

PART 8

Catalyst or Culprit?

We present a comprehensive, collaborative, in-depth and multi-faceted analysis of a very complicated and controversial but potentially defining technology theme in one document. Our work reflects: 1) Description of the Gen-AI tech stack across Infrastructure, Platforms and Applications (IPA); 2) Roadmap for evolution of this technology and comparisons to prior cycles, including our prognosis of seats vs consumption pricing; 3) Investable ideas for each phase of the cycle across Software, Internet, Semiconductors, Hardware, Business Services, IT Services and Industrial Tech, including a new five factor framework for screening Software stocks; 4) Proprietary IT buyer survey with findings on expected Gen-AI adoption; 5) Assessment of practical Gen-AI applications in multiple sectors based on interviews with marquee companies conducted by GS sector analysts; 6) The latest update from GS Macroeconomics team with a [Gen-AI adoption tracker](#); 7) Market sizing of Cloud revenues, expected to reach \$2 Trillion by 2030 – with Gen-AI reaching 10-15% of spend and 8) Detailed discussions with cutting edge private companies, leading Silicon Valley VCs and prominent academics.

Kash Rangan

+1 415 249-7318

kash.rangan@gs.com

Goldman Sachs & Co. LLC

Gili Naftalovich

+1 917 343-4286

gili.naftalovich@gs.com

Goldman Sachs & Co. LLC

Matthew Martino

+1 212 902-0695

matt.martino@gs.com

Goldman Sachs & Co. LLC

Toshiya Hari

+1 646 446-1759

toshiya.hari@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

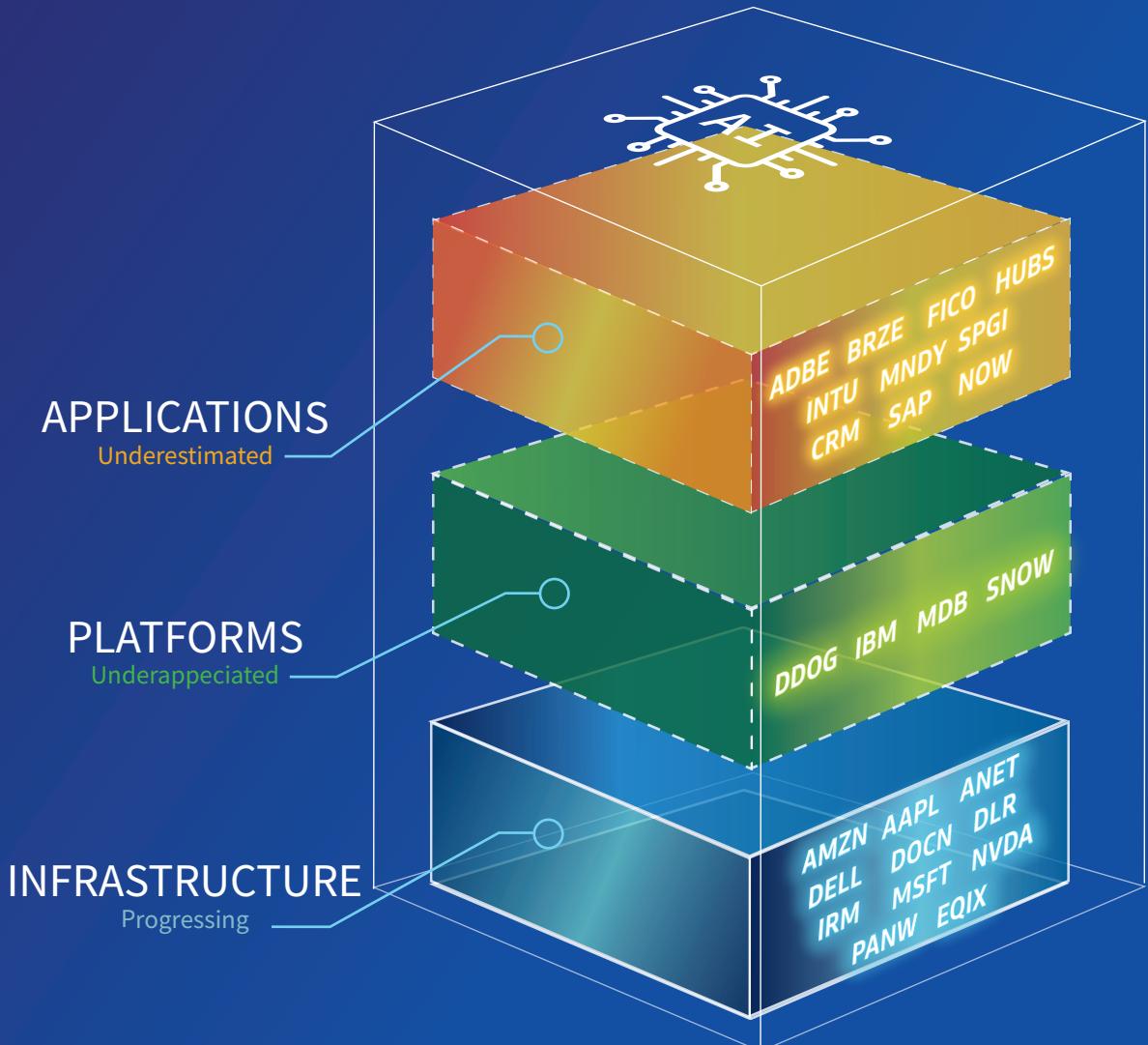
Goldman Sachs & Co. LLC

Mohammed Moawalla

+44 20 7774-1726

mohammed.moawalla@gs.com

Goldman Sachs International



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.

CONTRIBUTING AUTHORS

Kash Rangan
+1 415 249-7318
kash.rangan@gs.com
Goldman Sachs & Co. LLC

Toshiya Hari
+1 646 446-1759
toshiya.hari@gs.com
Goldman Sachs & Co. LLC

Mohammed Moawalla
+44 20 7774-1726
mohammed.moawalla@gs.com
Goldman Sachs International

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

Chris Hallam
+44 20 7552-2958
chris.hallam@gs.com
Goldman Sachs International

Bonnie Herzog
+1 212 902-0490
bonnie.herzog@gs.com
Goldman Sachs & Co. LLC

Jacob Staffel
+1 713 276-3526
jacob.staffel@gs.com
Goldman Sachs & Co. LLC

Ben Miller
+1 917 343-8674
benjamin.miller@gs.com
Goldman Sachs & Co. LLC

Patty Kanada, CFA
+1 212 934-4218
patty.kanada@gs.com
Goldman Sachs & Co. LLC

Max Gamperl
+1 415 249-7311
max.gamperl@gs.com
Goldman Sachs & Co. LLC

Greyson Sklba
+1 212 357-2534
greyson.sklba@gs.com
Goldman Sachs & Co. LLC

Tommie Reerink
+1 212 357-2470
tommie.reerink@gs.com
Goldman Sachs & Co. LLC

Karishma Raghuram
+1 212 357-6408
karishma.raghuram@gs.com
Goldman Sachs & Co. LLC

Gili Naftalovich
+1 917 343-4286
gili.naftalovich@gs.com
Goldman Sachs & Co. LLC

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

Adam Hotchkiss
+1 212 902-3941
adam.hotchkiss@gs.com
Goldman Sachs & Co. LLC

Mark Delaney, CFA
+1 212 357-0535
mark.delaney@gs.com
Goldman Sachs & Co. LLC

Salveen Richter, CFA
+1 212 934-4204
salveen.richter@gs.com
Goldman Sachs & Co. LLC

Kate McShane, CFA
+1 212 902-6740
kate.mcshane@gs.com
Goldman Sachs & Co. LLC

Henry Dane
+1 212 357-6378
henry.dane@gs.com
Goldman Sachs & Co. LLC

Alex Vegliante, CFA
+1 212 934-1878
alex.vegliante@gs.com
Goldman Sachs & Co. LLC

Pierre Riopel
+1 212 934-4505
pierre.riopel@gs.com
Goldman Sachs & Co. LLC

Carolyn Valenti
+1 212 357-2790
carolyn.valenti@gs.com
Goldman Sachs & Co. LLC

Morgan Leung
+1 212 934-4683
morgan.leung@gs.com
Goldman Sachs & Co. LLC

Michael Smith
+1 212 357-2136
michael.s.smith@gs.com
Goldman Sachs & Co. LLC

Uzair Merchant
+44 20 7774-7645
uzair.merchant@gs.com
Goldman Sachs International

Matthew Martino
+1 212 902-0695
matt.martino@gs.com
Goldman Sachs & Co. LLC

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

Michael Ng, CFA
+1 212 902-8618
michael.ng@gs.com
Goldman Sachs & Co. LLC

George K. Tong, CFA
+1 415 249-7421
george.tong@gs.com
Goldman Sachs & Co. LLC

Chris Shibutani, M.D.
+1 617 204-2437
chris.shibutani@gs.com
Goldman Sachs & Co. LLC

Carly Davenport
+1 212 357-1914
carly.davenport@gs.com
Goldman Sachs & Co. LLC

Nishad Patwardhan
+1 332 245-7726
nishad.patwardhan@gs.com
Goldman Sachs India SPL

Joshua M. Frantz, CFA
+1 917 343-4384
joshua.frantz@gs.com
Goldman Sachs & Co. LLC

Anmol Makkar
+1 212 934-5814
anmol.makkar@gs.com
Goldman Sachs India SPL

Anna Wu
+1 415 249-7235
anna.wu@gs.com
Goldman Sachs & Co. LLC

Will Bryant
+1 212 934-4705
will.bryant@gs.com
Goldman Sachs & Co. LLC

Zorayda Montemayor
+1 212 357-6403
zorayda.montemayor@gs.com
Goldman Sachs & Co. LLC

Table of Contents

PM Prelude	4
Adoption cycle expectations	12
Framework for investing in this new paradigm	23
Top Picks - Companies best positioned to benefit	26
Evolution of the tech stack: From compute to cognition	40
Promise of the Platform layer highly underappreciated today	52
Relevance of Application layer most debated and underestimated	70
Infrastructure layer the furthest along in tilt towards Gen-AI	97
Formation of ecosystem to support broadening out of use cases	120
Valuation and Key Risks	142
Disclosure Appendix	150

PM Prelude

Gen-AI could catalyze Cloud revenues of \$2 trillion in 2030. Investor sentiment has swung nearly 180° from the onset of the excitement around Generative AI back in early-2023. Back then it was widely posited that this new evolution would be massively disruptive. Now, the questions are: “What happened?”, “Why are we not seeing more meaningful changes?”, and “Where is the disruption?” Why are we not seeing meaningful impact to Software despite the massive Capital spending cycle undertaken by the Hyperscalers? Is Gen-AI a culprit in that it has taken up bandwidth in slowly-growing and crowded IT budgets? Is all this at the expense of Software industry growth, which is set to decelerate for the 3rd consecutive year? Or could it finally be a catalyst for Software growth and re-acceleration? The success and leadership positions of Microsoft and Nvidia at the Infrastructure layer of the Generative AI stack are well-deserved and well-appreciated. However, Application software companies in particular (CRM, ADBE, INTU, NOW, TEAM and WDAY, amongst others), and Platform software companies to a lesser extent (SNOW, DDOG and MDB, among the larger ones), not only have a narrow Generative AI valuation premium embedded but are valued at varied levels of meaningful discounts to intrinsic value, in our view. We believe Application and Platform software companies will ultimately be able to garner incremental value as Generative AI progresses through the IPA framework - that is, moving from the Infrastructure layer to the Platform layer and subsequently, to the Application layer. As a result we estimate Cloud revenues across IaaS, PaaS and SaaS will reach \$2 Trillion by 2030 (GSe) – With Gen-AI accounting for 10-15% of this spend. In keeping with prior thematic reports, we also highlight our intrinsic value analysis in Exhibit 1, where we attempt to determine the EV/S multiple of a software company in a zero-growth environment (no net new growth) - akin to a bond with a fixed yield.

Exhibit 1: Intrinsic Value Analysis

Ticker	Rating	Price				Zero-Growth Scenario - 3% WACC				Zero-Growth Scenario - 4% WACC				Zero-Growth Scenario - 5% WACC			
		Current	Target	U/D	Revenue	EV / Sales	Share Price	Maintenance	Growth	EV / Sales	Share Price	Maintenance	Growth	EV / Sales	Share Price	Maintenance	Growth
ADBE	Buy	\$558	\$640	15%	\$18.89	15.5x	\$706	127%	-27%	11.6x	\$531	95%	5%	9.3x	\$426	76%	24%
ADSK	Sell	\$255	\$225	-12%	\$5.35	13.7x	\$364	144%	-44%	10.3x	\$275	108%	-8%	8.2x	\$221	86%	14%
ALTR	Sell	\$91	\$71	-22%	\$0.60	10.3x	\$76	83%	17%	7.7x	\$58	62%	38%	6.2x	\$48	50%	50%
BASE	Sell	\$18	\$18	-1%	\$0.17	10.7x	\$44	273%	-173%	8.0x	\$34	205%	-105%	6.4x	\$28	164%	-64%
BSY	Neutral	\$50	\$50	-1%	\$1.20	14.2x	\$51	101%	-1%	10.6x	\$37	76%	24%	8.5x	\$29	61%	39%
CFLT	Neutral	\$21	\$28	31%	\$0.73	12.7x	\$29	132%	-32%	9.5x	\$21	99%	1%	7.6x	\$16	79%	21%
CRM	Buy	\$265	\$315	19%	\$33.95	14.2x	\$516	195%	-95%	10.6x	\$387	146%	-46%	8.5x	\$310	117%	-17%
DDOG	Buy	\$117	\$150	28%	\$2.01	14.4x	\$98	83%	17%	10.8x	\$74	62%	38%	8.6x	\$59	50%	50%
DT	Buy	\$50	\$62	23%	\$1.30	13.2x	\$69	139%	-39%	9.9x	\$52	105%	-5%	7.9x	\$43	84%	16%
ESTC	Neutral	\$106	\$106	0%	\$1.16	12.8x	\$153	145%	-45%	9.6x	\$115	109%	-9%	7.7x	\$92	87%	13%
GTLB	Buy	\$47	\$80	71%	\$0.54	13.8x	\$57	122%	-22%	10.3x	\$43	92%	8%	8.3x	\$35	73%	27%
INFA	Buy	\$24	\$38	56%	\$1.55	12.0x	\$59	225%	-125%	9.0x	\$44	169%	-69%	7.2x	\$34	135%	-35%
INTU	Buy	\$621	\$765	23%	\$14.75	9.8x	\$561	91%	9%	7.4x	\$419	68%	32%	5.9x	\$333	54%	46%
IOT	Buy	\$42	\$44	5%	\$0.85	14.8x	\$27	64%	36%	11.1x	\$20	48%	52%	8.9x	\$16	38%	62%
MDB	Buy	\$247	\$325	32%	\$1.59	13.1x	\$274	111%	-11%	9.8x	\$205	83%	17%	7.9x	\$163	66%	34%
MNDY	Buy	\$264	\$340	29%	\$0.68	15.7x	\$278	106%	-6%	11.7x	\$215	80%	20%	9.4x	\$177	64%	36%
MSFT	Buy	\$414	\$515	24%	\$218.31	13.5x	\$438	105%	-5%	10.1x	\$328	79%	21%	8.1x	\$261	63%	37%
NOW	Buy	\$828	\$940	13%	\$8.47	15.6x	\$719	87%	13%	11.7x	\$541	65%	35%	9.4x	\$434	52%	48%
ORCL	Neutral	\$139	\$137	-1%	\$0.96	14.1x	\$237	159%	-59%	10.6x	\$171	119%	-19%	8.5x	\$131	95%	5%
PCOR	Buy	\$62	\$73	18%	\$0.89	13.8x	\$98	162%	-62%	10.3x	\$74	122%	-22%	8.3x	\$60	97%	3%
PD	Neutral	\$20	\$24	20%	\$0.89	11.6x	\$52	253%	-153%	8.7x	\$39	190%	-90%	6.9x	\$31	152%	-52%
RBRK	Buy	\$35	\$48	37%	\$0.89	12.4x	\$135	416%	-316%	9.3x	\$102	312%	-212%	7.4x	\$82	249%	-149%
RNG	Neutral	\$34	\$34	1%	\$2.16	7.5x	\$169	384%	-284%	5.7x	\$123	287%	-187%	4.5x	\$96	230%	-130%
SNOW	Buy	\$115	\$220	91%	\$2.62	12.3x	\$114	98%	2%	9.3x	\$86	74%	26%	7.4x	\$70	59%	41%
SWI	Sell	\$13	\$12.50	-1%	\$0.75	10.8x	\$43	272%	-172%	8.1x	\$31	204%	-104%	6.5x	\$24	163%	-63%
TEAM	Buy	\$160	\$230	44%	\$3.71	13.2x	\$214	135%	-35%	9.9x	\$162	101%	-1%	7.9x	\$130	81%	19%
TWLO	Neutral	\$61	\$67	10%	\$4.10	7.0x	\$172	279%	-179%	5.3x	\$128	209%	-109%	4.2x	\$102	167%	-67%
U	Neutral	\$18	\$22.50	26%	\$2.03	11.3x	\$47	246%	-146%	8.5x	\$35	184%	-84%	6.8x	\$27	147%	-47%
WDAY	Buy	\$260	\$300	15%	\$6.98	13.1x	\$380	145%	-45%	9.8x	\$284	109%	-9%	7.9x	\$226	87%	13%
WEAV	Neutral	\$11	\$13	19%	\$0.16	8.5x	\$23	219%	-119%	6.4x	\$17	164%	-64%	5.1x	\$14	131%	-31%
ZM	Neutral	\$71	\$72	2%	\$4.50	12.0x	\$199	380%	-280%	9.0x	\$155	285%	-185%	7.2x	\$129	227%	-127%
Average					12.5x		177%	-7%	9.4x		133%	-33%	7.5x		106%	-6%	

Price targets are for a 12-month time frame, Priced as of market close on 08/23/2024.

Source: Goldman Sachs Global Investment Research, Company Data, FactSet

GS AI framework for Software stocks highlights MSFT, NOW, CRM, ADBE, INTU,

SNOW, DDOG, MNDY, and MDB. MSFT is best positioned, but we expect the investable universe in Software to broaden. When evaluating which companies can successfully evolve from the cloud era to the age of AI, we identified five key characteristics: **1) Large TAM opportunity** **2) Data quality** **3) Platform agility** **4) Product evolution** and **5) Capital allocation.** Overall, we like MSFT longer term as near-term headwinds to GMs from an elevated Capital spending cycle make way for operating leverage in the years ahead. Near term as Gen-AI moves into the Platform phase, we re-iterate our Buys on SNOW, MDB and DDOG as we see potential for upside to 2H24 estimates for cRPO, NER and revenue growth. Intermediate term and even longer term as Gen-AI moves into the Application phase in C25 and beyond, we retain high conviction on our Buy ratings on CRM, NOW, ADBE, and MNDY as new products start to mature, thereby driving upside to cRPO, net new ARR and NER expectations. If these metrics materialize, we could finally start to see the re-acceleration of revenue growth in 2H25 and into C26. We see WDAY, ESTC, DT and GTLB as showing early indications of being on the same growth trajectory.

Exhibit 2: Our top software picks through the lens of GS Five Factor Framework for AI-era winners

	IaaS	PaaS					SaaS			
	MSFT	MDB	DDOG	SNOW	NOW	salesforce	CRM	ADBE	MNDY	INTU
 Large TAM opportunity	\$2 trillion Cloud market by 2030, \$135bn+ Gen-AI TAM	\$120bn+ DBMS TAM	\$50bn+ Observability Market	\$340bn+ platform TAM in CY28	\$275bn TAM in CY26	\$128bn TAM	\$293bn TAM by 2030, \$4bn+ Gen-AI TAM	\$100bn TAM by 2026		>\$300bn TAM
 Data quality	Over 400mn commercial, paid users on O365	NoSQL database ensures quality transactional business data	Exabytes of mission-critical observability data from all three layers of tech stack	Exabytes of data under management	System of action sitting on top of orgs existing data and systems	Large concentration of enterprise data, Mulesoft integration advantage	>40mn paid Creative Cloud subs, +9bn images created with Firefly	Adding proprietary data across work management as well as emerging products (i.e.: CRM)		Wide array of data points, with over 500K attribution per small business, & years of high-quality data from its ~100mn customers
 Platform agility	Gen-AI investments going toward unified infrastructure	Document model flexible enough to accommodate various data types, new services	Platform-first; unified data model supports accelerated development	Unified platform	CMDB asset repository; Common data model drives engineering velocity	In-house integration tools + investing toward common data architecture	Unified data platform in Creative Cloud + Re-architected Digital Experience platform	Platform built on common low-code/no-code base		Platform built on a unified data architecture; acquisitions strongly integrated
 Product evolution	GitHub Copilot, Azure AI Services, Office 365 Copilot, LinkedIn Copilot, Dynamics Copilot	Introduction of Atlas Vector Search	LLM Observability, Bits AI, Watchdog AI	Cortex AI, Snowpark Container Services, Unistore	Now Assist, Pro+ SKUs	Data Cloud Einstein Copilot	Firefly, GenAI Studio	Internal efficiencies driving product roadmap with continuous product rollouts		Incorporation of Intuit Assist across the product suite
 Capital allocation	>\$75bn in cash + equivalents, elevated CapEx cycle	CapEx light business model	25%+ FCFM, CapEx light business model	25%+ FCFM, CapEx light business model	30%+ FCFM, CapEx light business model	Re-allocation of capital toward Gen-AI, CapEx light business model	40%+ OM, CapEx light business model	Re-allocation of capital toward Gen-AI, CapEx light business model		~40% OpM, CapEx light business model

Source: Goldman Sachs Global Investment Research

Multiple growth catalysts ahead for the industry. After three consecutive years of progressive growth deceleration (average growth across our coverage progressing from 29% to 18% to 13% in C22/23/24 GSe), the stage is set for potentially re-accelerating growth based on three factors; **1) Interest rate reductions on the horizon.** This could lower the hurdle rate for several IT projects, in admittedly mature but still relevant categories such as CRM, HCM, ERP, E-commerce, Databases and Analytics. The dynamic could also increase buyer confidence among SMBs that have been most affected by high cost of capital and stringent borrowing conditions; **2) More certainty about economic policies as we get past US elections,** which have been an overhang on

spending and **3)** More tangible progress with Gen-AI products as the Fall season opens the doors on several important Software user conferences such as Oracle Cloud World (September), Salesforce Dreamforce (September), Workday Rising (September) and Adobe MAX (October). These events could prove to be positive catalysts at a time of uncertainty around Gen-AI progress beyond the Infrastructure layer.

Findings from our IT Buyer survey present a picture of resilience for Gen-AI initiatives.

IT decision makers responding to our survey indicate that they expect 9% of their IT budgets to be potentially allocated to Gen-AI in 3 years, up from 7% indicated in our earlier January 2024 survey. Intentions for those planning on adopting Microsoft Copilot edged up from 13% in our January 2024 survey to 15%. Finally, while not a material uptick, the percent of Cloud workloads expected to be allocated to Gen-AI edged up 40bps to 10.4% versus January 2024.

Going through transition points like this can be very unnerving. Killer applications, like Salesforce (that drove cloud computing), took time to achieve mainstream status. We are mindful of the possibility, albeit very low, that Generative AI may not achieve its full potential. In that case, Software companies could get revalued higher as there will be less contention for IT budgets, the displacement risk is alleviated, and we go back to where we were prior to the pandemic, which is not a bad place to be. There is a much greater probability that the Generative AI opportunity is indeed real and that Software Applications and Platform companies are able to re-invent and therefore re-accelerate growth - especially as interest rates start to come down ([GS Macroeconomics Research](#)). This is the view we are taking. Either way, with or without Generative AI, Software Platforms and Applications companies are in a very good position to deliver attractive returns for investors over the next several years.

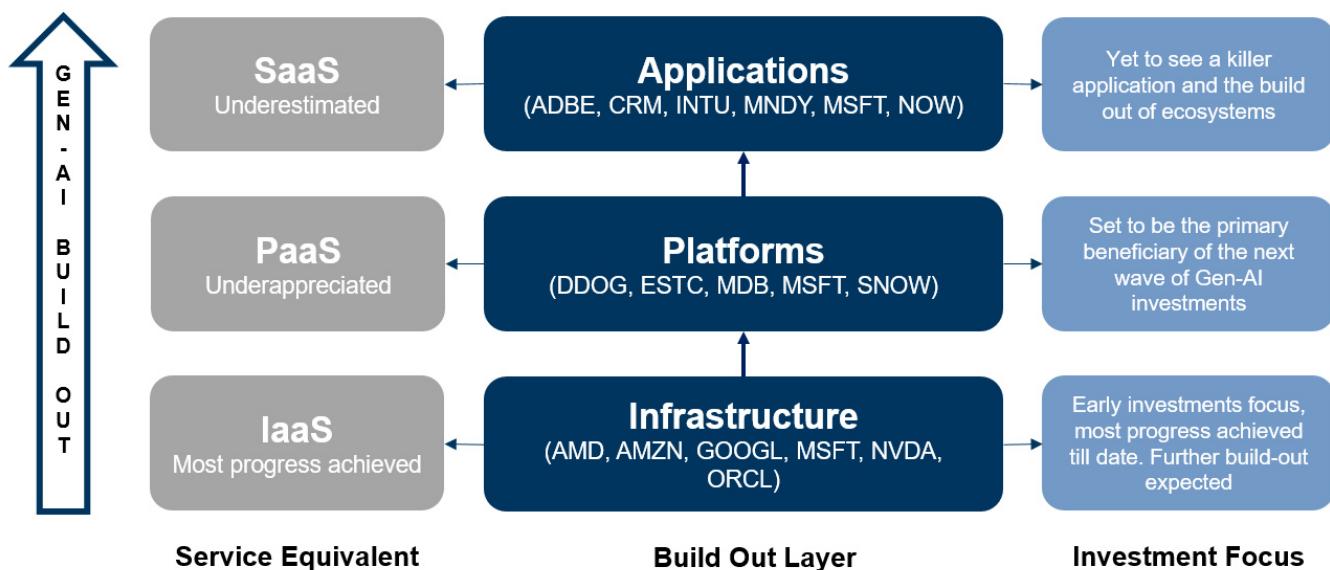
Gen-AI ROI hard to quantify – as was with Cloud in the early stages. Quantifying near term return on investment for a technology that has potentially profound long-term implications can be very hard. We reflect on the 2013-2016 time-frame when Microsoft was spending aggressively on CapEx to build out Azure. At one point, gross margins for Azure were negative. That made a reasonable investor wonder if this was all worth it. Was Microsoft playing defense or offense? It's easy to see in hindsight that Microsoft was playing offense but that clarity was not as obvious then. When there is a generational shift in technology with profound long-term implications, it would be futile to pass judgment based on near-term cost and return economics. Just as the Cloud revitalized the growth of the Software industry, longer term we believe Generative AI will revitalize the growth of the Software industry yet again.

Maturity path of Gen-AI applications could be similar to Cloud. We recall a decade ago when cloud-based apps were not nearly as robust as on-premise applications. The functionality was simply not at parity. However, what we have today is a completely different picture: cloud applications have overtaken the functionality of leading on-premise applications. Similarly, Generative AI applications and platforms that may look less mature today could be poised to become increasingly mature/mainstream over the next several years. This is predicated on the underlying Infrastructure doing its job and costs coming down, which is what we believe will occur. Once the training of models has been done accurately enough, the inference cycle (when decisions are

made) will pave the way for a software applications cycle. Currently, we estimate 25% (and rising) of Gen-AI use cases across the installed base of NVIDIA Accelerated Computing and a majority of Microsoft Azure AI Services revenue stemming from inferencing. Admittedly, despite these stats, we are not yet seeing the vibrance of the applications ecosystem. Even so, we are in a better position today than a year ago.

Traverse path of Gen-AI stack has similarities to Cloud. Although CRM SaaS Applications from Salesforce ultimately emerged as the killer apps, that was evident only in the rear view mirror. Salesforce paved the way with its Sales and Service applications. Then came the unboxing of these applications, which exposed the underlying platform for broader commercialization. Customers could build their own applications. Then, finally, as applications and platforms needed to scale, pervasive and highly scalable Infrastructure had to be built out, leading to the emergence of AWS, Azure, GCP and subsequently Oracle. To some extent, the Generative AI cycle is progressing along a comparable path where one could argue that Chatbots are the Apps that catalyzed the cycle. We are starting off with infrastructure and that should lead to growth in platforms which can help manage all the data and facilitate processing by applications. Although some apps have emerged as the first examples of scale applications – such as Microsoft Office Copilot, GitHub Copilot, GitLab Duo, Adobe Creative Applications, Salesforce Einstein Copilot, and ServiceNow Pro+ (to a lesser extent) - they are early in achieving mainstream status.

Exhibit 3: I-P-A Framework for the Gen-AI build-out



Source: Goldman Sachs Global Investment Research

We believe value will accrue to Software, Platforms, and finally Applications. The last 18 months or so have been mostly about the Infrastructure build out. This has resulted in a declaration of sorts that this cycle will be all about hardware and that software will be commoditized. We do not believe so. There was a similar temptation at the advent of the Cloud cycle to conclude as such when AWS created a massively scalable Infrastructure layer at first and then Platforms, including databases and

development tools. This led many investors to conclude that AWS would commoditize Software and specifically, the much-vaunted Application layer. In fact, multiple cloud-based databases and Platforms rose to prominence, such as Microsoft Cosmos DB, AWS RDS, Snowflake, Databricks and MongoDB, to name a few. Furthermore, the build out of AWS Infrastructure facilitated a whole host of SaaS Applications, just simply too numerous to list here. All of this was made possible due to the rollout of the underlying cloud Infrastructure. Similarly, we believe that the build out of Gen-AI infrastructure (evidenced by the Hyperscalers' CapEx ramp) could lead to the revamping of existing applications and birth of new Gen-AI applications.

CapEx intensity could be alleviated over time. The significant step-up in CapEx spending - expected to reach \$215bn in CY24 (GSe) across AMZN's AWS, MSFT, GOOGL, META and ORCL, vs \$133bn and \$125bn in CY23/CY22, respectively - has rightfully raised many questions over whether high compute costs will impede the mainstreaming of Gen-AI. After all, the compute costs of Gen-AI training are somewhat bound by the model parameters (number of attributes, size of datasets, etc.) stemming from the core Transformer architecture. There is not much getting around it so far. The impact to Microsoft's gross margin is expected to be significant at ~200 bps in F25 (GSe) and could see further declines, if not similar magnitude. Our analysis concludes that Microsoft Gen-AI revenues (\$5-6 billion annualized) have scaled rather swiftly relative to Azure, which took ~7 years to get to comparable levels. Though CapEx intensity is up sharply overall, it is still roughly where we were in the Azure cycle at comparable revenues. At a time when it felt that cloud compute costs were increasing at a concerning rate, technologies like virtualization, serverless, containers, Kubernetes, orchestration, cooling, etc. provided much needed relief.

Compute cost efficiencies could improve. Although hard to quantify and predict the exact timing, we identify a number of trends that may emerge and reduce Gen-AI's elevated compute costs: **1)** Rising competition among LLMs to produce better models that extract more efficiencies from hardware, **2)** Emergence of medium and small language models where compute costs scale inversely at a disproportionate rate versus model size, **3)** Vertical domain-specific models, **4)** Price performance improvements expected in GPUs, **5)** New sources of competition in hardware (example being Groq's LPUs), **6)** Shift in AI from learning to inferencing, the latter being several orders of magnitude less compute-intensive, **7)** Improvements in networking interconnect and finally, **8)** Shift in the Gen-AI activity from infrastructure (training) to platforms and onto applications (inference).

Better to look at incremental Gross Profit dollars vs Gross Margin percentage. We also believe it's important to look at incremental gross profit dollars from Gen-AI and not just impact to Gross Margin percentage. When cloud computing drove a material increase in the buildup of compute infrastructure some 10-12 years ago, investors grew concerned that the 85-90% gross margin structure of incumbent software companies would get impacted, which did happen. However, the cloud was largely additive to gross profit dollars despite low gross margin percentages. This was true even for large incumbents such as MSFT, ORCL, SAP, ADBE, INTU, etc. While the cloud transition could be characterized as replacement of on-premise compute for cloud compute (that

was still powerful enough to increase the TAM), Gen-AI is less about substitution of compute and more about augmenting productivity through compute. That's a new source of gross profit dollars, albeit at lower margins today (but incremental nonetheless).

Gen-AI likely to drive architectural change to the Software stack. Just as the cloud reshaped the Software industry and, in the process, drove significant architectural changes to every layer of the stack, we believe Generative AI is poised to drive significant change. The impact to the Infrastructure layer is already very well understood. What is not yet clearly defined and has yet to play out is the impact to the Platform and Application layers. We do believe that big changes are likely to happen. But at this stage, it is very difficult to be specific. We can expect architectural changes to the Platform layer if the data being collected is to be useful to Applications sitting on top. The so-called middleware layer which brings together the data and allows Applications to work freely is still being defined. The middleware layer played a significant role in on-premise applications and a somewhat significant role in Cloud Applications. The data layer looks like an important place to be in the evolution of the Platform layer. This is where Vector Search, Vector Databases and Retrieval-augmented generation (RAG) technologies look to be important enablers.

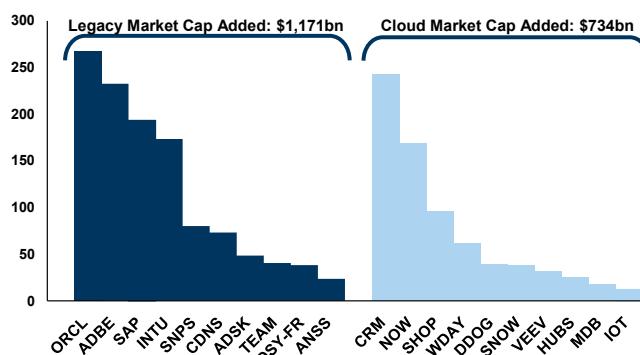
Cloud Apps potentially synergistic with Gen-AI. Enterprise Applications have always been about data, processes, execution, and analysis. That will not change with Generative AI. What will change is how this new cognition cycle perpetrated by Generative AI will offer new engagement avenues for end users to unlock value from software and offer an ability to triage problems in ways that could not have been done before. If we pose a very thoughtful question to a chatbot that sits alongside a classic application, it will give a well-informed answer. That does not circumvent the need for traditional Applications. But rather, it augments it. That is why incumbent software Applications and Platforms companies have a good opportunity to benefit from Generative AI. However, they cannot take their data incumbency for granted. The data has to be easily accessible in order for training models to work effectively. At the same time, a new opportunity is opening up for net new Software ideas which can be completely built on a new Generative AI stack. There is a risk to the incumbency of Software Applications, if new ideas and products can meld their way into end-user workflows and make themselves very fluid and seamless, thereby generating interactions. This can then provide meaningful basis for training, which could be disruptive. This can then provide meaningful basis for more established app companies to provide the training of these models, which could be disruptive to their value proposition longer term. This would represent a more bearish scenario for incumbents since training is a less exciting category to be relegated as it may lower the prospects of differentiation. Unless the incumbent Software company is able to execute on Generative AI and make bots/Copilots seamlessly work with their existing silos.

Incumbents created significant shareholder value throughout the Cloud cycle. We saw staggering scale of investor value creation among incumbents such as Microsoft, Adobe, Intuit, Autodesk, Oracle, and SAP over the past 12-14 years. Although it all happened at different points in time, there was a lot of aggregate value creation (\$1.2

trillion) versus Cloud-native Software companies (\$750 billion) since the Great Financial Crisis. The incumbents reinvented their products and business models. Let's not forget that in this current cycle, the incumbents are in a much stronger position than they were at the start of the cloud cycle. For one, many of the capital investments are being made by the incumbents. They have distribution as well. Though we acknowledge that is not a guarantee for success since the product still has to be good. With increasingly compelling products that can be packaged into their existing product offerings, incumbent software companies have a strong position to benefit in this cycle while creating net new opportunities for ideas and businesses.

Exhibit 4: Cloud cycle suggests there is room for significant market value accretion for both legacy and upstart companies

Market cap added by ten largest legacy software names vs. ten largest cloud natives (2010-Today)

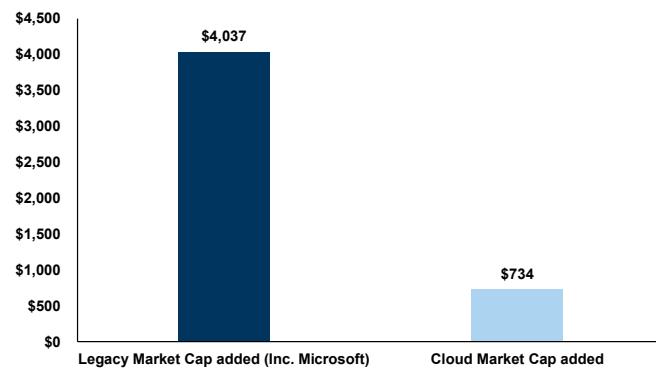


\$bn, ex. MSFT and ex. Security Software

Source: Goldman Sachs Global Investment Research, FactSet

Exhibit 6: If we include MSFT in the analysis, legacy companies have added around 5.5x of market cap when compared to their cloud native peers

Market cap added by top 10 legacy names (inclusive of MSFT) and top 10 cloud names (2010-Today)

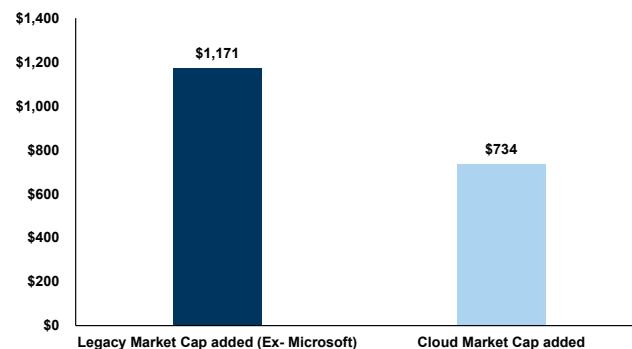


\$bn, incl. MSFT and ex. Security Software

Source: Goldman Sachs Global Investment Research, FactSet

Exhibit 5: While both added significant value, the largest legacy companies added over 1.5x the amount of market cap as cloud natives

Market cap added by ten largest legacy software names vs. ten largest cloud natives (2010-Today)



\$bn, ex. MSFT and ex. Security Software

Source: Goldman Sachs Global Investment Research, FactSet

Gen-AI infusion is likely to be gradual and support defensible and profitable growth. In our view, most Enterprise Applications will become AI enabled in the long

run. A decade ago, the big question among investors was whether a Software company had the ability to make a transition to SaaS. Today, that's not a question anymore. Most everything that has been developed in the last 10 years is cloud-based Software. The distinction between a cloud-based Software company and non-cloud is not relevant. Similarly, we think Generative AI will be broadly and deeply diffused into the Software industry by way of a myriad of products. It may not be a relevant question if Software has AI in it or not. Rather, the question is how does the industry monetize Generative AI? Clearly, IT budgets are not going to grow at the same rate as ARPU uptick implied by a lot of the Generative AI pricing premium vs. core Software functionality. Mass adoption of Generative AI over a narrow period of time is not likely given this above dynamic. We believe Software companies will start to embed more and more of the premium AI functionality into the highest end SKUs that they have to offer at a higher price. As a result, that infusion of Generative AI is likely to be graceful and gradual, thereby allowing companies to be able to create more defensible and profitable growth in the future.

