce), and **Meta** (digital advertising & spatial computing). **In Software, we prefer Microsoft** (leader in broad enterprise workflows), **Cloudflare** (leader in edge compute), and **Accenture** (AI business transformation).

James Schneider, Ph.D.
+1(212)357-2929
jim.schneider@gs.com
Goldman Sachs & Co. LLC

Eric Sheridan
+1(917)343-8683
eric.sheridan@gs.com
Goldman Sachs & Co. LLC

Gabriela Borges, CFA
+1(212)902-7839
gabriela.borges@gs.com
Goldman Sachs & Co. LLC

Luya You
+1(212)902-5297
luya.you@gs.com
Goldman Sachs & Co. LLC

Anmol Makkar
+1(212)357-1366
anmol.makkar@gs.com
Goldman Sachs & Co. LLC

Noah Naparst
+1(917)343-6395
noah.x.naparst@gs.com
Goldman Sachs & Co. LLC

Emma Huang
+1(212)902-7229
emma.huang@gs.com
Goldman Sachs & Co. LLC

Goldman Sachs does and seeks to do business with companies covered in its research reports. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision. For Reg AC certification and other important disclosures, see the Disclosure Appendix, or go to www.gs.com/research/hedge.html. Analysts employed by non-US affiliates are not registered/qualified as research analysts with FINRA in the U.S.

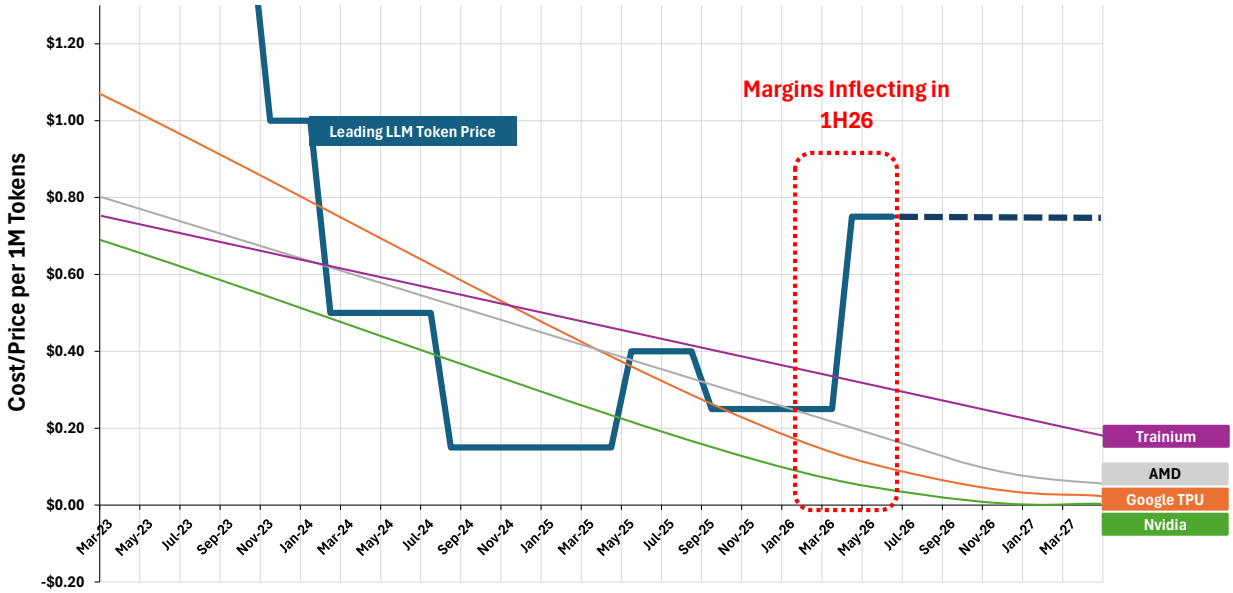
The Goldman Sachs Group, Inc.

Executive Summary

We See AI Unit Economics and Margins Inflecting as Agentic AI Usage Takes Off

- **Token explosion meets margin inflection.** We believe Agentic AI could drive a step-function change in token consumption, just as token economics are beginning to improve. Although the industry is compute-capacity constrained in 2026, our inference price vs. cost analysis curve suggests that unit economics of tokens are set to improve. Leading LLM token prices have now started to stabilize - but underlying compute cost per token across Nvidia, AMD, Google TPU, and Trainium continues to fall significantly faster. **In other words, the Agentic AI explosion in token consumption we model by 2030 may not only reflect a demand story, but also a margin expansion and profitability story across the AI value chain.**

Exhibit 1: Token economics are beginning to inflect, as compute cost declines continue and token prices stabilize

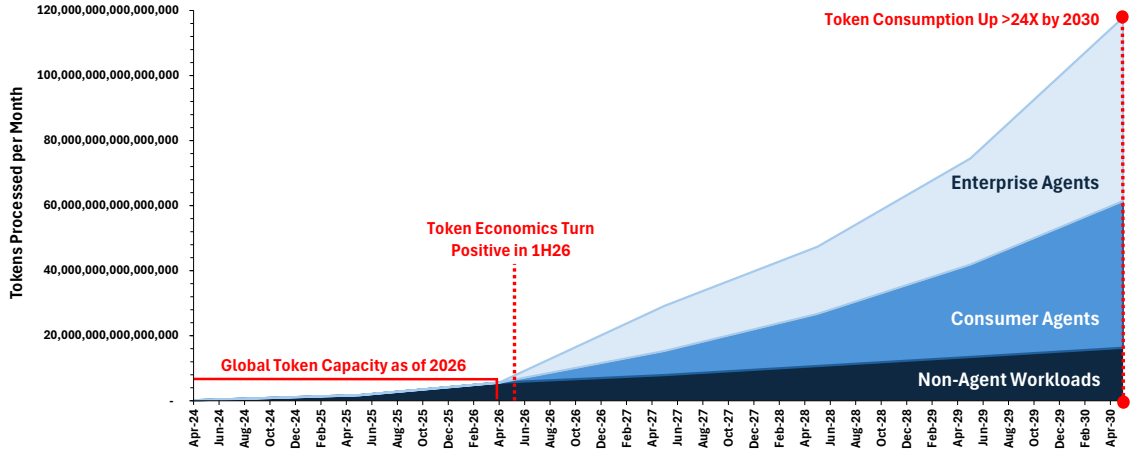


Source: Data compiled by Goldman Sachs Global Investment Research

Sizing the Consumer Agent Opportunity

- **We estimate consumer AI agents can lift global token consumption 12X by 2030.** Consumer AI queries are already a large and growing market, but the mix is shifting quickly as AI overviews and LLM agentic queries take share from traditional search. At ~23 billion AI queries per day by 2030, up from ~5 billion in 2025, we estimate up to 30% will be directed to agents across search, shopping, travel, email, and other personal productivity functions. This would add 60 quadrillion tokens per month by 2030, or 12X current global token consumption as of 2026. **Our simulations of basic consumer agents suggest the largest token step-up will come when agents move from user-initiated tasks to persistent, "always-on" background activity that continuously monitors context and acts when needed.**

Exhibit 2: By 2030, we estimate consumer and enterprise agents could push token consumption 24X above today's estimated global capacity.



Source: Data compiled by Goldman Sachs Global Investment Research

Sizing the Enterprise Agent Opportunity

- **We estimate enterprise AI agents can lift global token consumption 24X by 2030 and 55X by 2040.** Enterprise adoption of agentic AI is still early: while surveys suggest 70–90% of enterprises are experimenting, less than one-quarter are scaling agents. Ultimately, we think the curve will be S-Shaped. To validate token intensity, we built simulated agents across AI-exposed occupations to estimate the minimum tokens required for an AI agent to replicate core workflows. We found that some agents can consume very high token volumes but remain relatively low cost if the workflow is mostly text-based (such as coding), but others can consume fewer tokens and carry much higher API cost because they require multi-modal processing (such as real-time calls and video). **This tension means we do not expect agentic ROI to be adopted evenly across all enterprise workflows.**

Investment Implications

- **For hyperscalers and model providers, we prefer the following stocks (covered by Eric Sheridan)**:
  - **AMZN (Buy, $325 12-m PT):** We continue to see visibility into returns as AWS revenues compound, supported by a reported $364bn revenue backlog, driven by both AI workloads and rising momentum around its custom silicon (Trainium, Graviton, etc.).
  - **GOOGL (Buy, 12-m $450 PT):** Alphabet is seeing momentum across its Cloud business and Search multi-modality, leveraging a full-stack approach as management continues to see AI repositioning the company for sustained growth.
  - **META (Buy, 12-m $830 PT):** Meta remains a leader in its core advertising business (significantly outpacing total digital ad industry growth) as the application of AI-related compute is driving momentum around engagement and ads monetization.

- **For semiconductor companies, we prefer the following stocks (covered by Jim Schneider):**
  - **AVGO (Buy, 12-m $480 PT):** As the market leader in custom computing, we see more hyperscalers (Google) and LLM model providers turning to Broadcom to deliver cost-optimized chip solutions tailored to their specific workloads.
  - **NVDA (Buy, 12-m $250 PT):** We believe Nvidia can retain its dominant market leadership in the medium term as it remains the leader in AI performance across a broad range of training and inference workloads.
  - **AMD (Buy, $450 12-m PT):** We see AMD's market position strengthening as the company scales its high-performance datacenter GPU offerings over the next two years. Importantly, we believe AMD is also poised for an increasing share of agentic AI workloads in the enterprise as it gains share in X86 server CPUs and the CPU attach rate increases.
- **For software and IT services companies, we prefer the following stocks**:
  - **MSFT (Buy, $610 12-m PT, covered by Gabriela Borges):** Copilot feedback is getting better and the E7 upgrade cycle may drive further acceleration in Microsoft 365. The most likely scenario may be an ecosystem where Copilot coexists alongside domain-specific agents and domain-specific app software, and the usage of one pulls through usage of the others reciprocally.
  - **NET (Buy, $250 12-m PT, covered by Gabriela Borges):** We expect Cloudflare to take outsized share of AI inference workloads because of its performance and cost advantages, in turn driven by its architectural network advantages and the sophistication of its *isolates* software.
  - **ACN (Buy, $300 12-m PT, covered by Jim Schneider):** We expect Accenture to see growing tailwinds from agentic adoption as enterprises increasingly move from AI pilots to scaled agent deployments, driving demand for integration, workflow redesign, governance, and change management.

We see unit economics inflecting for Agentic AI: Implications for margins, ROI, and CapEx

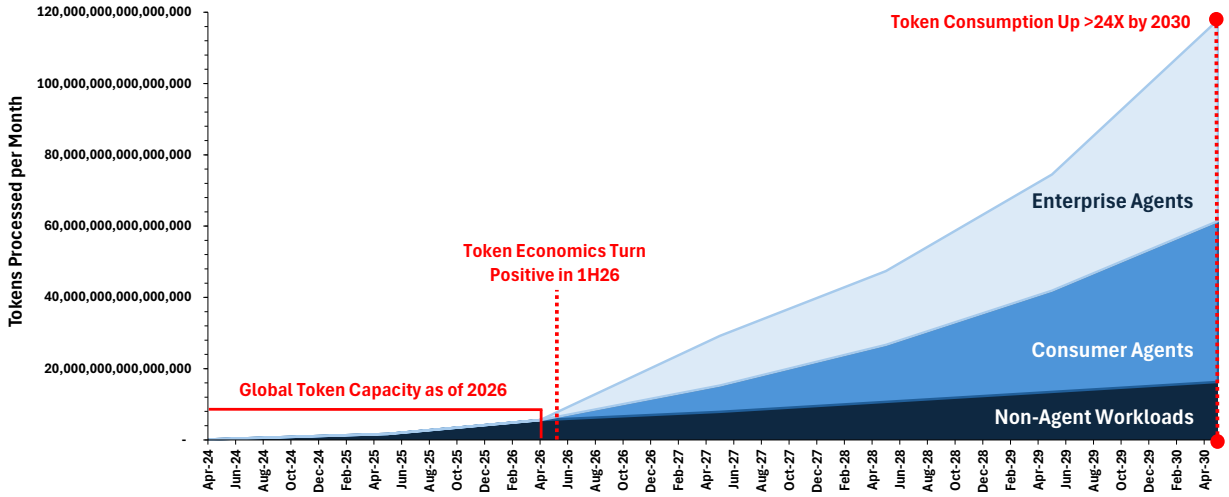
Token volume explosion may finally give way to a positive margin inflection for the hyperscalers

The Bottom Line: Agentic AI is likely to drive a step-function change in token consumption, just as token economics are beginning to improve for hyperscalers and LLM providers. By 2030, our bottom-up framework (built on our simulated agents that validate real-world token intensity) suggests that global token demand could rise by 2,400% versus 2026 levels, with consumer agents potentially reaching ~60 quadrillion tokens per month and enterprise agents reaching ~56 quadrillion tokens per month (or 278 quadrillion by peak adoption in 2040), assuming adoption broadens.

More important, our inference token price vs. cost curve suggests that the underlying unit economics of tokens are set to improve significantly for hyperscalers and LLM providers. Leading LLM token prices have declined rapidly, but have now started to stabilize or even increase in some cases. At the same time, our calculated all-in token compute costs for hyperscalers and LLM providers – powered by Nvidia, Broadcom, AMD, and Marvell appear to be falling faster. Simply put – the dramatic increase in token consumption we have modeled by 2030 may not only reflect a demand and revenue narrative, but also a margin expansion and profitability story across the AI value chain.

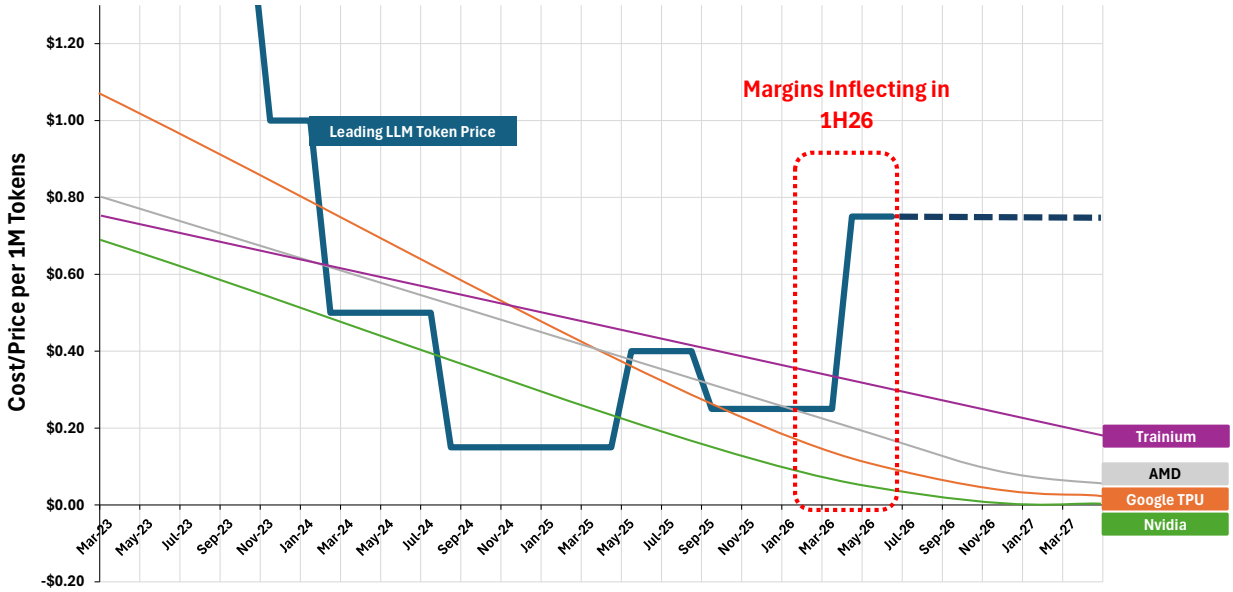
In the first phase of the AI cycle, investors largely viewed compute and tokens as a cost driver: more usage meant more inference load, more accelerators, more power, and more CapEx. But the shape of the inference price vs. cost curve suggests that this trend is changing. Although leading LLM token prices have declined meaningfully over time they have now started to stabilize or even increase in some cases. At the same time, our calculated all-in cost of compute per token for Nvidia, Google TPU (Broadcom), AMD, and Trainium (Marvell) continues to fall rapidly and more consistently. If token prices stabilize at levels higher than token costs, this implies that an increase in agentic AI adoption could produce positive margin expansion, not just revenue growth. **In other words, we now believe the industry could be moving from a phase where inference economics were uncertain and potentially dilutive to margins to a phase where token growth increasingly drops through at attractive incremental margins.**

Exhibit 3: Agentic AI may drive the next wave of token demand
By 2030, we estimate consumer and enterprise agents could push token consumption 24X above today's estimated global capacity.