Adoption cycle expectations

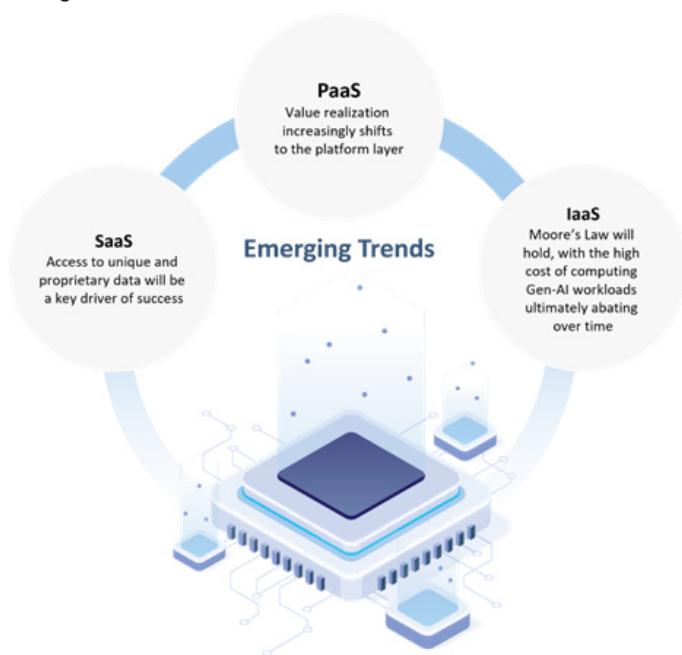
The gold rush we expected in 2024 has not yet materialized, as the adoption of Gen-AI solutions over the last year has proven sluggish versus our initial expectations. While we have only seen a number of applications that have truly offered Gen-AI native solutions, most have simply offered ways of engagement with foundational models, or extended use cases of AI, ML and natural language processing (NLP). Still, we see industry-level innovation progressing toward a Gen-AI oriented tech stack that will solidify this technology's viability over time. We expect IT spending to lean toward Gen-AI, with ~10-15% of the \$2 trillion in software spend expected in 2030 likely to be allocated to the new technology, with a broader application-layer TAM of ~400bn (to be reached at scale). The near-term constraints on IT budgets coupled with the need for more comprehensive Gen-AI solutions have limited any broad-based adoption. The large computational burdens and significant CapEx requirements also hinder companies' ability to ensure such investments can meet the elevated hurdle rates of the high interest rate environment.

While greenshoots like creative and developer tools have garnered momentum in the application layer, we expect expanded use cases to emerge on the back of: **1)** easing supply constraints for high-power computing components and their associated costs (e.g., energy), **2)** a broader ecosystem of platform solutions, and, most importantly **3)** the surfacing of a "killer use case", enabled by the fruition of a full Gen-AI tech stack.

Expect Gen-AI to begin materially contributing to sectors' growth in 2H25 as key platform components develop. While we dive into each of the three aforementioned drivers in this note, we largely see 2024 as the year when companies firm up their Gen-AI strategies across the tech stack. Therefore, we now expect Gen-AI to drive material industry growth contribution by 2H25. In 2023, we saw companies largely debating how they can securely and effectively incorporate Gen-AI, with capital allocated to the aggregation of data in hopes of building independent LLMs. We expect

such efforts to shift to smaller, more specialized multi-modal models that sit on top of the contextual understanding that larger foundational models provide. This change shifts large foundational models' role to one of providing the base level of contextual knowledge, but not serving as the core knowledge hub in which an applications output is predicated on. As the crux of the data differentiation lays within the refined proprietary data sets, we believe the following trends will emerge: **1)** companies' access to unique and proprietary data will be a key driver of success, **2)** the value realization increasingly shifts to the platform layer, where the Gen-AI tech-stack increasingly centralizes around a data procurement center and **3)** the complexity of inferencing is reduced via more defined data workflows. Whether leveraging GPUs or alternatives, we believe Moore's Law will hold, with the high cost of computing Gen-AI workloads ultimately abating over time.

Exhibit 7: As the crux of the data differentiation lays within the refined proprietary data sets, we believe the following trends will emerge:



Source: Goldman Sachs Global Investment Research

Killer application remains uncertain. Though the mix of training and inferencing is likely to skew towards inferencing over time, we note that Gen-AI needs an app that offers unique value to the end user to drive ubiquitous adoption. This can then underpin the value proposition of the technology and catalyze expansion. In the case of Gen-AI, we are still waiting for a market-ready use case that can produce results at scale for different end user domains, be it front, middle, or back office. While the pace of Gen-AI technology continues evolving, it is not quite to the point where killer applications are in clear manifestation today. What we experienced since ChatGPT's launch in November 2022 and the subsequent months was the temptation to extrapolate the initial wave of investment in infrastructure as the 'end all be all' with respect to innovation and mainstreaming. Given software platforms need to evolve in the direction of Gen-AI to sustain competitive relevance, we believe such innovation is happening and are just early in the adoption curve.

- **Microsoft's O365 Copilot may evolve to be the key Gen-AI use-case.** Based on the aforementioned criteria we see Microsoft's O365 Copilot holding promise and may very well become that use-case. Offering an extensive portfolio of apps functionality, Microsoft and O365 have an unparalleled lens into various points of organizations' workflows. This is likely to give the company an advantage as it pertains to the quality of the model output and therefore its relevance in driving productivity. The company's broad distribution channel, sitting at the crux of 40% of frontline works' workflows via its +400mn commercial install base also allows Microsoft to get broad adoption once the product has been more refined. We acknowledge that the adoption curve has taken longer to formulate than we initially expected but still recognize the fast-paced innovation in bringing this market and the pace of the updates released since the initial launch (i.e., through more specialized offerings - O365 Copilot for Finance, Sales, etc.). The evolution of the application is also likely to occur at a faster rate than the broader industry (as it fleshes out each layer of the I-P-A framework) given Microsoft's solutions across the tech-stack, adding to Copilot's advantages.

Post Infrastructure cycle, investments starting to seep into Platform layer before enabling industry-wide adoption in Application layer. With an expectation that there will be a digestion period between investments and deployment, we believe we are in Phase II of the investment cycle versus the initial phases of industry-wide deployment. We also note that tech cycles typically follow a pattern:

Infrastructure-Platform-Application. At first, the cycle is about the infrastructure buildup. Then we have platforms that allow the best use of that infrastructure while providing building blocks to construct next generation applications. The final leg is the buildup of applications. We believe that we have seen crowding in the foundational model layer of the Gen-AI tech stack, with all major cloud providers + private companies (Anthropic, Hugging Face, OpenAI, etc.) likely holding their prominent roles as this technology solidifies. While we see numerous avenues for viable application use cases to form, we believe a greater formation of the platform layer is needed before we see more material Gen-AI contributions in the SaaS layer. We discuss our view on how the new Gen-AI tech stack will shake out in later in this report.

Gen-AI - a Catalyst or Culprit?

What Gen-AI means for broader software through 2H25. While Gen-AI reinvigorated interest in software in 2023 after a year of optimization, this euphoria shifted as the broader macro remained uncertain and budget expansion remained limited. While the level of innovation around Gen-AI was a catalyst to bring organizations to the bargaining table after a period of optimizations (which we believe was evident in the better 4Q buying environment), the higher-for-longer interest rate environment thus far in 2024 has largely led Gen-AI spend to come at the expense of budgets that might have otherwise been spent on supporting software growth. We estimate that ~\$20bn will be spent on Gen-AI initiatives across both public and private peers (including Microsoft, Amazon, Alphabet, and OpenAI) that would have otherwise been allocated across our coverage. Gen-AI projects also circumvent the same scrutiny and high IRR barriers that enterprise software has to clear, given its nascence and disruption potential. Recalling the IPA

framework previously outlined, we don't see the application layer being fully formed yet, leading to hesitation around broader use-cases and shifting companies' focus to the infrastructure and platform layers in the near term.

- **Anticipate a duration gap between customer conversations, backlog and revenue contribution.** We believe that as the platform layer grows increasingly accommodating of Gen-AI needs and the killer application unlocked by Gen-AI becomes increasingly visible, this will lead the way to heightened urgency for both SaaS and PaaS solutions. In the meantime, we see Gen-AI contributing to pipeline growth, as we are starting to see with Microsoft, ServiceNow and Adobe. Therefore, we anticipate a duration gap between customer conversations, backlog and revenue contribution. What we experienced in 2023, and particularly 4Q, was the temptation to extrapolate the initial momentum related to growing AI revenues in the infrastructure layer as a signal that others across the stack, from ServiceNow, Salesforce, Adobe, etc. should also be able to call out Gen-AI specific revenues shortly thereafter. This led to a tangible effect on valuations for companies that had versus had-not moved rapidly to bring out Gen-AI products and capabilities. While this dynamic has largely dissipated as investor focus is back on overall top-line cadence given the air pocket with overall spending in 1Q, we see the market returning to this dynamic as we move through the IPA evolution – which we see being similar to that of the cloud migration.

Industry conversations spanning Healthcare to Consumer Staples and Retail+Hardlines suggest palpable engagement with Gen-AI tech; Tangible applications forming.

- **Viewing the Generative AI revolution through a healthcare lens (covered by Salveen Richter and Chris Shibutani).** Overall, we see value in Generative AI's potential to create high-quality content at a faster rate, which can increase labor productivity and drive new insights while providing cost savings. As it relates to potential use cases and advantages in healthcare, we highlight: synthetic data generation for drug development and diagnostics, generating designs for novel drugs, personalized medicine, diversity and equity in healthcare, manufacturing and supply chain efficiency, and approval and launch materials.
 - **We expect that Big Tech will continue to look to healthcare for optionality as a growth vertical and use their expertise and resources to support innovation from healthcare companies in the space.** Historically, companies such AMZN (AWS), MSFT (Azure), and GOOGL (Google Cloud), have stepped in as cloud service providers, essentially helping healthcare companies with productivity gains relative to using on-premise infrastructure for data analytics, managing unstructured data, while saving on CapEx by using computing resources as needed. In therapeutics, NVDA is involved both with software and hardware, providing supercomputers/chips that are crucial for tasks such as AI-powered imaging and analysis and foundation models. For providers, Generative AI can help automate tasks such as drafting clinical notes from patient conversations, integrating these in electronic health records and drafting insurance authorizations for productivity gains - one

player here is Microsoft's Nuance for automated clinical documentation, which incorporates GPT-4.

- **Regarding the risks,** much has been written about the risk of Generative AI models providing inaccurate information without realizing it (called "hallucinating"), even citing plausible-looking sources that do not exist to support its conclusions. **Such hallucinations could hinder uptake in life sciences given the impact on human health (e.g., making diagnoses) and, even if used as a tool, clinicians may require a clear trail of source material that the model used to arrive at its decision.** Thus, humans may need to develop efficient methods on how to reliably detect misinformation in AI, as manual verification would undermine AI's convenience/speed advantage.

■ **The use of AI across European Financial Services (covered by Chris Hallam)**

continues to be a focus in investor and corporate discussions. We estimate the potential AI adoption uplift to ROE at c.200bp (excluding the upfront investment required). We highlight: 1) The AI impact is most clearly driven by numerous modest cost gains, with the cost opportunity c.3x the revenue opportunity. 2) Gains are far from evenly distributed, as differences in bank cost and revenue make-up lead to a c.180bp spread within our coverage between those seemingly better placed to capitalize on the AI opportunity vs. those less well positioned. 3) Almost every bank is already undertaking AI initiatives, and for many this builds on long-standing endeavors in big data analytics and machine learning. 4) Given the early stage of many of these technologies, pricing mechanisms are still in their infancy and some may prove to be more or less economic over time. Investment flexibility (and budget) will therefore be key for banks to pursue the right initiatives for their business. 5) The latest AI product offerings from leading vendors focus on customer service, risk management and fraud detection. We see the greatest AI opportunity among relatively capital-light and cost-intensive business models where greater process and cost efficiencies can provide a meaningful uplift to returns. These capital-light business models are typically higher ROTE, with greater through-cycle organic capital generation and larger excess capital positions. These businesses should therefore also find it easier to self-fund AI investment; the adoption of AI across European Banks should accordingly serve to reinforce the returns of industry leaders relative to peers.

- **Across European Banks we currently see five broad use cases for Gen-AI.**
 - 1) Enhanced coding efficiency, 2) Data extraction: Synthesize data and content from large datasets or documents, 3) Chatbots, 4) Automation, 5) Human/AI collaboration
- **The focus of large technology incumbents** (KeyBank, Scotiabank, Banco de Crédito e Inversiones', HSBC, Mr. Cooper's, BNY Mellon) is largely on: **1) Customer conversations & recommendations, 2) Process enhancement & modeling, and 3) Controls and anomaly detection**
- **Financial firms, and banks in particular, (Barclays, HSBC, ABN Amro, FINRA) who are working with newer AI players are focused on leveraging**

AI in order to improve: 1) Risk management, 2) Fraud detection, and 3)

Improved services.

- **Spotlight on Generative AI Use Cases.** Within the broader field of AI technologies, the past year has undoubtedly seen a step change in Generative AI, particularly the range of use cases and potential beneficiaries. While there are certainly a large number of potential use cases for 'traditional AI' (Machine Learning, Big Data Analytics etc), we are also beginning to see some Gen-AI use cases emerging. Overall, how does this frontier technology suit Banks? On the positive side, generative large language models are currently well suited to: i) textual analysis and creating summaries, ii) content creation, and iii) knowledge mining (e.g. finding common patterns or information across diverse media). However, they are not yet well suited to: i) numerical calculation and representation, such as creating financial tables or complex risk models; ii) real-time information, given large language models are limited by content accessible up to their training date; and iii) direct interaction with customers, where new content or advice is developed by AI without any human oversight, as generative AI has the potential to hallucinate. Therefore, at this stage, we expect banks to focus in on a select number of Gen-AI use cases, rather than aiming to apply the technology as broadly as possible across the bank.

■ Consumer Staples (covered by Bonnie Herzog) companies are actively exploring how Gen-AI can help tailor customer solutions, sharpen innovation, increase speed-to-market, and ultimately drive incremental volume & revenue

- While early days, many consumer staples companies have begun to explore potential use cases for Gen-AI. This includes companies such as KO that are piloting AI-driven initiatives to improve their offerings and speed-to-market, including an AI-based price pack channel optimization tool across several markets and a separate AI-driven initiative to push personalized SKU recommendations to retailers. KO is encouraged by the YTD results and believes it is just scratching the surface of what is possible with AI. PEP has also announced plans to execute an aggressive data and AI strategy against its core capabilities, with the goal of improving their consumer insights, strengthening their forecasting capabilities, and ultimately delivering better execution/performance from the plant to the shelf – which ultimately should drive faster growth and productivity. In HPC, EL's collaboration with MSFT has allowed it to embed AI into its system to drive faster speed-to-market and local relevance, as well as improve the customization of its messaging at scale. CL expects Generative AI to be significantly disruptive in areas like innovation and content creation. To that end, CL is funding various test case initiatives to see which have the potential to be scaled, with the goal of integrating AI into CL's broader 2030 strategy. For its part, CLX recently deployed an AI-enabled digital core into all of its businesses, which uses a signal sensing technology to sense trends after combing through vast amounts of social media and marketplace data, then leverages Generative AI to create digital prototypes of innovative product ideas that get immediate consumer feedback in a virtual marketplace. While still early, CLX already finds AI has helped reduce the cycle time on innovation ideas by 50%. Overall,

consumer staples companies are broadly optimistic about the long-term potential of Generative AI and are committed to being at the forefront of AI exploration, learning and implementation.

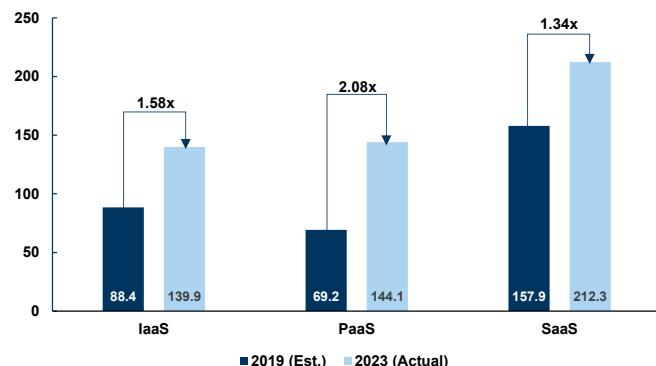
- **Key anticipated benefits and risk of deploying Gen-AI in Retail and Hardlines (covered by Kate McShane).** Companies are targeting improvement in employee productivity and cost savings in the near term and improved customer service longer term. Some companies plan to use tools for generating marketing copy for articles and website. All such generated material have manager oversight reviewed for accuracy to reduce risk around biases and errors. As companies get advanced and put up the necessary guardrails, the anticipated benefits would move to faster delivery of business insights. Companies are leveraging tech partners like Snowflake, Salesforce, Bloomreach, Microsoft, OpenAI, Google Gemini and Cordial who are enhancing their tools to allow clients to take advantage of GenAI and reduce risk exposure by limiting the dataset the AI can train from. They noted a heavy lean towards inference (pre-built model) in their Gen-AI deployments versus training (building).
 - Key risks are **1)** Proprietary information being leaked – Mitigation: Guard rails of what can be uploaded, structured data classification preventing shareability outside the organization; **2)** Incorrect information potentially discrediting reputation– Mitigation: Human monitoring; **3)** Fraudsters leveraging Gen-AI for more sophisticated phishing campaigns – Mitigation – Continuous training. For financial authorizations, leveraging Wet Signatures is one way to mitigate risks; **4)** The degree of data quality required is more critical in this space; and **5)** The need to be purposeful in investments with an eye on ROI.

Expect Gen-AI to account for 10-15% of Cloud spend in 2030

Cloud presents a \$2T TAM opportunity by 2030, bolstered by Gen-AI. We estimate that the Cloud Software TAM (IaaS, PaaS, and SaaS) could approach \$2T by 2030 (CAGR: 22%, 2024 – 2030). Our analysis suggests Gen-AI spending could constitute 10–15% of the Cloud Software market (\$200-\$300bn) as spending gradually extends beyond the Hyperscalers and Foundation Model providers to become more pervasive across all three layers of the tech stack.

- **Our methodology to arrive at our TAM approximation** was partially informed by an analysis of Gartner's Software market forecasts for IaaS, SaaS, and PaaS. Specifically, Gartner's under/over-estimation of each respective market's reported spending in a given year relative to its expectation 5-years prior (Exhibit 8). For instance, we measured Gartner's forecasting accuracy for 2023 reported spending across the IaaS, SaaS, and PaaS markets relative to their initial expectation in 2018. Our analysis found that in more recent years, Gartner's 5-year-out estimates across all three layers of the tech stack underestimated reporting spending anywhere from 10-50%+, with the PaaS layer showing the highest upward revision. We believe this helps to convey the strong value proposition that Enterprise Software provides as organizations continue to invest against digital transformation initiatives, a trend we expect to be amplified by the ubiquitous deployment and adoption of Gen-AI.

Exhibit 8: Actual spend in each layer of the tech stack supersedes 2019 forecasts by at least 34%

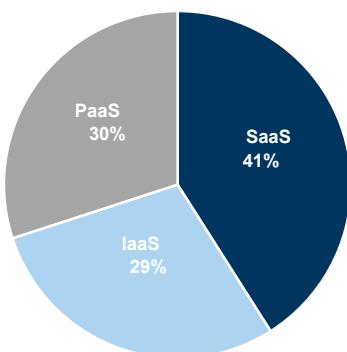


Source: Goldman Sachs Global Investment Research, Gartner

While both the growth and size of the TAM opportunity through 2030 may screen optimistic, the following informs our view: **1)** Digital transformation and cloud modernization remain durable secular tailwinds (with only ~30% of workloads moved to the cloud, based on [GS survey](#)) that can continue to fuel healthy double-digit Cloud Software growth, complemented by our expectations for a less restrictive monetary policy environment into late-2024 and beyond, **2)** The ongoing infrastructure build-out and increasing product velocity at the platform and application layers will support a broadening out of Gen-AI spending and adoption that can complement our view for a +22% Cloud Software CAGR (reaching 10-15% of total Cloud spending) and **3)** We highlight historical precedent, noting that the Cloud market increased more than 2.5x from 2019 (\$192bn) to 2023 (\$496bn), representing a +26% CAGR. We expect Gen-AI to have a comparable, if not greater, impact on the Software market relative to the cloud computing cycle.

We ascribe the following TAM figures to the various layers of the tech stack – IaaS: \$580bn, PaaS: \$600bn, and SaaS: \$780bn ([Exhibit 9](#)). This implies an equal relative weighting to IaaS and PaaS (~30% each), with SaaS garnering the remaining 40%. We also remind investors that our IPA framework implies Gen-AI tailwinds will not occur simultaneously. The infrastructure layer is poised to be the initial beneficiary, as we are already observing with the AI revenue ramp from the Hyperscalers. This should be followed by the Platform and Application layers, respectively. An inherent tethering exists between PaaS and SaaS – where PaaS solutions are needed to support the emergence of a killer application but the value in the Platform layer can't compound until more compelling applications emerge - an idea we address in detail in more detail in the application layer section of this report.

Exhibit 9: The split of cloud's 2030 TAM opportunity (~\$2tn) across the layers. We expect Gen-AI could garner 10 - 15% of the spend.

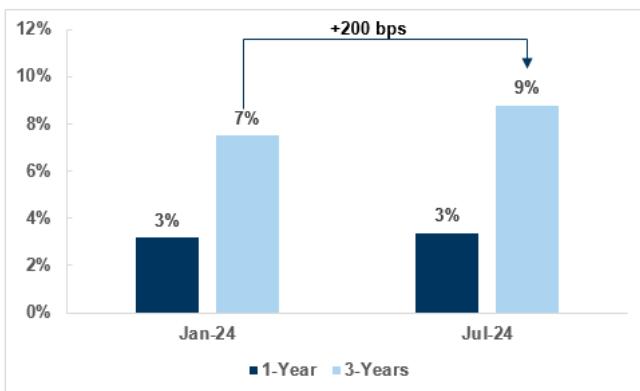


Source: Goldman Sachs Global Investment Research, Gartner

CIO survey data bodes favorably for Gen-AI spending. Using cloud migration's early days as a gauge could imply a sizable opportunity for Gen-AI. As a case study, we use our early CIO survey data showing the percentage of workloads in the cloud. Respondents indicated that in December 2014, only 4% of their workloads were in the cloud, with expectations for 9% workload presence in three years. Three years later (December 2017), workloads in the cloud had quintupled to 19%, far outpacing initial expectations (9%). We believe that this accelerated growth can be attributed to: **1)** Organizations gaining incremental comfort with the cloud and its implications (e.g. TCO reduction, faster innovation cycles, increased resiliency), and **2)** Continued investments by the Hyperscalers to advance their cloud capabilities, with other (albeit smaller) cloud providers emerging, thereby giving credence to the cloud movement. While acknowledging Gen-AI's nascent, we believe tangible proof-points exist indicating more durable future adoption, even as initial excitement has retreated over the last few months on the lack of tangible use-cases present in the market across the Platform and Application layer today. As shown in Exhibit 10, Exhibit 11, and Exhibit 12, respondents expect to continue increasing their Gen-AI exposure over the next 3-years versus the next 1-year, with the 3-year expectations showing positive revisions versus our prior survey. With catalysts for adoption likely centered on companies modernizing their data estates, more tangible use-cases emerging, and an industry-level focus on addressing key pain points (e.g. security concerns), these initial 3-year expectations could prove similarly conservative to early cloud migration expectations.

Exhibit 10: Future IT budget allocations could incorporate more Gen-AI-based spending

Question: "What percent of IT budgets could be allocated to Gen-AI in the next 1 year? 3 years?"



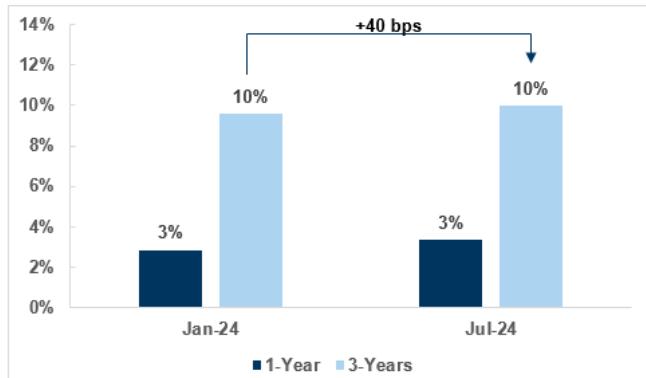
Source: Goldman Sachs Global Investment Research

Exhibit 11: Gen-AI cloud workload intentions are slightly up on a 3-year basis

Question: "What percent of your cloud workloads do you expect to be Gen-AI in 1 year? 3 years?"

Exhibit 11: Gen-AI cloud workload intentions are slightly up on a 3-year basis

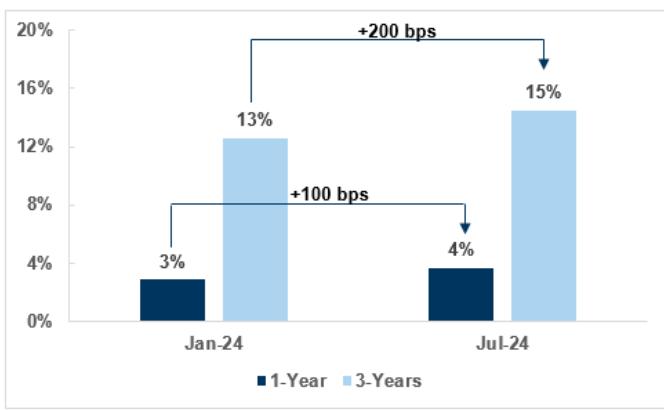
Question: "What percent of your cloud workloads do you expect to be Gen-AI in 1 year? 3 years?"



Source: Goldman Sachs Global Investment Research

Exhibit 12: Intentions around Copilot usage showed the most notable change on a 1-year, 3-year basis

Question: "If you plan on using Office Copilot what percent of the install based could be on office Copilot in the next 1 year? 3 years?"



Source: Goldman Sachs Global Investment Research

The progress of Gen-AI: In this section, we aim to offer a framework for how Gen-AI use-cases could evolve as the industry builds on the early adoption of applications like developer tools and image creation towards more ubiquitous productivity gains (enabled by the transformation of the underlying IT stack). To better understand Gen-AI's second phase of adoption, we work to understand how we moved from the proof-of-concept phase to economically solving business use-cases, and how we expect to get to greater accuracy and Gen-AI more integrated across the tech stack.

■ **Early Cycle: Focus on large foundational models and infrastructure build out.**

The first phase of the Gen-AI buildout involved most organizations focusing on proof-of-concept for the Gen-AI. The data input used by most foundational LLMs was broad-based web data. The initial version of the technology aimed to solve use-cases like coding and content creation - where there is a plethora of ready-available data. These early versions helped increase knowledge workers' efficiency. Accuracy and autonomy were not the parameters in focus, and human intervention remained necessary. At the same time, many companies focused their

investment efforts towards the infrastructure buildout for the new Gen-AI tech stack. Following the pioneers (Anthropic, Hugging Face, OpenAI, etc.), big tech started developing their own foundational LLMs and making them generally available.

- **Mid Cycle: Development of specialized models to solve business-specific problems.** As mentioned before, we believe we are currently in the second phase of the adoption cycle. This phase is marked by companies looking for business cases that can be efficiently resolved by Gen-AI. Simultaneously, many platform players will look to re-purpose their platforms to better suit and augment this new technology. Sans platforms, the applications can't be effectively built, deployed and maintained. The aim for most organizations will be to minimize human supervision of tasks and develop their own proprietary models to avoid model hallucinations for their own use cases. This will require large amounts of good quality, proprietary data along with Gen-AI friendly platforms to train and deploy these solutions. At the same time, expectations around associated costs versus material returns will need to be managed. With many industries exploring custom solutions, the definition of success will be case dependent. In this past year, we have seen the likes of MSFT, INTU, CRM launch their own Copilots for specific use-cases. This is expected to continue throughout this phase with Copilots gaining increased contextual understanding.
- **Late Cycle: Moving towards higher accuracy and Gen-AI native ecosystems.** Right now, the results produced by a model are the average of the data it is trained on. The third phase will involve moving the results from the center of the bell curve towards the right side of the curve in terms of result accuracy. The expectation from the AI models will be to augment human knowledge creation. Fine-tuning SLMs and LLMs will require high quality data and more data processing with data management systems coming in to focus. This phase might also see the formation of new software development cycles and new frameworks as the tech stack moves from computation to cognition. It will redefine productivity and change how we interact with IT systems. We will see the most advanced Copilots and Gen-AI become an integral part of various workflows for organizations.

Exhibit 13: Framework for how Gen-AI applications and investments can evolve

	Generative AI adoption		
	Early Cycle	Mid Cycle	Late Cycle
Objective	Proof of concept	Business use cases	Knowledge creation
Investment Focus	Infrastructure	Platforms/Applications	Ecosystem of Applications
Type of models	LLMs	SLMs, LLMs, Multi-modal models	Highly accuated models
Type of input data	Aggregated web data	Proprietary Data	High quality data
End-user Impact	Increase efficiency	Augment user workflows through seamless integration	Workflow processes built on Gen-AI native applications
Use cases	Developer tools, content creation	Basic copilots + Gen-AI integrated across more workflows	Advanced copilots, new application solutions

Source: Goldman Sachs Global Investment Research

Framework for investing in this new paradigm

Five factor framework to identify AI-era winners. When evaluating the companies we believe will successfully evolve from the cloud era to the age of AI, we identified five key characteristics: **1) Large TAM opportunity** **2) Data quality** **3) Platform agility** **4) Product evolution** and **5) Capital allocation.** While expecting investors to look for quantitative metrics to evaluate this opportunity, given how early we are in this cycle, we lean on more intangible and qualitative aspects that may precede more quantifiable metrics of early success. As a result, we leverage the following to frame our perspective: Knowledge we have on companies' current IT architectures, the evolution of their platform and solutions over time, recent management/company annotations around Gen-AI strategy, and industry conversations. We leverage these perspectives to expound on each of the five parameters below.

- **Large TAM opportunity.** We look for companies who we believe will be best positioned to capture IT budget share as C-suite budget allocation tilts toward Gen-AI solutions. We look for companies who can tap into categories of spend and expand their wallet share over time. This can occur by exhibiting some of the characteristics outlined in the other parameters of this framework. For companies that are well positioned (via data moats, platform agility, innovative product evolution and/or strategic investment ability) and can consistently execute, this can result in compelling multi-year, high revenue and profit growth business models. We estimate that Gen-AI presents a \$400bn incremental TAM opportunity to the global enterprise application software market (up from \$150bn estimate published in our

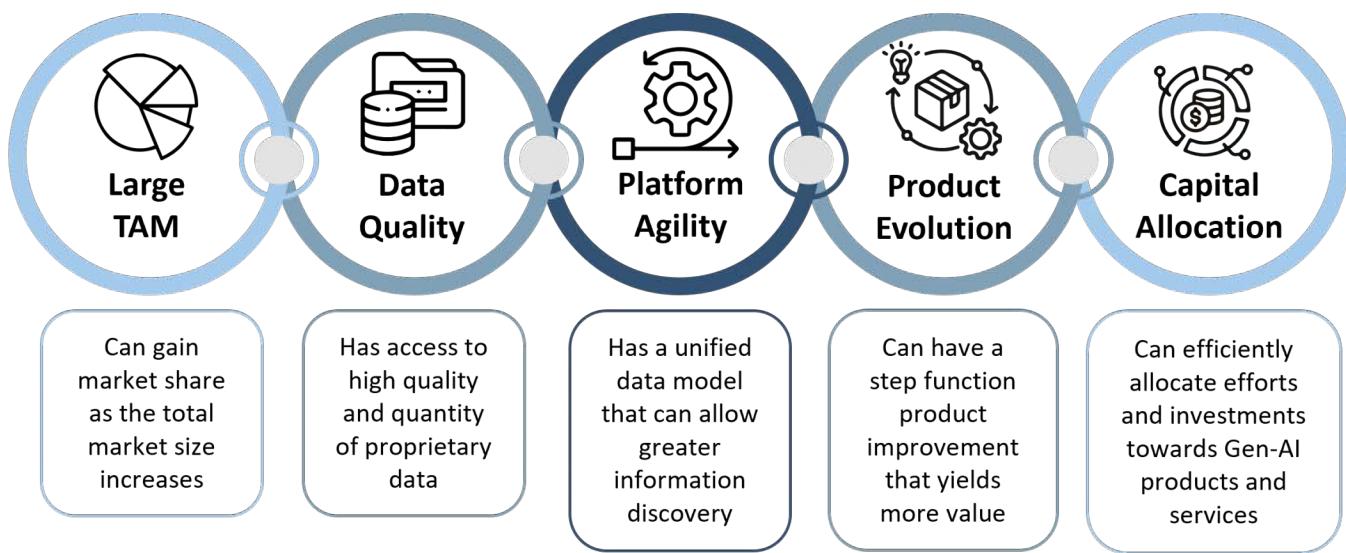
Gen-AI report from last year ([link](#)). We favor companies who can offer a platform of solutions in key categories of spend within IT budgets, (such as data management, CRM, HCM, and ERP), and whose users have direct engagement of the data through innovative use-case applications. This can likely support seat expansion, cross-selling, and upselling while building pricing power.

- **Data quality.** With data as the powerhouse that fuels LLMs, being a system of record can prove to be a key component of platforms' ability to succeed in the age of Gen-AI. Companies whose customers entrust them with their information are likely to possess access to proprietary data and a large pool of categorized data that can enrich the context and generated output used to serve their customers' needs. Although concerns arise pertaining to the platforms' rights to leverage this data, we remind investors of companies' ability to utilize data anonymously through privacy tools like masking. The curated data stored on the platform can bring AI to the forefront of end-user experiences without needing to access larger, more generalized foundation models for every inquiry. This can maintain the operating costs of these model, increase their relevant context, and therefore achieve differentiation and pricing power over time. We see such data having the potential to drive productivity as it enables users to extract value via data-driven insights and actionable tasks.
- **Platform agility.** Companies with an agile IT architecture with a unified data model are better positioned to leverage their data moat, enable the flow of data and generate more accurate, reliable and insightful responses from the data that sits within their ecosystem. Similar to how the industry is migrating from fragmented data storage architectures toward more collective data platforms, companies who best embody that philosophy internally are likely to maintain their market relevance. Some companies in our coverage (ServiceNow, Intuit, Datadog, Procore, Monday.com) have always had this philosophy while others, like Adobe, Microsoft, and Salesforce, have invested in adopting such an architecture over time. We see such a system effectively working as a centralized data procurement center that will allow for greater information discovery and be utilized for operational intelligence, risk management, product innovation, and efficiencies across knowledge workers. Years of investments in a unified, integrated data platform will enable better self-learning and cross-references without needing as heavy of an investment in compute whereas data silos are likely to add friction and cost to this process (where expenses are already high given supply constraints).
- **Product evolution.** We often gauge the level of innovation implied by companies' product roadmaps and evaluate that against the announced product releases. By discerning the level of innovation, we gain confidence in companies whose new releases gear toward transforming their value proposition over time. We favor evolution towards step-function improvements that can yield value creation versus incremental product enhancements that can quickly become table stakes - especially in an S-curve shift. As we move through this adoption curve, Gen-AI native use-cases that can fundamentally improve users' workflow and business decisions are of higher value than simpler Gen-AI add-ons (like API plug-ins with an NLP interface). Effectively, we are looking for innovators over implementers. Product

releases from Microsoft, Salesforce, Datadog, Snowflake and GitLab over the course of the last 12 months exhibited these companies' progress towards leaning into Gen-AI platforms.

- **Capital Allocation.** Keeping in mind the various technical investments required for Gen-AI, from capacity, data, architectural and UX, and in light of the macro cycle, with interest rates higher for longer, companies who can dedicate efforts and resources to Gen-AI are likely to move faster toward a Gen-AI strategy than those with strategies or initiatives that deviate resources elsewhere. In our view, management teams that limit dispersion across various initiatives and prioritize their AI product roadmap are likely to move faster, find product-market fit and therefore, be able to fine-tune their pricing model and capture both mind and wallet share at an accelerated pace. While we recognize Gen-AI's evolution still carries risks, we don't evaluate companies' path forward solely on this metric but rather as a complement to the other four factors laid out above. For example, we remain constructive on Adobe's push to add new functionality across both its Creator, Document and Marketing solutions. We see this complementing their down-market motion and believe their data moat, platform common data architecture, and rapid product innovation all complement these efforts.

Exhibit 14: GS' Five Factor Investment Framework for Gen-AI era winners



Source: Goldman Sachs Global Investment Research

Emerging competitors likely to have more complex criteria. We acknowledge that new entrants who provide native Gen-AI platform solutions are likely to emerge - like Anthropic and OpenAI. We expect such companies will be evaluated under a different framework. Given the nascence of Gen-AI adoption, these companies are fundamentally in a different stage of the maturity curve. We expect the framework that investors apply to these companies and opportunities to be a combination of both our 5-factor Gen-AI framework and 8-factor framework for software (Exhibit 15). Some of the most likely evaluated parameters are: **1)** understanding of one's value offering, **2)** the budget category it will tap into, **3)** business model evaluation (both revenue growth and path to profitability), and **4)** end-market exposure (SMB, enterprise, etc).

Exhibit 15: GS Framework for Investing in Software

GS FRAMEWORK FOR INVESTING IN SOFTWARE		
8 KEY FACTORS	TANGIBLE	INTANGIBLE
TAM	Top-down sizing Bottom-up sizing	Is pricing sustainable? Is it a feature or a real market?
SECULAR THEMES	Customers spending more on DX, UCaaS, CPaaS, CCaaS AI and IoT initial use cases	Is it a pull-forward of demand or a permanent shift? Is it a feature or a real secular tailwind?
ENTRY/EXIT POINTS	Technical Analysis Sector historical valuation parameters	When to buy? Rotation or short-term mis-execution? When to sell? Valuation ahead of fundamentals?
LT FRAMEWORK RETURN	Long-term revenue/margins scenarios Compounded return outcomes based on long-term scenarios	Which scenario is more likely? What are possible risk-reward scenarios?
UNIT ECONOMICS	Lifetime value (LTV) calculation Customer acquisition cost (CAC) calculation, LTV/CAC	How does competition impact customer lifetime? How does upsell and cross-sell impact CAC and cost to serve?
PLATFORM/ BEST OF BREED	Diversified revenue mix Best of breed gaining market share	Are there actual synergies between revenue streams? Can the best of breed "cross the chasm"?
COMPETITIVE MOAT	Revenue per R&D dollar Size of install base	Are there disruptive technologies? Why is the install base demanding bigger price discounts?
ESG	Carbon neutral = data center usage + carbon offsets Company becoming a large platform	What are the social impacts of AI products? What are the regulatory or security breach risks?

Source: Goldman Sachs Global Investment Research

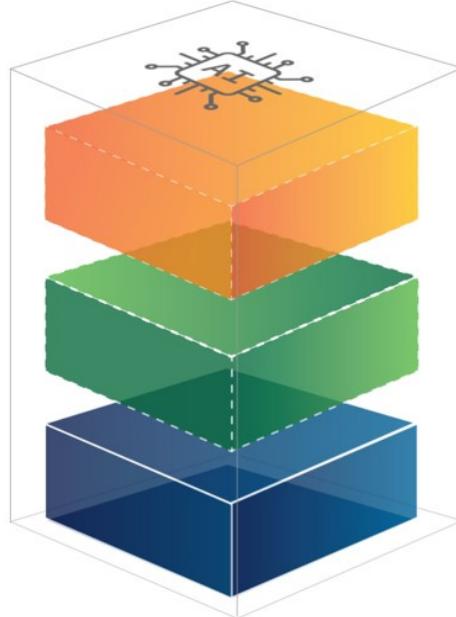
Top Picks - Companies best positioned to benefit

While we acknowledge that not many incumbents typically succeed in sustaining their technological (and therefore strategic) relevance through cycles, learnings from the on-premise to cloud migrations and the data-centricity of this cycle give us confidence that a company who screens well against the aforementioned framework is well-positioned to hold their market relevance in the face of competing Gen-AI-native solutions. In the below section we provide our top picks and rational as to how they fit within the framework laid out in the previous section.

Exhibit 16: GS' top Gen-AI picks across the tech stack Software, Hardware, and IT Services

Key Buys Across the AI Tech Stack

APPLICATION			
ADOBEE (ADBE) 12m PT: \$640 Upside: 15%	BRAZE (BRZE) 12m PT: \$62 Upside: 38%	FAIR ISAAC (FICO) 12m PT: \$1,879 Upside: 8%	HUBSPOT (HUBS) 12m PT: \$626 Upside: 24%
INTUIT (INTU) 12m PT: \$765 Upside: 23%	MONDAY.COM (MNDY) 12m PT: \$340 Upside: 29%	S&P GLOBAL (SPGI) 12m PT: \$561 Upside: 12%	
SALESFORCE (CRM) 12m PT: \$315 Upside: 19%	SAP (SAP) 12m PT: \$261 Upside: 19%	SERVICENOW (NOW) 12m PT: \$940 Upside: 13%	
PLATFORM			
DATADOG (DDOG) 12m PT: \$150 Upside: 28%	IBM (IBM) 12m PT: \$220 Upside: 12%	MONGO DB (MDB) 12m PT: \$325 Upside: 31%	SNOWFLAKE (SNOW) 12m PT: \$220 Upside: 90%
INFRASTRUCTURE			
AMAZON (AMZN) 12m PT: \$230 Upside: 30%	APPLE (AAPL) 12m PT: \$275 Upside: 21%	ARISTA NETWORKS (ANET) 12m PT: \$390 Upside: 10%	
DELL (DELL) 12m PT: \$160 Upside: 43%	DIGITALOCEAN (DOCN) 12m PT: \$42 Upside: 11%	DIGITAL REALTY (DLR) 12m PT: \$175 Upside: 15%	EQUINIX (EQIX) 12m PT: \$870 Upside: 5%
IRON MOUNTAIN (IRM) 12m PT: \$120 Upside: 6%	MICROSOFT (MSFT) 12m PT: \$515 Upside: 24%	NVIDIA (NVDA) 12m PT: \$135 Upside: 4%	PALO ALTO NETWORKS (PANW) 12m PT: \$376 Upside: 7%

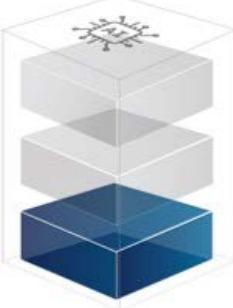


Price targets are for a 12-month time frame.

Source: Goldman Sachs Global Investment Research, Factset

Early cycle: IaaS in focus - In order to support the high-compute needs of Gen-AI workloads, a handful of well-capitalized companies were able to build the capacity pool for the industry. We expect companies such as MSFT, NVDA, AWS, GCP and smaller, private companies who have strong footholds in the large foundational models (OpenAI, Anthropic) to maintain their value proposition given the high capital requirements underpinning these services provide a relatively high competitive moat. Through an array of investments (direct inventory, minority investments, etc.), these companies have enabled the training of large foundational models and support the fine-tuning and inferencing that will be required to percolate demand to the application layer. Despite potentially adding pressure on their margin profiles, we expect unit economics to improve as we move through the adoption curve. Similar to the cloud computing cycle, we expect the distinction between new/core workloads to abate over time, with Gen-AI evolving into the common instance over time.

Exhibit 17: GS' top Gen-AI picks in Infrastructure layer

INFRASTRUCTURE			
AMAZON (AMZN) 12m PT: \$230 Upside: 30%	APPLE (AAPL) 12m PT: \$275 Upside: 21%	ARISTA NETWORKS (ANET) 12m PT: \$390 Upside: 10%	
DELL (DELL) 12m PT: \$160 Upside: 43%	DIGITALOCEAN (DOCN) 12m PT: \$42 Upside: 11%	DIGITAL REALTY (DLR) 12m PT: \$175 Upside: 15%	EQUINIX (EQIX) 12m PT: \$870 Upside: 5%
IRON MOUNTAIN (IRM) 12m PT: \$120 Upside: 6%	MICROSOFT (MSFT) 12m PT: \$515 Upside: 24%	NVIDIA (NVDA) 12m PT: \$135 Upside: 4%	PALO ALTO NETWORKS (PANW) 12m PT: \$376 Upside: 7%

Price targets are for a 12-month time frame.

Source: Goldman Sachs Global Investment Research

- **Microsoft** (*Covered by Kash Rangan*): With rapidly scaling Gen-AI solutions across the infrastructure, platform, and application layer of the tech stack and an ambitious investment strategy, Microsoft screens well across the five factors of our Gen-AI investment framework, remaining one of our most compelling Gen-AI top picks. From a TAM perspective, Microsoft brings a dominant footprint with scale and breadth that is a clear differentiator vs. peers. We see Microsoft successfully gaining momentum from this in multiple arenas. Azure AI Services revenue is rapidly scaling, now at a \$5.4bn+ run rate (or 7% of F4Q24 Azure revenue run rate) and on pace to become the fastest business in company history to reach \$10bn+. In Office, we project a TAM of \$135bn+ for Commercial Copilot and \$18bn for Copilot Pro (their consumer offering). Beyond these marquee offerings, Microsoft's opportunities include their Windows install base (with Windows Copilot already available on 225mm devices), applications like Dynamics, AI PCs, and more. Microsoft's data moat is similarly enviable. With a commercial install base of 400mm+ users, they are a key player in the business productivity space, with access to a wide variety of information and frontline worker data not aggregated on any other platform (similarly replicated across GitHub, LinkedIn, Dynamics, etc.). On the infrastructure side, the scale of Microsoft's platform allows users to easily leverage their data across a variety of native offerings deeply integrated with Azure AI like Cosmos DB, Fabric, and Power BI. The growing iterations of Office Copilot (M365 Copilot for Finance, M365 Copilot for Sales, and more) is a solid example of pace and scope of Microsoft's product innovation. Lastly, we are constructive on the company's investment philosophy that's supporting the elevated CapEx spending in FY24/FY25 and is predicated on building agile, broad-based architecture that Microsoft can direct toward wherever tangible demand builds without significant changes to their margin profile. As AI innovations and revenue opportunities become clear, we believe Microsoft will be able to invest quickly and boldly.
- **Amazon.com** (*Covered by Eric Sheridan*): Amazon Web Services (AWS) is the

world's largest public Cloud hyperscaler with AI revenue already at a multi-billion dollar run-rate. We believe that AWS is well positioned to benefit from the rising adoption of Gen-AI by enterprises across all three layers of the stack (infrastructure, platform, application). At the infrastructure layer, AWS offers the broadest selection of NVIDIA instances and has developed its own chips for training (Trainium) and inference (Inferentia) that offer competitive price performance. At the platform layer, Amazon SageMaker allows model builders to manage their data and deploy their models directly on AWS, with SageMaker increasingly becoming the standard for more models. Amazon Bedrock is another platform service and offers the largest selection of Gen-AI models to enterprises (incl. Claude 3.5, Llama 3.1, Mistral Large 2 and more) along with a robust suite of tools that allow organizations to evaluate and fine tune these models for more specific use cases. At the application layer, the adoption of AWS' coding assistant Amazon Q continues to rise. We are constructive on the appeal of AWS' integrated offering for enterprises and believe that this strategic focus on breadth and flexibility will be an important strategic differentiator and accelerate the adoption and scaling of Gen-AI solutions by enterprises, with this supporting our positive view of AWS' competitive positioning in an AI-first Cloud Computing landscape.

- **Nvidia** (*Covered by Toshiya Hari*) (**on the CL**): Nvidia designs Datacenter GPUs that are being widely adopted for both Training and Inferencing workloads across a wide range of customers, including, but not limited to U.S. Tier-1 CSPs, consumer internet companies, traditional enterprises, and governments across the globe. We believe Nvidia is positioned to benefit from the proliferation of Gen AI, as we see the company maintaining its competitive advantage built on its 1) ability to innovate across multiple facets of the data center including CPU, GPU, Networking and software, 2) large GPU installed base (which, in turn, incentivizes developers to build applications to Nvidia's GPU), and 3) ability and willingness to out-R&D its peers. From a valuation perspective, we note that the stock currently trade at an NTM P/E multiple relative to our Semiconductor coverage universe that is depressed vis-à-vis its own history. We are Buy-rated on NVDA and the stock is also on the Americas Conviction List.
- **Apple** (*Covered by Michael Ng*): Apple designs, manufactures, and markets personal technology devices and sells a variety of related services. Its long history and track record of (1) designing category-defining and innovative products (e.g., Mac, iPhone, iPad, Apple Watch, AirPods, iPod); (2) protecting digital privacy; and (3) delivering premium services & experiences have contributed to an unmatched brand strength. Apple's brand loyalty has resulted in a growing installed base of users that provide Apple with visibility into revenue growth by reducing customer churn, lowering customer acquisition costs for new product and services launches (e.g., Apple TV+, Apple Fitness, Apple Watch), and encouraging repeat purchases, such as phone upgrades. We are Buy-rated on AAPL as we believe that the market's focus on slower product revenue growth masks the strength of the Apple ecosystem and associated revenue durability & visibility. Apple's installed base growth, secular growth in services, and new product innovation should more than offset cyclical headwinds to product revenue, such as a reduced iPhone unit demand due to a

lengthening replacement cycle and reduced consumer demand for the PC & tablet category. Valuation is attractive relative to AAPL's historical multiple — both on an absolute & relative basis — and compared to key tech peers. The majority of gross profit growth over the next 5-years should be driven by Services, which should mark an inflection point in the Services investment narrative and support AAPL's premium multiple. The durability of Apple's installed base and the resulting revenue growth visibility from attaching more Services and Products is what underpins the recurring revenue — or Apple-as-a-Service — opportunity.

- **Arista Networks** (*Covered by Michael Ng*): As the leading branded provider of switches to US hyperscalers, ANET is well positioned to capitalize on the ongoing growth in data, the continued digital transformation driving workloads from on-premise to public and hybrid-cloud, and the growing demand for higher bandwidth, faster speed, and lower latency. ANET has the most revenue exposure to cloud spending among our coverage universe with 43% of its 2023 revenue from its Cloud Titans segment (META was 21%, MSFT was 18%) and 21% from tier 2 cloud and service providers. In addition, ANET is leveraging its dominant position in data center switches to other networking solutions, particularly in enterprise (e.g., campus switching, wireless, routing, telemetry). ANET should deliver strong double-digit revenue and EPS growth over the several years reflecting strong growth in data center and expansion into enterprise, as well as continued R&D and S&M investments particularly as ANET builds out its enterprise go-to-market capabilities.
- **Dell** (*Covered by Michael Ng*): Dell Technologies (DELL) designs, manufactures, and sells information technology infrastructure and related software for servers, storage, and personal computers. Dell is a market leader in the server and storage market, and a provider of PCs and related peripherals and services for consumers & enterprises. In our view, DELL's diversified portfolio positions the company to benefit from (1) AI server demand; (2) an inflection in industry PC trends driving growth in CSG segment revenue, which should also benefit from product innovation in higher-end PCs and greater exposure to commercial PCs; (3) strong long term growth in the ISG segment, with DELL's market leading position in enterprise servers & storage and AI-enabled portfolio of solutions positioning the company to be an IT partner-of-choice, which should help it maintain or gain share across key enterprise categories; and (4) shareholder friendly capital return programs, including at least 80% FCF return of capital to shareholders.
- **Iron Mountain** (*Covered by George Tong*): Iron Mountain is delivering robust 20%+ growth in data center revenue and signing an increasing mix of hyperscaler clients, driven to a large extent by Gen-AI deployments, with a significant expansion strategy in its data center business globally. IRM is additionally generating 30% organic growth in its asset lifecycle management business supported by a component price recovery and rapid server refresh cycles in connection with Gen-AI adoption, as well as double-digit growth in digital solutions. Meanwhile, IRM's revenue management initiatives in its legacy records & information management business are driving high-single-digit price realization, lifting overall storage organic revenue growth against a backdrop of flat-to-slightly up storage volumes. We expect revenue management and attractive growth in data center, ALM and digital solutions

driven by secular tailwinds including Gen-AI to drive healthy valuation upside in IRM shares.

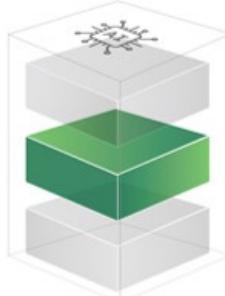
- **Digital Realty and Equinix** (*Covered by James Schneider*): We think DLR and EQIX stand as early to mid-cycle Gen-AI beneficiaries as we see a broadly positive demand read-across for our datacenter coverage - and we believe Buy-rated Digital Realty and Equinix are well positioned to benefit from the current elevated demand environment, coupled with constrained supply given the limited power utilities are able to deliver to new datacenter projects. Digital Realty, Equinix, and the broader market are seeing a shift toward server configurations with higher power densities, driven by generalized GPU-centric workloads as well as a nascent mix of workloads tied to Generative AI. We expect this trend to continue in the long run, and believe Digital's leading mix of large-scale datacenters which support high power densities (including capabilities such as water-based cooling) is likely to be the most future-proofed against this trend. Externally, the outsourced datacenter market has tightened materially over the past year, driven by a convergence of factors including strong secular demand from hyperscaler customers, a sudden increase in AI-driven compute demand, and constrained medium-term supply dynamics tied to limited power availability from utilities in key high-demand locations. Although Equinix is seeing some level of demand crosswinds from a segment of its enterprise customer base focused on rationalization, we expect market tailwinds to carry the day in the near term. On the margin, we prefer DLR given greater exposure to larger wholesale blocks of capacity in locations where supply/demand dynamics are tightest.
- **Palo Alto Networks** (*Covered by Gabriela Borges*) – Palo Alto Networks is generating \$200mn in ARR from its AI product suite including Artificial Intelligence Operations, XSIAM, and additional product launched earlier this year. The company's Precision AI bundle which combines current ML techniques with Gen-AI for predicting evolutions for zero day attacks increasingly allows Security Operations Centers (SOCs) to respond to threats in real-time while reducing manual intervention. In addition, Palo Alto Networks' recent acquisition of Dig Security for Data Security Posture Management (DSPM) positions the company to benefit from data posture investments by enabling strong data security and governance for AI-driven processes in cloud environments. By securing data pipelines and enforcing data access controls, Palo Alto Networks can mitigate risks associated with the training and operation of Gen-AI models, reducing the chance of data leaks thus building a more secure AI ecosystem and ensuring regulatory compliance (e.g., GDPR, HIPAA, etc.).
- **DigitalOcean** (*Covered by Gabriela Borges*): DigitalOcean acquired Paperspace in July 2023 and has focused on expanding its AI platform over the last 12 months, including additional investments in CapEx and its product roadmaps. AI offerings today include IaaS for companies developing/substantially customizing a model, PaaS for full life-cycle support, and GPU Droplets which provide fractional access to a single GPU within a virtual machine. DO investment levels screen as extremely modest relative to the CapEx intensity of the hyperscalers and specialized cloud providers. However, management has discussed its customer base focuses on AI

extenders, who are enhancing foundational models with their own custom data; and AI consumers, who are building applications that consume AI. This customer base does not require the same magnitude of highly specialized hardware for intensive computing tasks as the foundational model builders, and thus management feels confident in its ability to grow the business even with a relatively modest level of AI investment.

Mid-cycle: PaaS gains momentum - Much of the value from Gen-AI has accrued at the infrastructure layer, predominantly benefiting from the training, fine-tuning, and inferencing of foundation models. The Hyperscalers have been the primary beneficiaries of this early spending. However, we believe it's only a matter of time until value shift up to the platform layer, which will provide the foundational technologies and tools to enable tangible use-cases of Gen-AI at the application layer. The ever-growing complexity prevalent in most enterprises IT environments is likely to remain a choke point on the broader path to AI-native applications, thus presenting an opportunity for both incumbent and emerging vendors to simplify the process.

Exhibit 18: GS' top Gen-AI picks in the Platform layer

PLATFORM	DATADOG (DDOG)	IBM (IBM)	MONGO DB (MDB)	SNOWFLAKE (SNOW)
	12m PT: \$150 Upside: 28%	12m PT: \$220 Upside: 12%	12m PT: \$325 Upside: 31%	12m PT: \$220 Upside: 90%



Price targets are for a 12-month time frame.

Source: Goldman Sachs Global Investment Research

- **Datadog** (*Covered by Kash Rangan*): Providing a holistic Observability offering via three scaled offerings (revenues >\$500mn each) - Infrastructure Monitoring, Application Performance Monitoring, and Log Management - Datadog's strong developer mindshare and real-time unified platform with exabytes of mission-critical data should favorably position the company against Gen-AI's opportunity. With Gen-AI expected to increase the velocity of data creation, Datadog's well-regarded products, solid market positioning, and Gen-AI centric development to-date (e.g., LLM Observability, Watchdog AI, Bits AI) should further differentiate the company as customers start leveraging Gen-AI. While still early, we highlight that 4% of Datadog's revenue comes from AI-native companies (with this figure steadily increasing 50bps QoQ since F3Q23). Furthermore, Datadog sports a strong cross-sell profile, with the percentage of customers using 4+ (49%), 6+ (25%), and 8+ (11%) products increasing 4 pp. YoY as of F2Q24. This strong cross-selling ability can support the existing install base adopting newer Gen-AI products. With a >\$50bn TAM, consistently operating at the Rule of 50+ (F24E: 51%), and with best-in-class unit economics (F2Q24 GAAP CAC: 1.25x), we think Datadog is uniquely positioned to efficiently capitalize on durable Gen-AI tailwinds while maintaining a top-tier financial profile. This should be aptly appreciated by investors

and could position Datadog to deliver durable future outperformance.

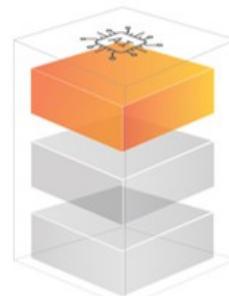
- **Snowflake** (*Covered by Kash Rangan*) (**on the CL**). In considering Platform-layer beneficiaries, we believe Snowflake screens favorably. Already possessing a best-in-class compute engine and data warehousing capabilities, exabytes of data under management, and an expansive TAM (\$340bn by C28, per the company) we remain encouraged by the company's emphasis on accelerated product innovation – best evidenced at their recent Data Cloud Summit. Highlighting this point are products like Cortex AI, Snowpark Container Services (SCS), and Unistore. Cortex AI is a fully managed service helping expedite the building and delivery of AI applications via utilizing LLMs. SCS is a fully managed service allowing users to run their containerized workloads within Snowflake, thereby ensuring the data can be processed in Snowflake's environment. Unistore can combine transactional (OLTP) and analytical (OLAP) workloads onto a single platform, allowing users to leverage their data for more immediate analytical purposes (with the value proposition of real-time analytics likely only increasing against a Gen-AI backdrop). We also believe that Iceberg Tables could prove an eventual tailwind. While customers storing their data in Apache Iceberg (versus Snowflake) could present a near-term revenue headwind (Storage: ~10% of revenue), the shift towards open table file formats could support more workloads coming onto Snowflake's platform as customers lean into the compute engine for core analytical use cases and AI enablement. Despite initial F25 Product Revenue guidance proved disappointing (22% versus Consensus (FactSet): 30%), and the company halved their F25 Operating Margin guidance in F1Q25 (6% 3%), we think that continued accelerated product innovation, and a stated focus on expanding profitability (GSe: 300bps YoY in F25+) could be key factors in the company delivering future outperformance. In considering all the aforementioned factors, we think Snowflake will prove an outsized beneficiary as Gen-AI benefits migrate to the platform layer.
- **MongoDB** (*Covered by Kash Rangan*): With a sizable TAM opportunity (>\$150bn (C28) per Gartner, 13% CAGR), strong developer mindshare, and a best-in-class NoSQL database offering via Atlas (>30% YoY growth since release), we think MongoDB can be an outsized PaaS Gen-AI beneficiary. Databases have represented one of the most consistent technologies across various computing cycles (e.g., internet, mobile, and cloud), and we expect this dynamic to continue with the Gen-AI cycle. Namely, as with other applications, Gen-AI applications will still require a database to store users' transactional data. Furthermore, with elevated growth rates in semi- and unstructured data slated to continue, NoSQL databases (like Atlas) – which are adept at handling these data types – are well poised to benefit. With companies likely prioritizing a holistic data strategy (versus structured data dependency) against an LLM / Gen-AI-centric future, MongoDB can leverage Atlas' demanding developer mindshare to support a continued attractive growth profile. Furthermore, we highlight that MongoDB has made Gen-AI investments (a key factor we consider), with Gen-AI functionality having been embedded across products like Atlas and Relational Migrator. While still in the early days of this cycle, we believe that MongoDB's opportunity to capture Gen-AI workloads from both new and existing customers can bode favorably for the company moving forward.

■ **IBM** (*Covered by James Schneider*): We think IBM stands as a mid to later-cycle Gen-AI beneficiary as organic investments in AI offerings (WatsonX) and use of open-source LLMs are compelling differentiators that continue to bear fruit and serve as a competitive edge in client engagements. IBM's inception-to-date AI bookings of \$2.0 bn (\$1.5bn in Consulting and \$500mn in Software) are broadly tied to transformative deals which seek to reduce overall operations cost and are expected to be recognized ratably over the longer term, along with Software tools (such as WatsonX Code Assistant and other products) designed to drive greater developer productivity. Reflective of clients' current business priorities, the bulk of AI bookings (\$1.5bn in Consulting) currently consists of early proof-of-concept projects as well as larger cost takeout deals tied to AI and automation as clients are primarily focused on leveraging AI to drive cost optimization over revenue growth. Looking internally, IBM has also been active in using Gen-AI in its own operations to drive cost synergies, where it cites cost reductions of between 50% and 80% in certain corporate functions. While the vast majority of clients remain in the exploratory phase of Gen-AI engagement, we believe IBM is well-positioned to capture a growing market of AI-led services demand as clients steadily progress towards broader AI transformation.

Later-cycle: SaaS finds its footing - Investors have been weighing the strategic relevance of application software companies since the emergence of Gen-AI given the simultaneous strength of companies such as Nvidia, OpenAI, Microsoft Azure, AWS, and others. This is reminiscent to the AWS debate in 2016, when the company's growing dominance in IaaS and PaaS was assumed to seep into the application layer and commoditize SaaS. Ultimately, their strength in these layers enabled a boost in software companies and had the adverse effect. We see this playing out in a similar fashion through this technological innovation cycle, with companies such as ADBE, MSFT, and CRM being able to maintain their competitive differentiation and strong value proposition through differentiated data, engaging user interfaces and comprehensive platforms solutions.

Exhibit 19: GS' top Gen-AI picks in Application layer

APPLICATION		
ADOBE (ADBE) 12m PT: \$640 Upside: 15%	BRAZE (BRZE) 12m PT: \$62 Upside: 38%	FAIR ISAAC (FICO) 12m PT: \$1,879 Upside: 8%
INTUIT (INTU) 12m PT: \$765 Upside: 23%	MONDAY.COM (MNDY) 12m PT: \$340 Upside: 29%	S&P GLOBAL (SPGI) 12m PT: \$561 Upside: 12%
SALESFORCE (CRM) 12m PT: \$315 Upside: 19%	SAP (SAP) 12m PT: \$261 Upside: 19%	SERVICENOW (NOW) 12m PT: \$940 Upside: 13%
		HUBSPOT (HUBS) 12m PT: \$626 Upside: 24%



Price targets are for a 12-month time frame.

Source: Goldman Sachs Global Investment Research

- **Adobe** (*Covered by Kash Rangan*): Considering our aforementioned five factor framework, we believe Adobe screens favorably against all key characteristics. Operating against a ~\$300bn TAM (2030), we estimate that Gen-AI presents an incremental \$4bn opportunity that Adobe is favorably positioned against. Adobe already possesses a best-in-class end-to-end platform, with Gen-AI's introduction possessing the ability to augment the platform's existing capabilities. With natural language processing interfaces lowering content creation barriers, Adobe's quick time-to-market with Gen-AI products – best evidenced in its Creative Cloud (e.g., Firefly – integrated from Photoshop to Premier; >9bn images created to-date) should further cement them as the de facto content creation solution. More broadly, we believe Gen-AI can underpin multiple growth further growth vectors across Document Cloud and Digital Experience, too, as this technology could complement the already high-quality data within Adobe's Customer Data Platform (CDP). Exhibiting a consistent history of product development over the last 30+ years, we believe investors are apt to appreciate Adobe's consistent product roadmap velocity moving forward. With the company approaching the Rule of 50 (F24E), prioritizing Gen-AI monetization could contribute to future revenue upside while still delivering a top-tier margin profile (Non-GAAP Operating Margins: >40%). Accordingly, given the factors described, we believe Adobe could be one of the most prominent Gen-AI beneficiaries in our coverage.
- **ServiceNow** (*Covered by Kash Rangan*): We believe ServiceNow is one of the best positioned application companies to benefit from the coming Gen-AI cycle. Built on a cohesive platform with huge amounts of differentiated, proprietary data, and a proven ability to execute, NOW is well positioned to augment its ~\$275bn TAM with a wide variety of Gen-AI tools. Likewise, as a system of record for both IT and business professionals and with a high level of engagement from users, NOW has unique access to comprehensive data across the enterprise. NOW's platform approach to application software gives the company the flexibility to leverage this data moat and quickly build compelling Gen-AI applications. Their ability to create compelling and rapidly scaled Gen-AI products is evident. Just three quarters after the launch of their Gen-AI SKU, Pro+, NOW is a rare example of a SaaS company that can quantify both significant demand and pricing power. NOW indicated Pro+ is seeing a 30% realized pricing uplift (vs. Pro's 25%) and is now the fastest growing product in company history. Signing eleven Gen-AI deals greater than \$1mm (vs. seven in 1Q24) and two deals greater than \$5mm, AI is becoming a meaningful contributor to revenue, with Now Assist NNACV doubling quarter over quarter. This rapid growth is fueled by quantifiable and material efficiencies. Among many examples, NOW has cited LTI Mindtree using Now Assist for ITSM to increase developer productivity by 30%, BT Group cutting case summaries and not review times down by 55%, a 30% improvement in mean time-to-resolution for service agents at a business services company, >80% improvement in self-service deflection for a large manufacturing organization, and internal software engineers accepting 48% of text-to-code generation. As Gen-AI NNACV contribution ramps, with ServiceNow not assuming an accelerated adoption curve vs. Pro, we believe Pro+ adoption can provide a multi-year tailwind to cRPO growth and complement our view for >20% Subscription growth through FY27 at a Rule of 50+.

- **Salesforce** (*Covered by Kash Rangan*): Across both the platform and application layers, we believe Salesforce is poised to execute well on the Gen-AI investment cycle which should support the company's long term growth narrative. The Company operates in segments of software that, collectively, represent one of the largest TAMs of the industry at almost \$300bn. As a system of record, Salesforce stores a significant amount of data across countless industries and customer sizes that can generate more insightful and actionable suggestions for customers. This concentration of data hosted within the Salesforce ecosystem, combined with Mulesoft's integration capabilities and efforts to migrate onto a common data architecture, leave us confident in their platform strength and agility. Similarly, with the launch of Data Cloud last summer, the company has reinforced its focus on delivering a platform that is powered by real time automation and intelligence with a simplified user experience. This should drive greater enablement of Salesforce's ultimate goal of surfacing the most relevant customer data. Generative AI's capabilities could simplify the growing complexity of the CRM offering, leading to lower adoption barriers (via improved onboarding, utilization, and engagement) and enhanced financial performance (namely better net revenue retention rates, net new logo additions, and increased deal sizes). Salesforce's prioritization and commitment to evolving their Gen-AI offerings is clear. The company has announced a slew of product announcements over the past few months across Salesforce's AI Day in June, Dreamforce conference in September, TrailblazerDX conference in March. The company has dedicated significant time and resources to their Gen-AI strategy, rapidly scaling Data Cloud, the Einstein product family, and doubling their Generative AI venture fund to \$500mn, all unambiguous efforts to prioritize their AI efforts. Ultimately, the potential for pronounced end-user productivity can set up the stage for an improving growth narrative.
- **Monday.com** (*Covered by Kash Rangan*): Monday.com is an underappreciated beneficiary of the coming AI cycle. With a 2026 TAM of \$150bn, Monday is investing in front of a massive market opportunity with natural end market extensions like HR, financials and others. As Monday increasingly becomes the platform of choice for SMBs and, increasingly, mid-market and enterprise companies, the company will benefit from scaling proprietary data in both core work management as well as emerging products like CRM, Dev, and Service. We believe that Monday's low-code no-code technology and intuitive interface will be an increasingly important differentiator as Gen-AI reduces the affect of technical differentiation and buyers increasingly place a premium on high-quality UI. Early AI investments are already showing signs of efficiencies. After early success deploying a third party chatbot for managing customer service tickets, Monday reported it had resolved around 50% of customer service tickets automatically and drove a significant increase in chat ticket volumes as customers benefited from its speed and accuracy. This will only enhance the projects Monday is bringing to market and will position their investments (like AI assistant, AI Automation Blocks, and Bots) for broader usage. Role-specific applications can increase end-user productivity, increase retention, and drive Net Expansion Rates (NER). These solutions should make it easier to customize and interact with Monday's products, driving further buy-in to the customizable platform and longer-term growth. As the company continues to scale investments and with a

clear capital allocation strategy focused on long-term product growth, we believe Monday is well positioned to aggressively invest ahead of this cycle.

- **Intuit** (*Covered by Kash Rangan*): Offering a comprehensive set of SMB solutions supported by an expansive trove of data to leverage, and a TAM >\$300bn (GSe), Intuit is well positioned to be an early Gen-AI beneficiary within the Application layer. The company has already exhibited a stated focus on making the appropriate investments to capitalize on this opportunity, with the company expecting to hire ~1,800 employees in F25 to advance AI-centric initiatives, such as enhancing Intuit Assist and creating new AI-native products. Intuit's platform harnesses unparalleled volumes of data on consumers and SMBS, with >500K customer/financial attributes per SMB and over 60K tax and financial attributes per customer. We take stock in the company's investments to provide a unified platform to leverage this data. The expansive data volume Intuit has access to, its years of investments in personalized user journeys (i.e., Turbo Tax), its expertise on how to augment human services with technology (Turbo Tax Live, QBO Live, etc.) are likely to put Intuit in a strong position to roll out Gen-AI services quickly that can immediately help optimize a user's experience (e.g., accelerated time to file, bigger refund). Furthermore, as Intuit works to further penetrate the more complex Assisted category, overlaying Gen-AI to help with more involved returns could result in accelerated adoption and provide a future upside lever for the top-line growth. With a stated focus on leveraging Gen-AI coupled with a best-in-class financial profile (F24E: Rule of 40+) we think Intuit could be a name that sees outsized Gen-AI contributions once the value more prominently shifts towards the Application layer.
- **HubSpot** (*Covered by Gabriela Borges*): HubSpot AI has a wide array of features today and has many more in development as the company has made AI one of the foundational building blocks of their R&D organization. Today, HubSpot AI can generate content (websites, emails, blogs, etc.), switch content into different formats, help build chatbots, drive sales forecasts, report on campaign performance, and more. After announcing HubSpot AI at Inbound last year, 2 months later 40% of enterprise customers had used HubSpot AI features and 20% of pro customers – speaking to the fast pace that HubSpot customers are willing to adopt new technologies. While HubSpot typically acts as a fast follower to functionality developed in the private ecosystem, in AI HubSpot has been more of a leader where they benefit from having a larger R&D budget and more unique data. HubSpot is incorporating AI across product tiers but 65% of features released in 1H24 were included in Pro and Enterprise tiers. Going forward we are positive on HubSpot's ability to innovate in AI in a way that further differentiates them from competition in a part of the market that typically can be more competitive.
- **Braze** (*Covered by Gabriela Borges*): Braze has been an early adopter of Gen-AI and has incorporated various AI and ML driven features in its products. These features in AI are a differentiator for Braze against both marketing companies in the startup space and legacy marketing clouds who are fundamentally not built in a way that supports nonlinear/dynamic customer journeys, limiting the level of personalization that Gen-AI can ultimately help deliver. Braze integrated GPT into its product in March 2022 and DALL-E in December 2022 to help marketers with content and

image generation — well ahead of competitors. Additionally, as enterprises focus on their data posture in order to take advantage of AI, we believe that AI has also increased the importance of how applications interact with data. Braze's key differentiation stems from its underlying architecture, which is vertically integrated across five layers of marketing functionality and allows for processing data in real time. The privacy landscape has also been shifting as providers look to limit the amount of spam and noise in consumer inboxes as well as protect consumer privacy. This is adding to the importance of personalized marketing campaigns and the use of first-party data — both areas where Braze specializes & can benefit from AI. We see upside to numbers and believe that Braze's technology differentiation can support durable growth over the medium term. Given Braze's premium price point and ability to power advanced use cases — we look to better understand how much share Braze can take longer term from the legacy marketing clouds.

- **SAP** (*Covered by Mohammed Moawalla*): We believe SAP is well positioned to benefit from the coming Gen-AI cycle driven by robust S/4 HANA product cycle momentum which remains in a sweet spot. With c.60% of the world's GDP touching a SAP system, as per the company, SAP's ERP system carry a both vast but valuable set of enterprise business data. This spans both the front/back-end and supply chains to further unlock key insights to drive business efficacy and productivity improvements. Given that SAP's AI capabilities will only be available in the cloud, Business AI is driving conversations with customers and so, increasing the pace of migration into SAP's cloud offerings contributing to both upsell/cross-sell opportunities. Management is committed to building its Business AI embedded capabilities, which span predictive models as well as Generative AI use cases while leveraging data from across SAP's entire portfolio. This, for example, includes SuccessFactors (HR data), Ariba (procurement data) as well as its cloud ERP and is underpinned by SAP's Business Technology Platform (BTP), which helps harmonise data and extract value from it. SAP has already begun to harvest anonymised customer data, with customers' permission, to create industry proprietary models. The company notes that AI is already playing a pivotal role, with 20% of all deals now including premium AI features. In SAP Concur, around 150,000 weekly users are already benefiting from embedded AI features in use cases ranging from intelligent document processing of matching invoices and validating receipts to analytics. In addition to this, the company will also offer a customized AI in which customers can develop their own use cases with a flexibility of choosing broad range of LLMs. Looking internally, SAP is leveraging Generative AI to drive cost synergies of its workforce with an estimated triple-digit-million efficiencies from embedding AI across the entire portfolio to come to fruition in FY25. Overall, we believe SAP is well positioned to capture the Gen-AI cycle underpinning our estimates of c.25% cloud revenue growth CAGR across FY24-28E.
- **S&P Global** (*Covered by George Tong*) (**on the CL**): S&P Global is a leader in AI-powered information services bolstered by its acquisition of Kensho, which deploys deep learning techniques such as natural language processing to solve real world problems in the financial services and other industries. We believe its Gen-AI investments in virtual assistants and data search functionalities will drive

strengthened medium-to-longer term revenue and margin performance across segments, with multiple Gen-AI capabilities launched over the past several quarters seeing strong traction. We look for high-single-digit revenue growth in SPGI's non-Ratings businesses, supported by market share leadership, pricing power, a wide competitive moat and secular tailwinds, such as the ongoing shift from active to passive asset management in Indices. Additionally, S&P's Ratings business is continuing to drive an upward estimate revision cycle backed by strong debt issuance volumes lifted by refinancing activity in high yield and leveraged loans and a pull-forward of issuance activity, with further upside potential to the company's Ratings revenue guide this year driven by approaching rate cuts and improving M&A volumes. We believe an improving cyclical outlook for Market Intelligence, rising AUM from fund flows into S&P-linked indices, the ongoing realization of merger revenue synergies and accruing benefits from Gen-AI will combine with debt issuance tailwinds to drive healthy upside in SPGI shares.