Source: Data compiled by Goldman Sachs Global Investment Research

Exhibit 4: Token economics are beginning to turn more favorable
LLM token pricing has fallen dramatically, but has now started to stabilize - while underlying compute costs continue to fall faster - creating room for positive margin inflection



Source: Data compiled by Goldman Sachs Global Investment Research

Agentic AI may also create a self-reinforcing economic flywheel as compute costs decline: (1) lower compute cost per token enables richer, more complex agents; **(2)** richer agents consume considerably more tokens through longer context, more loops, more validation, and more persistent monitoring; **(3)** higher utilization improves the economics of AI infrastructure and better economics allow providers to keep investing in model quality and distribution. This flywheel is very different from the prevailing market narrative that AI usage will simply drive an increasing and unsustainable cost burden.

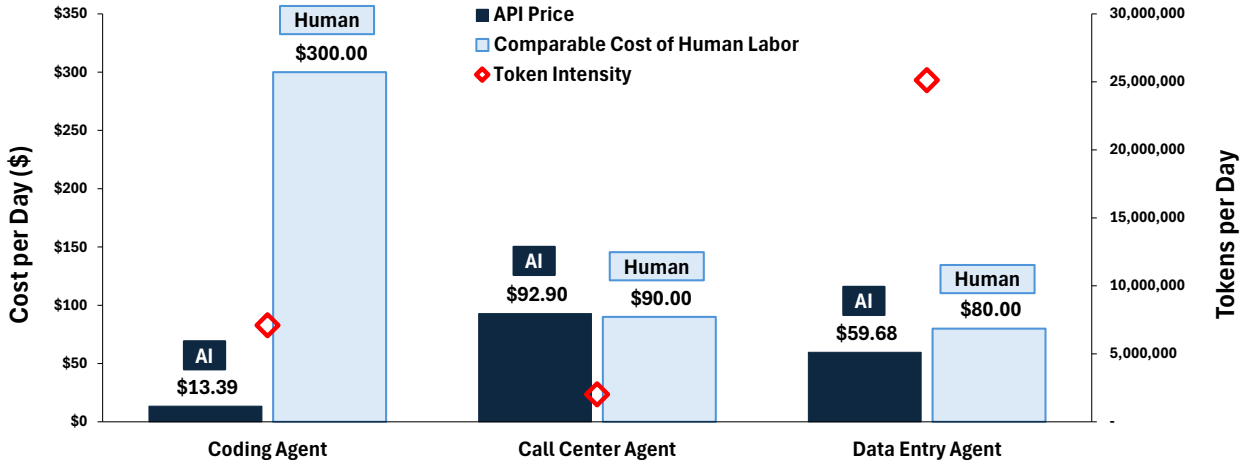
However, there continues to be a significant risk that this positive margin inflection is not guaranteed across all AI workloads. Competition could still force token prices down faster than compute costs, especially for more commoditized text-only chatbots.

But even assuming some level of pricing pressure, the overall price vs. cost trend suggests that the industry has significant room for economic improvement as accelerator efficiency, model optimization, routing, caching, and utilization continue to scale. **The key investment conclusion is not that every token will be profitable, but that the marginal economics of agentic AI may improve at the same time that token volumes accelerate. That combination is what makes the agent economy potentially much larger, more profitable, and more durable than a simple extrapolation of today's chatbot usage would imply.**

Agent ROI: Falling token costs pull more enterprise use cases into the money

The Bottom Line: The margin story is not just about cheaper tokens; it is about cheaper tokens expanding the set of enterprise AI use cases that can deliver positive ROI to the ultimate AI end consumer: enterprises. To test that, we translate our simulated agent workloads into estimated cost per task and compare those costs against the human labor they could augment or replace across coding, call center, and data-entry workflows. The key conclusion is that even workflows that consume millions of tokens per day can still compare favorably against human labor costs, particularly as token prices continue to deflate.

Exhibit 5: Not all agents are currently cost effective against human labor



Source: Data compiled by Goldman Sachs Global Investment Research

However, we do not expect this ROI to unlock evenly across all enterprise workflows yet. We estimate **a coding agent could consume roughly 7mn tokens per day at only ~$13/day**, which helps explain why software development has already seen faster agent adoption: the workflow is high-value, largely text-based, and the tooling ecosystem is comparatively mature. By contrast, we estimate **a call center agent could consume roughly 2mn tokens per day but cost ~$92/day** if it relies heavily on real-time voice, making full voice-based automation materially more expensive and operationally complex than outsourced human labor presently.

Still, the direction of travel is increasingly favorable. As token costs fall, the breakeven threshold for enterprise agents moves lower, pulling more workflows into positive ROI over time. This is the core economic bridge between our bottom-up token estimates and enterprise adoption: agentic workflows may be token-intensive, but if the cost per token declines faster than the complexity of the workflow rises, then many agents can generate attractive returns well before they reach full autonomy. **For investors, this means the enterprise agent opportunity should not be evaluated only through today's product maturity or near-term deployment friction. It should be evaluated through a declining cost curve, where each step down in compute costs expands the set of workflows that software, services, and infrastructure providers can economically automate and generate positive ROI.**

Investment Implications

- **For hyperscalers and model providers,** the rise of consumer agents and agentic computing represents a growing driver of compute demand while also creating an opportunity for value unlock, marking a shift where value lies in distribution and monetization of intent/utilization across both consumer and enterprise landscapes. Overall, operators remain supply constrained in their ability to meet current/forward compute demand (both internally and externally) and continue to invest in the necessary infrastructure to support an evolving computing landscape and broad-based AI adoption. As capex intensity remains elevated (with GOOGL & META raising FY2026 capex estimates and AMZN mgmt reiterating their strategy of maintaining elevated capex coming out of Q1'26 earnings), we expect investors to increasingly look for evidence of scope/visibility for returns, against the broader market debate if large-scale AI usage can become economically attractive to justify the capex cycle. **We prefer the following stocks:**
  - **AMZN (Buy, $325 PT):** We continue to see visibility into returns as AWS revenues compound (growth reaccelerated to +28% YoY in Q1; new GSe for ~low 30s % YoY growth in FY26 & FY27) supported by a reported $364bn revenue backlog (not including an additional $100bn for a deal announced in Q2), driven by both AI workloads and building momentum around its custom silicon (Trainium, Graviton, etc.).
  - **GOOGL (Buy, $450 PT):** Alphabet is seeing momentum across its Cloud business (+63% YoY in Q1 with backlog nearly doubling QoQ to ~$460bn) and Search multi-modality, leveraging a full-stack approach as mgmt continues to see AI as repositioning the company for sustained growth.
  - **META (Buy, $830 PT):** Meta remains a leader in its core advertising business (significantly outpacing total digital ad industry growth) as the application of AI-related compute is driving momentum around engagement & ads monetization (ads creation, targeting, measurement/attribution, etc.). Going forward, the company remains focused on applying its growing compute capacity and scaling 1P AI models (Muse Spark being the first iteration) to building out additional monetization opportunities around agentic commerce, AI business tools for SMBs/creators and new consumer AI applications.
- **For Semiconductor companies,** we see a clearly positive impact from ongoing CapEx spending from hyperscalers and LLM providers. We believe that falling token costs enabled by merchant GPU and ASIC leaders will make token-intensive use cases economically viable and hence increase the addressable compute market as volume elasticity more than offsets lower compute costs. Just as important, the positive margin inflection we see ahead for hyperscalers and LLM providers means that their margins can improve significantly - thus creating significantly more headroom for increased CapEx and making today's elevated infrastructure investments sustainable. **We prefer the following stocks:**
  - **Broadcom (Buy, $480 PT):** As the market leader in custom computing, we see more hyperscalers (Google) and LLM model providers turning to Broadcom to

deliver cost-optimized chip solutions tailored to their specific workloads.

- **Nvidia (Buy, $250 PT):** We believe Nvidia can retain its dominant market leadership in the medium term as it remains the leader in AI performance across a broad range of training and inference workloads.
- **AMD (Buy, $450 PT):** We see AMD's market position strengthening as the company scales its high-performance datacenter GPU offerings (MI450 and MI5XX) over the next two years. Importantly, we believe AMD is also poised to an increasing share of agentic AI workloads in the enterprise as it gains share in X86 server CPUs and the CPU attach rate increases.

- **For software and IT services companies,** we believe the margin story is more nuanced but we continue to see longer-term tailwinds. Lower token costs make it easier for software vendors to embed agents into existing products without significantly impacting gross margins, while also allowing them to price around outcomes, productivity, or units of work rather than seats alone. This supports our argument that agentic AI can expand software TAM: if the cost of delivering an automated workflow falls while the value of the completed work remains tied to labor substitution or productivity gains, software companies can capture a spread between falling AI delivery costs and the much larger value of the task being automated. Meanwhile, we expect IT Services companies to benefit as agents shift AI consumption from standalone tools to enterprise-wide, integration-heavy workflow transformation – increasing demand for integration, governance, and managed orchestration to levels not yet seen with standalone tools.
- **We prefer the following stocks:**
  - **MSFT (Buy, $610 12-m PT, covered by Gabriela Borges):** Copilot feedback is getting better and the E7 upgrade cycle may drive further acceleration in Microsoft 365. The most likely scenario may be an ecosystem where Copilot coexists alongside domain-specific agents and domain-specific app software, and the usage of one pulls through usage of the others reciprocally.
  - **NET (Buy, $250 12-m PT, covered by Gabriela Borges):** We expect Cloudflare to take outsized share of AI inference workloads because of its performance and cost advantages, in turn driven by its architectural network advantages and the sophistication of its *isolates* software.
  - **Accenture (Buy, $300 PT):** We expect Accenture to see growing tailwinds from agentic adoption as enterprises increasingly move from AI pilots to scaled agent deployments, driving demand for integration, workflow redesign, governance, and change management.

The Consumer Agent Landscape

What does the consumer landscape look like today?
Since our October 2025 note where we framed five investor debates on the AI landscape, we see trends of increased consumer utilization and a more robust product innovation landscape that continues to reflect progress of evolving consumer habits against the backdrop of an evolving computing landscape. In our view, this shift is characterized as a transition from consumers using AI for conversation and tool-based functions (prompting chatbots for information retrieval/action) and towards a potential end state where LLM agents and AI systems may play a more involved role in consumers' lives day-to-day, given the autonomy to execute multi-step workflows directly within a user's computing environment without being prompted by the user.

To level-set, consumer AI adoption continues to be broad-based and accelerating, as users increasingly turn to generative AI products with queries (away from traditional search) and interact with AI applications more frequently due to embedded features within consumer-facing products, as companies seek to implement more features in existing products and interfaces (i.e. shopping assistants such as AMZN's Rufus) while developing and launching AI-native products.

Monthly active users (MAUs) continue to scale across leading AI platforms (both on web and in-app). We would also note that consumer preference for AI platforms continues to broaden and diversify as well, both as a function of operators pushing product/platform developments that capture user intent and as consumers become more comfortable with AI, finding incremental value in specific platform strengths. Gemini has established itself as the second-largest application by leveraging Google's extensive distribution channels, showing substantial growth in both user acquisition and engagement.

Given the spectrum of consumer-facing AI applications and use cases available today, we identify three emerging categories of AI products: a) conversational chatbots, b) embedded tools, & c) agents. We delineate these by increasing levels of interference/autonomous functionality performed by an LLM, with decreasing reliance on instruction/prompts from the user:

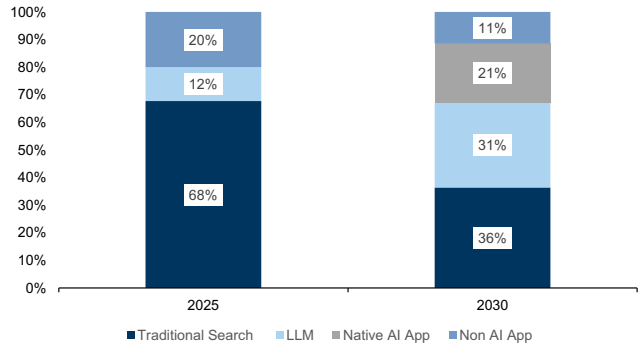
1. **Chatbots.** Today, chatbots (i.e. ChatGPT, Gemini, Claude) represent the lowest-friction entry point for consumers - tools that the user accesses voluntarily and prompts to create an output, typically as a medium for discovery and with the goal of identifying information.
2. **Embedded tools.** Embedded tools, such as Gemini's Personal Intelligence system, sit one layer deeper in the stack and are adopted by users who seek to use AI in a more interactive role in their lives and still operate on a set of instructions from the user. These tools operate autonomously/independently (beyond the generalist chatbot function) given a set of instructions, coordinating with integrated apps/personal data history to execute on and suggest tasks.
3. **Agents.** The third and most autonomous degree today, agents are characterized by minimal direct user input, running persistently in the background across devices or applications.

Where are we going?

In our view, the key enablers to broader consumer agent adoption are a) mitigation of user friction and value realization on the consumer end & b) operators overcoming supply constraints to meet compute demand, both at current levels and at scale. As we have framed previously, historical technology cycles (Web 1.0 – first iteration of Internet, Web 2.0 – mobile/interactive computing) show that durable consumer adoption takes hold when technology fundamentally reduces user friction.