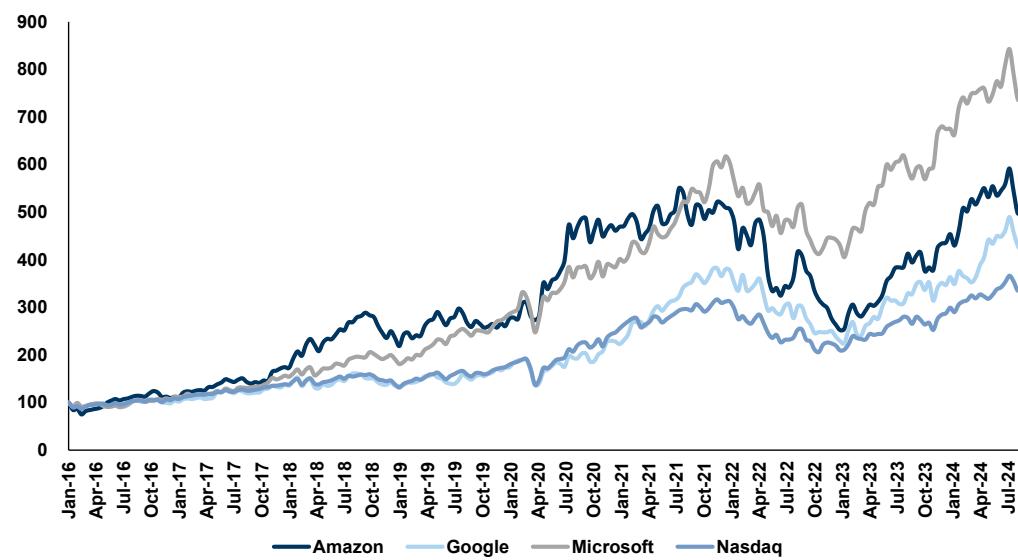
- **Fair Isaac** (*Covered by George Tong*): Fair Isaac's Software platform ARR is growing 30%+ as financial institutions continue to adopt the modern platform offerings and increase their utilization under a consumption-based revenue model. FICO is in the early innings of developing Gen-AI capabilities that mitigate risks involving safety, accuracy, operations and legal/ethics, and center on developing AI that is explainable and ethical, with transparency on how data is used and which LLMs are used, and where the user is in control of the AI. Over the past year, FICO has filed 16 new patents related to Gen-AI capabilities. Initial Gen-AI prototypes run on OpenAI's ChatGPT and Anthropic's Claude, and include 1) FICO Advisor, which provides advice, guidance and assistance on FICO Platform capabilities, 2) Simulation, which varies inputs and re-runs simulations across use cases, and 3) Enterprise Optimization, which helps with general optimizations and interacts with optimization-based apps. Meanwhile, FICO's Scores business is sustaining robust 20% y/y growth led by significant increases in mortgage revenue due to strong price realization. A strong growth outlook for the Software business, coupled with healthy price realization in Scores backed by continuous innovation, serves as a driver for healthy valuation upside at FICO.

Evolution of the tech stack: From compute to cognition

Rethinking the IT tech stack through the IPA framework. Investors have often brought up the view that the investment in companies' infrastructure is the 'end all be all' with respect to innovation, with little to no capital flowing through to the adjacent layers of the tech stack - i.e., that the same vendor that provides the hardware or infrastructure necessary for Gen-AI will also be the ones to offer PaaS/SaaS solutions. We largely disagree with this view as we see tech cycles following an IPA pattern. At first, the cycle is about infrastructure build out. Then we have platforms that allow the best use of that infrastructure while providing building blocks to build next generation applications. The final leg is the build out of applications.

Recall the implosion of demand for software coincident with the Y2K cycle that declared that software was dead. The reality was that a platform shift was underway between client-server and the internet architecture. Admittedly, the industry struggled to grasp what that killer application for the next tech shift was going to be. It only became clear following the IPOs of Google and Salesforce what next generation technology could do for the end user, and therefore what the sought-after killer applications would be (search and business applications). Again, the value creation shifted from infrastructure to platforms to applications. Similarly, we recall that at in the 2016 timeframe the existential threat ascribed to enterprise software from AWS' rise, which many assumed was going to destroy and disrupt platforms and applications. In fact, just the opposite occurred. Today, Microsoft Azure, Amazon Web Services, and Google Cloud Platform are the very foundations upon which multitudes of cloud-native platforms and application companies have sprouted and scaled driving significant shareholder value ([Exhibit 20](#)).

Exhibit 20: Hyperscalers have delivered outsized returns versus the Nasdaq since 2016

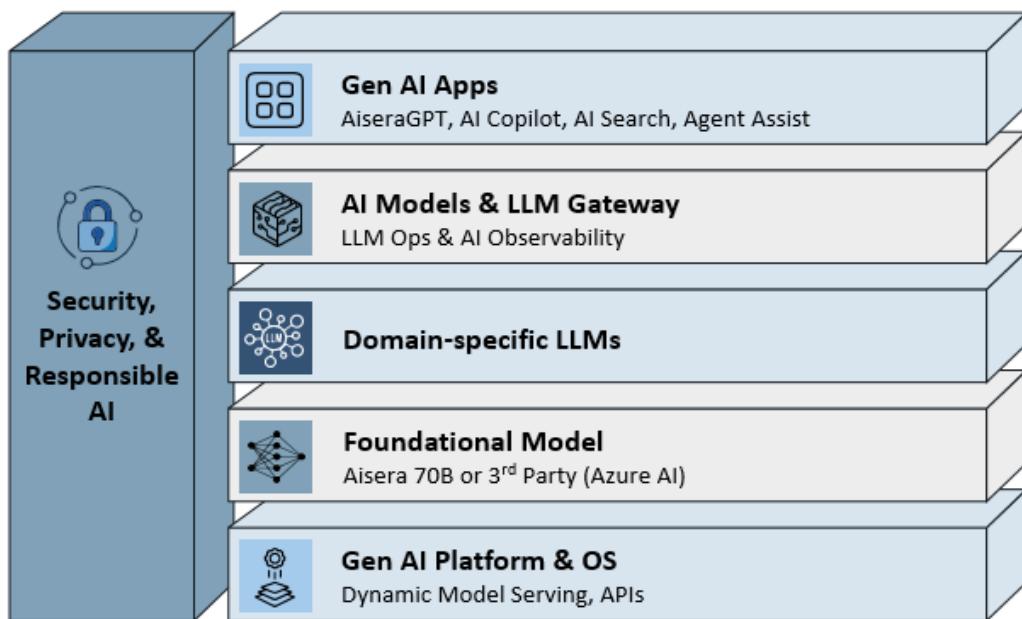


Source: Goldman Sachs Global Investment Research, FactSet

Similarly, in this cycle we're starting off by assuming that software is dead, being disrupted by hardware. The cyclical impact from high rates, contributing to disappointing

1H24 performance and outlook (our coverage down 8% YTD) only strengthens this bear argument. The fact that incumbents that led prior tech cycles rarely drove leadership in subsequent cycles also makes it tough for investors to gain comfort in. However, we return to the basics. Admittedly, this requires some patience for migration from the infrastructure to platforms to applications evolution to transpire. A big difference in this cycle of Gen-AI is that while the incumbents in the prior client server cycle were slow to adopt the cloud, the application software incumbents in this cycle, be it Salesforce, SAP, Oracle, ServiceNow, Adobe, Intuit or Workday, are developing Gen-AI products with a heightened sense of urgency through partnerships with various LLM providers. While brand-new startups that take full advantage of the completely new Gen-AI tech stack replete with GPUs, LLMs, and agent building platforms are favorably positioned to build successful businesses, data remains the one big consideration. Highlighting how the IT tech stack could look when accounting for Gen-AI developments, we highlight ([Exhibit 21](#)), originally published by Aisera. Aisera is a Gen-AI native company focusing on providing Gen-AI based solutions for, amongst other things, IT, HR, and customer service. Given their skew towards being a Gen-AI native company, this exhibit could illustrate how future tech stacks for companies could look.

Exhibit 21: Aisera's Gen-AI-centric tech stack serves as example of a Gen-AI native IT architecture



Source: Goldman Sachs Global Investment Research, Company data

Product velocity from data platforms Databricks and Snowflake add to our conviction that the structural shift of Platform architecture is underway. Thinking through what can enable the emergence of a step function change in how we leverage AI, we are forced to think in-depth on the changes that may need to be brought through the tech stack. The advent of Gen-AI will likely significantly influence the technology stack as we know it, with various changes and innovations likely to occur across different layers and components. We point to Databricks' launch of a full-service LLM development platform (Mosaic AI) and Snowflake's competing solution, Cortex AI, that supports RAG, fine-tuning and unstructured data analysis as early proof points that

platform players are leaning into the Gen-AI enablement opportunity. We do not expect Gen-AI to be just another layer in the tech stack; it is a transformative force that impacts every aspect of the stack, from the hardware used to the types of applications developed. It necessitates new infrastructure, tools, and approaches, while offering unprecedented opportunities for innovation and efficiency. As the field evolves, we can expect these changes to become even more pronounced, further integrating AI into the fabric of technology and business processes. For this reason, we see the industry migrating from the fragmented data storage architectures toward more collective data platforms.

Investments, new workflow processes and moats required before a Gen-AI native application materializes. Before outlining how we see Gen-AI ultimately reshaping the current IT architecture, we offer our thoughts as to what solutions will be table stakes in enabling Gen-AI's evolution. We draw parallels to both the cloud migration cycle and the introduction of the iPhone, where the innovation from a hardware perspective – i.e. improved processors, or the promise of shared cloud computing – was the impetus for major technological shifts in the platform and application layers. As solidifying the right IT architecture will prove paramount for future innovations to stem out from, we see the industry evolving beyond this over the coming years. We break this down into the investments, processes and moats that we expect to evolve, while drawing parallels to that of the prior cycles.

- **Investments:** We believe all organizations that will want to maintain technological relevance will need to invest in a range of solutions to best incorporate Gen-AI into their tech stack. Depending on their Gen-AI strategy – with some likely to opt for on-device applications, trained in-house, etc. – some investments are likely to be more prevalent than others. Nevertheless, we highlight the need for:
 - **Vector Databases / Search:** Vector databases index data uniquely in the form of vector embeddings, which are numerical representations of the data versus text. The compressed numerical representation of words, images and other forms of data helps optimize storage and enhances querying capabilities due to the closer proximity of vector embeddings in the database versus traditional databases. These embeddings have semantic capabilities providing richer meaning and context to search queries by leveraging the concept of similarity. In the world of Gen-AI vector databases play a pivotal role, considering that language models are built using large volumes of unstructured data, by providing greater context to the data. Vector databases reduce latency, improve model performance and enhance output relevance. Pinecone is a well-known vector database company that has robust partnerships with [Microsoft Azure](#), [Datadog](#) and others.
 - **High-performance computing:** As in prior technology curves, new applications of hardware offerings spark new ways of engagement before value realization occurs up the stack. In the case of Gen-AI, GPUs hold the right parameters to enable the high-frequency, complex training needed for foundation models that has led us to where we are today. As we move through the cycle, we expect an evolution of hardware solutions, too. We see

a scenario where chips are formed around Gen-AI use cases and that can present a viable alternative to GPUs – which were formed to solve for increased capacity needed for high-complex, high-resolution graphic instances. As GPUs emerged to enable improved video and graphic processing, we see chips being built for the intended use case of specific Gen-AI applications as they become more defined - which has already been somewhat witnessed by OpenAI's discussions with Broadcom ([here](#)).

- **Evolution of Workflow Processes:** As the aforementioned investments are founded on the premise of changing the way users interact and engage with their data and software, we expect new workflows to emerge, like:

- **RAG (Retrieval Augmented Generation) Tools:** Trained on a plethora of data sources, LLMs generate generic prompts and especially fail to incorporate the most up to date responses. A technique known as Retrieval Augmented Generation (RAG) helps bridge the gap by bringing more context to a query and incorporating dynamic data updates. Despite the numerous LLM options available, enterprises are often conflicted between building and training their own models or fine-tuning an existing model. Both options are expensive, time-consuming and more importantly lack enterprise relevant specialization. RAG offers the optimal solution and vector databases serve as the backbone for Retrieval Augmented Generation workflow tools. RAG architecture layers an information retrieval system on an existing LLM that gives users control over the data. In an enterprise context this gives enterprises the ability to limit the natural language processing of the main LLM to enterprise data only as opposed to all publicly available data. These contents are typically stored in vectorized databases which make the retrieval process easier. Leveraging RAG increases the accuracy and relevance of an LLM by providing it with access to the most trusted, current, contextually relevant data sources. Elastic is seeing some success with its [Elasticsearch Relevance Engine \(ESRE\)](#), an early pioneer of this technology combining RAG with enterprise vector search capabilities.
- **Reinforcement Learning from Human Feedback (RLHF):** is a technique that involves fine-tuning a pre-trained model using an iterative process that relies on human feedback. This method is particularly useful because it leverages preexisting foundation models that are easily accessible (on open source communities like Hugging Face) and can be adapted for specialized use cases using human feedback loops to improve model performance. This strengthens the efficiency of foundation models, augments human values (particularly useful for chat prompts/ virtual assistants) and ensures better safety since it incorporates human feedback. Fine-tuning models eliminates the need for developers to build their own LLMs.
- **Trust Layer:** Given the iterative nature of foundation models and organizations' growing data governance concerns, we see a growing need for the following techniques to be embedded as part of a new Gen-AI stack: **1) Grounding:** providing only the context needed for a user's prompt, **2) Masking:**

encrypting the proprietary information in any given prompt and **3)** Zero-Retention: ensuring that there is no stored version of the written prompt that can be iterated on at a later point. Such techniques are expected to sit between any given foundation model and a fine-tuned proprietary data set to monitor prompts and oversee the generated output. As they insulate an organizations' data from being fed into larger foundational models, we see the implementation of such services as critical to expanding the comfort associated with deploying such models and driving broader adoption of Gen-AI.

- **MLOps:** Machine Learning Operations leans on the principles of the DevOps life cycle. MLOps specifically focus on the early stages of the development cycle of a Gen-AI application which involve data collection and model development / enhancement. MLOps plays a crucial role in an AI enabled app development journey, despite using pre-trained large language models, developers perform several steps before leveraging a model. The machine learning life cycle requires the enterprise/business data to be collected and cleansed first followed by training and testing of the model, fine-tuning to meet enterprise specifications before finally being deployed to production. MLOps tools help in model metadata management, workflow orchestration, pipeline versioning and finally model deployment and monitoring performance in production. Amazon SageMaker is a popular MLOps tool.

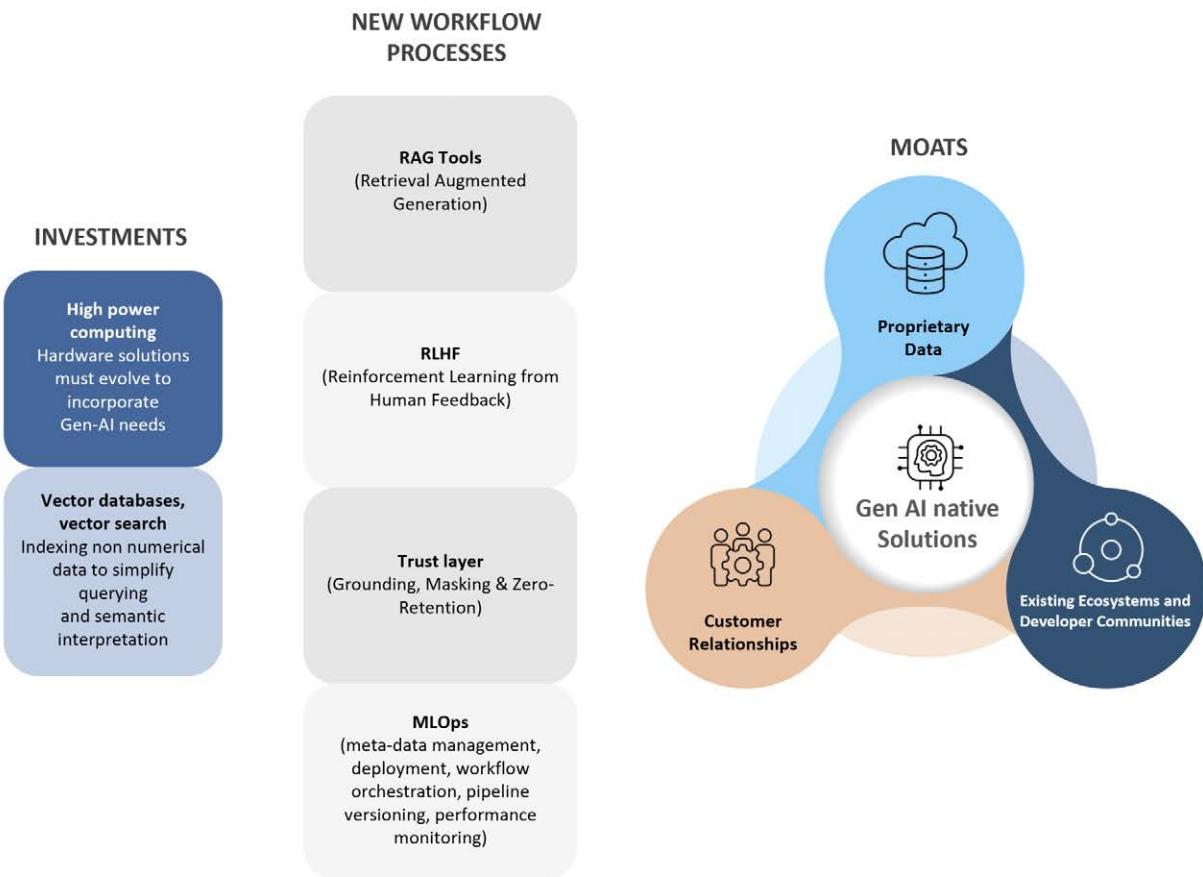
- **Moats:** Along with the investments and new workflow processes, many incumbents might have a significant edge in the development of Gen-AI native apps. Some factors which will help out some names will be:

- **Agile Data Structure and Proprietary Data:** Any Gen-AI application will be based on a foundational model and sometimes, specialty models that will sit on top or be used instead of foundational models. For this reason, models need to be trained and built on high quality data sets. For any company to build out a good model, it will need agile data structures so that the model produces accurate and reliable results. Such architectures have the following characteristics:
 - **Modularity:** Modular data structures and data sets will help a company easily append, remove or migrate data. This will support the iterative development which is of utmost importance for AI models.
 - **Scalability:** As the size of the data increases, the framework and process need to adapt to the change in scale. The performance of the model should not be compromised for the scale of the data.
 - **Interoperability:** As the origins and the type of data vary, it should not create a problem for the tools and processes to run.
 - **Metadata Management:** High quality data means good quality and quantity of data. Along with that, a rich data set is also only useful when there is good context around the data. Metadata should be well managed for tools to make more sense of the data.
- **Existing ecosystems and developer communities:** Incumbents like

Salesforce, ServiceNow, and Adobe will have an ecosystem of applications whose efficiency and productivity will be enhanced by Gen-AI. Seamless integration of the new tools into old applications have always yielded the best results. Users of the apps will already be proficient in the app usage, and they will be more appreciative of the Gen-AI add on to the existing applications versus having to get reacquainted with a new application. Similarly, developer communities also have proficiency navigating around established frameworks. This will lead to faster adoption of the Gen-AI apps by incumbents.

- **Customer relationships and understanding of workflow needs:** The most important step towards the creation of an impactful Gen-AI app will be understanding the users' requirements. Existing names have strong, long-standing customer relationships, and they can leverage this to understand the user better and come up with the best possible solutions. Furthermore, incumbents also have the capability to tailor their solutions to customer-specific needs.

Exhibit 22: Investments, newly required workflow processes and moats for the Gen-AI tech stack development



For the exclusive use of CHRISTINA.LIU@GS.COM

Source: Goldman Sachs Global Investment Research

Security Software (Covered by Gabriela Borges)

In contrast to a key debate in application software broadly, Security Software companies are seeing a more meaningful near-term tailwind from Gen-AI adoption driven by factors including a measurable increase of AI-generated cyberattacks (i.e. Gen-AI acting as a “force multiplier” of traditional breaches) as well as Security vendors looking to monetize Gen-AI through tangible ongoing product cycles (e.g. next-gen Data Loss Prevention, Data Security Posture Management, etc.).

We expect Security vendors to emerge as beneficiaries across both the infrastructure and application layers as well as within data posture investments as security continues to trend higher as a percentage of total budgets.

- **Infrastructure:** Data center build-outs for supporting new compute/storage are driving additional need for third-party and/or white-box data center firewalls in both

IaaS and private data centers. We also see an emerging need for AI-infrastructure firewalls which embed security on chipset DPUs for traffic controls at the chip level. Palo Alto Networks specifically called out that growing traffic volumes are driving the need for higher throughput across a greater number in form factors in both public and private data centers.

- **Application:** New methods for securing large volumes of new code are required as CISOs increasingly prioritize the integration of Security earlier in the application developments cycle and as AI generated continues to scale across organizations. [NYU](#) and [Stanford](#) have published studies showing AI-generated code is generally less secure than perceived and Snyk has noted that >50% of developers commonly encounter security issues in AI code suggestions. As such, Cloud Security vendors increasingly cater their products beyond security analysts to developers more broadly.
- **Data:** An emerging product category called Data Security Posture Management (DSPM) allows organizations greater visibility into the various tools are being deployed as well as context around potential exposure and/or regulatory/data compliance risks. We are seeing a number of technology acquisitions with focus on driving differentiation through the accuracy on data classification including: Palo Alto Networks' acquisition of Dig Security, CrowdStrike's acquisition of Flow Security, Wiz's acquisition of Gem Security, and a number of other private company funding rounds, with the focus on data classification accuracy as the key differentiator. Moreover, organizations are increasingly leveraging next-generation DLP products as guardrails to stop data leakage into Gen-AI tools with broader use cases including data classification, policy control, & loss prevention.

Exhibit 23: Security products to benefit across infrastructure, application, and data posture investments

Category	Product	Description	Select Vendors
Infrastructure Layer			
Data center build to support new compute/storage	Data Center Firewalls / Zero Trust Network Access (ZTNA)	Mix of third-party and white-box data center firewalls in IaaS or private data centers; or ZTNA for replacing DMZ inbound traffic	PANW, FTNT, CHKP, ZS
Embedded security on chipset DPUs (data processing units)	AI-infrastructure Firewall	Traffic protocols programmed at the chip level	PANW, CHKP
Application Layer			
Writing secure code / Dev/SecOps	Code VM	AI applications need to be built securely; as GenAI generates more code, code needs to be checked for vulnerabilities	Snyk (private)
Protecting application once deployed	Cloud Workload Protection Platform (CWPP)	Advanced CWPP functionality to account for real-time communications with embedded third-party LLMs	Wiz (private), PANW, CRWD, S
Data Posture			
Visibility into what GenAI tools are being used	Data Security Posture Management (DSPM)	Similar to CSPM when SaaS apps inflected - lights up GenAI usage within the organization	PANW, CRWD
Guardrails to stop data leakage into GenAI tools	Data Loss Prevention (DLP)	Classic DLP, but updated for broader set of use cases	ZS

Source: Company data, Goldman Sachs Global Investment Research

Vendors are increasingly showcasing the application of Gen-AI in driving better Security efficacy at greater scale with the longer term goal being the automation of the Security Operations Center (SOC).

- **Separating signal from the noise** is one of the biggest challenges of modern SOCs, as an expanding threat landscape, talent shortages, & other complexities

across hybrid work environments increasingly burden the effectiveness of manual processes for responding to security alerts. As such, security platforms see an opportunity to consolidate security vendors by integrating Security functions from EDR/XDR to SIEM, Cloud, and beyond, and by leveraging Gen-AI to effectively demonstrate the platform value proposition (which historically had been more challenging despite the historical cycle of platform vs. point solution as a result of spending cycles). Early case studies show that Generative AI implementations can drastically reduce the time to respond to threats and increasingly working towards an at-least partially autonomous SOC. At the RSA conference in May 2024, Microsoft articulated its view that AI in Security will introduce a structural shift to typical pendulum swings between point products and platforms: with AI, signal intelligence across end products becomes critical for data analysis and remediation, creating a structural reason to move to platforms that will compound over time.

- **Security co-pilots**, similar to other co-pilot applications, will allow users to apply natural language for product usage, lowering hurdle the onboarding of Security analysts and reducing some talent shortages which currently pose a bottleneck in further product adoption. Microsoft shared early statics on the effectiveness of its Security Copilot including early career security practitioners being 26% more efficient and 45% more accurate when using the Security Copilot, and experienced practitioners are 22% more efficient and 7% more accurate. In addition to Microsoft, a number of Security vendors in our coverage have announced Copilots as part of the product integrations including CrowdStrike with Charlotte AI, SentinelOne with Purple AI, and Palo Alto Networks with its Stata Copilot product.

The role of IT Services in supporting this transformation

Acknowledging the undertaking that is modernizing the IT architecture, we discuss the role IT services companies will have in driving the tech-stack transformation needed for Gen-AI. For large organizations, IT Services will help companies consider how to best break down data silos and improve its data management systems to implement Generative AI applications. We expect this process to entail the following stages over a multi-year horizon: Strategy Assessment, Technical Proof of Concept, Modernizing Data to Prepare for AI Transformation and Executing the Broader AI Transformation.

What's the business case for hiring an IT Services company to do Generative AI?

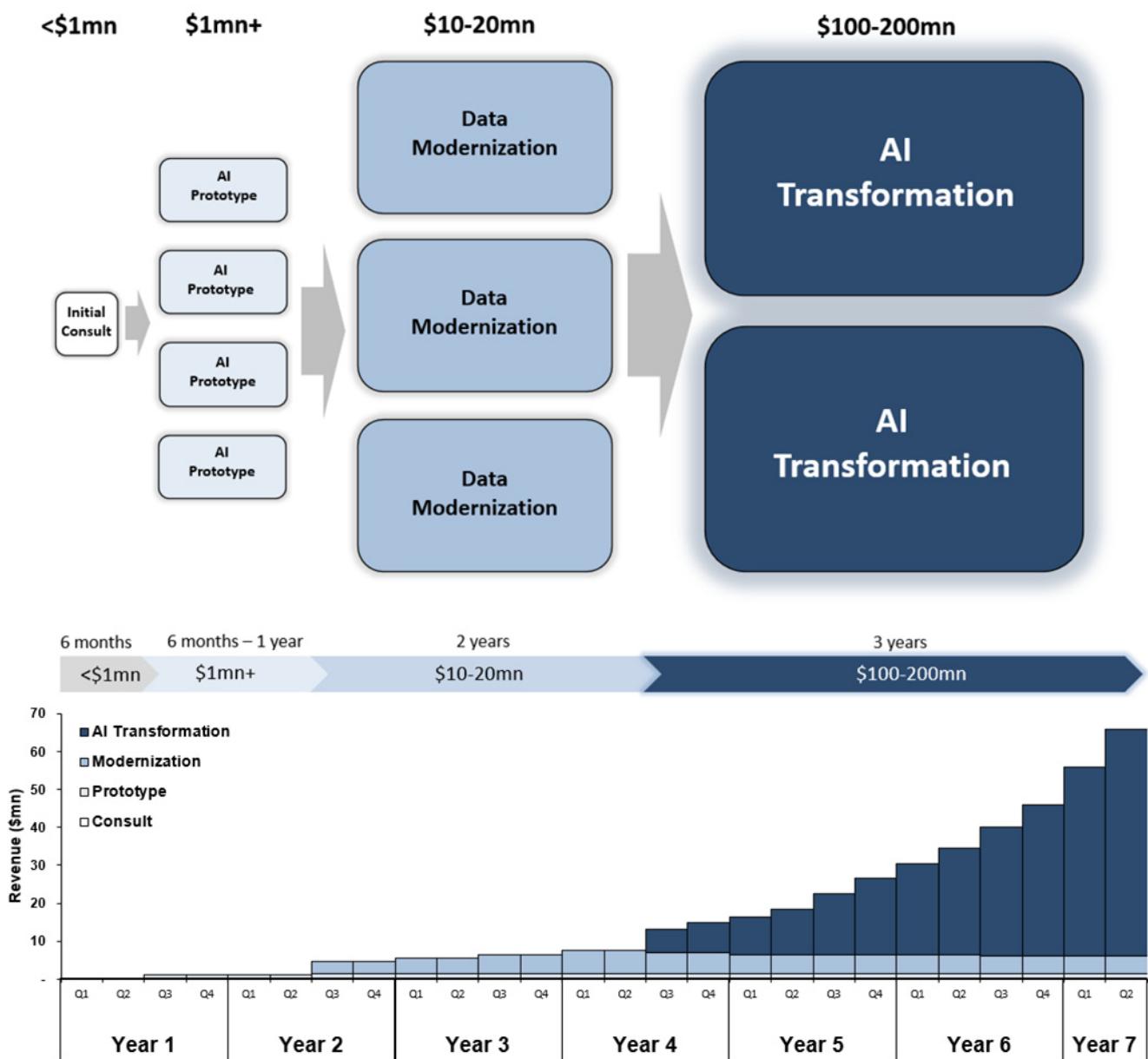
When considering how to implement Generative AI in a company, a corporate IT organization must choose whether to perform the work internally (using its own employees or a captive IT services entity), outsource the work to a software vendor (who likely has one or more point product offerings), or outsource the work to an IT Services vendor (who likely is using one or more third-party LLMs). To be sure, there will be some companies (in technology industries, for example) with significant internal expertise in data and analytics who will choose to do this work internally, as well as others (with large captive services organizations) who may decide to use internal headcount. However, we think there is a series of reasons why companies would choose IT Services firms for their Generative AI work:

- **Industry-level and organizational expertise:** IT Services providers take a

specialized, industry-specific approach to deliver solutions, and often have knowledge of a specific company (and business processes) as a result of a commercial relationship developed over years or decades which other third-party vendors (including many horizontal software vendors) lack.

- **Ability to choose among and adapt smaller LLMs to drive lower cost solutions:** In many cases, the largest general-purpose LLMs developed by OpenAI, Meta, Google, and others may be too large and costly to train for a targeted set of data within a corporation. By being able to evaluate and source multiple LLMs for a specific task and develop smaller, targeted versions for limited-scope applications, IT Services companies are well positioned to deliver solutions that can be executed with as little as 10%-15% (8X - 10X less) of the computing cost of larger models.
- **Expertise in integrating multiple applications and systems across the enterprise:** At their core, truly transformational enterprise AI projects will integrate insights across multiple parts of the organization, leveraging many databases tied to different applications, running on disparate systems and even different public/private clouds. Because of their ongoing knowledge and integration work across these different systems within a corporate organization, IT Services vendors are poised to capture a significant portion of many larger-scale projects.

Exhibit 24: Early AI engagements can lead to cascading AI transformation projects over the next 2-7 years



Source: Gartner, Goldman Sachs Global Investment Research

The Corporate AI journey: Ramp will be meaningful, but the flywheel takes time to spin. Although the largest hyperscale cloud providers (Amazon, Microsoft, Meta, and Google) have aggressively invested CapEx to satisfy both their own and their customers' AI needs, we believe that most corporations are at the very beginning of their AI transformation journey. This journey will ultimately entail an initial assessment of the company's capability and strategy for AI, as well as several smaller-scale "proof of concept" prototypes to prove the efficacy before any larger effort is undertaken. Importantly, if AI is to fulfill its promise of "transforming" insights from data stored within a company's IT estate, the company's applications and databases must first be modernized before they can be accessed and integrated into a dataset that can be easily

processed by an LLM. This implies that in many cases, a significant amount of preparatory work must be done to overcome an organization's "technical debt" to enable the AI transformation in the first place. In other words, the modernization of data and applications within an enterprise is a necessary - but not sufficient - condition for deploying AI at the enterprise level. Based on our checks, we present a notional sequence of project work that might take place in a typical organization on the road to a broader transformation:

- **Strategy Assessment:** For the vast majority of medium-size and large corporations, there is a lack of awareness about what AI can do, its limits, and what the best applications might be within an organization. To help shape this strategy, we believe many companies have commissioned early strategy engagements, which might be ~6 months in duration with a total project size of \$0.5 - \$1.0 million, to help inform the view of the C-suite.
- **Technical Proof of Concept:** Once an organization has a sense of the types of goals it can eventually achieve as part of its AI strategy, it is likely to commission a number of small technical proofs-of-concept or "testbed" projects. Typically, these projects start with a smaller number of the company's business and engineering experts, augmented by a lean external team of consultants who are well-versed both in the current commercially available LLMs and the client's industry. These projects will typically define a goal, perform closed-end analysis on a limited set of data, and deliver initial results and recommendations about whether to proceed with a broader implementation - thus providing a "roadmap" for future work. A handful of these programs might take 6-9 months to complete, at an average project size of \$0.5 - \$1.0 million per prototype. We believe the vast majority of corporate clients are currently at this stage.
- **Modernizing Data to Prepare for AI Transformation:** Once a set of broader AI objectives are identified for the organization, the most difficult work begins. In most cases (we would estimate over 90%), multiple applications and databases need to be modernized to prepare for LLMs to use and interact with the required data across the enterprise. As a result, a number of larger engagements must be launched to prepare for broader transformational projects. We estimate that on average, each of these smaller preparatory programs might take ~18 months with a typical size of \$10 - \$20 million, depending on the scope of the work required.
- **Executing the Broader AI Transformation:** At the point when the broader application and data stack is fully modernized, the most important modernization work can begin. We believe the earliest examples of such broader modernization include initiatives for content summarization and insights across the enterprise, automated customer relationship management, front-line client service, and so on. We think this type of engagement can command project sizes in the \$100 - \$200 million range with a duration of 18-24 months. Although we believe few corporations have reached this stage, there is anecdotal evidence of a handful of larger transformational deals reaching this size threshold.

Motivation and urgency around Gen-AI will catalyze tech-stack innovation. We believe most corporations will need significant help in developing a broad AI strategy.

The vast majority have significant technical debt that must be overcome before they can take advantage of AI on a broader scale. Thus, even with substantial motivation, management buy-in, and a business case based on firm ROI metrics, it will take most organizations several years to execute on this vision. The broader point here is that the biggest revenue opportunity or “payoff” for IT Services vendors is likely to come several years into the AI “hype cycle.” We believe the translation of AI work to services revenue will likely be significantly faster than it was for the cloud nearly a decade ago - mainly because we believe the motivation level and urgency for companies to participate in the AI transition is significantly higher.

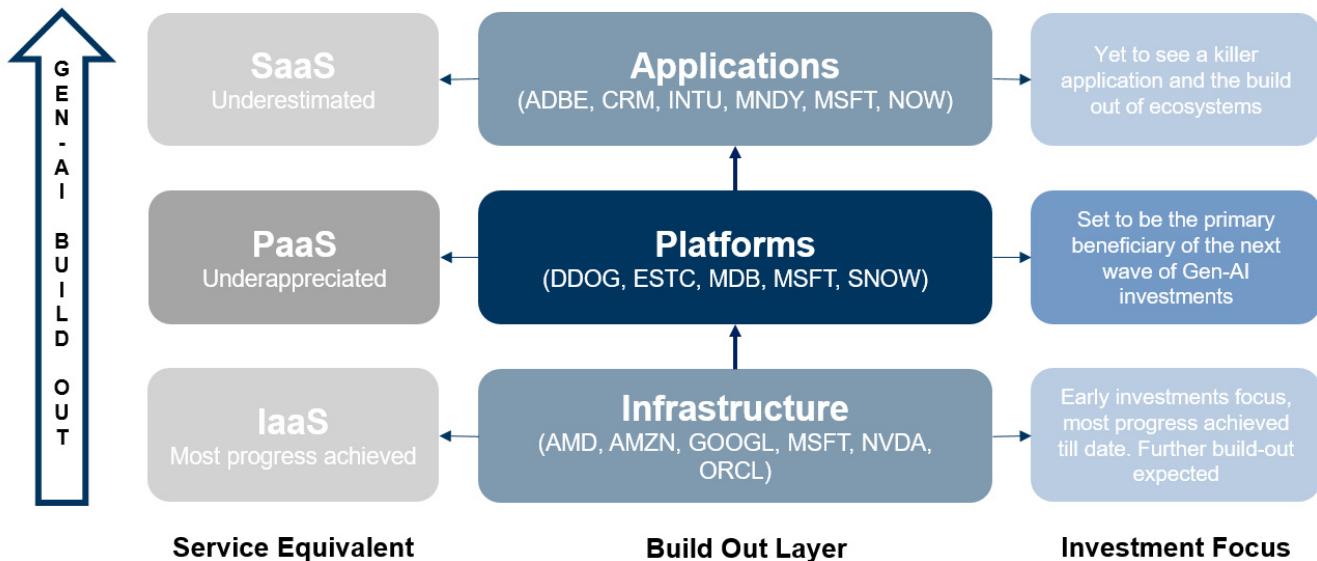
Promise of the Platform layer highly underappreciated today

Platform layer poised to be next beneficiary of AI investment cycle. Much of the value from Gen-AI has accrued at the infrastructure layer, predominantly benefiting from the training, fine-tuning, and inferencing from foundation models. The Hyperscalers have been the primary beneficiary of this early spending. However, we believe it's only a matter of time until value shift up to the platform layer ([Exhibit 25](#)), which will provide the foundational technologies and tools to enable tangible use cases of Gen-AI at the application layer. The ever-growing complexity prevalent in most enterprise IT environments is likely to remain a friction point on the path to broader AI-native applications, thus presenting an opportunity for both incumbent and emerging vendors to simplify the process. Below we highlight the key aspects of the platform layer, both old and new, that we believe will be required to enable Gen-AI at-scale. In particular, we highlight: **1)** The continued relevance of databases in the Gen-AI technology cycle, **2)** The potential for data platforms to emerge as a primary destination for AI app enablement, **3)** The role of end-to-end data management platforms in preparing an organization's data for Gen-AI, **4)** The importance of IT monitoring tools as new AI models are developed and IT complexity grows in lock step with Gen-AI adoption, **5)** The foundational role Vector Databases and Search will play in enabling RAG-based AI applications, and **6)** The advantages and disadvantages of various architectural approaches to enabling LLM-based solutions (Prompt Engineering, RAG, Fine-Tuning, and Pretraining).

Expect crux of value realization to occur in Platform layer as we move from fragmented data storage architectures toward collective data platforms. Ultimately, the cognitive component of this cycle is the main draw, with the industry moving toward a greater convergence of data with simplified user-experiences. We see the main change occurring in the platform layer as the industry moves further away from the fragmented data storage architectures toward more collective data platforms. With Generative AI, data will be: **1)** leveraged more easily than ever before and **2)** embedded into the core value proposition of software. The new components of the tech stack that enables such technologies are unlocking operational data to a greater number of knowledge workers whereas previously workers were only exposed to historical/reported data. This should allow for greater understanding of real time dynamics as they unfold, an element which we think has strong business use cases

over time in the application layer. We see the technological additions or extensions to the current architecture effectively working together to build a centralized procurement center for data that will allow for greater information discovery and be utilized for operational intelligence, risk management, product innovation, and efficiencies across organizations. In the context of the IPA framework, we see the value that can be created in the platform layer being the most underappreciated today.

Exhibit 25: Believe the platform layer currently remains the most underappreciated aspect of the Gen-AI cycle



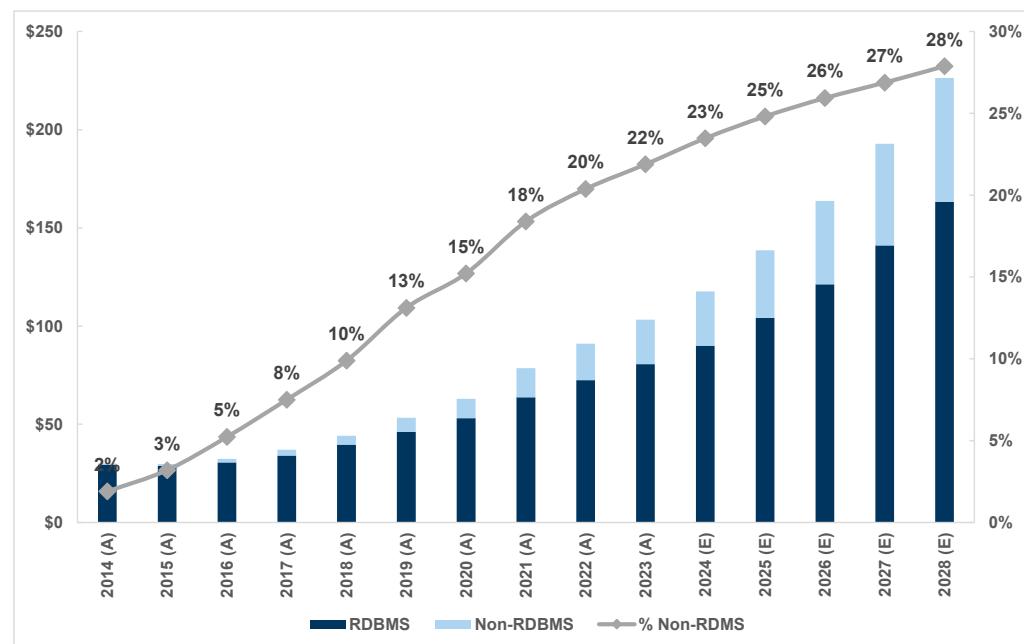
Source: Goldman Sachs Global Investment Research

Key components of the modern data stack

Databases will retain relevance in new technology cycle. Databases are one of very few technologies that have remained highly relevant through the last few computing cycles, whether that be internet, mobile, or cloud. This is unlikely to change with Gen-AI, in our view. The reason is clear: Gen-AI native applications will still require an underlying database to store transactional information as users interact with these new services. We expect the debate to center on whether we see growing standardization on newer architectures such as NoSQL (document, graph, etc.) or a continuation of current market dynamics where both relational and non-relational models coexist. NoSQL databases have gained favor in the market in recent years due to their efficacy in handling unstructured and semi-structured data, increasingly paramount in a world where data relationships are more complex. Take a social media platform with millions of concurrent users, for instance. We would point to the volume of unstructured and semi-structured data generated by the application, whether user profiles, posts, comments, likes and multimedia content. In this instance, a non-relational database that can store this information in a document format (e.g. MongoDB or CosmosDB) would be more suitable than a relational database given the volume and variety of data structures. Relational models have historically been better suited for applications with well-defined relationships. Think of an Airline Reservation Systems (ARS) that manages interrelated data such as flight schedules, seat availability, customer information and reservations. Relational databases are a natural fit as they support high transactional integrity -

ensuring all database operations are completed and data remains consistent. The structured and related nature of the data also lends itself to the format of relational models (rows and columns).

- **NoSQL may be better suited for AI-native apps...** With many AI-native apps needing to rely heavily on the processing of text, images and audio, we believe non-relational databases could emerge as one of the primary destinations for AI applications. The database requirements for next-gen applications will predominantly center around unstructured data processing, scalability, and data model flexibility. These characteristics tend to be present in document models which offer a schema-less design that contrasts the more rigid architecture of relational models that work with rows and columns. The incorporation of vector embeddings that sits alongside customers' operational data could create more of a draw for enterprises that already maintain portions of their database environment in non-relational architectures, especially as standalone vector database options emerge (Pinecone, Milvus). As such, we do envisage a scenario in which NoSQL providers capture an outsized portion of new workloads relative to the historical trend line observed in the database market, which has already been favorable as it has scaled from 2% of database spending in CY14 to 22% in CY23, per Gartner ([Exhibit 26](#)). MongoDB's document model, along with competing offerings from Azure (CosmosDB) and AWS (DocumentDB) should be considered as primary beneficiaries. For example, OpenAI leverages CosmosDB to run its ChatGPT service.
 - **...but don't rule out relational.** While NoSQL databases (document, standalone vector) may have a leg up when it comes to net new applications, we can't overlook the significance of enterprises' current relational database footprint that can continue to benefit incumbents as they evolve their offerings to meet new market demands. We still expect relational databases will play a role considering: 1) Many organizations still run their mission-critical apps on relational databases, which are very difficult to migrate off of, 2) Relational models are highly performant in storing critical structured data (e.g. patient records, financial transactions, inventory catalogs); there will be significant opportunities to modernize longstanding legacy apps by infusing AI. Oracle, for instance, released JSON Relational Duality Viewer that enables customers to view stored relational data as simple JSON documents, also working with its AI Vector Search feature. In addition, PostgreSQL (#4 relational database, per DB-ENGINES) introduced the PG Vector extension in Aug. 2023, providing the core database with traditional vector capabilities; according to Retool's State of AI report (2024), PostgreSQL is the most frequently used vector database (21.3%).

Exhibit 26: Non-Relational could accelerate share gains due to unstructured data's relevance to Gen-AI

Pre-RDBMS spending is excluded from the chart but is reflected in the Non-RDBMS % of Total

Source: Gartner, Goldman Sachs Global Investment Research

Data platforms could become a natural destination for AI app development. Data warehouses, data lakes and lakehouse architectures represent online analytical processing systems (OLAP) that are complementary to database technology (such as online transaction processing, known as OLTP), combined representing the broader database management systems market. These data platforms were purpose-built with the ability to manage petabytes of enterprises' structured, semi-structured and unstructured data that are critical to businesses' operations. Ultimately, we expect these data platforms (OLAP systems) to play a significant role in enabling Gen-AI applications due to their extensive repositories of mission-critical enterprise data that is uniquely suited for the customization of LLMs. Before unpacking their role in enabling Gen-AI, it's worth breaking down the distinction between data lakes and data warehouses as well as the hybrid lakehouse architecture.

- **Data Warehouses** are best suited for handling structured data from systems of record, including data types such as customer profiles, sales transactions, financial information, or inventory catalogs. Data warehouses are most effective for running analytical queries (leveraging SQL commands) on high-quality datasets to extract insights and build visualizations to augment business intelligence and reporting.
- **Data Lakes** are best suited for storing diverse data types in their raw format, including text, video, and images as well as structured data. They tend to be more cost-effective as a data repository; warehouses require data to be transformed into a unified format before (ETL) or after (ELT) being loaded in, a compute-intensive process. Since Data Lakes are better suited to handle diverse data types, they tend to be best positioned for data science and ML, big data analytics, and data engineering workloads.
- **Lakehouses** combine the best of both by providing the raw storage and structured

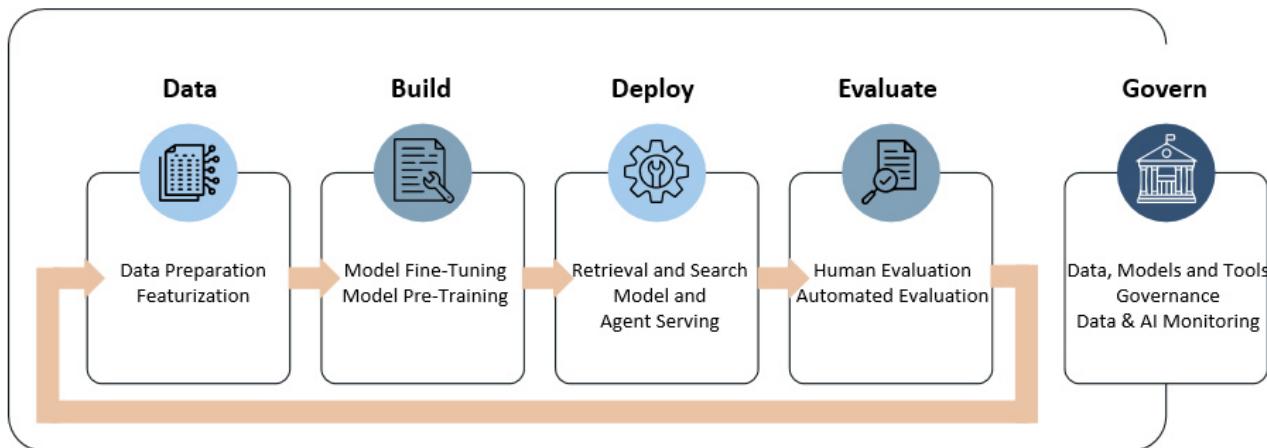
data formats, as well as a processing engine that supports ML workloads via Apache Spark and SQL commands for structured data analysis.

Early examples of data platforms leaning into the Gen-AI opportunity. Over time, we expect companies such as Snowflake and Databricks to broaden their platform tooling to democratize the development of LLM-based solutions, helping customers unlock more value from the data that resides both inside and outside their respective platforms. Databricks' Mosaic AI already holds the promise of providing a unified platform for the entire AI development lifecycle, from data preparation to development, deployment and monitoring, and evaluation ([Exhibit 27](#)). Snowflake is similarly making strides with the launch of Cortex AI, which offers a broad array of services that includes interfacing with proprietary data through natural language, developing RAG-based applications, and fine-tuning Foundation Models ([Exhibit 28](#)). Snowpark Container Services (SCS) will provide organizations with access to GPU-powered compute and enable the development and deployment of AI applications securely within the Snowflake ecosystem. While many of these services remain in private or public preview and enterprises more broadly have remained in the evaluation stage for Gen-AI (consistent with our IPA framework), we believe those platforms providing the core capabilities to simplify organization's ability to extract new value from troves of mission-critical enterprise data will emerge as primary beneficiaries of the next wave of the investment cycle. We would also highlight Scale AI's emerging role at the data layer, providing its core data engine to apply supervised fine-tuning, reinforcement learning from human feedback (RLHF) and model evaluation (via Scale Evaluation) to pretrained models.

Exhibit 27: Databricks' Mosaic AI provides end-to-end platform for LLM-Based / AI-native application development

Mosaic AI

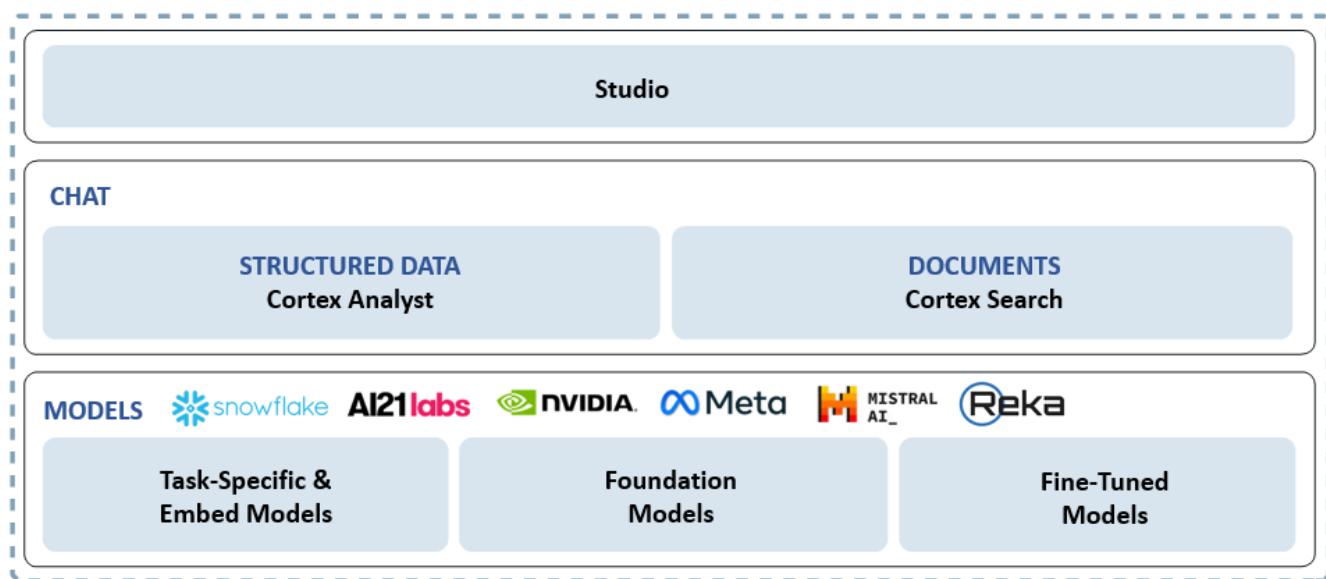
Build Production Quality Compound AI System



Source: Databricks , Goldman Sachs Global Investment Research

Exhibit 28: Snowflake's Cortex AI services provide multiple avenues to extract value from enterprise data

Cortex AI



Source: Snowflake, Goldman Sachs Global Investment Research

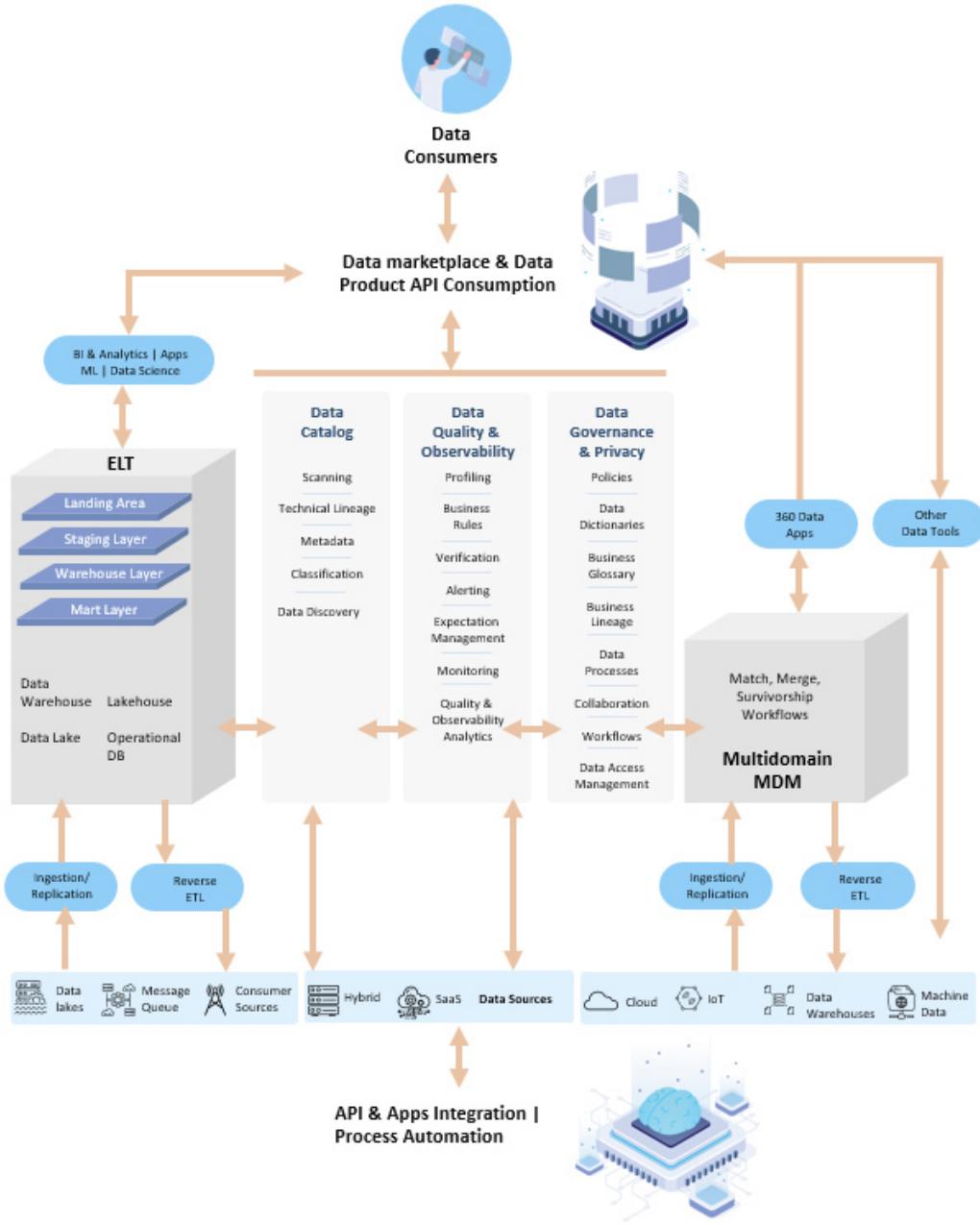
Strong data foundation essential to Gen-AI enablement. While the adage that there is no AI strategy without a data strategy has become an increasingly commonplace saying, we believe there is merit to the phrase. Organizations that seek to leverage or implement AI in their applications will require robust data management processes. Consider the issue of accuracy and bias in language models. For enterprises to reap the benefits from Gen-AI, whether through internal productivity gains or new revenue streams, they must have a solid data foundation to ensure the information fed into the AI models is of high-quality to minimize hallucination and bias. We believe this will ultimately amplify the importance of data management platforms due to their innate ability to qualify and standardize data across many source systems. Necessary steps in this process include adopting services that aid with data cataloging, data quality and governance, amongst others, to ensure language models underpinning next-gen applications are performant enough for mission-critical business use cases such that hallucinations become statistically insignificant. As a basic example, we'd point to the varying ramifications of ChatGPT incorrectly answering a consumer prompt (e.g. which restaurant to go to in a vacation destination) vs. an enterprise-grade Copilot fine-tuned on proprietary data providing an unfounded output to the end user, which has practical business and regulatory implications.

- **CIOs likely to prioritize services supporting data quality, cataloging, and governance.** We highlight a few key data management services we expect to be increasingly strategic as more enterprises seek to develop and adopt LLM-based solutions ([Exhibit 29](#)): **1)** Data Quality will help ensure the integrity of AI model training data through data profiling, standardization, and cleansing, **2)** Data Governance can help organizations with regulatory compliance and data privacy via data masking and access controls, **3)** Data ingestion (e.g. ETL or ELT) supports the development and orchestration of data pipelines for structured and unstructured

data alike, which we expect to play an important role in the modern data stack as enterprises push down mission-critical data into LLMOps and inferencing platforms (e.g. Databricks, Amazon Sagemaker) for AI enablement, and **4) Data cataloging** aids organizations in efficiently creating an inventory of all data assets, which can streamline the data preparation process for model training.

- **External survey of Chief Data Officers underscores need for data management.** With many organizations still managing the complexity of disparate tools and data silos across many source systems, we expect CIOs to prioritize the implementation of modern data management platforms to ensure organizational readiness for Gen-AI (e.g. Informatica, Collibra, Fivetran). For instance, we note that in a [survey](#) conducted by Informatica (over 600 Chief Data Officers, Jan. 2024), 42% of data leaders cited data quality as the main obstacle to Gen-AI adoption and 40% cited data governance and privacy. Further, 58% of respondents indicated that they need five or more data management tools to manage their data estates. These results underscore the importance of adopting an end-to-end data management platform that can better prepare organizations' data to be leveraged for Gen-AI. We expect over time these companies will enhance their strategic relevance by incorporating AI functionality to automate routine tasks and free up data practitioners for more strategic work, akin to Informatica's recent launch of [CLaIRE GPT](#).
- **Highlighting private players in the data management ecosystem.** On the private company side, we highlight dbt labs' role in the data transformation process, with close ties to data platforms such as Snowflake and Databricks. Through dbt Cloud, the company can ensure data consistency for critical business metrics, improve data quality by running automated data tests and provides its Discovery API to help customers gain visibility into metadata that can serve to enrich LLMs. These processes represent an important step in data preparation for downstream data tools and platforms that enable AI application development. We also highlight Collibra, a data intelligence platform with an emphasis on data quality, data cataloging and data governance. The company's Feb. 2024 announcement of [AI Governance](#) stands out, which helps customers remain compliant with legal and privacy policies, mitigate data risks, improve model performance and ROI, and accelerate time to production.

Exhibit 29: Illustrating the components of a modern data architecture



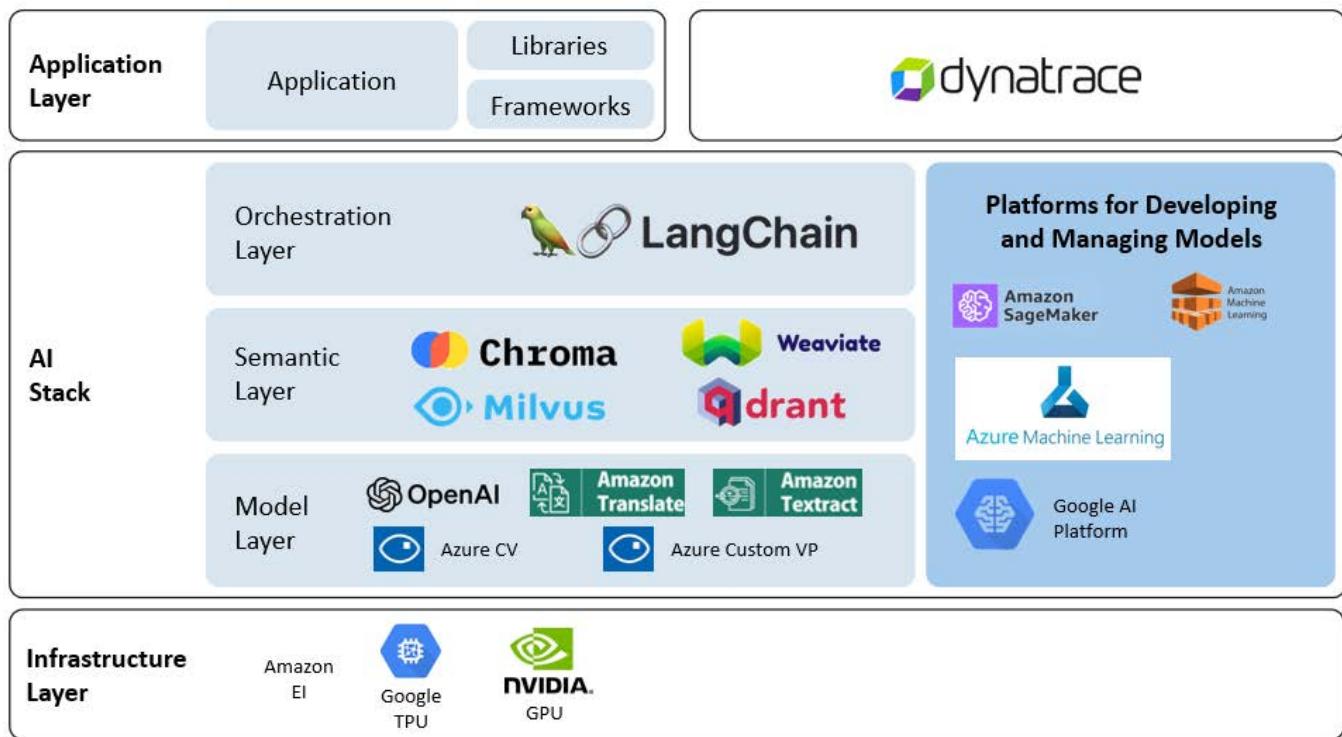
The evolving role of monitoring tools

AI model adoption will amplify the importance of IT monitoring tools. Another critical aspect of the platform layer is observability, which includes the monitoring of an organization's IT environment and applications. We expect the secular tailwinds of cloud modernization and digital transformation to remain a boon to the broader market growth rate, though Gen-AI can further necessitate the adoption of best-in-class observability platforms due to the additional complexity it will inject into customers' IT environments in addition to driving accelerated data creation as AI-native applications come to market. Observability as a category has historically been effective in adapting to emerging market trends and new tech cycles, akin to the growing relevance of cloud-native

vendors such as Datadog and Dynatrace as technology shifted from on-prem to the cloud. This was perhaps the most significant inflection point as cloud adoption drove the need for full-stack observability, which unifies IT monitoring, APM and log management on a single data platform. As organizations shifted to cloud, the number of accessible software tools, compute availability and the rate of app development all increased exponentially, requiring a platform that can connect into and monitor constantly evolving infrastructure and application requirements that on-prem environments and tools were not accustomed too. We see Gen-AI's impact on observability as less disruptive and more complementary to best-in-class platform providers when compared to prior tech cycles as this time around they possess the necessary architectures to maintain their relevance. Consider the snowball effect of thousands of enterprises adopting AI models to power their next-gen applications – this will translate into more compute instances, more applications and more data. This is the complexity that cloud-native, full-stack observability platforms were created to manage.

- **Gen-AI to present new monetization vectors for Observability platforms.** One significant market shift in Observability centers on the expected proliferation of small and large language models that will power SaaS applications. While we remain in the experimentation phase for AI application development, we expect a palpable growth opportunity will emerge around LLM Observability in the not-so-distant future. Just as the infrastructure and application layer require continuous monitoring to gauge system health and performance, AI models will need much of the same. We see observability platforms tapping into their core competency in full-stack monitoring to extend their value proposition to AI models themselves, particularly as more enterprises lean into RAG and fine-tuning for AI app enablement. Datadog and Dynatrace have already released products dubbed LLM Observability and AI/ML Observability, respectively. These services will provide the ability to track an AI models success rate in generating an accurate response, the time it takes to fill a request (latency), abnormal usage patterns, model drift (hallucination) and token costs. As mentioned prior, AI model monitoring is one aspect of a broader opportunity, which will likely extend to other aspects of the tech stack that enable these applications such as Vector Databases and GPU-powered infrastructure ([Exhibit 30](#)).

Exhibit 30: AI Observability will prove critical to all layers of the tech stack as enterprises build and deploy AI applications

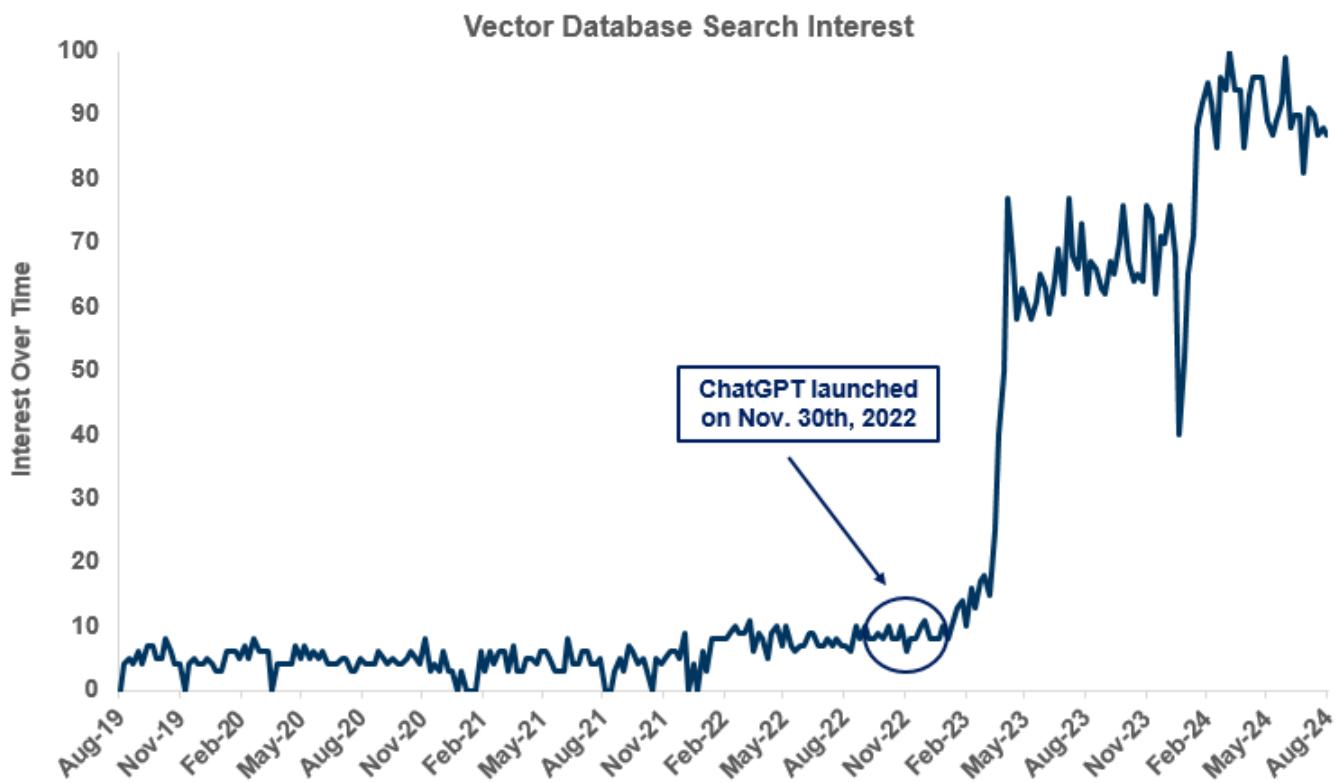


Source: Dynatrace, Goldman Sachs Global Investment Research

New components of the modern data stack

Vector Search to act as the retrieval engine for AI applications. As discussed above, databases will remain a foundational technology for AI native applications. However, traditional databases alone will not be enough. Vector Search will underpin many of the use cases emerging in applications today, such as Virtual Agents and Copilots. Vector Search indexes data uniquely in the form of vectors embeddings, which are numerical representations of the data rather than text. These embeddings allow for AI models to better capture the semantic meanings and relationships in unstructured data (e.g. words, phrases, images, videos). When an AI model queries an index based on a specific end-user prompt, Vector Search enables AI models to compare the query to these vector embeddings and pull in the relevant context to address the prompt ([Exhibit 32](#)). Vector Search is underpinned by Nearest Neighbor algorithms which pull in the most relevant (or semantically similar) information pertaining to the query, even if it is not an exact match. This differs from traditional search which relies on exact matches of keywords. For instance, an e-commerce site may leverage vector search for image-based product search, which will enable customers to find similar products based on uploaded images. While Vector Databases have been around for more than a decade, the technology has gained renewed interest ([Exhibit 31](#)) due to its applicability to Gen-AI workloads, particularly with language and multi-modal models. LLM-based applications with natural language interfaces requiring interaction with a knowledge base will ultimately require some form of retrieval capabilities, underscoring the importance of Vector Search and Vector Databases.

Exhibit 31: Search interest for Vector Database spikes post-ChatGPT launch; Interest remains at record highs



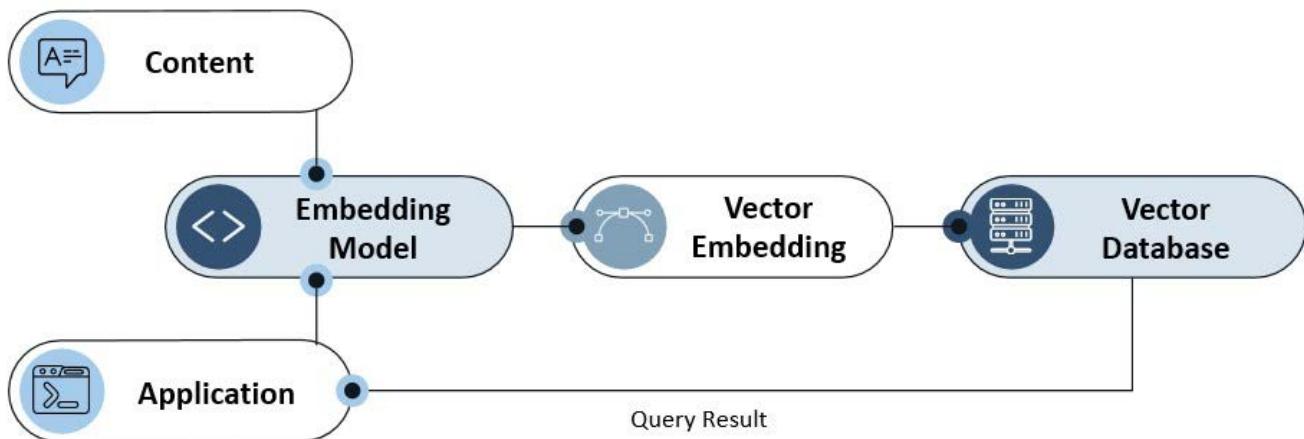
Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Source: Google Search Analytics, Goldman Sachs Global Investment Research

- **Understanding the algorithm(s) underpinning Vector Search.** We note that there are typically two distinct approaches to Vector Search, 1) Hierarchical Navigable Small World (HNSW), also known as Approximate Nearest Neighbor (ANN) and 2) Exhaustive KNN. The distinction between these two approaches centers on latency, scalability, accuracy, and computational costs. HSNW algorithms tend to be less computationally intensive and provide better query performance against larger datasets, with a resulting tradeoff of less exact results (on a relative basis). With KNN, the algorithm runs the vector query (prompt input) against all known datapoints in the index, making it computationally more intensive (and expensive) and slower from a query performance standpoint, but provides the benefit of producing more exact results. While both approaches have merit for certain use cases, we believe that the ANN algorithms are likely to see broader-based adoption for the purposes of information retrieval applications due to an adequate compromise between faster performance and lower computing costs vs. a slightly less accurate response output. ANN algorithms will be best suited for real-time applications with larger datasets, whereas KNN will play a role in applications with smaller datasets where exact nearest neighbor responses are required. One example for KNN would be classifying medical conditions based on patient data. Ultimately, we expect that most database providers will offer services for both HSNW and KNN depending on the customers use case.

Exhibit 32: Vector Databases gaining renewed relevance due to importance of retrieval capabilities for Gen-AI applications

Process Explained: Convert data in vectors using embedding model —> index and store vectors in Vector Database

The Rise of Vector Databases

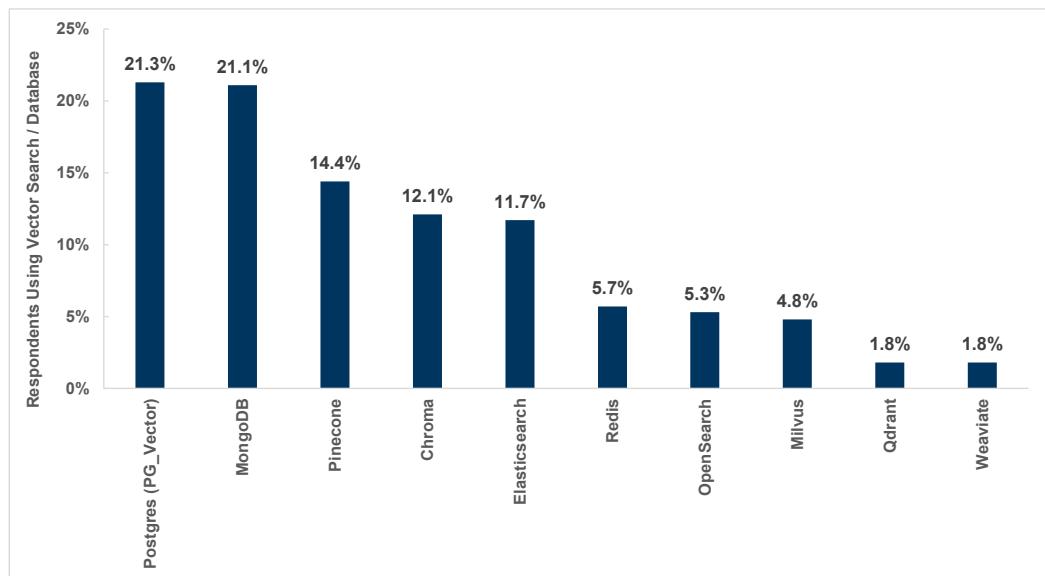
Source: Goldman Sachs Global Investment Research, MongoDB

- **Comparing Vector Search tooling to standalone Vector Databases.** As database players have quickly introduced new tooling and services to enable Gen-AI applications, particularly vector capabilities, we note that a diverging trend has manifested between standalone Vector Databases like Pinecone and Milvus (which saw strong initial excitement post-ChatGPT launch) and database providers such as MongoDB and Azure (CosmosDB) providing Vector Search as a native component embedded in the core platform. Incumbent database providers quickly launched Vector Search tooling with an appreciation that this is likely to become one of the primary technologies underpinning AI applications. We believe OpenAI's most recent acquisition of Rockset, a Vector Database provider, is validation of the strategic relevance of this technology. OpenAI intends to leverage Rockset's retrieval infrastructure to power its future product suite ([link](#)). In existing database architectures (particularly non-relational) that provide Vector Search, customers can store vectors alongside operational data, providing unified data management and reducing the need to manage multiple databases (e.g. tool sprawl). A few advantages of standalone Vector Databases include: **1)** Purpose-built for vector operations, which can make them more performant for certain workloads and **2)** Provide potential scalability advantages in instances where large vector collections are required. We believe ultimately the selection criteria for leveraging Vector Search compared to a Vector Database will center on the price, performance, and incremental complexity to adopt (or lack thereof).
- **Database architecture shifts historically benefit incumbents.** According to Retool's State of AI report (2024), respondents indicated that PostgreSQL's PG Vector was the most used Vector Database at 21.3%, while MongoDB was a close second at 21.1% and Pinecone a distant third at 14.4% ([Exhibit 33](#)). Despite Pinecone being founded in only 2019, we believe their success to date is validation of the quality of its technology despite lacking the distribution footprint and

developer mindshare of PostgreSQL and MongoDB, which rank at #4 and #5 in DB-ENGINES complete rankings. The lack of a killer application to date suggests that the current landscape and preferences could evolve further, with no clear winner(s) yet to emerge. The advantage could arguably reside in developer mindshare, distribution footprint and platform simplicity considering newer database architectures and companies have historically been hard-pressed to disrupt the incumbents' stranglehold on the database market given their scale advantages and ability to replicate emerging tech trends; consider Amazon and Microsoft's successful diversification into non-relational architectures such as document, graph, and time-series. The Hyperscalers and large-tech (IBM, SAP) still control >80% of the database market ([Exhibit 34](#)), with MongoDB the only recent standalone operational database to scale to >\$1B in ARR (ex. data platforms such as Databricks and Snowflake, [Exhibit 35](#)). The Hyperscalers have already released Vector Search functionality, including Azure's Vector Database in CosmosDB and Vector Search for DocumentDB (AWS).