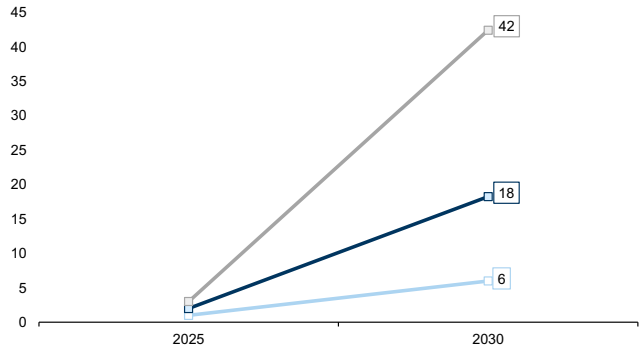
We take the opportunity to frame forward consumer AI adoption through the lens of daily query volume and shifting query share mix as consumer computing habits evolve – with traditional search engine volume growth moderating and LLM query volume accelerating as consumers increasingly utilize generative AI products/tools over time. Our base-year view reflects a market where traditional search still dominates query volumes, but AI has already introduced incremental demand. Looking to 2030, we assume continued growth in total consumer queries, driven by a mix shift towards LLM mediated and AI-native interactions, partially offset by ongoing compute costs (albeit decreasing).

Exhibit 6: While Traditional Search comprises the majority of query volume today, we expect queries through LLM applications (both browser chat and native app) to accelerate over the next 5 years and gain outsized share.
2025 vs. 2030 Query Market Share



Source: Data compiled by Goldman Sachs Global Investment Research

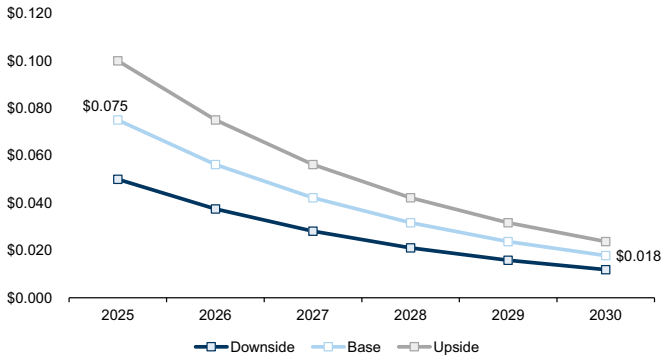
Exhibit 7: We expect daily queries to LLM agents to grow from 2bn in 2025 to 11bn in 2030 at a ~40% CAGR, and frame downside/upside scenarios.
Total daily queries to LLMs (both web and native app)



Source: Goldman Sachs Global Investment Research

As per-query costs decline, absolute compute demand rises as AI expands the surface area of consumer interaction.

Exhibit 8: As compute costs/pricing dynamics scale, we also expect cost per query (across a range of bull/bear cases) to reduce materially over time.



Source: Data compiled by Goldman Sachs Global Investment Research

Consumer Agents: Sizing the Token Impact

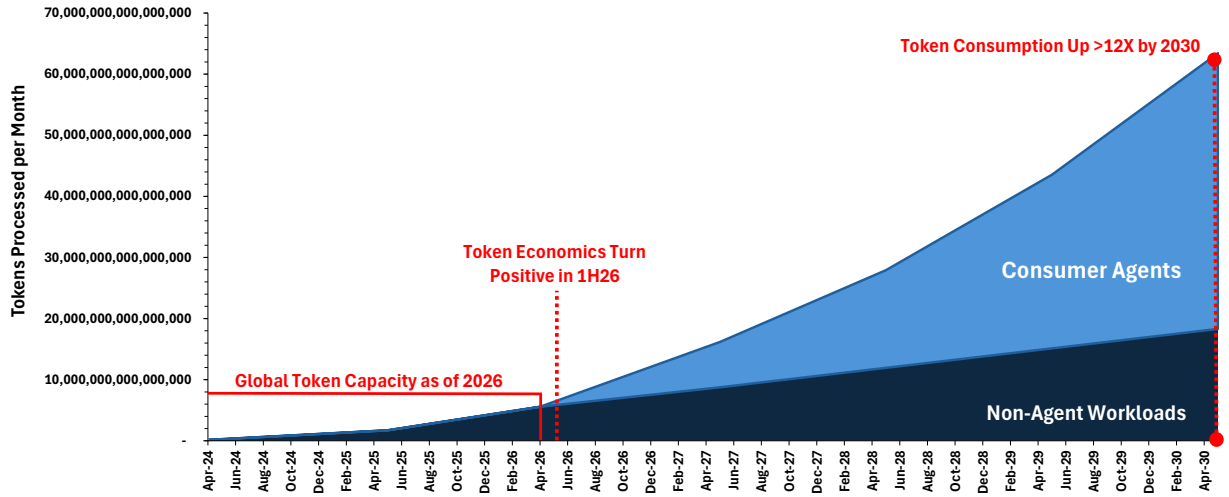
Consumer agents could expand AI demand from answering questions to running everyday workflows

The Bottom Line: We estimate that consumer AI agents could become a significant driver of token volumes as consumer AI behavior shifts increasingly from episodic chatbot sessions to on-demand agents and always-on personal assistants (e.g. OpenClaw). Assuming broad adoption of these agents by 2030, we estimate that consumer agent workloads could reach ~60 quadrillion tokens monthly and global token consumption could rise ~12X from current levels.

We estimate the potential for ~23 billion AI queries per day by 2030 (up from ~5 billion queries per day in 2025), with up to 30% of consumer queries directed to LLM agents across search, shopping, travel, email, and other personal productivity functions. This would result in an additional ~60 quadrillion tokens processed per month by 2030, or ~12X the current global token consumption as of 2026 (Exhibit 9).

In short, we believe the key consumer token multiplier is not simply "more chatbot usage", but instead a major shift in usage patterns. Today, most consumer AI usage is still episodic: a user asks a question or gives a prompt, receives an answer from the LLM, and exits. We see evidence of this behavior in the average tokens consumed per query, which as of 2025 was an average of 1,715 tokens per query (equivalent to a ~3-5 minute session with an AI chatbot). However, in an agentic consumer model, the AI tool plans, executes, monitors, and acts on the user's behalf across a range of daily tasks. **We estimate the single largest token step-up is likely to come as agents move from user-initiated tasks to persistent or "always on" background activity, where the agent continuously monitors context and decides when action is needed.**

Exhibit 9: Consumer agent workloads could dramatically increase token consumption, leading to 10x token consumption by 2030.
Consumer token growth is driven by broader user reach, higher daily frequency, and the shift from chat sessions to both on-demand and always-on agents.

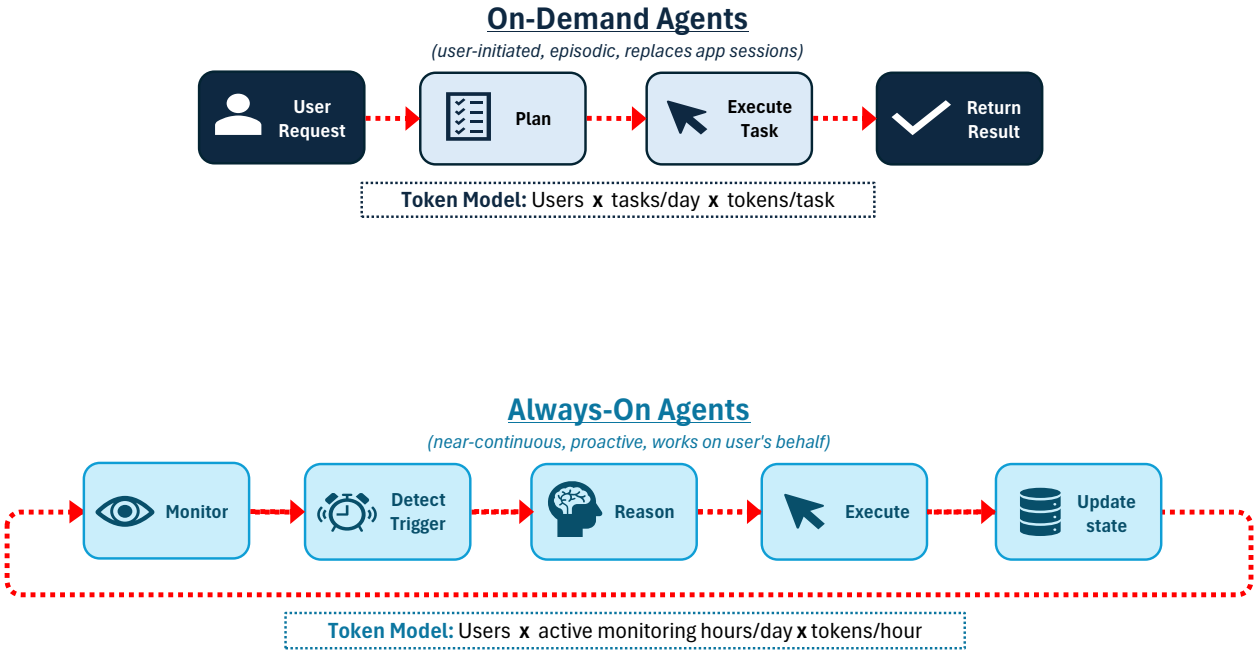


Source: Data compiled by Goldman Sachs Global Investment Research

Consumer agents: two distinct models, divergent token demand

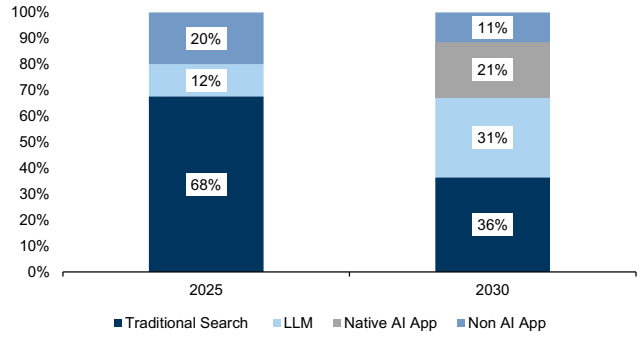
To frame a realistic forecast of the consumer AI agents in coming years, we segment consumer agents into "on-demand" and "always-on" use cases - and today's agentic products already appear to follow this line. Tools such as OpenAI Operator, Claude Code, and other browser-based agents are predominantly **on-demand**. This means users must initiate a task - and from there, the agent then plans, takes actions, loops through execution, and finally stops once the task is complete and approved. By contrast, we estimate the next wave of consumer agents may look more like persistent email monitors, digital life assistants, or schedule managers that perpetually run in the background and act only when needed (seen with OpenClaw). We believe understanding the token consumption model of this latter **always-on** consumer agent category matters most because enterprise agents are also persistent by design, having to operate continuously across business workflows rather than waiting for discrete user prompts.

Exhibit 10: We segment consumer agents into two models: On-demand and always-on
On-demand agents execute discrete user-initiated tasks, while always-on agents continuously monitor context



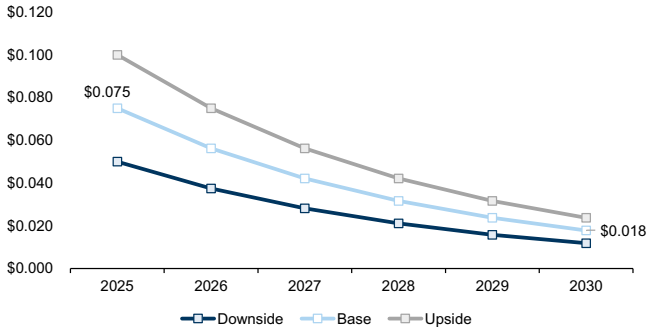
Source: Goldman Sachs Global Investment Research

Exhibit 11: We expect queries through LLM applications (both browser chat and native app) to accelerate over the next 5 years and gain outsized share.



Source: Data compiled by Goldman Sachs Global Investment Research

Exhibit 12: As compute costs/pricing dynamics scale, we also expect cost per query (across a range of bull/bear cases) to reduce materially over time.



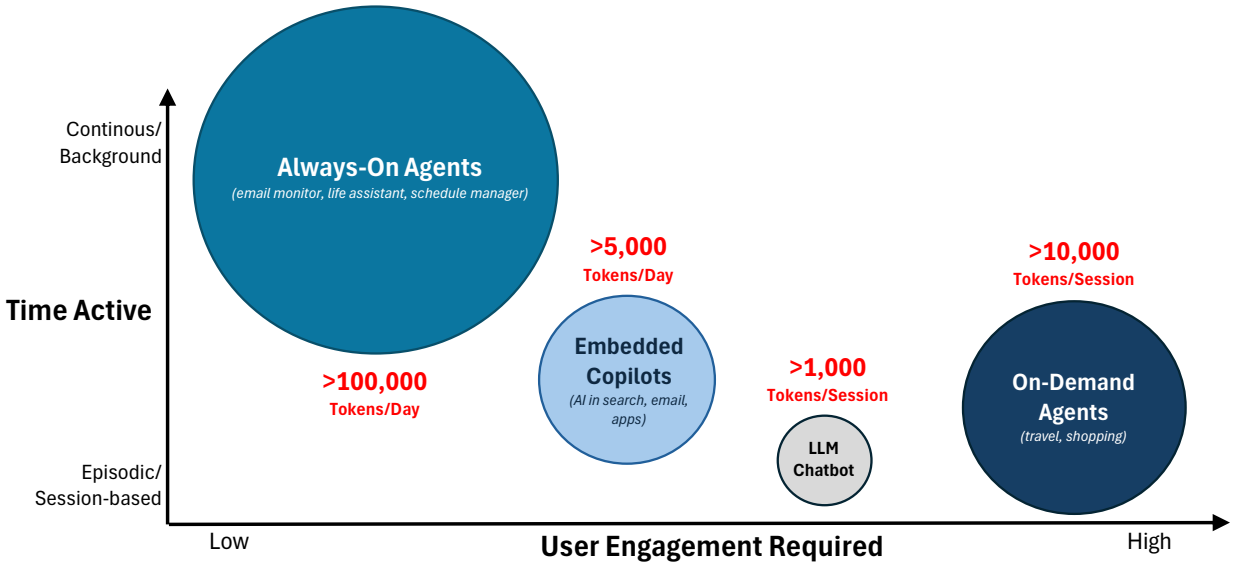
Source: Data compiled by Goldman Sachs Global Investment Research

Inside the agent: why agents consume significantly more tokens

To ground our consumer token forecast, we build simulated agents that show how token usage accumulates across a full workflow rather than a single chat response. The travel booking agent flow (Exhibit 14) illustrates the core mechanism: an agent parses intent, clarifies missing inputs, searches external systems, ranks options, incorporates user feedback, validates availability, and loops before completing the task. The accompanying tables translate that logic into token counts by step, showing why even simple consumer agents can become materially more token-intensive as they add context, tools, validation, and repeated decision loops (Exhibit 15).