Exhibit 33: Retool State of AI Survey: Most frequently used vector Database / Search offering

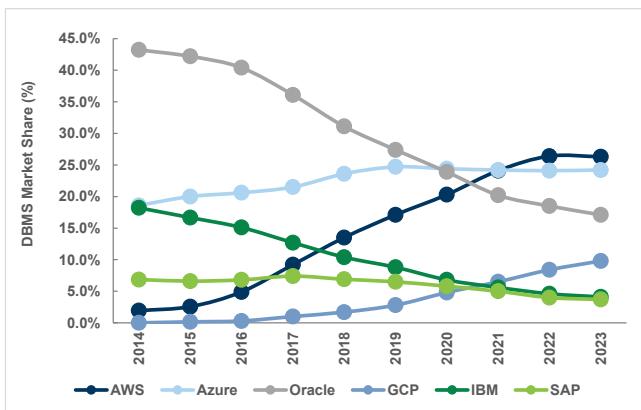
63.6% of respondents reported using a Vector DB in June 2024, up from 20% in prior survey



The responses from the Retool survey were gathered from 730 people in April 2024; 1-99 employees (59.3%), 100-999 employees (28.1%), 1000+ employees (12.6% of respondents)

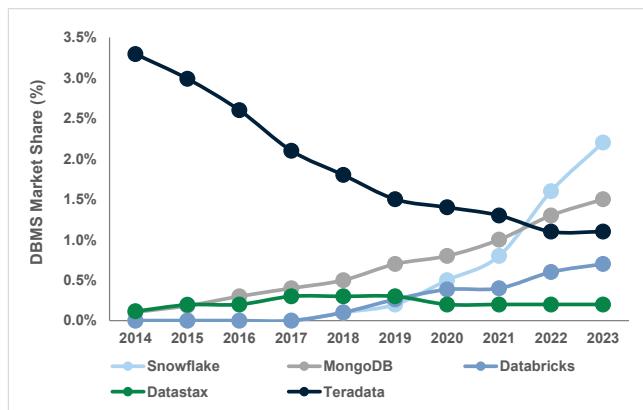
Source: Retool

Exhibit 34: 'Big Get Bigger' dynamic in Database market reflects Hyperscalers' ability to stay relevant as new technologies emerge



Source: Goldman Sachs Global Investment Research, Gartner

Exhibit 35: Despite Hyperscalers' success in DBMS, pure-plays can scale with differentiated technology (SNOW, MDB, Databricks)...
We anticipate an open debate over the standalone success of pure-play Vector Database providers vs. platforms



Source: Goldman Sachs Global Investment Research, Gartner

Addressing the architectures best suited to enable Gen-AI

As we consider the path towards AI-native applications, we believe there is merit in breaking down the various architectural patterns that have emerged to enable LLM-based solutions ([Exhibit 36](#)). Namely, this includes **1) Prompt Engineering, 2) Retrieval Augmented Generation (RAG), 3) Fine-tuning and 4) Pretraining**. As a rule of thumb, Prompt Engineering tends to be the simplest and least computationally intensive method of enabling Gen-AI, whereas Pretraining tends to be most expensive and complex (think Foundation Models). In the below section, we walk through the various architectures, the advantages, disadvantages and use cases most suitable for each. Understanding the technology underpinning new applications can help investors assess the strategic positioning and moat of emerging AI businesses. Further, we believe the platforms enabling the development of AI applications will emerge as primary beneficiaries of the nascent AI tech cycle, akin to the already exponential ramp in AI revenue attributed to the Hyperscalers for LLM training and inferencing.

Exhibit 36: Architectural patterns best suited to enable Gen-AI applications and their associated complexity

Evaluating Gen-AI architectures across seven key parameters

Gen-AI Architectures	Prompt Engineering	RAG	Fine-Tuning	Pretraining
Development Effort	Low	Moderate	Medium-High	Very High
Customization	Low	Medium	High	Very High
Data Requirements	Low	Medium-High	High	Very High
Maintenance	Low	Medium	High	Very High
Time to Implement	Short	Medium	Long	Very Long
Cost	Low	Moderate	Medium-High	Very High
Adoption Level	Very High	High	Medium	Low

Source: Goldman Sachs Global Investment Research

Prompt Engineering to be widely adopted but lacks domain-specificity. Prompt engineering, otherwise known as in-context learning, involves tailoring the question prompt to elicit the most accurate (or desired) results from a pretrained LLM without altering or augmenting the model itself. The benefit of prompt engineering lies in its relative simplicity to set up given it requires no specialized model training or context from external data sources. Consider the example of an AI chatbot developed for an application - through prompt engineering, developers can specify the primary task of the LLM (e.g. provide information or make recommendations), influence the style of the response generated and improve the relevance of the output. These parameters can be adjusted according to the nature of the application. Prompt engineering has broad applicability across a variety of use cases, including customer support, law, education, and online travel, to name a few. However, just as pretrained foundation models are becoming increasingly commoditized, we believe that enterprise-grade SaaS applications leveraging a prompt engineering architecture will lack the differentiation necessary to find viable commercial success in the market given the continual disruption risk (as it lacks domain-specific, proprietary knowledge). Even so, prompt engineering will play an important role for less demanding (albeit important) tasks that don't require supplemental context such as product review summarization or personalized product suggestions. As the least cost-prohibitive method of operationalizing Gen-AI in applications, we believe organizations will leverage prompt engineering to automate more routine tasks and services (both internal and external) to drive down costs.

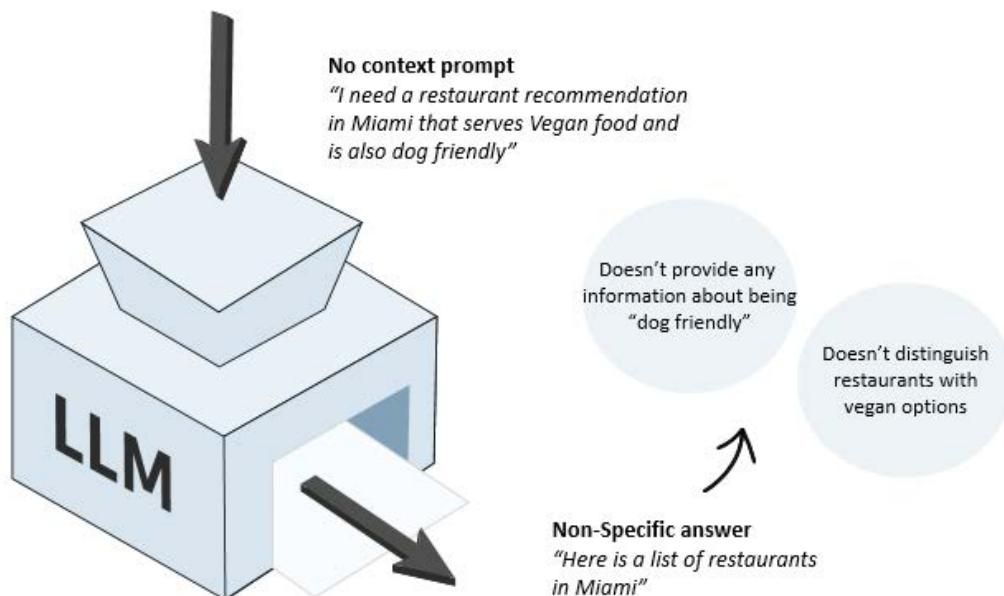
Retrieval Augmented Generation (RAG) augments the capabilities of pretrained models with proprietary data. Retrieval Augmented Generation is definitionally comparable to Vector Search. The simplest way to understand RAG is that it combines the best-in-class retrieval capabilities of Vector Search with the contextual relevance from generative models such as LLMs ([Exhibit 37](#), [Exhibit 38](#)). Said differently, Vector

Search acts as the retrieval engine that can source contextually relevant raw data but lacks the capabilities to convert this information into a digestible format for the end-user, for which a generative model is required. We believe RAG architectures will be a critical element to enabling AI-native applications, which already underpins some of the earliest use cases of Gen-AI such as customer service chatbots and consumer search engines (e.g. Perplexity AI). While there are multiple ways to enable AI-native applications, the benefit of the RAG architecture is rooted in its relative simplicity (more complex and resource-intensive than prompt engineering but less so vs. fine-tuning or pretraining) through a combination of proprietary enterprise data, Vector Search and Foundation Models (OSS or closed models).

- **RAG's ability to source external data will prove critical for enterprise-grade applications.** The concept of RAG is best illustrated by drawing a basic example between two highly relevant Gen-AI consumer products, ChatGPT and Perplexity AI. With ChatGPT, responses are generated based on the preexisting data used to train the model up to a certain time period (e.g. pretrained model). Consumer search-engine Perplexity AI, alternatively, leverages RAG to enhance its generative model by retrieving real-time information from external sources to improve the relevance and context of the response. We believe RAG architectures will be most suitable for use cases in which customers seek to take advantage of the generative capabilities of LLMs, while augmenting it with non-sensitive (at least initially), proprietary information that can enable use cases such as clinical trial analysis in healthcare, personalized chatbots in e-commerce or online travel, and anomaly detection in manufacturing. In addition, RAG architectures are very well-suited for applications with real-time data requirements due to external retrieval capabilities that fine-tuned and pretrained models lack. It's important to note that since RAG architectures rely on data retrieval from external sources, data privacy and security will remain a consideration for enterprises seeking to adopt the technology (e.g. PII, financial data, health records).
- **Database providers are already taking the initial steps to simplify RAG adoption.** With many organizations still evaluating the prospects of both modernizing their critical applications and building new AI-native applications, database providers are already taking steps to simplify RAG adoption ahead of what we expect will be a significant investment cycle (consistent with our IPA framework). We highlight MongoDB's AI Applications Program (MAAP), announced on May 1, 2024, which helps organizations quickly develop and deploy modern apps infused with Gen-AI at enterprise scale. MongoDB will provide large customers with strategic advisory, professional services and access to an end-to-end tech stack in partnership with leading AI companies such as AWS, GCP, Azure, Anthropic, LangChain, Llamaindex and Fireworks.ai. Similarly, private database player DataStax has introduced DataStax Langflow, providing developers with a visual framework that simplifies the experimentation, design and testing of Gen-AI applications. While these examples are not exhaustive, we believe that investments into a supporting ecosystem at the platform layer to simplify AI app development will serve to reduce the friction associated with bringing AI apps to market, especially as tangible use cases emerge.

Exhibit 37: LLMs without RAG lack context of proprietary enterprise data; Reliance solely on base model training data limits domain-specific use cases

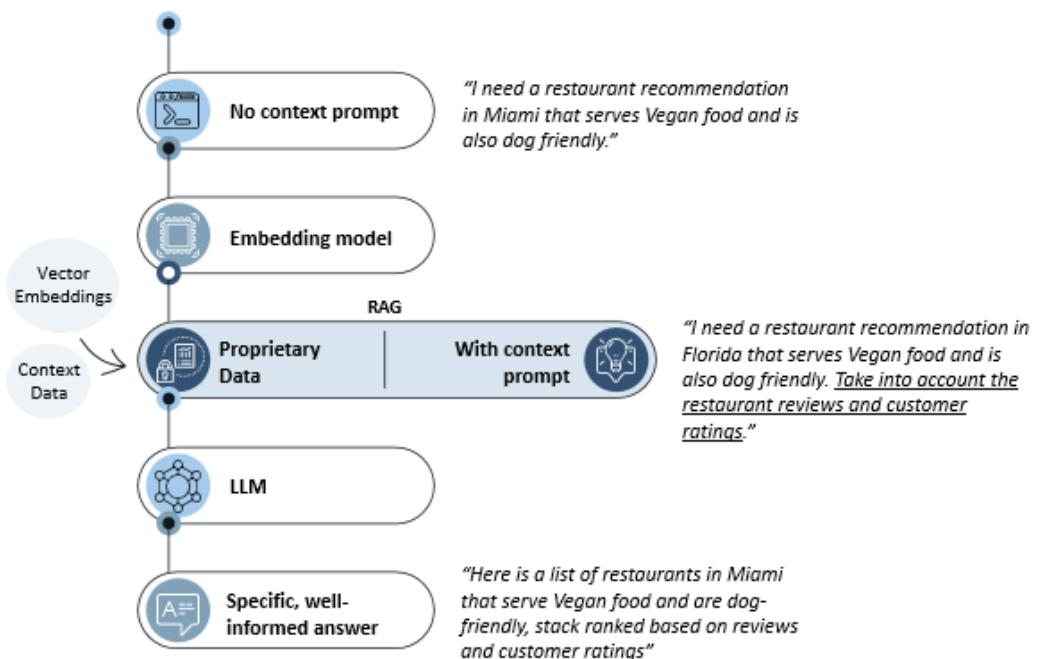
LLM Not Augmented with RAG



Source: Goldman Sachs Global Investment Research

Exhibit 38: LLMs augmented with RAG support retrieval of proprietary and real-time information from 3rd-party sources, allowing enterprises to tap into differentiated data to augment LLMs

Augmenting an LLM with RAG



Source: Goldman Sachs Global Investment Research

Fine-Tuning supports high levels of customization and model ownership, proving attractive to privacy-sensitive organizations. Fine-tuning an LLM builds upon the basic principle of the RAG architecture, which is to provide factually and contextually accurate responses to the prompt. At its core, fine-tuning enables organizations to train

existing Foundation Models (e.g. ChatGPT-4) on proprietary data to improve their efficacy in executing against domain-specific tasks. The important distinction relative to RAG beyond improved performance on fine-tuned tasks and more precise control over model behavior (e.g. tailored to local languages or internal business processes) is the ability for enterprises to keep sensitive data within their own IT environment. This is consequential for large organizations that retain significant IP and sensitive customer records, whether from a regulatory, privacy or security perspective. While possessing notable privacy advantages to RAG, we note that fine-tuning is more time-consuming and resource-intensive. Fine-tuned models are also limited to the data they are trained on, whereas RAG architectures can support real-time information retrieval. The simplified process for fine-tuning is as follows: 1) Select pretrained LLM, 2) Collect and label the dataset for training the LLM, 3) Set up training environment with GPUs and deep learning framework (PyTorch, TensorFlow), 4) Initiate fine-tuning process, and 5) Evaluate model performance and deploy into production. We believe fine-tuning will emerge as one of the primary methods to enable Gen-AI applications due to an organization's ability to "own" the model (maintaining data privacy) and tailor the model based on domain-specific data. We expect fine-tuning to have a high degree of relevance for organizations in regulated verticals such as public sector, healthcare, and financial services, amongst others.

Pretraining relevant to Foundation Model providers but resource and cost

constraints will restrict broader adoption. Pretraining an LLM is often the most complex and resource-intensive approach to enabling Gen-AI due to the need for large-scale data preprocessing and running long-form GPU training jobs. Pretraining has historically been tied to the FM providers such as OpenAI (ChatGPT), Anthropic (Claude), Google (Gemini), Meta (Llama), and AWS as well as a new class of smaller language models developed by companies such as Databricks (DBRX) and Snowflake (Arctic). The prior Gen-AI architectures referenced in this report, be it prompt engineering, RAG, or fine-tuning, all leverage existing FMs and augment them with varying levels of customization to enable specific use cases. Developing a language model from scratch typically requires a distinct application that is less suited for fine-tuning. For instance, if an organization: **1)** has a proprietary dataset that is unique to existing pretrained LLMs, **2)** requires domain-specific knowledge embedded in the core model, **3)** wants control over the training data or **4)** seeks to avoid the inherent biases from third-party pretrained models.

■ Growing instances of smaller, pretrained models developed for commercial

use. Even as FM providers have constituted the bulk of pretraining to date, often in an effort to drive technology advancement (e.g. GPT-1 to GPT-4o), referenceable case studies are emerging that suggest pretraining smaller language models is a viable path for organizations with the appropriate resources (data, budget), technical expertise and specific requirements. For instance, Snowflake recently developed its Arctic model in 3 months for <\$3mn (1/8th the training cost of peer models), while comparing favorably against competing models in the MMLU benchmarks. Importantly, the Arctic model was developed to excel in areas highly relevant to its end-customers, including SQL generation, coding and instruction following. Arctic boasted a better performance/cost ratio for inferencing against the mentioned

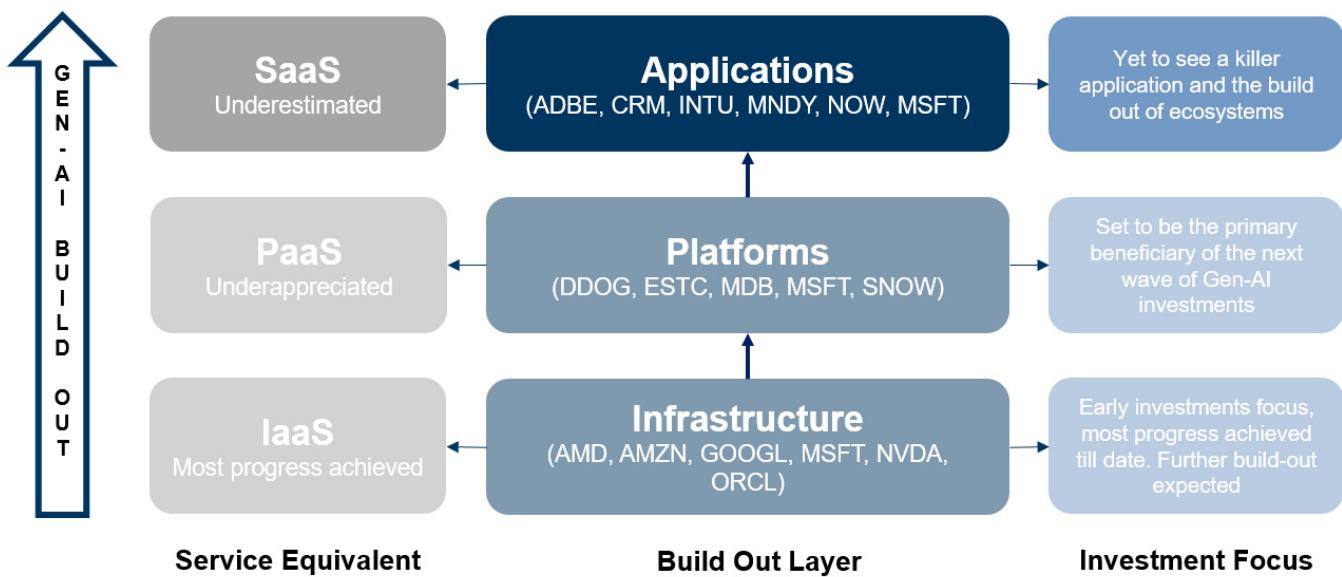
parameters vs. many leading models, including DBRX, Mixtral 8x7B, and Llama-3 8B (and 70B). Snowflake's integration of the Arctic family models across its burgeoning AI and ML product portfolio can ensure it delivers higher quality and cost-effective results. Databricks also released its industry-leading DBRX model, which can be leveraged as a base model when building custom models through the Mosaic AI training stack, amongst additional use cases across the platform.

- **Expect narrower adoption rates for pretraining, largely with small language models.** As full-service AI enablement platforms emerge to democratize pretraining for Gen-AI application development, we believe adoption rates should pick up for a narrower set of organizations. We believe the most significant opportunity in pretraining resides with smaller language models. Contrasting with LLMs such as ChatGPT, small language models are trained on substantially reduced data sets, which makes them less computationally burdensome for both training and inferencing. While the cost-effectiveness, flexibility, and lower development times associated with both RAG and fine-tuning are likely to prove more attractive for a wider set of organizations, we note that pretraining an SLM can be effective in instances in which a highly specific capability is required.

Relevance of Application layer most debated and underestimated

Long-term potential of the application layer is one of the most underappreciated areas of the Gen-AI buildup (Exhibit 39). We recall that investors debated application software as an investable area dating back to the decade following the GFC and even as far back 2003-2008 timeframe following the dot-com collapse. When the cloud went mainstream in 2010, it was followed by worries that application software would get commoditized. Ultimately the SaaS industry grew at a 30% CAGR from 2010 (Gartner) and is currently the largest segment of the cloud industry, producing a few dozen investable ideas in the last decade. We expect a similar dynamic to unfold over the coming years with Gen-AI, with SaaS companies leveraging the Hyperscalers' infrastructure instead of competing with them. We expect companies such as ADBE, MSFT, CRM, INTU and others to be able to leverage both their differentiated data and the capabilities offered by Gen-AI infrastructure and platforms. As the platform and infrastructure layers evolve, we look for the emergence of a "killer application" that will drive ubiquitous adoption. This can unlock \$400bn revenue opportunity at scale.

Exhibit 39: We believe the application layer remains the most underestimated aspect of the Gen-AI cycle



Source: Goldman Sachs Global Investment Research

Expect Gen-AI applications to unlock previously under-utilized cloud capabilities.

Enterprise Applications have always been about data, processes, execution and analysis. Those things will not change with Generative AI. What will change is how this new cognition cycle perpetrated by Generative AI will offer new engagement avenues for end users and offer an ability to triage problems in ways that could not have been done before. If we ask a very thoughtful question of a chatbot that sits alongside a classic application, it will give a well-informed answer. That does not circumvent the need for traditional applications. But it rather, it augments it. We believe the accessibility of Gen-AI applications via a command prompt could fundamentally change how users get the most out of their cloud applications and ultimately favor app companies with complex functionality. Cloud SaaS applications have grown increasingly complex and harder to navigate, with layers of difficult-to-navigate formatting, tools bars, and hidden features. We point to ERP and CRM as examples - the vast majority of users use only a fraction of the application's capabilities. For new users, it takes time to learn the application's advanced features and users must essentially adopt the program's configuration. With mature Gen-AI use cases, we believe users will be able to have applications conform to their needs and significantly lessen the learning curve necessary to get started. Lowering UX barriers with Gen-AI tools should help unlock large amounts of functionality and will be a clear catalyst to wider Gen-AI adoption.

- **We envision a Gen-AI bot or prompt sitting alongside application softwares' UI and helping navigate complex queries and commands.** Such AI bots could help get more productivity out of application software, which codifies a treasure trove of business processes and workflows. In this scenario, we would equate it to an airplane's controls being taken over by a copilot or even an autopilot. This gives the flexibility for the pilot (the power user) to be able to take over at any given point in time and access advanced functionality if needed. By having a bot work alongside a

complex application with significant functionality, application software companies could not only re-assert their pricing power, but also enjoy the benefits of potentially higher retention and therefore lifetime values.

Though current CRM Gen-AI applications are in their infancy, we see large long-term opportunity. We view the CRM market poised to realize significant value from Gen-AI features due to the potential for productivity boost for frontline professional workers. Aided by a domain specific copilot, customers could get more usability despite the growing complexity of CRM applications driven by multiple UIs, tabs, tool bars and data silos. We believe Gen-AI tools should be able to democratize access and lower barriers of entry, which will ultimately provide more value to the end user. It's important to realize that CRM software has become mission-critical tool for most enterprises and is now the largest application software segment in the Software industry. ZoomInfo, a go-to-market data intelligence platform for optimizing digital selling, is early evidence of customers' willingness to spend on lead-generating benefits for sales reps: in just two months of selling Copilot, they have more than \$18mn of Copilot ACV and 1,000 logos. We see Gen-AI-enabled CRM applications providing meaningful boosts to smart pipeline generation, tighter prospect targeting and more accurate closes, thereby driving greater sales rep productivity. Salesforce boasts Sales AI features that can draft sales emails, gives personalized close plans, automatically capture and sync customer data, mine conversation data, and more. Our diligence indicates that early adopters of Gen-AI are already seeing a boost to sales productivity in a relatively short span.

- **See potential for ARPU lift coupled with consumption bent.** While the seat-based business model in the CRM industry has limited avenues for growth in mature end markets such as sales and service, we see ARPU uplift from the potential for greater productivity as a decent offset. As the level of interactions with Gen-AI co-pilots starts to increase, we see augmentation of core seats with a consumption-based aspect. This augmentation of pricing will be closely aligned with the incremental value being realized by users as they interact more with Gen-AI Apps. Furthermore, market segments such as Marketing automation and E-commerce which are less hinged to seat growth but more towards consumption before the advent of Gen-AI, will see more interactions as agent workflows increase the level of automation in rote tasks. This again is likely to enhance the pricing power of CRM Applications. All this begs the question of whether Gen-AI will become so good that it wipes out large portions of seat-based professional workers. We don't believe that to be the case for a few reasons: **1)** Adoption of Gen-AI, especially given the maturity at this stage, is likely to play out over time in phases; **2)** More productivity, if realized, will mean a race for adoption where users who benefit from AI will stand to significantly outperform those that don't; **3)** Most software markets, especially, CRM, have been pressured by 3 years of economic uncertainty due to high rates that discretionary hiring and related software spending has been pressured, which we see as potentially alleviating once we get rate cuts. As growth agenda regains footing after years of depressed activity levels, we could see hiring, especially in the front office start to pick up.

Incumbents' data + knowledge of workflow processes a key advantage; Should

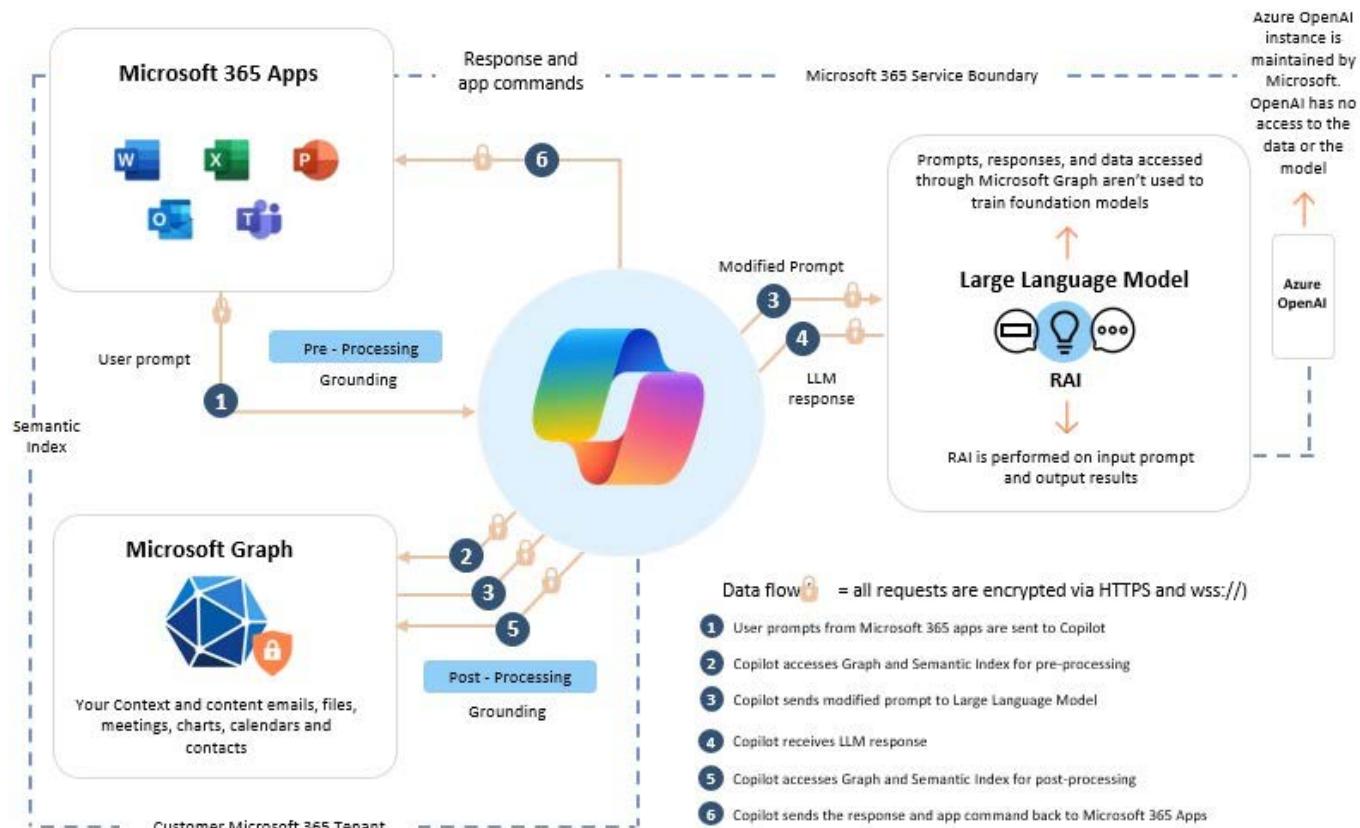
not be taken for granted. Incumbent software Applications and Platforms companies have a good opportunity to benefit from Generative AI. However, they cannot take their data as an important source of incumbency for granted. The data must be easily accessible for training models to work effectively. At the same time, a new opportunity is opening up for net new Software ideas which can be completely built on a new Generative AI stack. There is a risk to the incumbency of Software Applications, if new ideas and products can meld their way into end-user workflows and make themselves very fluid and seamless, thereby generating interactions. This can then provide meaningful basis for more established app companies to provide the training of these models, which could be disruptive to their value proposition longer term. This would represent a more bearish scenario for incumbents since training is a less exciting category to be relegated as it may lower the prospects of differentiation.

Uncertainty of killer application underpins fears; We expect several use-cases to emerge. While applications such as Office Copilot, Adobe Firefly, ServiceNow Pro+, and Salesforce Einstein have begun to show potential, there is still considerable room for innovation. While Gen-AI applications are beginning to emerge and understand the context of users' work, we have yet to see any application offer enough value to be table-stakes for enterprises, with perhaps the pointed exception of coding suggestions tools. For front, middle, and back office functionality, we are still waiting for a "must-have" application that will justify the material cost of Gen-AI compute. We also note the chicken-or-the-egg dilemma around the idea that platform solutions need a tangible use case to scale but applications need solutions that can be provided by larger scale providers such as security, governance, etc. to manifest. Ultimately, we believe the platform layer will need to be fleshed out, with the new data workflows crystallized before more ubiquitous adoption occurs on the application layer.

- **Developer software is an early Gen-AI success; highlights long-term potential of the technology.** Developer applications have become one of Gen-AI's first scaled and productivity-driving use-cases, one of the closest approximations to a "killer app" we have today. Companies like GitLab, Microsoft (with GitHub), and Atlassian are rapidly innovating and scaling developer products that are beginning to become material drivers of growth and retention. The success of these early products reinforces our belief that once productivity benefits are apparent, adoption of killer Gen-AI applications will be widespread. Examples of quantifiable productivity benefits are everywhere, from Atlassian estimating the average TEAM engineer has saved 1-2 hours/week with their documentation search tool to Microsoft increasing developer productivity up to 55%. Likewise, these tools have scaled fast. Just over two years since GA, over 77k organization have adopted GitHub Copilot, now accounting for >40% of GitHub revenue growth. Broadly, we believe that as developer software is one of the first examples of a scaling Gen-AI application, advancements in the developer space indicate where Gen-AI applications will evolve. Examples like GitLab's Duo Workspace, an effort to build AI that proactively engages in the workflow instead of being purely reactive, and Atlassian's Rovo Agents, that can be programmed to take actions, are indicative of a potential future shift in the application market towards more autonomous Gen-AI tools or agents.

■ **Microsoft's O365 Copilot may evolve to be the key Gen-AI use-case.** We view Microsoft's early Office 365 Copilot investments ([Exhibit 40](#)) as promising and may evolve to become the industry's key Gen-AI use-case. Microsoft and O365 have an unparalleled lens into organization's workflows across Outlook, OneDrive, Word, Teams, and more. As Microsoft's AI solutions continue to learn from employee usage and behavior in each application (like writing style, locations of useful documents, etc.), we believe M365 Copilot will be well positioned to assist in a wide range of organizations' workflows. However, these applications serve very different functions and it will take time for Copilot to work seamlessly across applications. In the meantime, we believe Microsoft's concentration on domain-specific applications (like M365 Copilot for Finance, Sales, etc.) will offer better near-term value. The company's broad distribution channel, sitting at the crux of 40% of frontline workers' workflows via its +400mn commercial install base, will allow Microsoft to get broad adoption once the product has become more refined. We acknowledge the adoption curve has taken longer to formulate than we expected, but the pace of innovation within Copilot gives us confidence in longer-term growth. The evolution of the application is also likely to occur at a faster rate than the broader industry (as it fleshes out each layer of the I-P-A framework) given Microsoft's solutions across the tech-stack, adding to Copilot's advantages.

Exhibit 40: M365 Copilot may emerge as the industry's killer use case



Source: Company data

■ **Creative AI tools gain traction; see room for growth within broader market.** We continue viewing creative software as one of the most compelling growth

opportunities enabled by Gen-AI. As the industry continues prioritizing digital tool usage, Gen-AI can broaden the user base to encompass a broader swath of creatives, such as those who may be newer entrants or have less technical knowledge. We point to Adobe's Firefly offering as helping underscore this point. Utilizing a natural language prompt interface, users can quickly create, and iterate on, content. This expedites the content creation cycle, while better enabling less technically-savvy and more recreational creators to help create meaningful content. Firefly's generation of >9bn digital assets from March 2023 to May 2024 helps underscore the aforementioned points. We have seen other players announce Gen-AI feature functionality, such as Figma (via their Make Design product) and Canva (in products such as Text-to-Image, Magic Write, Magic Edit, and Translate). For Adobe, Firefly is already said to be driving improved retention, indicating the realized value of the product. As the need for timely and relevant content increases, the inclusion of Gen-AI within content creation tools is likely to become more of a table stakes expectation, versus offering substantial differentiation. As such, the players leveraging Gen-AI will be favorably positioned given the broader implied reach they have amongst both nascent, and more experienced, creatives.

- **Usage largely being prioritized over monetization given the early innings opportunity.** Given the early days, companies have understandably focused on driving adoption and usage versus meaningful monetization. Adobe includes a set number of Generative Credits per month with a user's subscription, with an offer to purchase more credits upon depletion. Generative Credits could prove a lever to sustained revenue growth longer term as it capitalizes on the engagement and value of the product. Ultimately, we expect enablers of creativity to prioritize the product velocity of Gen-AI tools as users increasingly look to incorporate such tools in their workflow processes.

Gen-AI in HCM can help realize the potential for increased employee engagement and productivity. HCM Software has come a long way from a back-office record keeping app to more of a front-line app that can help with employee engagement, productivity and growth. But we believe there is a lot of work that can be done to create even better employee experiences with Gen-AI. We believe there is room for Gen-AI apps such as ServiceNow's Employee Workflows to help internally source the right skills or talent for a specific project. Such AI tools can help create internal knowledge networks that are not just static databases with stoic titles and responsibilities. Rather, they can be lit up with updates that provide contextual information gleaned from posts and articles, which can help get users to access the right experts as opposed to spinning wheels in unproductive effort. Many of an organization's policies, procedures, and even some employee data/context is stored in unstructured data format in employee handbooks, resume PDFs, etc. This type of unstructured data is where Gen-AI can help extract insights through Chatbot interfaces. Most users access HCM software fairly infrequently and, when used, is used only briefly. A copilot can help answers nuanced questions in tax policy/reporting, track vacation days, or interact with stored resumes with a level of granularity that can't be obtained from public websites like LinkedIn. Workday, notably, does not charge for their AI features like content generation,

document analysis and correction, personalized work experiences, and more. While it's still early to put a price wrapper for Gen-AI since a lot of the emerging Gen-AI functionality is yet to crystallize, we believe at a minimum that HCM software companies will be able to justify pushing such features into premium tiers and increase renewal rates. We can see 'freemium' approaches yielding better monetization as opposed to marketing blitzes trying to sell the audience on 'features'.

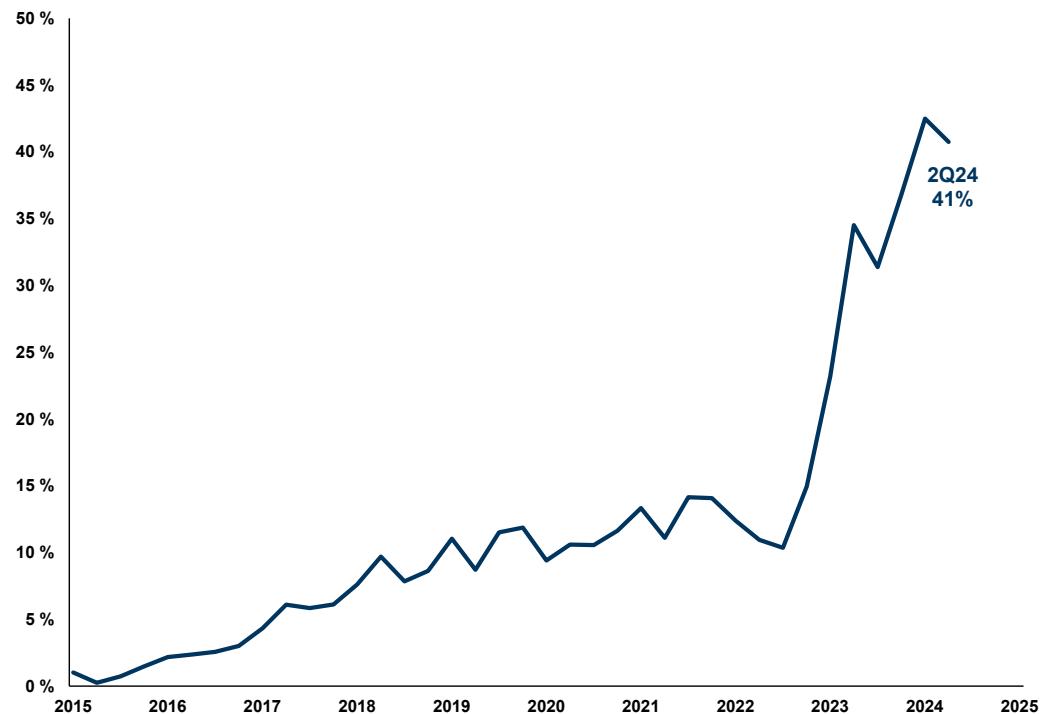
Business model implications; outsized revenue opportunity likely to support minimal margin impact

During the cloud buildout, software companies had a reasonably clear benefit to customers: cloud offered a lower TCO, regular software updates, and the trading of CapEx to OpEx. Essentially, software applications were substituting compute for compute and a business process for another business process. The processes for both were the same, but companies had to rewrite everything for cloud and mobile architectures. In contrast, Gen-AI is less about a replication process, but augmenting and improving human productivity. It allows the user to get an answer or solve a problem more quickly and creatively. Therefore, adoption of Gen-AI applications allows customers to trade GM for lower OpEx (as software is typically expensed in COGS). If done correctly, the roughly \$2tn enterprise IT industry would pale in comparison to the productivity dollars that could be unleashed by Gen-AI.