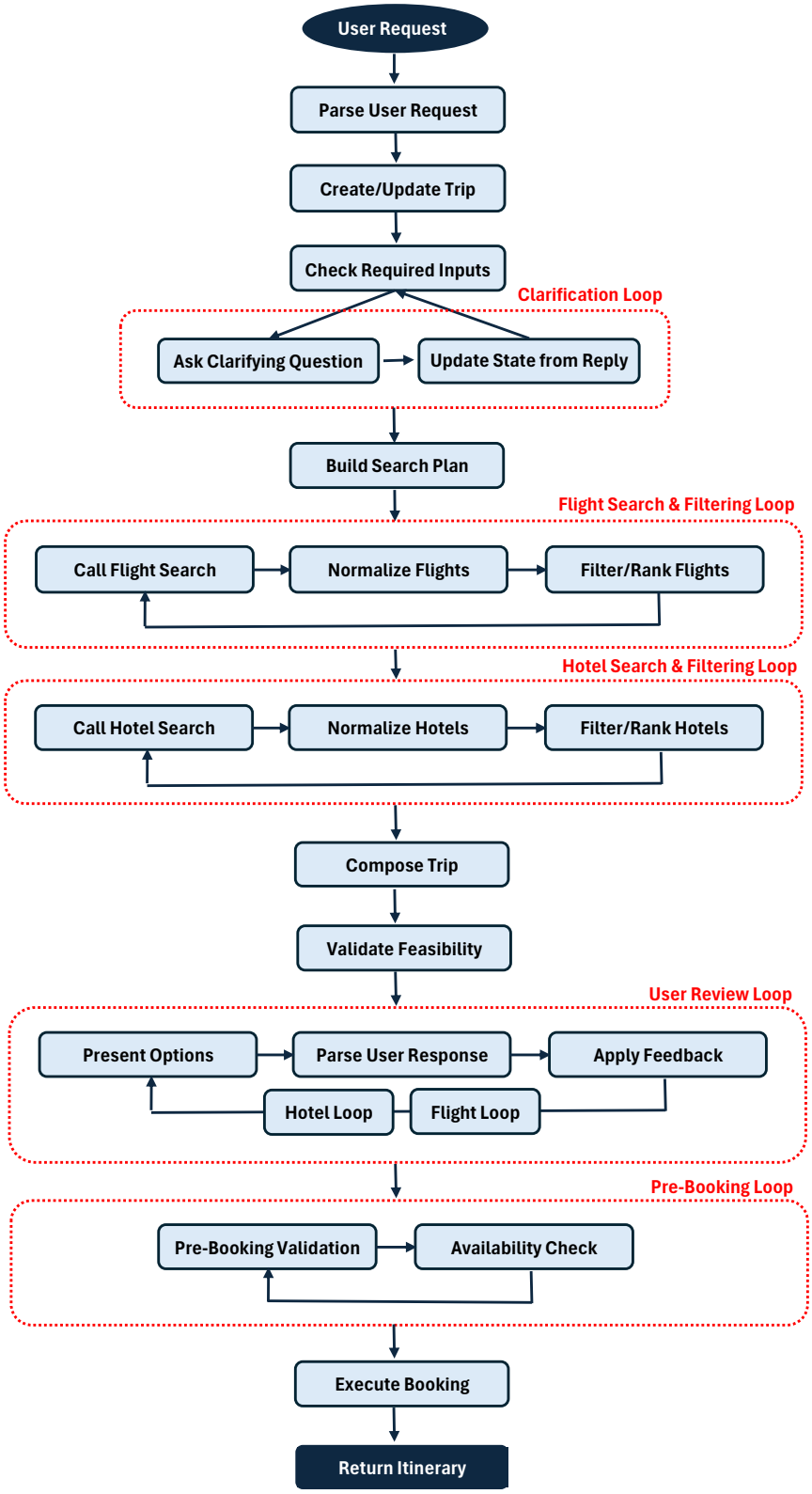
We clearly observe through our simulated agents how token demand rises materially as agents shift from discrete, user-initiated tasks to persistent background operation. Based on the agents we tested, LLM chatbots consume roughly 1,000 tokens per session, while embedded copilots consume >5,000 tokens per day, or several times more per user before accounting for broader adoption. Embedded copilots can reach >100,000 tokens per day, highlighting that the biggest driver of token demand is not just task complexity, but how continuously the agent remains active.

Exhibit 13: Always-on agents could become the highest-intensity consumer AI workload



Source: Goldman Sachs Global Investment Research

Exhibit 14: Travel agent booking flow
Simulated workflow shows how a single user request expands into repeated tool calls, filtering, feedback, and pre-booking checks



Source: Data compiled by Goldman Sachs Global Investment Research

Exhibit 15: Detailed email monitor agent build and token intensity

Email Monitor Agent				
Action	**Model Reasoning**	**Input Tokens**	**Output Tokens**	**# Loops**
Parse hourly objective	Medium	180	60	
Load inbox state/rules/context	Low	150	40	
Check inbox for new emails	External Tool	External Tool	External Tool	
Normalize/sort new email batch	Medium	220	50	
Hourly scan loop total		13200	1200	24
Read email body/context	Medium	320	40	
Classify email (spam, etc.)	Low	220	70	
Extract info (deadlines/action)	Medium	260	90	
Decide next action	Medium	180	69	
Update inbox state/memory	Low	140	40	
Email triage loop total		56000	15450	50
Retrieve deeper thread	Medium	500	80	
Draft reply	High	650	220	
Check tone/policy/consistency	Medium	250	70	
Save draft/queue response	External Tool	External Tool	External Tool	
Reply loop total		16800	4440	12
Detect scheduling intent	Medium	300	90	
Check calendar	External Tool	External Tool	External Tool	
Reason availability	High	450	140	
Draft scheduling reply	High	500	180	
Create/update cal hold	External Tool	External Tool	External Tool	
Scheduling loop total		4750	1600	5
Generate daily inbox summary	Medium	400	150	
Pre-booking loop total		400	150	1

Model	Input Tokens	Output Tokens
Total Tokens	91,150	22,840
API Cost	$0.02735	$0.02741

Travel Booking Agent				
Action	**Model Reasoning**	**Input Tokens**	**Output Tokens**	**# Loops**
Parse user request	Medium	250	90	
Create/update trip state	Low	180	60	
Check required inputs	Low	130	35	
Ask clarifying questions	Medium	140	35	
Update state from reply	Low	130	40	
Clarification loop total		270	75	1
Build search plan	High	320	110	
Call flight search	External Tool	External Tool	External Tool	
Normalize flights	Medium	1750	180	
Filter/rank flights	Low-Medium	500	120	
Flight loop total		2250	300	1
Call hotel search	External Tool	External Tool	External Tool	
Normalize hotels	Medium	1750	180	
Filter/rank hotels	Low-Medium	500	120	
Hotel loop total		2250	300	1
Compose trip packages	Medium	650	180	
Validate feasibility/budget	Low	380	100	
Present top options	High	600	250	
Interpret user response	High	180	60	
Apply feedback to state	High	160	50	
Human review loop total		940	360	1
Pre-booking validation	Low	420	120	
Availability check	Medium	250	70	
Pre-booking loop total		670	190	1
Execute booking	External Tool	External Tool	External Tool	
Return final itinerary	Medium	300	180	

Model	Input Tokens	Output Tokens
Total Tokens	8,590	1,980
API Cost	$0.00716	$0.00869

Source: Data compiled by Goldman Sachs Global Investment Research

Level Setting 2026

We see two key milestones in the Enterprise Agent landscape over the past two years: 1) Salesforce announced Agentforce in August 2024, which raised the ceiling for AI implementations (at least in theory) in customer experience from text-based RAG chatbots to more sophisticated processes; and 2) Anthropic announced Claude Cowork in January 2026, which added a knowledge worker wrapper to its coding tool. In practice, enterprise adoption of agentic AI is early. Headline survey data suggests upwards of 70–90% of enterprises are now experimenting with AI agents, but functional penetration is limited (McKinsey; PwC). Fewer than one quarter of firms are scaling agents and adoption is typically confined to one or two workflows. These are generally bounded, high-ROI use cases: customer support triage, IT operations, sales enablement, and internal knowledge workflows. (Our conversations with WRITER showcase their success with targeting very specific workflows). Additionally, these are often under the banner of "copilots" rather than fully autonomous agents.

In our view, the key enablers are organizational rather than technical. Advances in foundation models, tool-calling, and memory architectures have removed much of the core technical challenges, such that cloud software vendors are shipping agents that increasingly can do more. The main area of focus in our industry conversations has shifted from model performance to orchestration; enterprises are less focused on whether agents can act, and more focused on how and when they should be allowed to act. As a result, today's deployments are overwhelmingly human-in-the-loop, workflow-embedded, and tightly permissioned, reflecting a bias toward augmentation rather than autonomy. There are two limiting factors:

1. **Trust at scale:** data governance, security boundaries, deterministic vs. non-deterministic controls, auditability, and change management Enterprises struggle most where agents must operate across systems of record, particularly when accountability for outcomes is diffuse.
2. **ROI visibility:** benefits often show up as productivity gains rather than hard cost take-out, which complicates budgeting and ownership. In addition, new highs on inference costs illustrate the importance of guardrails on cost.

In short, agentic AI is technologically ready, but economically and institutionally constrained.

While we stop short of predicting an inflection in the next 12 months, we do believe enterprises will see gradual adoption in agent use cases as risk management frameworks improve and POCs move to production. The next phase of enterprise adoption is likely to center on domain-specific agents embedded deeply within vertical workflows (finance, customer support, etc.) where success criteria are well-defined and error costs are understood; we model examples of these in the following section. Over time, we believe that enterprises that invest early in agent governance, observability, and exception handling will unlock higher levels of autonomy, shifting agents from task execution to goal-level orchestration.

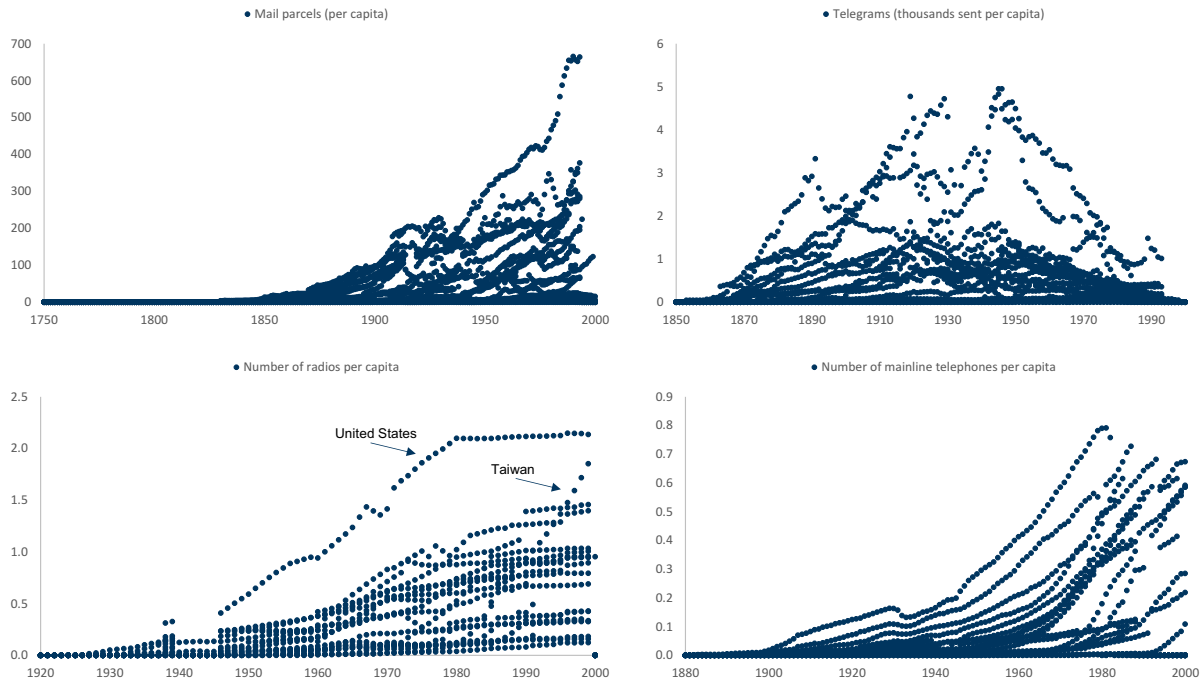
What will adoption look like over the next 10 years?
We introduce an adoption curve for enterprise agentic AI to inform investors' forecasts of compute demand, software demand, and IT budgets, which in turn serve as inputs for the companies across our TMT coverage. The two questions we focus on are: what is peak adoption, and how long does it take to get there? To guide our forecasts, we leveraged Comin and Hobijn's 2009 dataset on Cross-country Historical Adoption of Technology (CHAT). This project compiles data on adoption of 101 technologies in 161 countries from 1800-2000. Many of these technologies are primarily consumer-driven, but there are 39 technologies that are inputs to enterprises and have a robust enough timeseries for robust analysis. We think understanding the diffusion of these prior technologies is a helpful guide to understanding how technologies become adopted over time. **Ultimately, we think the curve will be S-Shaped and agentic demand 20x-25x what it is today, using these historical analogues and what we know about the technology today**.

1. **What is peak adoption?**

- **Peak adoption is more complicated than simply measuring the percent of enterprises using agents.** Some technologies work this way; Comin and Hobjin's data on number of telephones per capita shows no country went above a 1:1 ratio, while data for radios shows ~2 is the natural ceiling. Given homes, offices, and some public places are the most likely places to have phones and radios, it makes sense that the ratio was roughly ~1-2 per person. Other technologies, however, do not have an obvious ceiling. The number of mail parcels sent per capita varied widely by country, and the range of 7-665 per person does appear constrained by an inherent physical constraint such as the number of households. Instead, economic factors mattered more, with richer countries generally having more mail sent per capita. Similarly, the number of telegrams sent clearly increased over time, was higher in wealthier countries, and declined during the Great Depression, meaning it was impacted by real-world economic factors. But the peak quantity sent (~1,000-~5,000) varied between countries the way mail penetration did, and did not appear governed by external factors.
- **There are ways to theorize/contextualize a logical upper limit.** In this context, we believe the most logical way to measure and forecast peak agentic penetration is the way software companies are increasingly pricing their offerings: **in terms of "units of work."** The two inputs here are 1) how much knowledge work there is to do at enterprises/in the economy; and 2) how much of it ends up being done by agents. The analysis in the following section is essentially the product of these two variables (with additional assumptions on incrementality/substitution), with **the result being ~35-40% peak workflow adoption, implying ~1.4 trillion labor-hours, ~280 quadrillion tokens, ~$220bn flowing to AI infrastructure annually at peak and an $5.4trn software TAM.**
- **Another conclusion from the data is that peak adoption of a new technology can be higher than prior technologies**. This is because new technologies do not necessarily replace existing technologies. The various forms of communication technologies (mail, telegrams, radios, telephones, and television) were not substitutes for each other; instead the overall TAM of communication/information

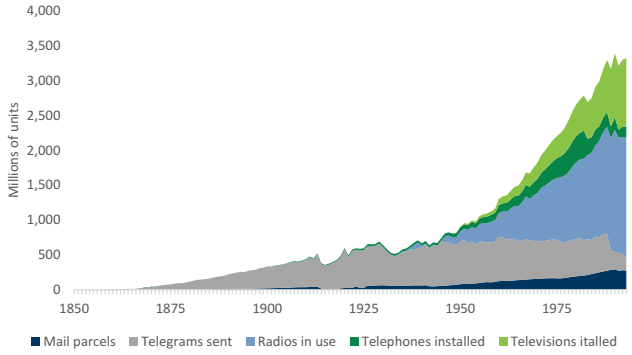
technology expanded. Applying this framework to agentic AI, agents do not necessarily need to substitute 1:1 for human labor; the amount of knowledge work that is done in the economy should expand due to agentic AI. Consider the example of a call center (also explored below). A simple framework that does not factor in an expanding pie would suggest that at a certain ratio of agent cost to human wages, firms will switch all their labor from human call center workers to AI agents (100% peak adoption). But as anybody who has ever been on hold with customer support can attest, there is unmet demand in the form of long wait times and customers who give up on waiting and hang up. Without resource constraints, wait times would be zero, and no customers would ever give up. AI agents can theoretically serve this unmet demand, which means the total amount of customer service provided increases. Even if firms decide to retain some of the savings of automation, some of the leftover budget becomes consumer surplus. The net of this means that firms become more efficient, more of the service is provided, and human labor is likely to evolve rather than be replaced. We have seen real-life examples of this: Navan (travel and expense management for enterprises) has >50% AI call deflection rates with high customer satisfaction. AI handles most of Navan's basic queries, and humans support customers with more complicated needs.

Exhibit 16: Some technnologies' peak adoption was gated by real-world constraints; others were more detached from physical variables
Dotted lines indicate historical adoption rates by country



Source: Comin and Hobijn, Goldman Sachs Global Investment Research

Exhibit 17: Advances in communications technology partially cannibalized the old technology, but generally expanded the TAM

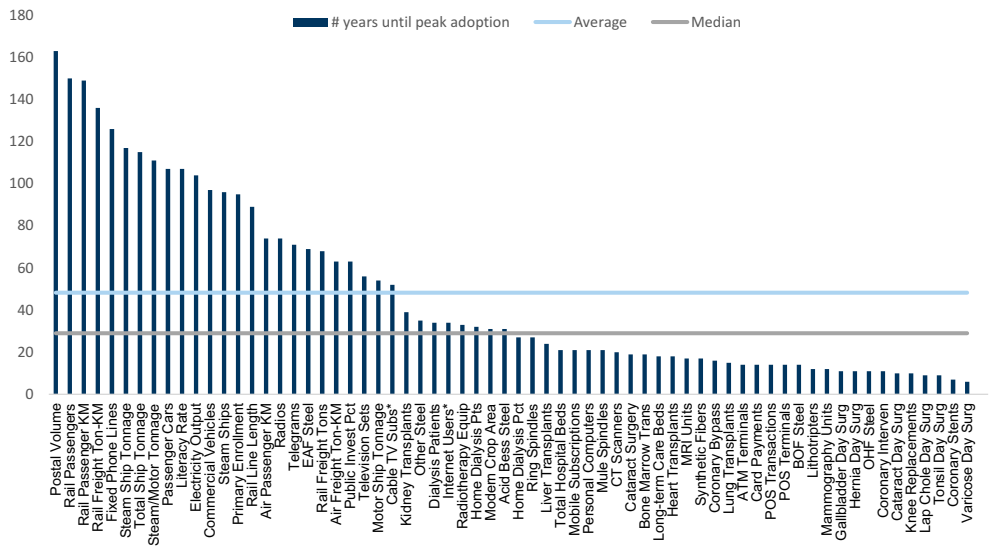


Source: Comin and Hobijn, Goldman Sachs Global Investment Research

2: How long does it take to get there?

- **The speed of adoption varies.** Rail, steamships, and landline phones all took 100 years or more to peak. Cable TV, ATMs, and new surgery techniques took 20 years or less. The median technology in the dataset took 29 years.
- **The shape of the curve varies**. Some technologies very quickly diffuse following their invention (and/or an up-front investment period), following a J-Curve (internet in the 1990s, the blast oxygen furnace method for steel production). Some follow an S-curve (at-home kidney dialysis, the telegram) where adoption starts slowly, then accelerates rapidly following a tipping point (often driven by network effects), before plateauing. Finally, some technologies lack virality or face significant capital constraints, such that the spread of the technology is more constant over time (postal volumes/electricity output). We can see elements of all of these factors for agentic AI, given: a) the significant capital constraints the sector faces implies **linear adoption**; b) token economies of scale/network effects from figuring out use cases with ROI implies an **S-curve adoption**; c) the rapid trial from enterprises and consumers ~3 years on from ChatGPT's popularization implies virality and hence **J curve adoption**. The J-curve scenario is the most bullish for demand, while the linear scenario implies a more tempered outlook.
- **The speed of adoption is independent of the shape of the curve.** For example, both at-home kidney dialysis and the telegram roughly followed S-Curves, but at home dialysis took ~25 years to reach peak penetration while telegrams took 70. Because agentic AI is a non-physical good, and because the peak is bounded by economics and not per capita constraints (e.g. number of households), the peak of demand might be so large that the technology could take a long time to defuse. On the other hand, the pace of change may be speeding up with physical mail diffusing for millennia before peaking (in 2006 in the US), telegrams/the radio each taking ~75 years, and the internet at ~75% penetration in 2025 according to the United Nations (link) after 36 years since the advent of the World Wide Web. We therefore think a <20 year time-to-peak is not unreasonable for agentic AI.

Exhibit 18: Technologies have varied widely in their speed of adoption, as measured by time to reach peak penetration/capita



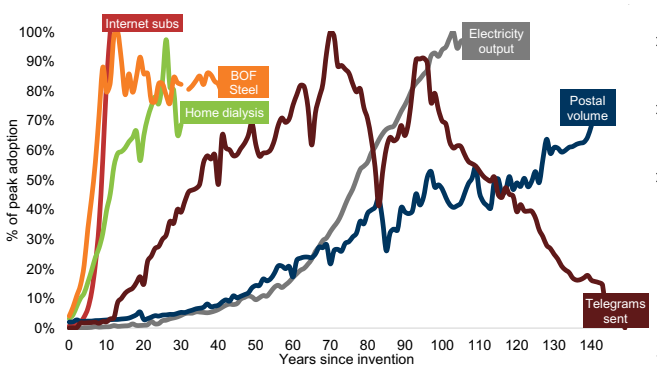
Source: Comin and Hobijn, Goldman Sachs Global Investment Research

Exhibit 19: We theorize 6 types of adoption curves for new technologies...

		Pace of Adoption		
		J-Curve	S-Curve	Gradual
Time to saturation	Not yet reached	Internet subscribers	At-home kidney dialysis	Electricity output
	Reached	Blast oxygen furnace steel	Telegrams	Postal volume

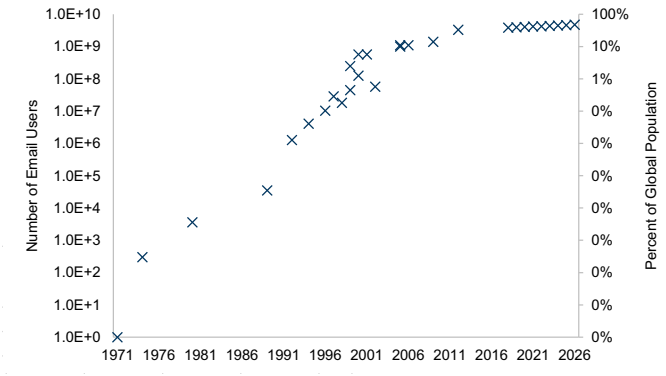
Source: Goldman Sachs Global Investment Research

Exhibit 20: ...which we observe in the historical data



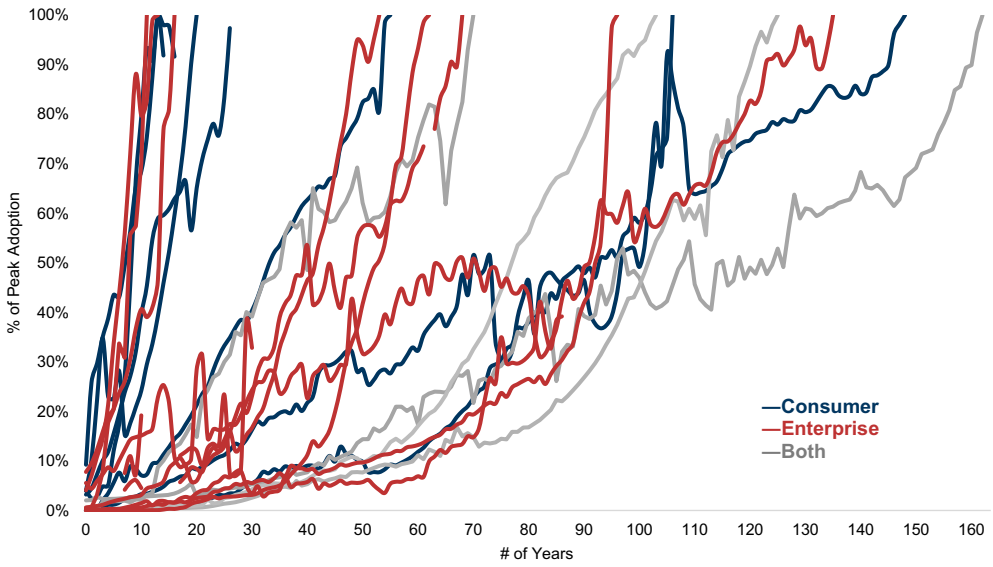
Source: Comin and Hobijn, Goldman Sachs Global Investment Research

Exhibit 21: Email followed a clear S-Surve, where adoption built gradually and more recently has leveled out



Source: Company data, Statista, EmailToolTester.com, World Bank, Goldman Sachs Global Investment Research

Exhibit 22: Consumer and enterprise technologies have had variable adoption curves historically



Source: Comin and Hobijn, Goldman Sachs Global Investment Research

Introducing our baseline enterprise agentic forecast:
We introduce two base case assumptions and welcome feedback to our baseline:

1. **We believe an S-Curve is most likely:** We believe we are in an initial trial/adoption period. As we near 2030, we expect a critical mass of firms will begin adopting agents at the steepening part of the curve followed by a leveling out as the laggards adopt. It will take time for the technology to diffuse given: 1) There are upfront costs outside of tokens that only need to be spent once, such as data hygiene and process management; 2) there are institutional barriers that need to be overcome; 3) enterprises are still figuring out the use cases, which in the short term will skew domain-specific and more likely to leverage SLMs than LLMs. Further out, though,

we expect data hygiene and process management to be less of a barrier as once an enterprise has leveled up its data warehouse strategy, new data created will by default live within an agent-friendly framework. We also think that as the use cases are figured out and workers get used to the technology, adoption will follow.

2. **We introduce a time to peak adoption of 15 years**. We expect the time to peak adoption to be faster relative to the median technology (26 years), with a peak of diffusion of 15 years. We see plausible arguments for faster/elongated time to peak adoption. On the side of faster adoption: a) newer technologies appear to be reaching mass adoption more quickly than before; b) a key input in building software (the cost of code) is going down enabling more experimentation/finding use cases faster; c) we know that startups are reaching milestones of $100mn and $1bn much faster than before because of their success in monetizing specific use cases; d) enterprises are acutely aware of the need to show progress with AI roadmaps to minimize the risk of market share loss to more tech-forward competitors. On the side of a longer path to peak adoption: a) the enterprise technologies that took longer to diffuse tended to be commodity technologies with wide applications like air freight electricity, or telegrams; b) the long tail of laggards may take a very long time to

Implications for our software coverage

We are more convicted in the Software TAM. We think that as companies move to sell units of labor or productivity, this creates a high ceiling of potential adoption/penetration which implies a significant value creation/capture opportunity for software companies. In practical terms, this means that over time they will be able to capture more budget from enterprises. There are already signs that the prior ceilings of enterprise application deal sizes are biased upward, including in our conversations with Sierra (customer experience) and Serval (IT help desk); and in survey work from Avenir that suggests enterprises would be willing to pay $2,400 per year for a general purpose intelligence agent (as a reference point, Microsoft 365 subscriptions approximate ~$680 per year at the middle E5 tier and ARPU for coding tools is already up 25x). Furthermore, AI natives will continue to find new white space to fill, while the supply-side impulse from agentic AI should create value across the ecosystem. That means that the budgets themselves should get bigger.

We look for single-stock ideas that are well positioned to capture share of agentic wallet.

To date, evidence is mixed. For example, ServiceNow took up its AI ARR target to $1.5bn exiting 2026 (from $1bn prior) in April 2026, but kept its organic guide the same. Salesforce guided to acceleration in 2H26 at its investor event in September 2025, but since that time, its organic business has slowed from 8.0% to 6.5%, implying that acceleration will be from a lower base. For Software stocks to work, we believe companies need to prove that AI is additive to their core businesses.

In order to successfully transition from a SaaS leader to an AI Software leader, we believe the following heavy lifting is necessary:

- Streamlining the core platform to minimize technical debt (as many SaaS leaders founded their technical platforms 10+ years ago and may have also added inorganic

assets during that time);

- Reset the organic roadmap for a much higher pace of innovation and evaluate where acquisitions make sense (we don't think it makes sense for R&D to be entirely organic when leaders can use their cash-rich balance sheets to acquire next-gen early stage technology assets that have proven product market fit);
- Fine tune monetization.

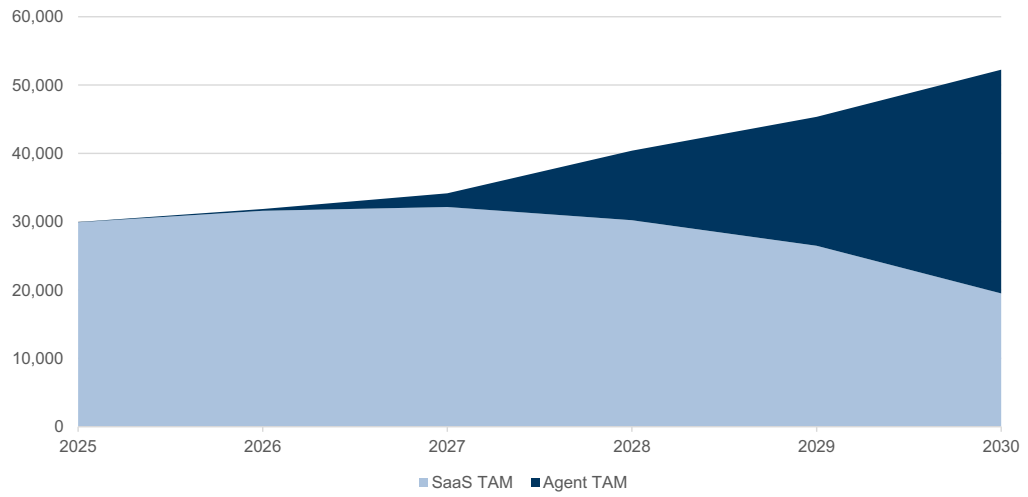
Some of our coverage is earlier on in this transition than others.

We are incrementally positive on the token providers. Our token forecast below and recent conversations make us more bullish on the token providers: in a world where supply/demand is even more constrained than in the prior 3 years, token providers should have more pricing power. This is echoed by comments from BaseTen at our private company conference that the supply/demand imbalance is materially worse than most industry participants perceive. Companies are overrunning initial inference budgets by orders of magnitude; one industry datapoint had inference costs in engineering now ~10% of headcount cost but could be on track to be on par with headcount costs in the next several quarters based on current trajectories.

Two stocks to highlight:

- **Microsoft**: Copilot feedback is getting better and the E7 upgrade cycle may drive acceleration in Microsoft 365. The most likely scenario may be an ecosystem where Cowork coexists alongside domain-specific agents and domain-specific app software. For example, Microsoft believes that the Cowork Excel plug in reinforces Excel's strategic value within the knowledge work stack – and our understanding is that using a Copilot license alongside Cowork allows for deeper MCP integrations and better user outputs.
- **Cloudflare:** Traffic on Cloudflare's network may change by a step function and we believe their product moat (rapid pace of innovation, control points of data, network and software architecture) should allow them to continue to compound their leading position over time. We expect outsized share of wallet in the inference compute TAM.