Demand is fundamentally easier to match with Gen-AI than with the original cloud buildout. During both the cloud cycle and the burgeoning Gen-AI cycle, companies were able to invest ahead of what would clearly become a massive market. However, the Gen-AI cycle is different in that infrastructure companies are able to match workloads with capacity in real time. While, for the cloud cycle, moving data and processes from on-prem to the cloud was a massive lift requiring months (if not years) of planning and whole teams to oversee the migration; the Gen-AI transition is often as simple as purchasing an extra subscription. As Gen-AI applications are structurally similar to cloud applications, it is an easier lift for application companies to push these features onto users. While many companies have significant work to do in terms of organizing their data to best take advantage of the full-extent that Gen-AI features can provide, Gen-AI applications can effectively be turned on at the switch of a button. Partially due to the (relatively) easy opt-in discussed above, AI buy-in from companies across industries and of all sizes has spiked. As evidenced by the mentioning of AI on earnings calls ([Exhibit 41](#), an admittedly rough proxy for enterprise-level interest), AI has jumped to the forefront of strategic conversations. While to a much smaller degree, several companies have been able to start attributing a level of revenue to these platform enhancements (NOW, CRM), despite just being in year 1.

Exhibit 41: A rough proxy for AI interest over time, companies citing AI spiked following the release of Chat GPT

Proportion of S&P 500 firms mentioning “AI” during quarterly earnings calls



Source: FactSet, compiled by Goldman Sachs Global Investment Research

Tight IT budgets, technological nascentcy needed before more material revenue attribution.

attribution. The higher-for-longer interest rate environment thus far in 2024 has largely led Gen-AI spend to come at the expense of budgets that might have otherwise been spent on supporting software growth. While Gen-AI projects can circumvent the same scrutiny and high IRR barriers that enterprise software has to clear, given its nascentcy and disruption potential, the focus is likely to remain on the IaaS/PaaS parts of the tech stack until 1) the macro backdrop improves and/or 2) the formation of the lower levels of the tech stack enables the expansion of use-cases to the application layer. Based on where we are today, both from a macro and technical point of view, we estimate that the spend on IaaS solutions is orders of magnitude higher than that of applications and platforms. This is underpinned by MSFT Azure/AWS/GCP all pointing to billions of AI spend, while app companies are still in the early stages with sub-\$1bn in ACV attributed to such services. Until either one of the aforementioned trends improve to make room in otherwise crowded IT budgets, we believe companies’ focus will remain on the infrastructure and platform layers in the near term.

Most organizations have meaningful work to do before widespread deployments.

Though companies across industries have been relatively successful in deploying early Gen-AI applications, we believe there is meaningful work to be done before organizations see significant productivity benefits from widespread deployments. In order to realize the full potential of any killer application, the industry must accomplish the following steps:

- **Modernizing the Infrastructure and Platform layers:** As outlined in the prior two sections, we believe significant investments and innovation are required at both the infrastructure and platform layers of the tech stack. While the infrastructure transformation is largely underway, as Gen-AI demand has continued to outstrip capacity, more capacity at cheaper prices will be a clear priority for the near-term.
- **Data organization and clean up:** Differentiated and specialized Gen-AI applications will require, crucially, differentiated and accessible data. For many best-of-breed and source-of-truth application companies, access to this data is not an issue. By virtue of being cloud companies for at least a decade, companies like CRM, INTU, NOW, WDAY, ADBE, and more have a substantial collection of data on business processes and transactions within their vertical. However, in order to actually train models on this data, it has to be clean and accessible. Similarly, in order for customers to use Gen-AI applications, customized for their specific business use-cases, they must also have clean, accessible data for training.
- **Growth and re-appropriation of IT budgets:** Our conversations have indicated that most companies have not yet allocated a separate budget to Gen-AI. This leads us to expect longer-term funding for AI applications to come from three sources: 1) Redirected budget from legacy workloads, 2) Efficiencies realized with AI elsewhere in the organization, and 3) A large collection of rapidly expanding Gen-AI native companies. As during the cloud cycle, we expect in the longer-term IT budgets will be redirected from legacy project to AI projects. It is difficult to determine at this moment what kinds of workloads may become obsolete, but we believe organizations will inevitably adjust alongside the new technological paradigm. At the same time, if Gen-AI lives up to its promise, Gen-AI customers will reap productivity benefits that can be redeployed back into purchasing this software. Finally, just as a significant source of cloud spend comes from cloud-native companies, we argue the AI-native companies will eventually become meaningful purchasers of this new software.

Pricing unlikely to be massive hurdle in LT value realization. We believe companies will be able to show clear pricing power with differentiated Gen-AI applications, while revenue benefits will more than offset the likely margin impact from higher computing costs. Thus far, managements' pricing strategies have been varied as they grapple with how to best monetize the new compute-intensive applications while balancing their profitability profile (a key metric for investors). Companies have approached this with five main strategies: **1) Additional AI SKUs, 2) Premium AI Pricing Tiers, 3) Consumption-based pricing, 4) Hybrid pricing, and 5) No direct monetization.** As applications begin to scale and reach maturity, we believe the most prevalent pricing method will be a premium AI pricing tier as customers become more comfortable with AI productivity benefits and application companies look to benefit from upselling.

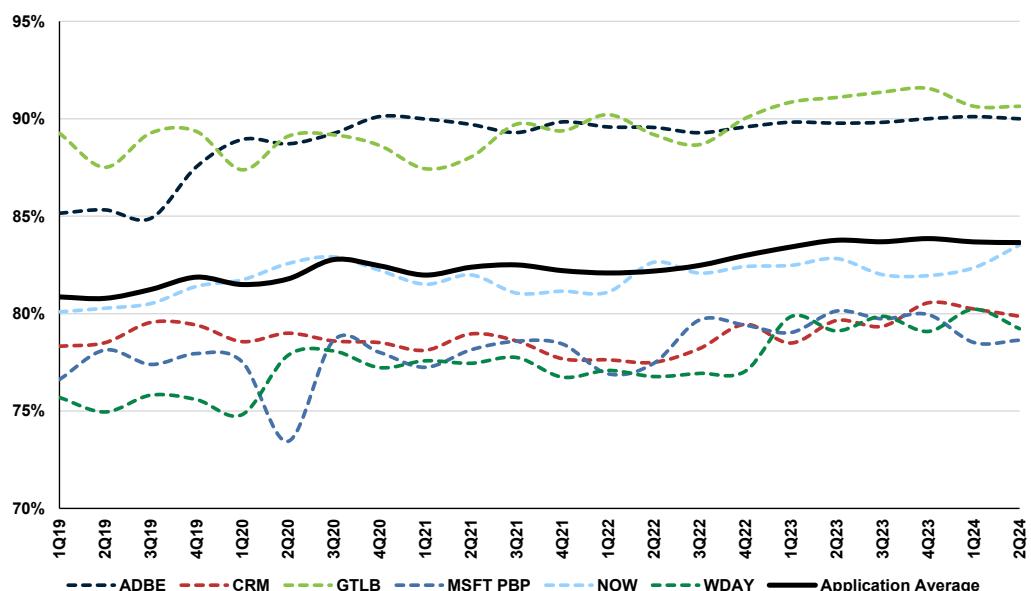
Impact to margin profiles likely to be contained vs other layers of the tech stack. As Gen-AI revenue contributions and associated costs flow through the income statement, we believe the impact to operating margins will be minimal. Even where compute costs may begin to suppress gross margins (as cloud computing for Gen-AI workloads comes at a higher cost), we expect management teams to offset gross

margin pressure through: **1) Pricing** (average price per app is \$20-30 per user/month), **2) Driving efficiencies in non Gen-AI initiatives and/or, 3) Getting leverage from existing GTM.**

■ **Margin compression at the Platform and Infrastructure layers is more pronounced.** While impacts at the application layer have been surprisingly limited, we believe there will continue to be significant impacts to gross margins at the infrastructure layer and, to a lesser degree, the platform layer. At the infrastructure layer, similar to the initial cloud buildout, companies are effectively trading gross margin for gross profit dollars. In our analysis of MSFT's CapEx ramp, we conclude that MSFT's projected CapEx ramp will have a 150bps impact to GM in FY25 and a 300bps of headwind in FY26. Despite these headwinds, our analysis suggests AI-specific gross profit dollars in FY24 are comparable to 8 years after the Azure launch. We believe this depreciation accumulation is similar for Alphabet and Meta (covered by Eric Sheridan) as they similarly ramp their AI-related CapEx, with depreciation expense likely to grow at a ~19% 3-year CAGR for Alphabet and a ~28% 3-year CAGR for Meta. At the platform layer, we note examples like SNOW, who lowered their full year operating margin guidance by 300bps in F1Q25, at least partially predicated on an increase in GPU-related costs for their AI initiatives.

Early success in balancing AI investments suggests profitability concerns may be overblown. While the infrastructure layer of the AI tech stack has indicated significant margin impact from the Gen-AI buildout, the impact to margins is less clear at the application layer. Few application software companies have called out a hit to margins from their new AI SKUs. Since application companies aren't spending as significant amounts of CapEx to build out their Gen-AI capabilities, instead using largely OpEx, the impacts from depreciation on gross margins are much less meaningful. That's not to say the impact has been zero. Many application software companies who were showing consistent gross margin improvement prior to Gen-AI have, in the face of ramping AI investments, instead plateaued. This holds particularly true during early Beta trials. Companies like CRM, ADBE, and INTU were running massive Gen-AI trials with meaningful inferencing costs while recognizing essentially zero revenue. Purchasing the higher-cost compute conceivably dragged on profitability. Considering the impact to margins was highest during Beta trials (with peak GPU capacity constraints and little to no AI revenue) and given pricing on Gen-AI solutions has been higher than we initially expected (more below), we were impressed with large application companies maintaining their gross margins (Exhibit 42). While that is not to say we don't see possible scenarios where application companies will see a gross margin drawdown, early success in balancing AI investments with operating efficiencies suggests early concerns about significant profitability hits may have been overblown.

Exhibit 42: Despite the launch of major Gen-AI products, application-layer gross margins have remained relatively flat



Source: Company data, Goldman Sachs Global Investment Research

Smaller models likely to become more prevalent

Foundation models are already showing signs of maturity and crowding. The functionality and breadth of these models are rapidly converging given the crowding and investments, particularly from the largest players like OpenAI, Anthropic, and Mistral. This is great for SaaS companies who can benefit from these significant investments to develop more unique and differentiated models. These more specialized models can sit on top of or be used instead of the foundational models.

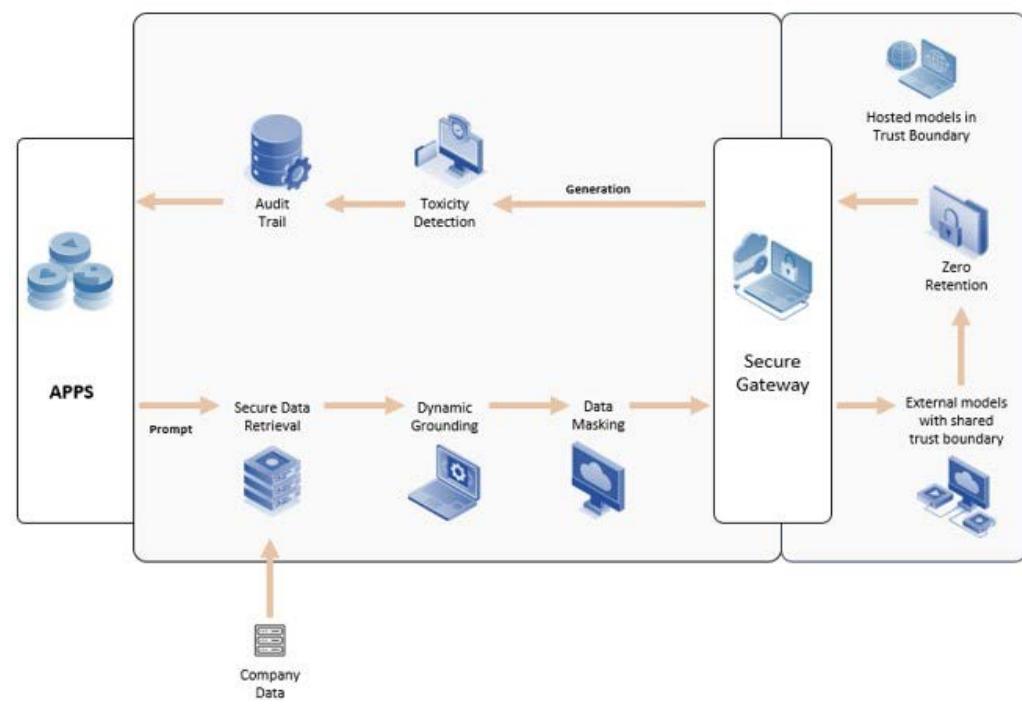
We expect efforts to shift to smaller, more specialized multi-modal models that sit on top of the contextual understanding that larger foundational models provide. This change shifts the role large foundational models play to one of that provides the base level of contextual knowledge, but does not serve as the core knowledge hub in which an applications output is predicated on. Less than two years following the release of ChatGPT on GPT-3.5, there are dozens of LLMs on the market from companies like Anthropic (Claude), Google (Gemini, LaMDA), Meta (LLaMA), Mistral, and more. Largely due to the current scale of models (GPT-4 has trillions of parameters), we believe the costs of training foundational LLMs have quickly become prohibitive to all but a small number of well-funded LLM companies and the world's largest tech giants (MSFT/AWS/GCP). At the cutting edge, there is limited ability to differentiate between foundational models beyond scale (which is largely a function of resources). Thus, most of the opportunity for differentiation lies elsewhere, within smaller, specialized models that can deliver task-specific performance without the cost of a larger model. For example, rather than querying the trillions of parameters that top-of-the-line LLMs now boast, ServiceNow can train its general-purpose model on specific ITSM tasks, knowledge base articles, or something similar, and create a highly performant,

domain-specific model. Following the IPA framework, as the platform develops, application companies will have better access to a wider array of data that will support increasingly customized and use-case specific SLMs.

Specialized models increase differentiation. Data differentiation is one of the key moats we see companies needing to sustain to successfully evolve from the cloud era to the age of AI. At the crux of this lies refined, proprietary datasets, and unique perspective/insights into an organization/users' workflow. The high degree of dispersion of enterprise data, scattered across various organizational silos will drive the need for a more cohesive data layer (such as Salesforce's Data Cloud). This can serve as a platform upon which specialized models can be leveraged and where data sitting across different silos can be accessed. Software providers are likely to leverage this inherent complexity to enhance the value proposition of their own LLMs via the proprietary data that flows through their ecosystem. Initial adoption hurdles based on privacy controls for data security and governance will likely be resolved via new workflow controls.

Salesforce's strategy ([Exhibit 43](#)) that utilizes masking, grounding and zero-retention is likely to serve as the standard for the industry and provide an example of how companies can utilize their data while maintaining privacy amongst their various customers. This also doesn't thwart the inherent need for the foundational models need to iteratively learn and adjust their output suggestions as sensitive data will be separated from the context ingested. This type of implementation can help avert commoditization as the emergence of numerous data models offers greater differentiation vs. ubiquitous plug-ins into more generalized foundation models such as Alphabet's Gemini models or OpenAI's GPT series.

Exhibit 43: Salesforce's SLM Gen-AI strategy serves as an example of utilizing data while maintaining privacy amongst various customers



Source: Salesforce, Goldman Sachs Global Investment Research

Smaller, specialized models likely to offer better unit economics. Smaller models offer a versatile way for organizations to embed use-case specific AI into their broader tech stack and, ultimately, save on inferencing costs. These specialized models are less expensive to train than their larger foundational counterparts, such as LLMs. This is due to having fewer parameters and simpler architecture. The magnitude of parameters of an LLM, for example, adds to the training expense, so SLMs trained on billions of parameters can be orders of magnitude cheaper. Snowflake, for example, trained its internal model, Arctic, on a budget of under \$2mn. This compares to the roughly \$100mn OpenAI stated was spent on training GPT-4. Once trained on a more limited set of parameters, inferencing costs can also lower. Companies like Aisera, which specializes in SLMs, boast costs that are 1/20th or 1/30th that of many of their LLM competitors. This change shifts the role large foundational models play to one of that provides the base level of contextual knowledge, but does not serve as the core knowledge hub in which an applications output is predicated on. As these models are less complex, they can run on more attainable hardware and demand less power. We believe the efficiency and accessibility of smaller models will help make AI applications more accessible for a wider variety of organizations and help drive the proliferation of Gen-AI applications.

Significant growth opportunity for both legacy cloud and AI-native companies

Value, so far, accruing largely to incumbents. At the beginning of the cloud cycle, there was great anxiety that cloud native applications companies could end up disrupting on-prem applications company. However, the strongest incumbents (MSFT, ADBE, INTU, ADSK, ORCL, SAP, etc.) managed to create significant amounts of market value alongside the cloud native companies (CRM, WDAY, SHOP, VEEV, etc.). Similarly, we see incumbency in the upcoming Gen-AI cycle as having value, notwithstanding investor concerns. The cloud cycle had the added complication of not only an architectural change, but also a business model change (from licenses + maintenance toward subscriptions) for incumbent on-prem companies. In this upcoming Gen-AI cycle, we at least have alignment of the business model as we continue to see a seat-based, subscription component to pricing. Despite potential of moving in the direction of consumption based pricing, we don't see this being as disruptive as prior cycles given the better understood strategic value of software. As said earlier, if incumbent software applications companies can harness their data effectively (admittedly this is not an easy ask) then they can create meaningful value.

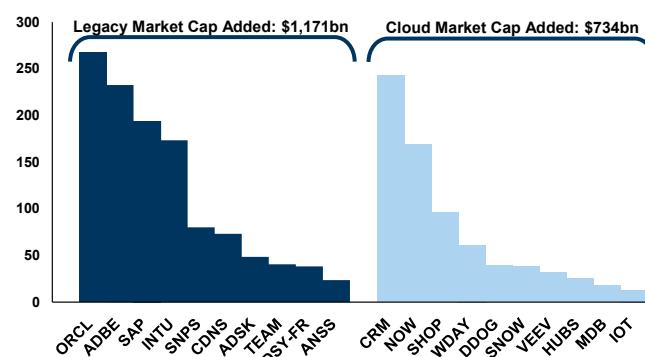
- **Legacy on-prem companies were able to capture significant portions of the Cloud market.** Looking at Cloud companies today who offered on-premise software solutions (MSFT, ADBE, ORCL, SAP, INTU), these companies were ultimately able to reposition themselves and grow their businesses at a faster revenue clip and scale to new levels through the last two decades, accruing significant investor value (Exhibit 44). Companies like Adobe, who embraced the cloud pivot as early as 2011, saw significantly less degradation in market share. We see incumbents SaaS application companies today having similar potential. We particularly favor companies whose management teams and operational architecture allow them to act with a sense of urgency. On the other hand, legacy players like Oracle and SAP

did eventually pivot into the cloud, albeit a bit later than companies like ADBE and INTU (who embraced the cloud as early as 2011 and saw significantly greater increases in market share).

- **Door still open for upstarts.** While we see room for legacy companies to execute in this market, new upstarts can create new TAMs in verticals or other horizontal applications which we cannot fully grasp or visualize. In the cloud cycle, upstarts got into areas of the software market that were previously not considered mainstream (like CRM or HCM). In doing so, they became specialists that took advantage of cloud architecture to build big businesses. During the cloud buildout, excluding Microsoft, legacy companies added over 1.5x the market cap as cloud native companies ([Exhibit 45](#)); including Microsoft, they added over 5.5x the market cap ([Exhibit 46](#)). That being said, cloud native companies still added over \$700mn of market value since 2010. Similarly, we see space for the next generation of startups to become specialists in applications domains that have not been attacked with a level of focus.

Exhibit 44: Cloud cycle suggests there is room for significant market value accretion for both legacy and upstart companies

Market cap added by ten largest legacy software names vs. ten largest cloud natives (2010-Today)

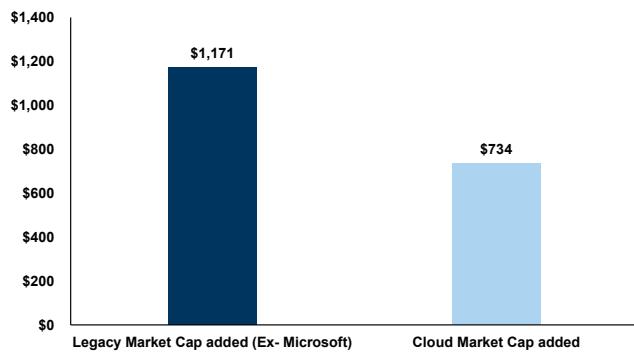


\$bn, ex. MSFT and ex. Security Software

Source: Company data, Goldman Sachs Global Investment Research, FactSet

Exhibit 45: While both added significant value, the largest legacy companies added over 1.5x the amount of market cap as cloud natives

Market cap added by ten largest legacy software names vs. ten largest cloud natives (2010-Today)

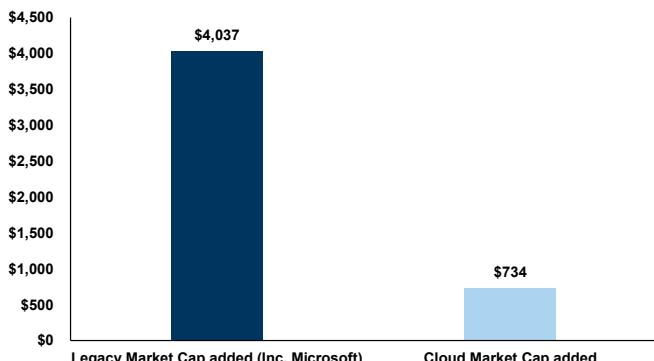


\$bn, ex. MSFT and ex. Security Software

Source: Goldman Sachs Global Investment Research, FactSet

Exhibit 46: If we include MSFT in the analysis, legacy companies have added around 5.5x of market cap when compared to their cloud native peers

Market cap added by top 10 legacy names (inclusive of MSFT) and top 10 cloud names (2010-Today)



\$bn, incl MSFT ex. Security Software

Source: Goldman Sachs Global Investment Research, FactSet

Application segment has a long and solid track record. We recall that investors debated application software as an investable area dating back to the decade following the GFC and even as far back as 2003-2008 following the dot-com collapse. When the cloud went mainstream in 2010, it was followed by worries that application software would get commoditized. Rather, SaaS grew at a 30% CAGR from 2010 (Gartner) and is currently the largest segment of the cloud industry, producing a few dozen investable ideas in the last decade.

See Gen-AI's application TAM reaching \$400bn

Making the case for Gen-AI's role in expanding software's application TAM.

Following a full year of per-seat AI pricing disclosures, our updated application TAM at scale is increasing to \$400bn ([Exhibit 48](#)) vs \$150bn prior. While the uplift was predicated on pricing of Gen-AI applications coming in higher than we initially expected, our analysis was built around three main drivers:

- 1. The number of Gen-AI applications each knowledge worker is likely to be granted.** We landed on 5-6 as we honed in on the applications that are likely to warrant the additional spend for more innovative features. We concluded that applications such as collaboration, productivity, HR, ERP (i.e., T&E), which are broadly adopted by most knowledge workers fit into this category given their value proposition and wide use. We also discerned that there are likely to be 1-2 role-specific solutions (i.e., CRM platform, developer tools) that may also garner investment to induce productivity. This led us to conclude that Generative AI tools can unlock value in ~5-6 apps per knowledge worker ([Exhibit 47](#)).
- 2. The average additional cost of each application. Our initial analysis assumed Gen-AI solutions will be priced at a \$10 premium/add-on.** Platforms that have evolved over the last five years, such as Creative Cloud, also showed a pattern of ~\$10 price increases, with Adobe increasing the price of its Teams bundle by \$10 in

2018 (vs the initial price of ~\$70/month set in 2014) and announcing an additional \$10 step-up in 2022. We point to Adobe's launch of Sensei, its AI platform tool, in 2016 as a catalyst for the product innovation that warranted such pricing boosts. This also led us to gain confidence in future Generative AI releases being able to garner at least \$10/month. Teams SKUs of Office followed a similar pattern. However, we update our forecasts in light of recent pricing announcements and growing revenue contributions. While Gen-AI applications will likely become cheaper over time, initial applications have largely been priced in the \$20-30/month range. This range reflects different enterprise controls and/or more performant model across different pricing tiers. We believe initial willingness to pay this premium by many first-mover enterprises means, similarly to the cloud cycle, these prices may become anchored. As we are seeing a wide range of initial approaches to monetization, including consumption pricing and implicit upselling, we believe regardless of monetization strategy, realized value will likely land around this target price.

- 3. The penetration across the ~1bn global knowledge workers.** As we expect Gen-AI to be the next wave of innovation within software, extending the utility model and enhancing productivity at the end user level, our base case assumptions are underpinned by the level of adoption seen by Microsoft for its commercial Office 365 suite, which has over 400mn paid commercial seats.

Exhibit 47: Estimate application-layer Gen-AI TAM is ~\$400bn based on 5-6 applications and an average price of \$25

Knowledge Workers:	1,000,000,000
Penetration:	30%
Discount Rate:	20%
Price per Month:	\$25.00
GenAI Applications:	5.5
TAM (billions)	396

Source: Goldman Sachs Global Investment Research

Exhibit 48: Sensitivity analysis supports our TAM estimates

Number of Applications	Price per Month					
	\$10.00	\$15.00	\$20.00	\$25.00	\$30.00	\$35.00
3	\$86	\$130	\$173	\$216	\$259	\$302
4	\$115	\$173	\$230	\$288	\$346	\$403
5	\$144	\$216	\$288	\$360	\$432	\$504
6	\$173	\$259	\$346	\$432	\$518	\$605
7	\$202	\$302	\$403	\$504	\$605	\$706
8	\$230	\$346	\$461	\$576	\$691	\$806

TAM estimates in billions

Source: Goldman Sachs Global Investment Research

Evolution of Application pricing

Likely to see growing use of hybrid pricing models (subscription + consumption).

Though our TAM analysis above is predicated on a per-seat model, we see a reality where the industry will ultimately coalesce around a strategy that leverages both seat based + consumption pricing. While companies will likely position their pricing tiers in a way that bundles AI into more premium tiers to drive upselling, consumption based pricing will likely reflect the customers' engagement with the platform and the value extracted. While base packages are likely to include a certain threshold of AI credits, or tokens, once they reach the allotted amount, they will pay more for additional interactions with the company's AI service. End-users will be able to experiment with small deployments of usage (typically via a number of prompts to the models) and have the flexibility to quickly scale up if value is realized. As customers increasingly incorporate such solutions into their workflows we expect there to be a growing mix

toward pre-commitments at a certain level of usage, which can continue to maintain software's predictable revenue streams.

- As we reach a more mature stage of the cloud cycle, where adoption of software applications has reached later stages, Gen-AI offers a value-realization component that should allow best-in-class platforms to garner wallet share gains should they drive compelling value realization. Historically, subscription made sense as a business model because it offered software companies consistent recurring revenues that could be priced dynamically. These revenues were less cyclical and companies could deliver growth with increasingly expanded capabilities and product portfolios.
- **Adobe Firefly can be an early leader of this philosophy.** Adobe has highlighted its intent to monetize Firefly via 'Price x Quantity + Value' strategy. As this plan includes a set number of Generative Credits for subscription plans, with users able to purchase additional credits upon depletion, this is a prime example of the hybrid pricing model we outlined above. Similar to Adobe's leading transition to a subscription-based model in 2013, we see Adobe potentially being one of the first application companies to embrace this revenue model. While Adobe was a pioneer of the subscription business model and some investors were skeptical, Adobe quintupled its subscription revenue base as a percentage of total revenue (10% to 50%) from F10 to F14, ultimately garnering a premium for the predictability it instilled in the model. Ultimately, the transition paid dividends - Adobe's revenue has grown at a 13% CAGR from F10 - F23, and 400% overall in that same time frame. It also allowed Adobe to innovate faster, helping them increase wallet share while expanding their user base. Soon after, subscription-based pricing became the de facto standard across software more broadly. We believe similar dynamics could transpire this cycle, as already somewhat evidenced by Firefly's popularity - in part due to the fact that it lowers the barrier to entry to less technical users, further expanding the potential user base. Having been a proven leader in the space, we believe Adobe's Gen-AI pricing approach could percolate throughout the software ecosystem and become the de facto pricing standard for Application companies.

Gen-AI pricing in market today. As it still remains uncertain how Gen-AI will influence the application layer (yet companies need to respond in real-time) we don't yet see a standard pricing approach from application companies today. While management teams have come out with initial pricing plans, they have caveated that the price point and approach may change as they evaluate market reception. Below we discuss the application companies have adopted five main pricing strategies we see having been adopted thus far, outlined below:

- **Additional Gen-AI SKUs:** Perhaps the most common pricing strategy at the application layer has been the introduction of an additional Gen-AI SKU or add-on. This strategy, typically done on a per-seat basis, benefits from clarity and simplicity. Such SKUs work seamlessly within existing SaaS pricing models and buyers can be clear about how much they will spend on AI. If AI products prove useful with clear productivity benefits, buyers can quickly expand AI seats across the organization.

However, there are drawbacks to this strategy. Priced too low can lead to customers' use of the AI tools and the associated costs quickly outstripping the incremental revenue. Alternatively, priced too high and buyers will be unwilling to allocate large amounts of finite IT budgets to the nascent products with unproven benefits. Examples of seat-based pricing [Exhibit 49](#):

- **GitLab (GTLB):** GitLab's initial pricing strategy for its AI solution - 'Duo' has been entirely seat-based. It includes tools like code explanation, test generation, code review, and more for an added price to a base subscription of either Premium (\$29/user/month) or Ultimate (\$99/user/month). GitLab Duo Pro costs \$19/user/month and GitLab Duo Enterprise costs \$39/user/month. Duo Pro is only available for Premium and Ultimate customers while Duo Enterprise is only available for Ultimate customers.
- **Microsoft (MSFT):** Microsoft has announced Office 365 Copilot will be \$30/user/month while their consumer version, Copilot Pro, will be priced at \$20/user/month. Similarly, GitHub Copilot will be priced at \$10/user/month for Copilot Individual, \$19/user/month for Copilot Business (50k organization as of F2Q24), and \$39/user/month for Copilot Enterprise. Copilot accounted for 40% of GitHub's revenue growth over the past 12 months. We estimate that GitHub Copilot is 5-10% of AI contribution to Azure revenue (GSe \$5.4bn run rate in F4Q24).

Exhibit 49: Published per seat pricing models for additional SKUs largely hover around \$20 to 30 per seat

Product	Tier	Price	Description
Anthropic: Claude.ai			
	Free	Free	<ul style="list-style-type: none"> - Talk to Claude on the web, iOS, and Android - Use Claude 3 Opus and Haiku
	Pro	\$20 / User / Month	<ul style="list-style-type: none"> - Higher usage limits - Priority bandwidth and availability
	Team	\$25 / User / Month	<ul style="list-style-type: none"> - Higher usage limits - Share and discover chats from teammates - Central billing and administration
GitHub Copilot			
	Individual	\$10 / User / Month	<ul style="list-style-type: none"> - For individual developers who want to code faster - Includes organization-wide policy management, can exclude specified files from Copilot
	Business	\$19 / User / Month	<ul style="list-style-type: none"> - Also includes Copilot chat, pull request summaries, knowledge bases
	Enterprise	\$39 / User / Month	
GitLab Duo			
	Duo Pro	\$19 / User / Month	<ul style="list-style-type: none"> - Code generation, code completion, code explanation, test generation, etc. - Only available to Premium and Ultimate customers
	Duo Enterprise	\$39 / User / Month	<ul style="list-style-type: none"> - Discussion summar, marge request summary, code review summary, vulnerability explation/resolution, model personalization, etc. - Only available to Ultimate Customers
Google Workspace			
	Gemini Business	\$20 / User / Month	<ul style="list-style-type: none"> - Gemini in Gmail, Docs, Slides, Sheets, and Meet - Enterprise-grade security and privacy
	Gemini Enterprise	\$30 / User / Month	<ul style="list-style-type: none"> - Translated captions in 15+ languages - Full access and usage of Gemini
Loom			
	Loom AI	\$4 / User / Month	<ul style="list-style-type: none"> - Message composer, auto CTA, auto titles, auto chapters, etc.
Microsoft Copilot			
	Microsoft Copilot	Free	<ul style="list-style-type: none"> - Search, create and edit images, use plug-ins - Can use in select M365 apps like Word, Excel, Powerpoint, etc.
	Copilot Pro	\$20 / User / Month	
	M365 Copilot	\$30 / User / Month	<ul style="list-style-type: none"> - Works alongside apps like Word, Excel, Powerpoint, Outlook Teams, etc. - Customize and extend Copilot in Copilot Studio - Enterprise-grade security, privacy, and compliance
OpenAI: Chat GPT			
	Free	Free	<ul style="list-style-type: none"> - Access to GPT-4o mini, limited access to GPT-4o
	Plus	\$20 / User / Month	<ul style="list-style-type: none"> - Access to GPT-4, GPT-4o, GTP-4o mini - 5x more messages for GTP-4o - DALL-E image generation, custom GPTs, data analysis, file uploads, etc.
	Team	\$30 / User / Month	<ul style="list-style-type: none"> - Also included unlimited access fot GPT-4o mini, higher message limits on GPT-4, GPT-4o - Admin console for workspace management - Team data excluded from training data
	Enterprise	N/A	<ul style="list-style-type: none"> - Unlimited high speed access to GPT-4, GPT-4o, DALL-E, web browsing, etc. - Admin controls, domain verification, analytics
Perplexity Pro			
	Free	Free	<ul style="list-style-type: none"> - Unlimited Quick searches, 5 Pro Searches / day - Standard Perplexity AI model
	Professional	\$20 / User / Month	<ul style="list-style-type: none"> - Unlimited Quick searches, 600 Pro Searches / day - GPT-4o, Claude-3, Llama 3, etc.

- **Premium AI Pricing Tier:** We believe the industry may move towards a premium pricing tier for Gen-AI. Though similar to an additional SKU, companies establishing Gen-AI as their most premium pricing offering explicitly benefit from upselling. This strategy also sells AI on a per-seat basis, with similar upside and risks to the "Additional Gen-AI SKU" strategy. However, establishing a premium tier restricts Gen-AI access to only customers willing to pay for a software company's highest priced offerings and requires more sales involvement given a decision needs to be made to opt into/migrate into a new plan. For application companies, this drive upsells and increase attach rates for additional functionality. We liken Premium Tier AI pricing to Microsoft Office's E5 SKU. While the predominant difference between E5 and its step-down tier, E3, is that it gives customers access to advanced security features, customers also get access to a lesser-used product Power BI Pro. In the future, Microsoft could theoretically incorporate Copilot into the next iteration of a 'E' SKU, effectively streamlining the selling motion for customers who want to provide Copilot natively across their organization. The potential downside of this strategy is that it elongates the adoption curve in the near term by limiting AI access only to customers who are willing to buy a software company's most premium application. The early days of AI are largely experimental; companies risk losing early exposure to customers who may be unaware of the AI products' productivity benefits. This is likely why the premium AI pricing tier strategy, so far, has largely been restricted to companies who serve enterprises and whose customers are already comfortable with the highest-tier offerings (like NOW). However, longer term, as these applications proliferate, this will become much less of a risk. We believe AI's productivity benefits will become common knowledge and enterprises will see AI adoption as less of an experiment and more of a business necessity.
- **ServiceNow (NOW):** ServiceNow released their Pro+ SKU in September 2023 with a 60% pricing uplift to their Pro SKUs. As of their 2024 Analyst Day, ServiceNow noted they are seeing a >30% realized pricing uplift to Pro Plus vs. Pro SKUs (after enterprise discounting, etc.). We see scope for the Pro+ adoption rate to outpace that of their Pro SKU which now accounts for 45% of the customer base since being introduced in 2018. With customers required to adopt Pro before Pro+, we believe this also presents an up-tiering lever for the remaining 55% of the customer base that is still on Standard.