Exhibit 23: We believe the software TAM will continue to grow, powered by a nascent agentic TAM



Source: Goldman Sachs Global Investment Research

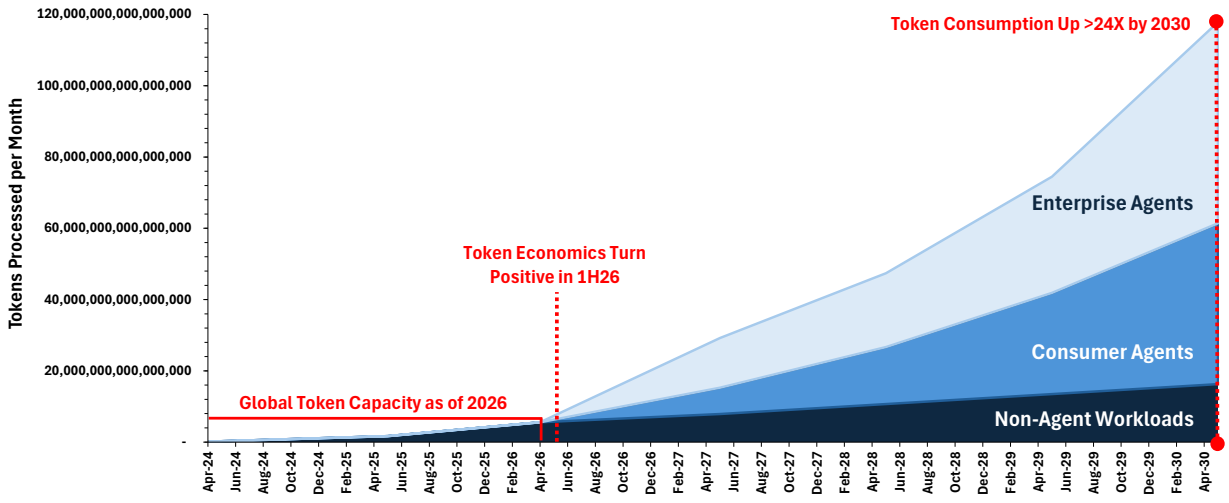
Enterprise Agents: Sizing the Token Impact

Enterprise agents could drive the greatest token consumption

The Bottom Line: By building model agents which simulate workflows across AI-exposed occupations, we estimate the minimum tokens required for an AI agent to replicate core workflows is likely to be far greater than basic chatbots given its always-on, persistent nature. Assuming ~37% knowledge-worker adoption by 2040 (peak adoption), we estimate global token consumption could rise by ~55X from current levels.

We estimate that ~37% adoption across global knowledge workers could drive total token consumption to ~55X current levels (or ~278 quadrillion tokens processed per month) by 2030, with enterprise workloads representing >70% of total usage. Although consumer agents have already reached a larger user base than enterprise agents in 2026, enterprise agents should be materially more token-intensive per user and eventually become the primary source of incremental token demand by 2030. Critically, this is because enterprise workflows require agents to perform considerably more complex and precise actions (e.g. monitor tasks, retrieve context, reason through exceptions, validate outputs, update systems, and escalate issues throughout the workday). Enterprise agents may also involve heavier multi-modal inputs (voice, images, documents, screen activity, application data, logs, and structured system records) in order to replicate the complex and varied workflows of real knowledge workers, which can materially increase token intensity versus text-only consumer prompts. **In short, the core difference is rigor: a consumer agent can often stop at a "good enough" answer, but an enterprise agent may need multiple loops to produce work that is accurate, auditable, and usable for real business processes.**

Exhibit 24: Assuming ~37% knowledge-worker adoption by 2040 (12% by 2030), we estimate global token consumption could be 55X current levels (24X by 2030)



Source: Data compiled by Goldman Sachs Global Investment Research

Where is AI being used in today's enterprise?

To estimate token intensity, we built model agents for the most exposed occupations (according to Anthropic's AI usage observations) in order to gauge

where enterprise token demand could scale first. Specifically, Anthropic defines "observed exposure" based on actual Claude usage patterns mapped to occupational tasks, rather than a purely theoretical assessment of whether a job could be automated (Exhibit 25). In our analysis, these occupations represent roles where users are already observed applying AI to meaningful parts of the workflow, including software development, customer service, data entry, administrative support, QA, security analysis, and other knowledge-work functions.

We built our own model agents to move beyond abstract "job exposure" and directly estimate tokens consumed by task steps, reasoning loops, validation checks, tool calls, and multi-modal inputs. We then use these bottom-up agent builds to triangulate how total enterprise token demand could scale through 2030 as adoption broadens across exposed occupations.

Exhibit 25: AI exposure by occupation (as confirmed by Anthropic usage data)

	Occupation	Est. Global Count	Est. Exposure	Leading Automated Task
1	Computer programmers	36,500,000	74.5%	Write, update, and maintain software programs
2	Customer service representatives	70,000,000	70.1%	Confer with customers to provide info, take orders, handle complaints
3	Data entry keyers	4,000,000	67.1%	Read source documents and enter data into systems
4	Medical record specialists	200,000	66.7%	Compile, abstract, and code patient data
5	Market research analysts and marketing specialists	8,500,000	64.8%	Prepare reports of findings, illustrating data graphically and translating complex findings into written text
6	Sales representatives	6,500,000	62.8%	Contact customers to demonstrate products and solicit orders
7	Financial and investment analysts	500,000	57.2%	Inform investment decisions by analyzing financial information to forecast business, industry, or economic conditions
8	Software quality assurance analysts and testers	9,000,000	51.9%	Modify software to correct errors or improve performance
9	Information security analysts	6,000,000	48.6%	Perform risk assessments and test data processing security
10	Computer user support specialists	17,000,000	46.8%	Answer user inquiries regarding computer software or hardware operation to resolve problems

Source: Anthropic, Goldman Sachs Global Investment Research

Why roles exposed to potential automation may not translate to rapid agentic AI adoption

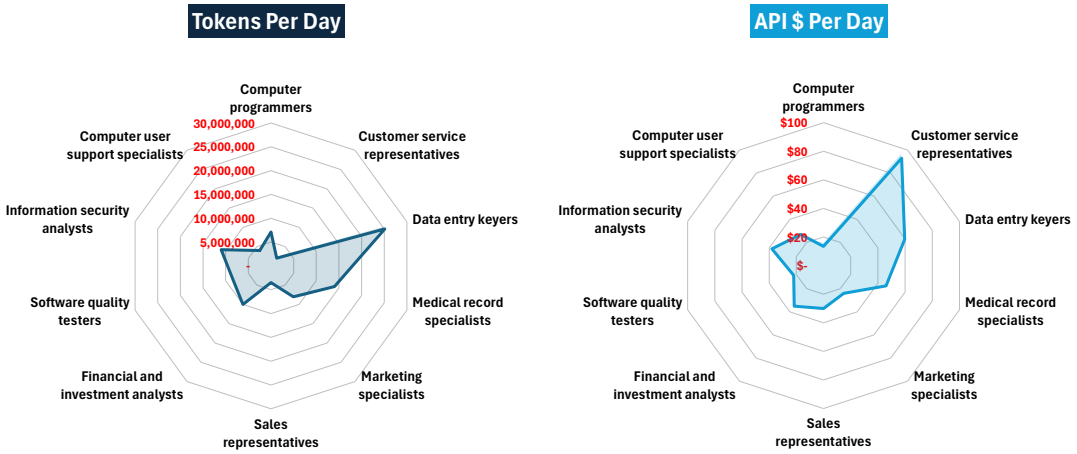
The Bottom Line: In the enterprise, we believe agentic AI adoption will depend less on token volume and more on the full API cost of replicating a user's workflow. The radial charts (Exhibit 26) show a key tension: some agents can consume very high token volumes but remain relatively low cost if the workflow is mostly text-based, while others can consume fewer tokens but carry much higher API cost because they require real-time voice, cadence, latency, and multi-modal processing.

This is clearest in the contrast between coding and call center agents. We estimate **a coding agent could consume roughly 7mn tokens per day at only ~$13/day**, which helps explain why software development has already seen faster agent adoption: the workflow is high-value, largely text-based, and the tooling ecosystem is comparatively mature (Exhibit 27). By contrast, we estimate **a call center agent could consume roughly 2mn tokens per day but cost ~$92/day** if it relies heavily on real-time voice, making full voice-based automation materially more expensive and operationally complex. As a result, we believe many customer-service workflows may initially shift toward text-first automation, with voice used selectively rather than continuously. Finally, we estimate **a data entry agent could consume roughly 25mn tokens per day**

and cost ~$59/day, reflecting a low-modality but high-volume workflow (Exhibit 28) that may still be considerably cheaper than the cost of a human per day.

Overall, we believe adoption speed by occupation will depend on four variables: token volume, API cost, modality mix, and implementation complexity. Text-heavy workflows with mature tools should scale first. Voice-heavy or deeply integrated back-office workflows may scale more slowly, even when the underlying tasks are highly exposed to potential AI automation. As a result, adoption may be uneven across occupations because the most exposed roles are not necessarily the cheapest to automate. Enterprises may first optimize existing workflows around current price points before full-scale human-equivalent agents become cost-effective. However, as model costs fall and performance improves, organizations may increasingly integrate agents into production.

Exhibit 26: Token intensity and API costs do not always move in lockstep



Source: Data compiled by Goldman Sachs Global Investment Research

Inside the agent: What simulating our own agents taught us about enterprise agent architecture

Our methodology consisted of building occupation-specific, minimally-functional agents on paper by decomposing each role into workflow steps, model calls, tool calls, validation checks, and retry loops. We believe this is an appropriate way to measure token intensity because enterprise agents will rarely run as a single end-to-end model, and our agents represent the likely minimum token intensity for such workflows.

What distinguishes an AI agent from a chatbot? A chatbot typically responds to a discrete user prompt; an autonomous enterprise agent must ingest context, decide what to do, retrieve relevant information, execute the task, check its own work, resolve errors, update systems, and produce a usable output. That means the agent is not a single model call, but a sequence of model calls across a workflow. Different steps may require different model capabilities. For example, a lower-cost model may classify or extract information, a stronger reasoning model may plan, debug, reconcile conflicts, or resolve exceptions, a code model may write or test software, and a vision or document model may parse images, PDFs, screenshots, or forms. The resulting architecture looks less like a chatbot and more like a workflow system with LLMs embedded across each decision point.

What drives token intensity for AI agents? One key driver of token intensity is the loop structure required to make the output reliable enough for enterprise use. For a software coding agent, loops are needed to read the codebase, generate an implementation plan, write code, run tests, debug errors, review security or performance issues, and then revise until the output passes functional tests. In the data entry agent, loops are needed to ingest documents, extract fields, compare against source systems, flag inconsistencies, correct errors, and validate that the final entry is complete. These loops increase but do not guarantee accuracy, as they simply create more opportunities for the agent to catch and correct mistakes before the output reaches production.

Bottom line: The token consumption curve can steepen quickly in complex workflows. A clean task may pass through the loop once; however, a messy task with incomplete inputs, failed tests, or additional human guidance may trigger multiple rounds of reasoning and validation.

Exhibit 27: Building a coding agent
Text-heavy, loop-heavy, but cost-efficient

Computer Programmer Agent				
Action	**Model Reasoning**	**Input Tokens**	**Output To**	**# Loops**
Parse daily objective	Medium	2000	500	
Load repo architecture/system context	High	45000	2000	
Load relevant codebase	High	55000	2000	
Plan implementation strategy	High	12000	3000	
Workday setup loop total		**114000**	**3000**	**1**
Retrieve relevant files/modules	Medium	30000	1000	
Reason about code changes across components	High	18000	3000	
Generate/update code	High	45000	12000	
Run build/tests/linter	External Tool	External Tool	External Tool	
Interpret compiler/test failures	High	12000	2000	
Revise code based on failures	High	25000	6000	
Update local coding memory/state	Low	5000	800	
Feature implementation loop total		**1620000**	**297600**	**12**
Reproduce bug/ingest logs/stack traces	Medium	20000	1500	
Inspect execution paths	High	35000	2000	
Hypothesize root cause	High	25000	2000	
Patch fix	High	28000	5000	
Re-run validation/targeted tests	External Tool	External Tool	External Tool	
Summarize debugging findings	Medium	8000	1000	
Debugging loop total		**2320000**	**230000**	**20**
Scan for technical debt/cleanup candidates	Medium	22000	1500	
Compare against style/standards/existing patterns	Medium	18000	1000	
Refactor code	High	32000	6000	
Update comments/inline docs	Medium	10000	2000	
Persist repo knowledge/state	Low	5000	500	
Maintenance/refactor loop total		**1044000**	**132000**	**12**
Ingest diff/PR context	Medium	25000	1000	
Review for bugs/security/regressions	High	20000	2500	
Propose improvements/fixes	Medium	12000	3000	
Apply edits/respond to review comments	Medium	18000	3000	
Update changelog/task state	Medium	6000	700	
Code review loop total		**810000**	**102000**	**10**
Inspect depency/API/schema changes	Medium	28000	1500	
Reason about compatbility/migration impact	High	16000	2000	
Implement integration updates	High	24000	5000	
Run integration/build verification	External Tool	External Tool	External Tool	
Document migration/rollout steps	Medium	10000	2000	
Integration update loop total		**312000**	**42000**	**4**
Generate end-of-day handoff summary	Medium	6000	2000	
Update backlog/next steps	Medium	8000	2000	
Write final technical documentation	Medium	14000	4000	
End-of-day wrap total		**56000**	**16000**	**2**

Total Cost	**Input Tokens**	**Output Tokens**
Total Tokens	6,276,000	822,600
API Cost	$8.79	$4.61

Source: Goldman Sachs Global Investment Research

Exhibit 28: Building a data entry agent
Text-heavy, low-modality, but token-heavy

Data Entry Agent			
Action	**Model Reasoning**	**Tokens**	**# Loops**
Load task queue/document batch	Medium	25000	
Load schemas/required fields/formats	Medium	40000	
Load validation rules/business logic	High	70000	
Initialize structured output templates	Medium	20000	
Shift setup loop total		155000	1
Ingest document	Medium	18000	
Run OCR/text extraction	Medium	22000	
Clean/normalize raw text	Medium	12000	
Segment into logical sections (tables, fields, notes)	Medium	10000	
Document ingestion loop total		6200000	100
Identify required fields/targets	Medium	8000	
Extract structured data fields	High	16000	
Map extracted values to schema	Medium	10000	
Handle missing/ambiguous fields	High	12000	
Generate structured JSON/table output	Medium	8000	
Data extraction loop total		5400000	100
Cross-check extracted values vs. source text	Medium	12000	
Validate against business rules/constraints	High	14000	
Detect inconsistencies/anomalies	High	10000	
Resolve via reasoning or flagging	Medium	12000	
Update validation state	Low	4000	
Validation loop total		5200000	100
Detect low-confidence fields/OCR errors	Medium	10000	
Re-run targeted extraction/parsing	High	16000	
Compare alternate interpretations	High	12000	
Select best candidate/merge results	Medium	8000	
Update structured output	Low	4000	
Error correction loop total		2500000	50
Retrieve external system records	External Tool	External Tool	
Match entity	Medium	14000	
Revolve duplicates/conflicts	High	16000	
Confirm final record linkage	Medium	8000	
System reconciliation total		1900000	50
Input structure data into target system	Medium	12000	
Verify successful write/API response	Medium	8000	
Log transaction/audit trail	Low	5000	
Systen entry loop total		2500000	100
Generate structure summary of document	Medium	10000	
Tag document type/category/metadata	Medium	8000	
Store in retrieval systems/index	External Tool	External Tool	
Update memory for similar future docs	Low	4000	
Post-processing loop total		1200000	100
Aggregate error rates/accuracy metrics	Medium	20000	
Identify recurring extraction failures	High	25000	
Update heuristics/prompts for next batch	High	30000	
Generate supervisor QA summary	Medium	20000	
End-of-shift wrap total		75000	1

Model	Tokens
Total Tokens	25,130,000
API Cost	$59.68375

Source: Goldman Sachs Global Investment Research

Valuation/Risks

MSFT (Buy): Our 12-month price target of $610 is based on a 28x P/E multiple to Microsoft's SNTM adjusted net income. Key downside risks include: longer ramp for internal silicon which can limit market share gains or gross margin expansion, greater investments in projects outside of expectations (e.g., non-Azure), key leadership changes, and more meaningful shift to custom software that could negatively impact its applications business.

NET (Buy): Our 12-month price target of $250 is based on 27x Q5-Q8 revenue. Key risks include: the timing of new productivity ramps and potential disruption in the sales organization as Cloudflare expands into enterprise, SSE traction, and inference demand potentially being met by hyperscalers or other compute solutions.

ACN (Buy): Our 12-month price target of $300 is based on 20x our one-year forward earnings estimates. Downside risks: 1) lower multiple from further demand slowdown, 2) notable deceleration in AI bookings, 3) increased weakness in key verticals.

AVGO (Buy): Our 12-month target price of $480 is based on 30X our normalized EPS estimate of $16. Key downside risks include: (1) slowdown in AI infrastructure spending, (2) share loss in custom compute franchise, (3) persistent inventory digestion in non-AI; (4) increased competition in VMware.

NVDA (Buy): Our 12-month price target of $250 is based on a 30X P/E multiple applied to our normalized EPS estimate of $8.25. Key downside risks include: (1) slowdown in AI infrastructure spending, (2) share erosion due to increased competitive intensity, (3) margin erosion due to increased competition; (4) supply constraints.

AMD (Buy): Our 12-month target price of $450 is based on a 30X P/E multiple applied to our normalized EPS estimate of $15.00. Key downside risks include: (1) slower than expected adoption of agentic AI; (2) slower than anticipated deployment of AMD GPUs related to the Meta deal; (3) share erosion of x86 architecture in enterprise AI and (4) lack of operating leverage.

AMZN (Buy): Our 12m PT of $325 is based on: (1) An equal blend of EV/GAAP EBITDA applied to our NTM+1 estimates and a modified DCF using an EV/(FCF-SBC) multiple applied to our NTM+4 estimates discounted back 3 years; (2) SOTP of EV/EBIT applied to North America and AWS segments, and EV/Sales applied to the International segment, all based on NTM+1 estimates.

GOOGL (Buy): Our $450, 12-month price target is based on an equal blend of (1) EV/GAAP EBIT applied to our NTM + 1 year estimates and (2) a modified DCF using EV/FCF-SBC multiple applied to our NTM + 4 years estimates discounted back 3 years.

META (Buy): Our $830 12-month price target is based on an equal blend of (1) EV/GAAP EBIT applied to our NTM + 1 year estimates and (2) a modified DCF using EV/FCF-SBC multiple applied to our NTM + 4 years estimates.

Disclosure Appendix

Reg AC

We, James Schneider, Ph.D., Eric Sheridan, Gabriela Borges, CFA, Luya You, Anmol Makkar, Noah Naparst and Emma Huang, 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.

Contributing Authors: James Schneider, Ph.D. Goldman Sachs & Co. LLC, Eric Sheridan Goldman Sachs & Co. LLC, Gabriela Borges, CFA Goldman Sachs & Co. LLC, Luya You Goldman Sachs & Co. LLC, Anmol Makkar Goldman Sachs & Co. LLC, Noah Naparst Goldman Sachs & Co. LLC, Emma Huang Goldman Sachs & Co. LLC.

Unless otherwise stated, the individuals listed in the Contributing Authors disclosure of this report are analysts in Goldman Sachs' Global Investment Research division.

GS Factor Profile

The Goldman Sachs Factor Profile provides investment context for a stock by comparing key attributes to the market (i.e. our universe of rated stocks) and its sector peers. The four key attributes depicted are: Growth, Financial Returns, Multiple (e.g. valuation) and Integrated (a composite of Growth, Financial Returns and Multiple). Growth, Financial Returns and Multiple are calculated by using normalized ranks for specific metrics for each stock. The normalized ranks for the metrics are then averaged and converted into percentiles for the relevant attribute. The precise calculation of each metric may vary depending on the fiscal year, industry and region, but the standard approach is as follows:

Growth is based on a stock's forward-looking sales growth, EBITDA growth and EPS growth (for financial stocks, only EPS and sales growth), with a higher percentile indicating a higher growth company. **Financial Returns** is based on a stock's forward-looking ROE, ROCE and CROCI (for financial stocks, only ROE), with a higher percentile indicating a company with higher financial returns. **Multiple** is based on a stock's forward-looking P/E, P/B, price/dividend (P/D), EV/EBITDA, EV/FCF and EV/Debt Adjusted Cash Flow (DACF) (for financial stocks, only P/E, P/B and P/D), with a higher percentile indicating a stock trading at a higher multiple. The **Integrated** percentile is calculated as the average of the Growth percentile, Financial Returns percentile and (100% - Multiple percentile).

Financial Returns and Multiple use the Goldman Sachs analyst forecasts at the fiscal year-end at least three quarters in the future. Growth uses inputs for the fiscal year at least seven quarters in the future compared with the year at least three quarters in the future (on a per-share basis for all metrics).

For a more detailed description of how we calculate the GS Factor Profile, please contact your GS representative.

M&A Rank

Across our global coverage, we examine stocks using an M&A framework, considering both qualitative factors and quantitative factors (which may vary across sectors and regions) to incorporate the potential that certain companies could be acquired. We then assign a M&A rank as a means of scoring companies under our rated coverage from 1 to 3, with 1 representing high (30%-50%) probability of the company becoming an acquisition target, 2 representing medium (15%-30%) probability and 3 representing low (0%-15%) probability. For companies ranked 1 or 2, in line with our standard departmental guidelines we incorporate an M&A component into our target price. M&A rank of 3 is considered immaterial and therefore does not factor into our price target, and may or may not be discussed in research.

Quantum

Quantum is Goldman Sachs' proprietary database providing access to detailed financial statement histories, forecasts and ratios. It can be used for in-depth analysis of a single company, or to make comparisons between companies in different sectors and markets.

Disclosures

Pricing information

Accenture Plc ($179.83), Advanced Micro Devices Inc. ($360.54), Alphabet Inc. ($385.69), Amazon.com Inc. ($268.26), Broadcom Inc. ($421.28), Cloudflare ($217.50), Meta Platforms Inc. ($608.75), Microsoft Corp. ($414.44), Nvidia Corp. ($198.45) and ServiceNow Inc. ($91.16).

The rating(s) for Alphabet Inc., Amazon.com Inc. and Meta Platforms Inc. is/are relative to the other companies in its/their coverage universe: ACV Auctions Inc., Airbnb Inc., Alphabet Inc., Amazon.com Inc., Angi Inc., AppLovin Corp., Booking Holdings Inc., Bumble Inc., Chewy Inc., Clear Secure Inc., Corsair Gaming Inc., Coursera Inc., Cricut Inc., DoorDash Inc., DoubleVerify Inc., DraftKings Inc., Duolingo Inc., Ethos Technologies Inc., Expedia Group, Fiverr International Ltd., Frontdoor Inc., GoodRx Holdings Inc., Grindr Inc., Groupon Inc., IAC Inc., Ibotta Inc., Instacart, Lyft Inc., Match Group, MediaAlpha Inc., Meta Platforms Inc., Netflix Inc., Nextdoor Holdings, Opera Ltd., Pattern Group Inc., Peloton Interactive Inc., Pinterest Inc., Playtika, Reddit Inc., Roblox, Snap Inc., Sportradar Group, Spotify Technology S.A., StubHub Holdings, Take-Two Interactive Software Inc., Tripadvisor Inc., Uber Technologies Inc., Unity Software Inc., Upwork Inc., Wayfair Inc., Webtoon Entertainment Inc., Xometry Inc., Yelp Inc., ZipRecruiter Inc.

The rating(s) for Accenture Plc, Advanced Micro Devices Inc., Broadcom Inc. and Nvidia Corp. is/are relative to the other companies in its/their coverage universe: ARM Holdings, Accenture Plc, Advanced Micro Devices Inc., American Tower Corp., Amkor Technology Inc., Analog Devices Inc., Applied Materials Inc., Broadcom Inc., Cadence Design Systems Inc., Camtek, Cognizant Technology Solutions, Credo Technology Group, Crown Castle Inc., EPAM Systems Inc., Entegris Inc., GlobalFoundries Inc., Globant SA, International Business Machines Corp., KLA Corp., Lam Research Corp., MKS Instruments Inc., Marvell Technology Inc., Microchip Technology Inc., Micron Technology Inc., NXP Semiconductors NV, Nvidia Corp., ON Semiconductor Corp., Qnity, Qualcomm Inc., SBA Communications Corp., SanDisk Corp., Seagate Technology, SiTime Corp., Synopsys Inc., TaskUs Inc., Teradyne Inc., Texas Instruments Inc., Western Digital Corp.

The rating(s) for Cloudflare, Microsoft Corp. and ServiceNow Inc. is/are relative to the other companies in its/their coverage universe: Adobe Inc., Akamai Technologies Inc., AvePoint, Braze Inc., Check Point, Cloudflare, Commerce.com Inc., CoreWeave Inc., CrowdStrike, Datadog Inc., DigitalOcean Holdings, EverCommerce Inc., Figma Inc., Fortinet, HubSpot Inc., Intuit Inc., Klaviyo, Microsoft Corp., Navan Inc., Okta, Oracle Corp., Palantir Technologies, Palo Alto Networks, SailPoint Inc., Salesforce Inc., SentinelOne, ServiceNow Inc., Shopify Inc., Snowflake Inc., Workday Inc., Zeta Global, Zscaler

Company-specific regulatory disclosures

Compendium report: please see disclosures at https://www.gs.com/research/hedge.html. Disclosures applicable to the companies included in this compendium can be found in the latest relevant published research

Distribution of ratings/investment banking relationships

Goldman Sachs Investment Research global Equity coverage universe

	Rating Distribution			Investment Banking Relationships		
	Buy	Hold	Sell	Buy	Hold	Sell
Global	50%	34%	16%	65%	60%	45%

As of April 1, 2026, Goldman Sachs Global Investment Research had investment ratings on 3,074 equity securities. Goldman Sachs assigns stocks as Buys and Sells on various regional Investment Lists; stocks not so assigned are deemed Neutral. Such assignments equate to Buy, Hold and Sell for the purposes of the above disclosure required by the FINRA Rules. See 'Ratings, Coverage universe and related definitions' below. The Investment Banking Relationships chart reflects the percentage of subject companies within each rating category for whom Goldman Sachs has provided investment banking services within the previous twelve months.

Price target and rating history chart(s)

Compendium report: please see disclosures at https://www.gs.com/research/hedge.html. Disclosures applicable to the companies included in this compendium can be found in the latest relevant published research

Target price history table(s)

Cloudflare (NET)

Date of report	Target price ($)	Closing price ($)
11-Feb-26	250.00	189.41
31-Oct-25	268.00	253.30
01-Aug-25	250.00	200.11
09-May-25	172.00	132.34
13-Mar-25	161.00	113.70
07-Feb-25	183.00	166.66
02-Jan-25	140.00	112.54
08-Nov-24	77.00	91.25
02-Aug-24	72.00	79.46
03-Jun-24	68.00	67.43
03-May-24	77.00	74.40
12-Feb-24	89.00	105.62
03-Nov-23	45.00	64.42
04-Aug-23	43.00	69.52

Microsoft Corp. (MSFT)

Date of report	Target price ($)	Closing price ($)
30-Apr-26	610.00	407.78
29-Jan-26	600.00	433.50
11-Jan-26	655.00	479.28
31-Jul-25	630.00	533.50
20-May-25	550.00	458.17
01-May-25	480.00	425.40
24-Apr-25	450.00	387.30
10-Oct-24	500.00	415.84
26-Apr-24	515.00	406.32
25-Oct-23	450.00	340.67
21-Jul-23	400.00	343.77
26-May-23	350.00	332.89

Nvidia Corp. (NVDA)

Date of report	Target price ($)	Closing price ($)
20-Nov-25	250.00	180.64
31-Oct-25	240.00	202.49
06-Oct-25	210.00	185.54
07-Aug-25	200.00	180.77
10-Jul-25	185.00	164.10
21-Nov-24	165.00	146.67
11-Oct-24	150.00	134.80
11-Jun-24	135.00	120.91
23-May-24	1,200.00	103.80
07-May-24	1,100.00	90.55
19-Mar-24	1,000.00	89.40
22-Feb-24	875.00	78.54
05-Feb-24	800.00	69.33
22-Nov-23	625.00	48.72
24-Aug-23	605.00	47.16
09-Jul-23	495.00	42.50
25-May-23	440.00	37.98

Amazon.com Inc. (AMZN)

Date of report	Target price ($)	Closing price ($)
30-Apr-26	325.00	265.06
13-Apr-26	275.00	239.89
06-Feb-26	280.00	210.32
13-Jan-26	300.00	242.60
31-Oct-25	290.00	244.22
03-Oct-25	275.00	219.51
01-Aug-25	240.00	214.75
21-Apr-25	220.00	167.32
07-Feb-25	255.00	229.15
01-Nov-24	240.00	197.93
02-Aug-24	230.00	167.90
15-Jul-24	250.00	192.72
01-May-24	225.00	179.00
02-Feb-24	220.00	171.81
10-Jan-24	200.00	153.73
27-Oct-23	190.00	127.74
12-Oct-23	175.00	132.33
04-Aug-23	180.00	139.57

Broadcom Inc. (AVGO)

Date of report	Target price ($)	Closing price ($)
05-Mar-26	480.00	332.77
12-Dec-25	450.00	359.93
25-Nov-25	435.00	385.03
13-Oct-25	380.00	356.70

Alphabet Inc. (GOOGL)

Date of report	Target price ($)	Closing price ($)
30-Apr-26	450.00	384.80
05-Feb-26	400.00	331.25
12-Jan-26	375.00	331.86
30-Oct-25	330.00	281.48

Broadcom Inc. (AVGO)

Date of report	Target price ($)	Closing price ($)
05-Sep-25	360.00	334.89
13-Aug-25	340.00	309.09
10-Jul-25	315.00	275.40
10-Jan-25	255.00	224.31
16-Dec-24	240.00	250.00
06-Sep-24	190.00	137.00
13-Jun-24	1,850.00	167.90
08-Mar-24	1,550.00	130.87
19-Jan-24	1,325.00	121.12

Alphabet Inc. (GOOGL)

Date of report	Target price ($)	Closing price ($)
14-Oct-25	288.00	245.45
24-Jul-25	234.00	192.17
10-Jul-25	225.00	177.62
25-Apr-25	220.00	161.96
11-Apr-25	205.00	157.14
05-Feb-25	220.00	191.33
13-Jan-25	215.00	191.01
30-Oct-24	210.00	174.46
14-Oct-24	208.00	164.96
24-Jul-24	217.00	172.63
08-Jul-24	211.00	189.03
26-Apr-24	195.00	171.95
15-Apr-24	185.00	154.86
31-Jan-24	171.00	140.10
10-Jan-24	164.00	142.28
04-Oct-23	154.00	135.24
26-Jul-23	152.00	129.27
17-Jul-23	140.00	124.65

ServiceNow Inc. (NOW)

Date of report	Target price ($)	Closing price ($)
23-Apr-26	163.00	84.78
06-Apr-26	188.00	102.42
29-Jan-26	216.00	116.73
11-Jan-26	205.00	141.80
30-Oct-25	1,250.00	186.94
24-Jul-25	1,200.00	199.24
24-Apr-25	1,150.00	187.71
10-Apr-25	1,050.00	156.70
10-Jan-25	1,200.00	204.91
24-Oct-24	1,050.00	191.32
10-Oct-24	1,000.00	188.93
25-Jul-24	940.00	165.76
25-Jan-24	910.00	153.34
16-Jan-24	800.00	145.51
27-Jul-23	700.00	111.98

Advanced Micro Devices Inc. (AMD)

Date of report	Target price ($)	Closing price ($)
24-Feb-26	240.00	213.84
07-Oct-25	210.00	211.51
06-Aug-25	150.00	163.12
10-Jul-25	140.00	144.16
05-Feb-25	125.00	112.01
10-Jan-25	129.00	116.04
01-May-24	175.00	144.27
31-Jan-24	180.00	167.69
18-Dec-23	157.00	138.90
07-Dec-23	137.00	128.37
01-Nov-23	125.00	108.04
14-Jun-23	137.00	127.33

Accenture Plc (ACN)

Date of report	Target price ($)	Closing price ($)
19-Mar-26	300.00	203.55
25-Sep-25	330.00	232.56
20-Jun-25	370.00	285.37
03-Jun-25	390.00	312.90
20-Mar-25	415.00	300.91
19-Dec-24	430.00	372.16
06-Dec-24	420.00	362.70
26-Sep-24	370.00	355.81
24-Jun-24	335.00	307.21

Meta Platforms Inc. (META)

Date of report	Target price ($)	Closing price ($)
30-Apr-26	830.00	611.91
13-Apr-26	840.00	634.53
29-Jan-26	835.00	738.31
30-Oct-25	815.00	666.47
14-Oct-25	870.00	708.65
31-Jul-25	830.00	773.44
10-Jul-25	775.00	727.24
01-May-25	690.00	572.21
11-Apr-25	685.00	543.57
30-Jan-25	765.00	687.00
13-Jan-25	688.00	608.33
31-Oct-24	630.00	567.58
14-Oct-24	636.00	590.42
01-Aug-24	555.00	497.74
08-Jul-24	522.00	529.32
25-Apr-24	500.00	441.38
15-Apr-24	555.00	500.23
02-Feb-24	500.00	474.99
10-Jan-24	414.00	370.47

Accenture Plc (ACN)

Date of report	Target price ($)	Closing price ($)

Meta Platforms Inc. (META)

Date of report	Target price ($)	Closing price ($)
27-Jul-23	384.00	311.71
17-Jul-23	335.00	310.62

Price targets shown in table(s) are unadjusted for corporate actions.

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 a subject company, public appearances and trading in securities covered by the analysts.

Distribution of ratings: See the distribution of ratings disclosure above. **Price chart:** See the price chart, with changes of ratings and price targets in prior periods, above, or, if electronic format or if with respect to multiple companies which are the subject of this report, on the Goldman Sachs website at https://www.gs.com/research/hedge.html.

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 information on the securities of covered companies referred to in this research may be obtained on request from Goldman Sachs (Asia) L.L.C. **India:** Further information on the subject company or companies referred to in this research may be obtained from Goldman Sachs (India) Securities Private Limited, Research Analyst - SEBI Registration Number INH000001493, 10th Floor, Ascent-Worli, Sudam Kalu Ahire Marg, Worli, Mumbai-400 025, India, Corporate Identity Number U74140MH2006FTC160634, Phone +91 22 6616 9000, Fax +91 22 6616 9001. Goldman Sachs may beneficially own 1% or more of the securities (as such term is defined in clause 2 (h) the Indian Securities Contracts (Regulation) Act, 1956) of the subject company or companies referred to in this research report. Investment in securities market are subject to market risks. Read all the related documents carefully before investing. Registration granted by SEBI and certification from NISM in no way guarantee performance of the intermediary or provide any assurance of returns to investors. Goldman Sachs (India) Securities Private Limited compliance officer and investor grievance contact details can be found at: https://www.goldmansachs.com/worldwide/india/documents/Grievance-Redressal-and-Escalation-Matrix.pdf, and a copy of the annual audit compliance report can be found at this link: https://publishing.gs.com/content/site/india-annual-compliance-report.html. **Japan:** See below. **Korea:** This research, and any access to it, is intended only for "professional investors" within the meaning of the Financial Services and Capital Markets Act, unless otherwise agreed by Goldman Sachs. Further information on the subject company or companies referred to in this research may be obtained from Goldman Sachs (Asia) L.L.C., Seoul Branch. **New Zealand:** Goldman Sachs New Zealand Limited and its affiliates are neither "registered banks" nor "deposit takers" (as defined in the Reserve Bank of New Zealand Act 1989) in New Zealand. This research, and any access to it, is intended for "wholesale clients" (as defined in the Financial Advisers Act 2008) unless otherwise agreed by Goldman Sachs. A copy of certain Goldman Sachs Australia and New Zealand disclosure of interests is available at: https://www.goldmansachs.com/disclosures/australia-new-zealand/index.html. **Russia:** Research reports distributed in the Russian Federation are not advertising as defined in the Russian legislation, but are information and analysis not having product promotion as their main purpose and do not provide appraisal within the meaning of the Russian legislation on appraisal activity. Research reports do not constitute a personalized investment recommendation as defined in Russian laws and regulations, are not addressed to a specific client, and are prepared without analyzing the financial circumstances, investment profiles or risk profiles of clients. Goldman Sachs assumes no responsibility for any investment decisions that may be taken by a client or any other person based on this research report. **Singapore:** Goldman Sachs (Singapore) Pte. (Company Number: 198602165W), which is regulated by the Monetary Authority of Singapore, accepts legal responsibility for this research, and should be contacted with respect to any matters arising from, or in connection with, this research. **Taiwan:** This material is for reference only and must not be reprinted without permission. Investors should carefully consider their own investment risk. Investment results are the responsibility of the individual investor. **United Kingdom:** Persons who would be categorized as retail clients in the United Kingdom, as such term is defined in the rules of the Financial Conduct Authority, should read this research in conjunction with prior Goldman Sachs research on the covered companies referred to herein and should refer to the risk warnings that have been sent to them by Goldman Sachs International. A copy of these risks warnings, and a glossary of certain financial terms used in this report, are available from Goldman Sachs International on request.

European Union and United Kingdom: Disclosure information in relation to Article 6 (2) of the European Commission Delegated Regulation (EU) (2016/958) supplementing Regulation (EU) No 596/2014 of the European Parliament and of the Council (including as that Delegated Regulation is

implemented into United Kingdom domestic law and regulation following the United Kingdom's departure from the European Union and the European Economic Area) with regard to regulatory technical standards for the technical arrangements for objective presentation of investment recommendations or other information recommending or suggesting an investment strategy and for disclosure of particular interests or indications of conflicts of interest is available at https://www.gs.com/disclosures/europeanpolicy.html which states the European Policy for Managing Conflicts of Interest in Connection with Investment Research.

Japan: Goldman Sachs Japan Co., Ltd. is a Financial Instrument Dealer registered with the Kanto Financial Bureau under registration number Kinsho 69, and a member of Japan Securities Dealers Association, Financial Futures Association of Japan Type II Financial Instruments Firms Association, and Investment Management Association of Japan. Sales and purchase of equities are subject to commission pre-determined with clients plus consumption tax. See company-specific disclosures as to any applicable disclosures required by Japanese stock exchanges, the Japanese Securities Dealers Association or the Japanese Securities Finance Company.

Ratings, coverage universe and related definitions

Buy (B), Neutral (N), Sell (S) Analysts recommend stocks as Buys or Sells for inclusion on various regional Investment Lists. Being assigned a Buy or Sell on an Investment List is determined by a stock's total return potential relative to its coverage universe. Any stock not assigned as a Buy or a Sell on an Investment List with an active rating (i.e., a stock that is not Rating Suspended, Not Rated, Early-Stage Biotech, Coverage Suspended or Not Covered), is deemed Neutral. Each region manages Regional Conviction Lists, which are selected from Buy rated stocks on the respective region's Investment Lists and represent investment recommendations focused on the size of the total return potential and/or the likelihood of the realization of the return across their respective areas of coverage. The addition or removal of stocks from such Conviction Lists are managed by the Investment Review Committee or other designated committee in each respective region and do not represent a change in the analysts' investment rating for such stocks.

Total return potential represents the upside or downside differential between the current share price and the price target, including all paid or anticipated dividends, expected during the time horizon associated with the price target. Price targets are required for all covered stocks. The total return potential, price target and associated time horizon are stated in each report adding or reiterating an Investment List membership.

Coverage Universe: A list of all stocks in each coverage universe is available by primary analyst, stock and coverage universe at https://www.gs.com/research/hedge.html.

Not Rated (NR). The investment rating, target price and earnings estimates (where relevant) are removed pursuant to Goldman Sachs policy when Goldman Sachs is acting in an advisory capacity in a merger or in a strategic transaction involving this company, when there are legal, regulatory or policy constraints due to Goldman Sachs' involvement in a transaction, and in certain other circumstances. **Early-Stage Biotech (ES).** An investment rating and a target price are not assigned pursuant to Goldman Sachs policy when this company has neither a drug, treatment or medical device that has passed a Phase II clinical trial nor a license to distribute a post-Phase II drug, treatment or medical device. **Rating Suspended (RS).** Goldman Sachs Research has suspended the investment rating and price target for this stock, because there is not a sufficient fundamental basis for determining an investment rating or target price. The previous investment rating and target price, if any, are no longer in effect for this stock and should not be relied upon. **Coverage Suspended (CS).** Goldman Sachs has suspended coverage of this company. **Not Covered (NC).** Goldman Sachs does not cover this company.

Global product; distributing entities

Goldman Sachs Global Investment Research produces and distributes research products for clients of Goldman Sachs on a global basis. Analysts based in Goldman Sachs offices around the world produce research on industries and companies, and research on macroeconomics, currencies, commodities and portfolio strategy. This research is disseminated in Australia by Goldman Sachs Australia Pty Ltd (ABN 21 006 797 897); in Brazil by Goldman Sachs do Brasil Corretora de Títulos e Valores Mobiliários S.A.; Public Communication Channel Goldman Sachs Brazil: 0800 727 5764 and / or contatogoldmanbrasil@gs.com. Available Weekdays (except holidays), from 9am to 6pm. Canal de Comunicação com o Público Goldman Sachs Brasil: 0800 727 5764 e/ou contatogoldmanbrasil@gs.com. Horário de funcionamento: segunda-feira à sexta-feira (exceto feriados), das 9h às 18h; in Canada by Goldman Sachs & Co. LLC; in Hong Kong by Goldman Sachs (Asia) L.L.C.; in India by Goldman Sachs (India) Securities Private Ltd.; in Japan by Goldman Sachs Japan Co., Ltd.; in the Republic of Korea by Goldman Sachs (Asia) L.L.C., Seoul Branch; in New Zealand by Goldman Sachs New Zealand Limited; in Russia by OOO Goldman Sachs; in Singapore by Goldman Sachs (Singapore) Pte. (Company Number: 198602165W); and in the United States of America by Goldman Sachs & Co. LLC. Goldman Sachs International has approved this research in connection with its distribution in the United Kingdom.

Goldman Sachs International ("GSI"), authorised by the Prudential Regulation Authority ("PRA") and regulated by the Financial Conduct Authority ("FCA") and the PRA, has approved this research in connection with its distribution in the United Kingdom.

European Economic Area: 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 within the European Economic Area.

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.

The analysts named in this report may have from time to time discussed with our clients, including Goldman Sachs salespersons and traders, or may discuss in this report, trading strategies that reference catalysts or events that may have a near-term impact on the market price of the equity securities discussed in this report, which impact may be directionally counter to the analyst's published price target expectations for such stocks. Any such trading strategies are distinct from and do not affect the analyst's fundamental equity rating for such stocks, which rating reflects a stock's return potential relative to its coverage universe as described herein.

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 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.

Certain transactions, including those involving futures, options, and other derivatives, give rise to substantial risk and are not suitable for all investors. Investors should review current options and futures disclosure documents which are available from Goldman Sachs sales representatives or at https://www.theocc.com/about/publications/character-risks.jsp and https://www.goldmansachs.com/disclosures/cftc_fcm_disclosures. Transaction costs may be significant in option strategies calling for multiple purchase and sales of options such as spreads. Supporting documentation will be supplied upon request.

Differing Levels of Service provided by Global Investment Research: The level and types of services provided to you by Goldman Sachs Global Investment Research may vary as compared to that provided to internal and other external clients of GS, depending on various factors including your individual preferences as to the frequency and manner of receiving communication, your risk profile and investment focus and perspective (e.g., marketwide, sector specific, long term, short term), the size and scope of your overall client relationship with GS, and legal and regulatory constraints. As an example, certain clients may request to receive notifications when research on specific securities is published, and certain clients may request that specific data underlying analysts' fundamental analysis available on our internal client websites be delivered to them electronically through data feeds or otherwise. No change to an analyst's fundamental research views (e.g., ratings, price targets, or material changes to earnings estimates for equity securities), will be communicated to any client prior to inclusion of such information in a research report broadly disseminated through electronic publication to our internal client websites or through other means, as necessary, to all clients who are entitled to receive such reports.

All research reports are disseminated and available to all clients simultaneously through electronic publication to our internal client websites. Not all research content is redistributed to our clients or available to third-party aggregators, nor is Goldman Sachs responsible for the redistribution of our research by third party aggregators. For research, models or other data related to one or more securities, markets or asset classes (including related services) that may be available to you, please contact your GS representative or go to https://research.gs.com.

Disclosure information is also available at https://www.gs.com/research/hedge.html or from Research Compliance, 200 West Street, New York, NY 10282.

© 2026 Goldman Sachs.

You are permitted to store, display, analyze, modify, reformat, and print the information made available to you via this service only for your own use. You may not resell or reverse engineer this information to calculate or develop any index for disclosure and/or marketing or create any other derivative works or commercial product(s), data or offering(s) without the express written consent of Goldman Sachs. You are not permitted to publish, transmit, or otherwise reproduce this information, in whole or in part, in any format to any third party without the express written consent of Goldman Sachs. This foregoing restriction includes, without limitation, using, extracting, downloading or retrieving this information, in whole or in part, to train or finetune a machine learning or artificial intelligence system, or to provide or reproduce this information, in whole or in part, as a prompt or input to any such system.