Exhibit 50: Potential premium AI SKU could perhaps look like Microsoft adding Copilot to their highest tier SKU

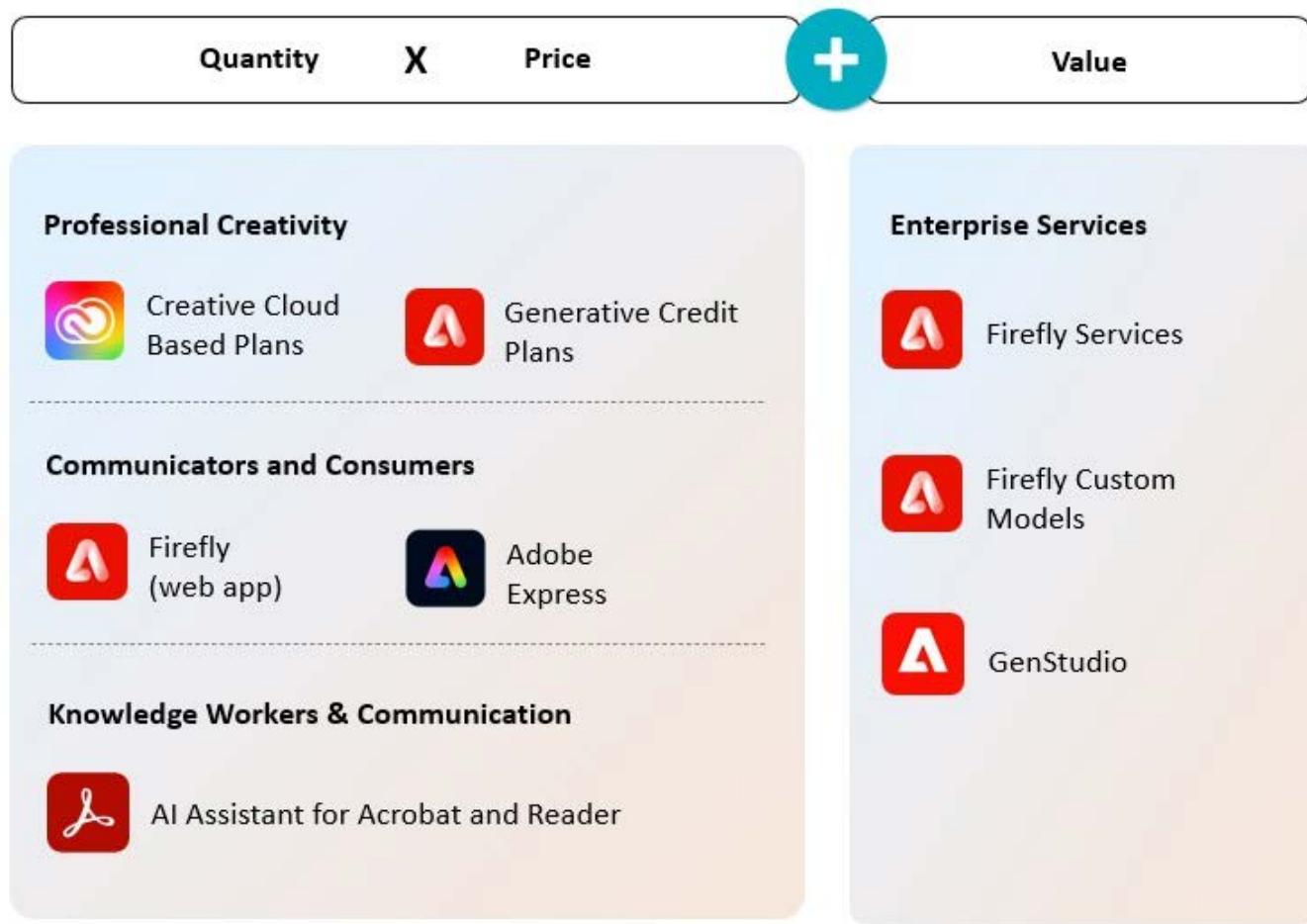
Office 365 E3	Office 365 E5	Hypothetical AI SKU
\$20.75 user/month	\$35.75 user/month	TBD
Everything in Office 365 E1 plus:	Everything in Office 365 E3 plus:	Everything in Office 365 E5 plus:
<ul style="list-style-type: none"> Install Microsoft 365 apps on up to five PCs or Macs, five tablets, and five mobile devices per user Message encryption Rights management Data loss prevention for email and files 	<ul style="list-style-type: none"> Advanced security and compliance capabilities Scalable business analytics with Power BI 	<ul style="list-style-type: none"> Office Copilot
Apps and Services Included:	Apps and Services Included:	Apps and Services Included:

Source: Company data, Goldman Sachs Global Investment Research

- **Consumption/token-based pricing:** Consumption based pricing for AI applications charges customers based on how much inferencing they do with their AI tools. This pricing method has not been common for application software names but reflects the current high costs of AI compute, in our view. Whereas much of traditional SaaS was essentially a zero-marginal cost business, AI requires very tangible compute costs with each inference. This may allow companies to adjust AI prices over time as the cost of compute comes down - as it has in prior technology curves. Ultimately, though, the pricing will reflect the customers' engagement with the platform and the value extracted. It also has the added benefit of allowing early adopters to try products out with minimal to no commitment and scale up if/when benefits are realized. The downside of this pricing strategy is it creates unintentional volatility between quarters since customers may dial up or down their usage. This is likely to introduce a fair amount of volatility in reported NER metrics even if the long term trends are intact. Examples of consumption pricing of AI applications:

- **Adobe (ADBE):** Adobe's launch of Firefly, its Generative AI platform, is one example of a consumption-oriented offering in the market today ([Exhibit 51](#)). While users need an existing plan with Adobe's Creative Cloud, the monthly subscription includes a limited allocation of Generative Credits. Users draw down their credits when they use Firefly to create content. Upon the depletion of the monthly allocation, users can buy additional Generative Credits via a Firefly paid subscription plan. We note that drawdowns typically occur after touch-point with Firefly, not per image/video/etc. For example, if a user requests Firefly to create an initial take of an image, then wants to fine-tune it to their liking (i.e., more realistic background, different font style, etc.), each request typically results in a use of a credit. Therefore, the increased reliance on Firefly is likely realized fairly quickly.

Exhibit 51: Adobe's consumption-based pricing demonstrates advantages of matching price to usage and value



Source: Company data

- **Hybrid Pricing:** A number of companies are also beginning to experiment with combining subscription and consumption pricing models. While charging for a Gen-AI specific SKU is the most analogous with current SaaS pricing models, it was designed for a zero-marginal cost product. Similarly, while consumption pricing pairs usage of the product perfectly with price, companies have little visibility into how expensive their usage of Gen-AI will be. Companies with hybrid pricing model

combine the simplicity of the premium SKU model with the downside protection of a consumption model. The downside of this model is its complexity, which makes it both difficult to set reasonable pricing and difficult for enterprises to understand how much they will be paying for the service.

- **Salesforce (CRM):** In July 2023, Salesforce announced the pricing of their AI SKUs, under the Einstein product umbrella, would be priced at a per-seat basis with a limited number of GPT credits. Customers can then purchase additional credits if desired. Sales GPT is included in Sales Cloud Einstein at \$50/user/month and Service GPT is included in Service Cloud Einstein at \$50/user/month.
- **No explicit pricing/value-add approach:** Companies who don't explicitly charge for AI tools are either hoping that AI features will encourage upselling to premium product SKUs or believe it will drive engagement and retention. Many organizations are facing the difficult reality that IT buyers likely do not have enough spare budget to spend on all of these new AI features. While we believe the long-term productivity benefits from Gen-AI will be significant, in the shorter-term, IT buyers have relatively fixed budgets and similar OpEx costs to last year. The mismatch between expensive new Gen-AI applications and the tight spending environment has led some companies to conclude it is better for long-term customer goodwill and product adoption to effectively roll out these offerings into their platform. This strategy also has the longer-term benefit of encouraging customers to keep their data on a given platform while they use AI features for "free". Arguably, access to differentiated data will be far more valuable in the long-run than the relatively small near-term revenue benefits other application companies are hoping for. However, the downside is clear. Gen-AI functionality is expensive and companies could be leaving revenue on the table for a more nebulous long-term uplift. Without a revenue uplift, companies' margins could take a hit. Examples of the value-add approach:
 - **Workday (WDAY):** Workday has stated they've built AI functionality into all of their applications, with 50 live use-cases as of F1Q25. These capabilities include content generation, document analysis and correction documents, personalized work experiences, and more. Such a strategy could boost win rates and renewal rates. Workday has reported that their competitive win rates and renewal rates going up is partially due to their AI capabilities.
 - **Monday.com (MNDY):** Monday.com is integrating AI across all areas of their tech stack, including AI generated templates, auto-generated actions items, thread summaries, and enhanced text extraction capabilities. Currently, Monday.com is focusing on providing value to customers and has not introduced additional pricing for these features; instead, the company is using AI products to support adoption and customer retention. However, longer term, the company has considered monetizing AI tools via either an add-on SKU or included as a premium AI tier.
 - **Intuit (INTU):** Intuit unveiled Intuit Assist ~1 year ago at no additional cost, with integrations planned to cut across all components of the platform, from TurboTax, Credit Karma, QuickBooks, and Mailchimp. Intuit Assist has already

shown promise, with the company noting strong usage across Quickbooks (1mn SMB's), TurboTax (24mn users), and Credit Karma (20mn users) to-date. We believe Intuit Assist can support retention and drive conversion as it bridges the data silos across Intuit's platform. Looking ahead, the company is prioritizing the integration of Intuit Assist (and Gen-AI more broadly) in both existing and new services, with the potential for new monetization avenues over time.

- **Zoom (ZM):** Zoom's AI companion and meeting summary features are examples of how companies might choose to offer new AI features for free that could be a useful lever for consumer retention. Similar to Workday, their goal is to provide customer value at no additional cost. Tools like Zoom AI Companion and Meeting Summary will be a key driver of both customer retention and converting customers from free to paid.

Seat pressures unlikely to impact TAM as Gen-AI to benefit the labor market

There is also a debate around how the proliferation of Gen-AI may impact the labor market. 'If productivity is improved significantly, will organizations operate at more lean levels? What does this mean for the need of software? Particularly seat-based models?' While we understand the perspective, we do not buy into the view that this technology will reduce the pie of knowledge workers who leverage technology + particularly software in their roles. We point to prior tech cycles, from as far back as the industrial revolution, as they provide proof points around the idea that the pie has always gotten bigger. New companies and industries have historically sprung up which would have been either cost prohibitive or technologically impossible before. We also believe that businesses tend to strive higher as prior goals may become more easily achievable - therefore leading to a continuously full work load for their labor force. Thus, while productivity benefits are likely to be realized, we expect that to augment human capacity and support GDP growth - not push employees out of the labor force.

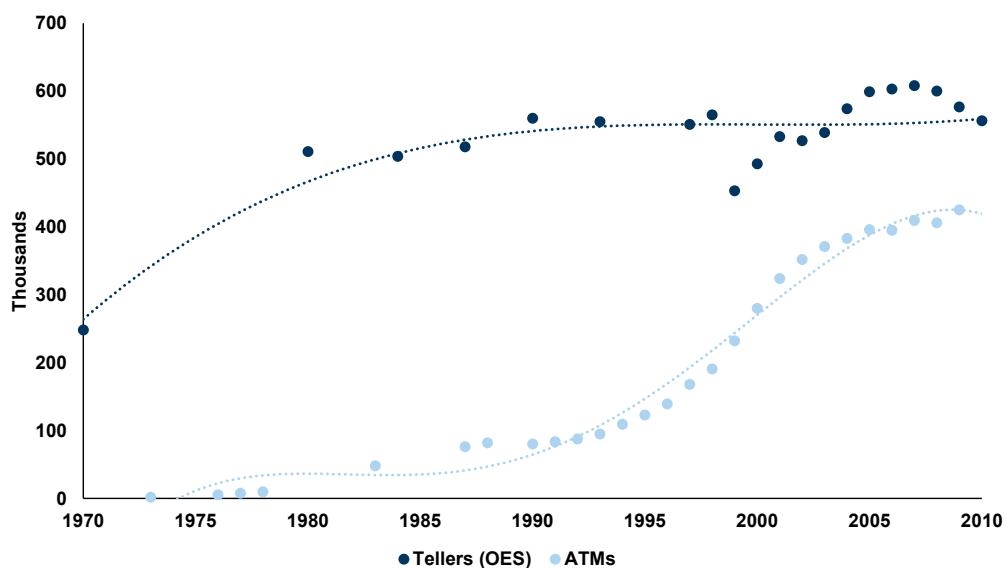
Low adoption has limited the labor market impact, but preliminary evidence suggests that AI is modestly raising labor demand while driving negligible job loss. GS Macro states that based on actual adoption of AI has only modestly increased so far, with only 5% of companies reporting use of Generative AI in regular production. And while adoption is higher in industries that we estimate will benefit the most from AI—including computing and data infrastructure, information services, and motion picture and sound production—and is expected to rise going forward, adoption remains well below levels necessary to see large aggregate productivity gains. Low adoption has limited the labor market impact, but preliminary evidence suggests that AI is modestly raising labor demand while driving negligible job loss, thereby creating a slightly positive impulse to net hiring. Finally, emerging evidence from early adopters points to large increases in labor productivity. While early estimates should be interpreted cautiously given selection and reporting biases, recent academic studies imply an average 25% increase in labor productivity (16% median) following AI adoption, with anecdotal company reports suggesting similarly large efficiency gains. The slow adoption pace

suggests that sizable macroeconomic impacts are still several years off.

Expect labor reshuffle toward Gen-AI skills; labor market to re-skill to meet market demand over time. While we may see some companies re-structure their workforce to orient their headcount around AI-enabling talent, we expect workers to re-skill over time to meet market demand. Intuit's July workforce reduction is an example of such a move which the company stated was designed to re-direct resources to Gen-AI-related investments rather than cutting costs in the face of AI efficiencies. We note that the company emphasized that headcount will actually be up by the end of the fiscal year in which they are making this transition, not down. Similarly, while a [survey by Upwork](#) reported 96% of C-suite leaders expect AI to boost worker productivity, 77% of employees report AI has increased their workload. This is consistent with Monday.com's commentary that they have seen a 2x increase in chat ticket volume after the implementation of Gen-AI reduced the time to resolve a ticket by 50%. From the same Upwork survey, 47% of employees said they have no idea how to achieve the productivity gains their employers expect with AI, so it appears we remain a ways off from broader labor force disruption.

Workers capabilities and productivity can be significantly expanded with Gen-AI applications. In similar ways that the internet and mobile platform shifts included a significant number of workers previously excluded from the global knowledge economy, we believe Gen-AI will expand the labor pool and improve productivity. Workers around the world will be able to reduce relatively mundane work and redirect their attention to more complex and people-oriented tasks. Similarly, workers will be able to lessen language and education barriers, opening the global economy to otherwise under-tapped potential. The transformation of the labor market is difficult to predict and often counter-intuitive. Take the invention of the ATM. Few would have imagined that 50 years after the first ATMs were installed, the amount of tellers employed in the US has more than doubled ([Exhibit 52](#)). Bank branches became cheaper to open, employees diversified across branches, and the job of a teller expanded beyond simply handing customers cash into a more service-oriented role. Examples like this are all over the labor market, with technological disruption actually expanding the role of the labor force. We believe that as Gen-AI proliferates and evolves, the expanded capabilities and time of the global workforce will create a self-reinforcing cycle that expands both the demand for Gen-AI products and the global economy writ large.

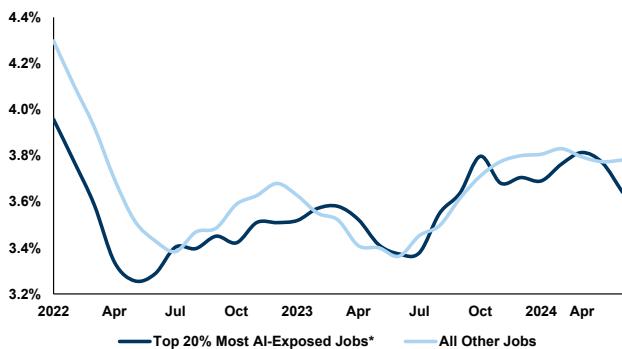
Exhibit 52: Despite predictions that the ATM would displace hundreds of thousands of bank tellers, total bank tellers employed in the US actually increased



Source: James Bessen "Learning by Doing" (2015), US Bureau of Labor Statistics, Bank of International Settlement

- **AI's impact on the labor market has been very modest thus far.** AI-related layoff announcements have remained muted over the past few months and the unemployment rate for the most AI-exposed positions has ticked down slightly in recent months relative to the broader unemployment rate. While we believe the labor market will have to adjust to the broader adoption of AI tools in the longer-term, there has been very little evidence so far of any significant AI-related labor disruption. Exhibit 53 compares the unemployment rates for workers that are most exposed to AI automation (defined as those in the top 20% most AI-exposed occupations) with the rest of the workforce. Unemployment rates across these groups have tracked each other closely since 2022, suggesting little job loss due to Generative AI thus far. Similarly, as Exhibit 54 shows, the share of AI-related positions declined in 2022 as tech companies pulled back on aggressive hiring after the pandemic, but stabilized in late-2022 following the emergence of ChatGPT and subsequently began to rise in 2023. While overall AI-related hiring demand remains subdued, the share of AI-related job openings in the IT sector has surged over the last year and a half, consistent with an increasing focus on AI in the technology industry.

Exhibit 53: AI exposed jobs have not seen an unusually large rise in unemployment

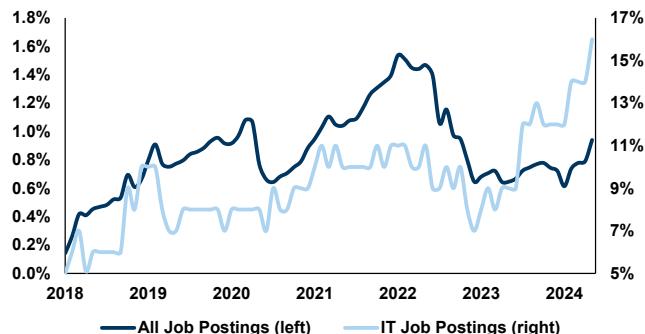


*Jobs for which an estimated 40%+ of task content is automatable by AI, benchmarked to February 2024 CPS

Source: Goldman Sachs Global Investment Research, Census Bureau, IPUMS

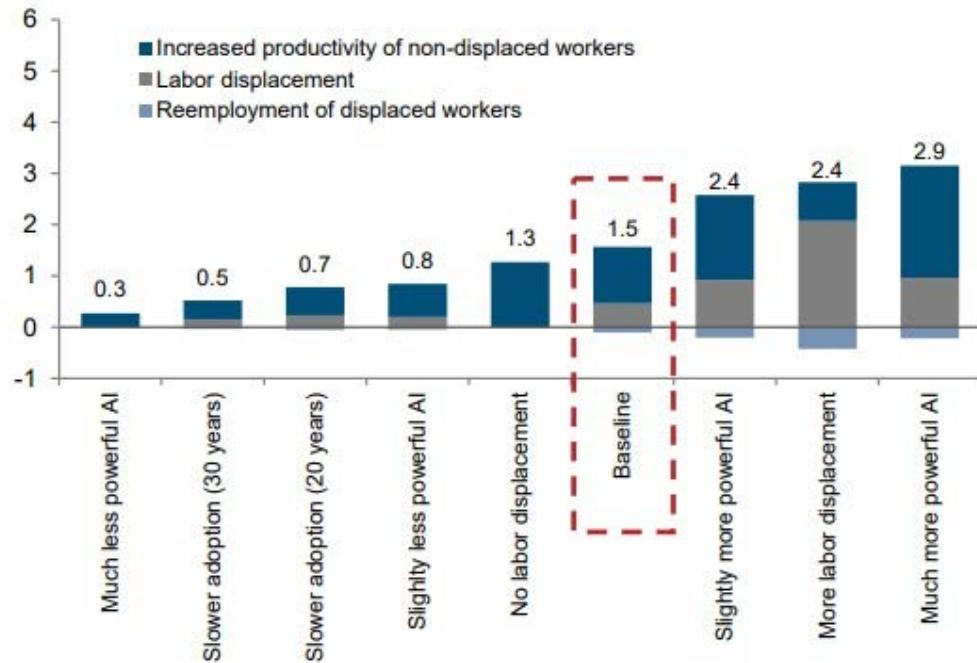
Exhibit 54: AI job openings have risen modestly as a share of overall jobs, but surged as a share of IT job openings

AI-Related Job Postings as a share of all job openings and IT job postings

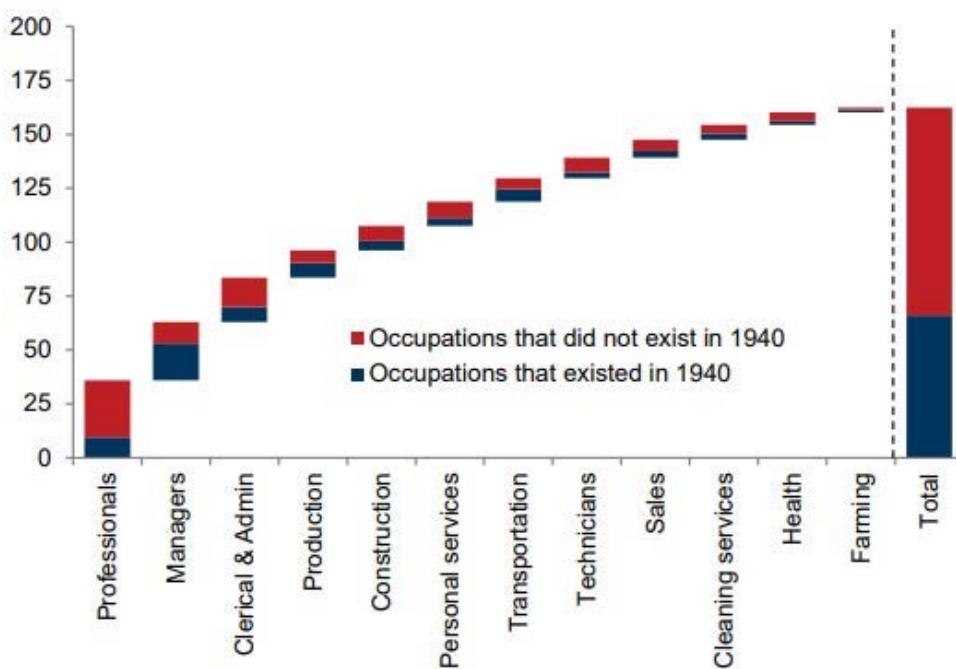


Source: Goldman Sachs Global Investment Research, LinkUp

- **History suggests technological shifts boost productivity, create jobs.** As we believe Gen-AI will have the ability to automate a large share of work tasks, Gen-AI should provide a boost to labor productivity and drive employment growth. GS macro baseline estimates imply as much as 15% cumulative gross upside to US labor productivity and GDP growth following widespread adoption of AI (Exhibit 55). Similarly, MIT economist David Autor and coauthors, using Census data, find 60% of workers today are employed in occupations that did not exist in 1940, with their estimates implying that the technology-driven creation of new occupations accounts for more than 85% of employment growth over the last 80 years (Exhibit 56).

Exhibit 55: GS Macro estimates AI will provide a significant boost to US labor productivity

Source: Goldman Sachs Global Investment Research

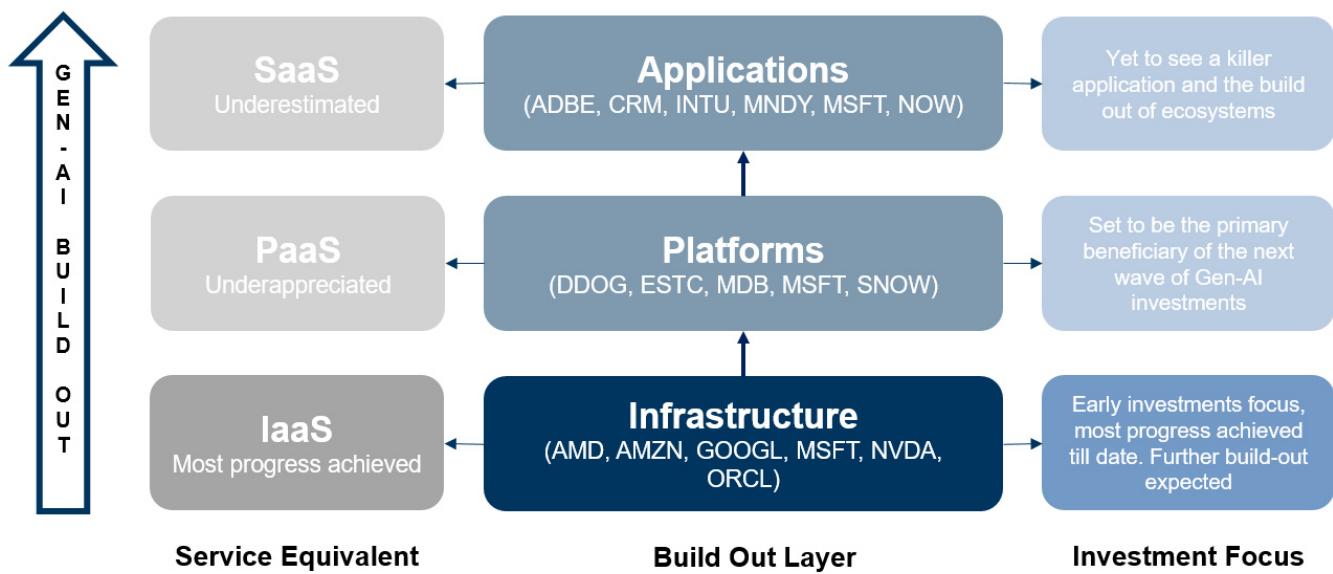
Exhibit 56: Technological creation of new opportunities is the main driver of employment and economic growth**Employment by new and pre-existing occupations, millions**

Source: Autor et al. (2022), Goldman Sachs Global Investment Research

Infrastructure layer the furthest along in tilt towards Gen-AI**Hyperscalers, Foundation Model providers to serve as AI enablers.** To support the

high-compute needs of Gen-AI workloads, a handful of well-capitalized companies were able to build the capacity pool for the industry. With the hyperscalers already possessing notable market share of the cloud cycle (C23 cumulative spend: \$193bn), we believe this leadership likely percolates into the Gen-AI era, too, as the high capital requirements underpinning Gen-AI services provides a durable competitive moat. We also think NVIDIA will remain comfortably entrenched given their exposure to highly sought after GPUs. The TAM expansion that Gen-AI can catalyze as companies work to migrate to the cloud to utilize this technology, and the architectural shift in the underlying tech stack leaves room for foundational model providers to become notable players, too. Through an array of investments (direct inventory, minority investments, etc.), these companies have enabled training of large foundational models and are supporting the fine-tuning and inferencing that will be required to migrate demand to the application layer. Despite adding potential margin pressure, we expect unit economics to improve as we move through the adoption curve. Similar to the cloud computing cycle, we expect the distinction between new / core workloads to abate over time, with Gen-AI evolving into the common instance.

Exhibit 57: Infrastructure layer has experienced the most value accretion from Gen-AI to date



Source: Goldman Sachs Global Investment Research

Addressing the current capital investment cycle

Clear parallels between cloud computing and Gen-AI cycles. Mirroring the cloud computing cycle where the Infrastructure layer was the initial beneficiary, a similar dynamic is transpiring with the Gen-AI cycle. We believe companies like Amazon, Google, Microsoft, and Oracle are favorably positioned to see the initial revenue tailwinds. This is supported via their exposure to building out and providing the required infrastructure to support broader Gen-AI proliferation to the Platform and Application layers. While investors may draw pause given Gen-AI's current CapEx efficiency (Gen-AI CapEx / Gen-AI Revenue), we believe that the initial capital outlay (albeit notable) is required to reap the eventual rewards. Furthermore, we remain encouraged given the

near-term ROI on this CapEx has proven more tangible than that seen during cloud's initial buildout. Accordingly, we believe this initial phase of the I-P-A cycle could follow a similar progression to cloud - where investors were initially hesitant to underwrite the CapEx requirements, but the eventual revenue benefits helped assuage those concerns.

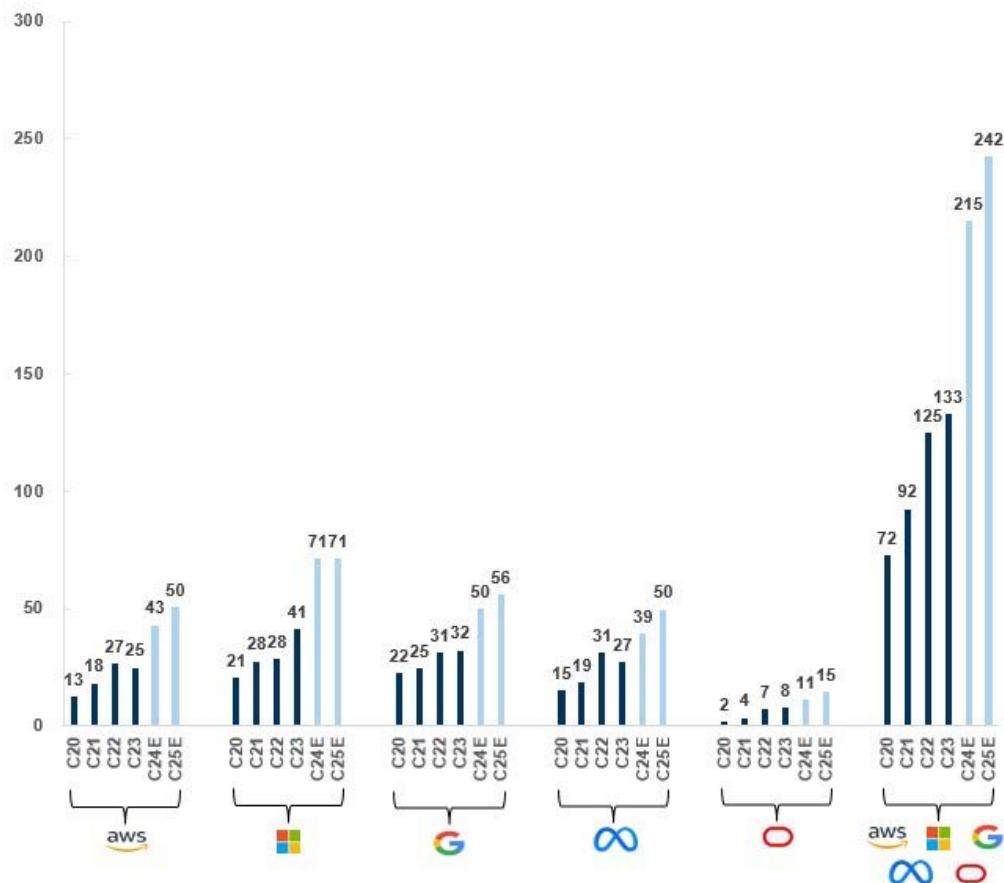
Cyclical differences - substitution versus augmentation. While appreciating the parallels between cloud computing and Gen-AI, we highlight a key difference between the cycles. The cloud computing cycle was substitution-focused – customers were trading their on-premise equipment for cloud-based infrastructure. This compares to the Gen-AI cycle, which has been, and will continue to be, augmentation focused. Namely, the augmentation of human capital which fosters enhanced productivity. This has been initially evidenced via Microsoft's Copilot and Adobe's Firefly.

Hyperscalers: The Capital Outlay and Margin Implications

Current CapEx cycle sets stage for emerging AI opportunity. We see the magnitude and duration of the current CapEx cycle as having both historical precedent and merit for justification given its current success. As it stands, we expect the largest players in the market (Alphabet, Amazon, Microsoft, Meta, Oracle) will each invest a heightened amount on CapEx in C24E / C25E, resulting in cumulative CapEx spend of >\$200bn in each of those years, respectively (Exhibit 58). For context, this compares to \$92bn / \$125bn / \$133bn C21 / C22 / C23. In laying the groundwork for the cloud computing cycle, the Hyperscalers poured billions into CapEx to develop the underlying infrastructure that eventually facilitated the remarkable growth in SaaS applications. These investments revolved around things like networking equipment, servers, and CPUs. In this evolution, companies are indexing their investments toward GPUs and energy, with the supply limitations of each driving a race to capture capacity (further fueling the group's elevated spending). Foundation Model providers require a heavy architecture for both training (as we continue to push the boundaries of performance) and inferencing (as these providers commercialize their offerings; i.e., ChatGPT, Perplexity). Supported by highly cash generative business like Google Ads (GCP), Microsoft O365 / PCs (Azure) and Oracle Database (OCI) the hyperscalers have a competitive moat, in addition to their expansive data center footprints and existing IT architectures.

Exhibit 58: Hyperscaler CapEx is expected to see material increases in C24 / C25 (GSe), largely driven by incremental AI investments

\$bn\$



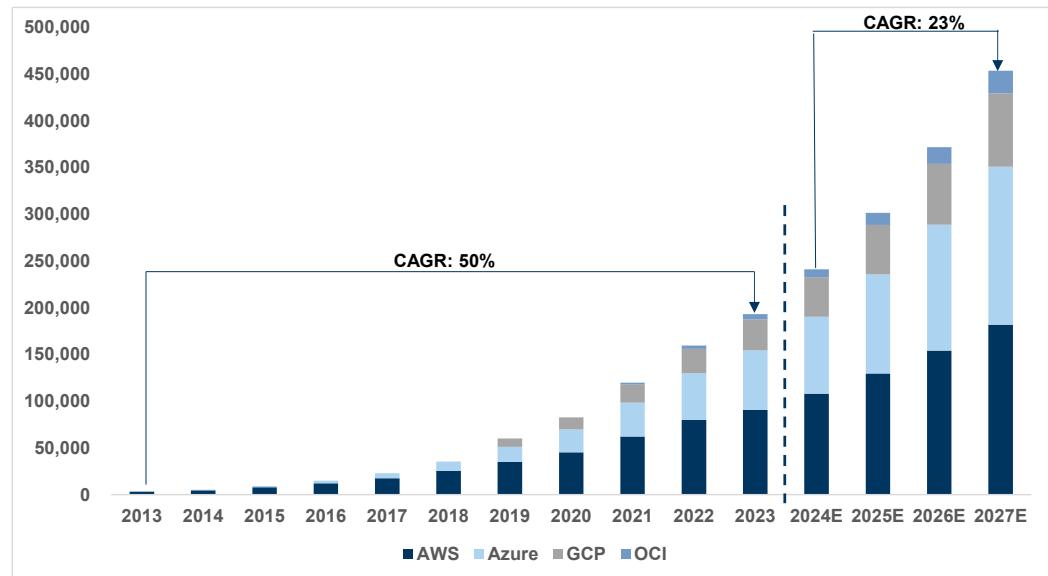
Microsoft CapEx is calendarized, includes Finance Leases

Source: Goldman Sachs Global Investment Research, Company Data

We've seen this initial investment in infrastructure (e.g., networking, equipment, servers, CPUs, etc.) yield substantial dividends to date as hyperscalers' cloud revenue grew at a +50% CAGR from 2013 to 2023 ([Exhibit 59](#)); on a go-forward basis, we estimate a >20% CAGR through 2027 even as aggregate Hyperscaler revenues approach >\$200B.

Exhibit 59: Hyperscaler revenue trends have continued showing durable growth despite an increasing scale

Vertical axis represents millions of dollars



Source: Goldman Sachs Global Investment Research, Company Data

Efficiency curve adds promise

Initial proof points suggest accelerating CapEx cycle supporting demonstrable ROI

over a longer arc; similar to cloud computing cycle. It's worth drawing a parallel to the initial transition from the client-server architecture to the cloud back in the early-2000's, which was initially met with much skepticism given the lack of killer applications. The IPOs of next-generation companies like Google and Salesforce in 2004 ultimately brought to the forefront the possibilities of the cloud model for both consumer and enterprise services. Accordingly, today we can very clearly observe the breadth of applications and platform services that were enabled and scaled from the initial cloud computing build-out. As laid out in the application layer section of this report, we remain on the hunt for "killer applications." What drives our conviction in the longer-term ROI of upfront CapEx is enterprises' extensive data repositories that, for the first time, can be more easily tapped into to create highly personalized and domain-specific applications that unlock a new level of productivity. We highlight Microsoft as an example of this, which has seen tangible demand for AI services within its Azure segment. This is evidenced by Gen-AI revenues exiting FY24 at a >\$5bn run-rate only five quarters after it started contributing to the segments' growth.

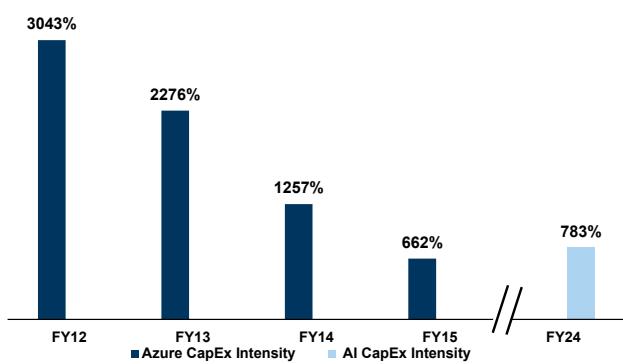
Hyperscaler monetization has outpaced cloud computing cycle despite ROI debate.

While we acknowledge the growing scrutiny over ramping CapEx investments versus the near-term ROI, Gen-AI is still displaying an accelerated monetization ramp relative to the cloud computing cycle. Further, spend remains largely concentrated with Foundation Model providers (training and inferencing). Exiting C2Q24 – roughly 6 quarters after ChatGPT's ascent to commercial popularity – Gen-AI revenues have outpaced that of cloud when compared on a relative basis. For instance, Microsoft

Azure's annualized Gen-AI run-rate of >\$5bn exiting F4Q24 compares to Azure's scale ~7 years after its initial commercial release (2010), with Azure not achieving this mark until C2Q17. This monetization has transpired elsewhere, too - Google disclosed that AI-related revenues have reached "billions of dollars" year-to-date on their C2Q24 earnings.

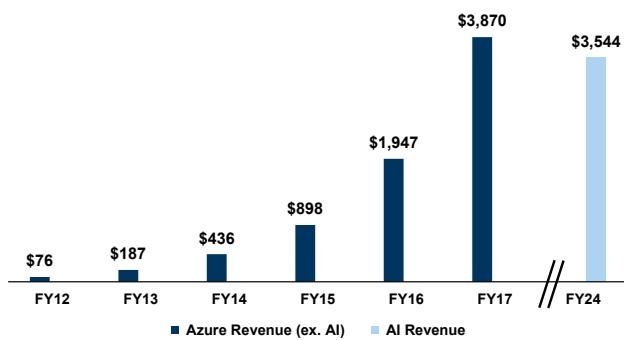
- **Analysis indicates Microsoft's spend on AI initiatives likely amounted to ~\$28bn in FY24.** Based on Microsoft's ~\$56bn in F24 CapEx, we estimate that ~\$28bn was directly tied to Gen-AI spending. Although the AI CapEx build out has scaled rapidly, Microsoft has successfully matched these investments with tangible revenue contribution (F4Q24: >\$5bn run-rate, GSe). Five quarters in, we see this playing out well as (based on our estimates) Microsoft's AI-related CapEx build out has a similar CapEx intensity (CapEx / Revenue) to Azure four / five years after it was commercially available (2010). We find that the AI-specific CapEx intensity has rapidly fallen over the last four quarters, now matching FY14 / FY15 of the cloud computing cycle ([Exhibit 60](#)). Revenue, however, is ramping at an accelerated pace ([Exhibit 61](#)), with AI-specific revenues reaching ~\$3.5bn FY24 (GSe) - F4Q24 run-rate: >\$5bn - comparable to year seven (FY17) of the cloud computing cycle.
- **AI ecosystem players also scaling rapidly.** We would also highlight our expectations for Gen-AI related revenue to be ~\$20bn in C24, highlighted by OpenAI achieving a \$3.4bn run-rate in May. Furthermore, on the hardware side, NVIDIA's Data Center revenue (of which we approximately 80 - 85% is attributable to AI) exceeded a \$90bn run-rate following F1Q25 as the company remains a crucial part of the Gen-AI value chain. This heightened cumulative spending behavior likely highlights customers' interest, and constructive view on Gen-AI's early value proposition, which should only accelerate against increased investments.

Exhibit 60: AI CapEx intensity (AI CapEx / AI Revenue) today is at similar levels to 4 - 5 years after Azure went GA



Source: Goldman Sachs Global Investment Research, Company Data

Exhibit 61: Five quarters in, Azure AI revenue is at a higher base than 7 years after Azure's launch



Source: Goldman Sachs Global Investment Research, Company Data

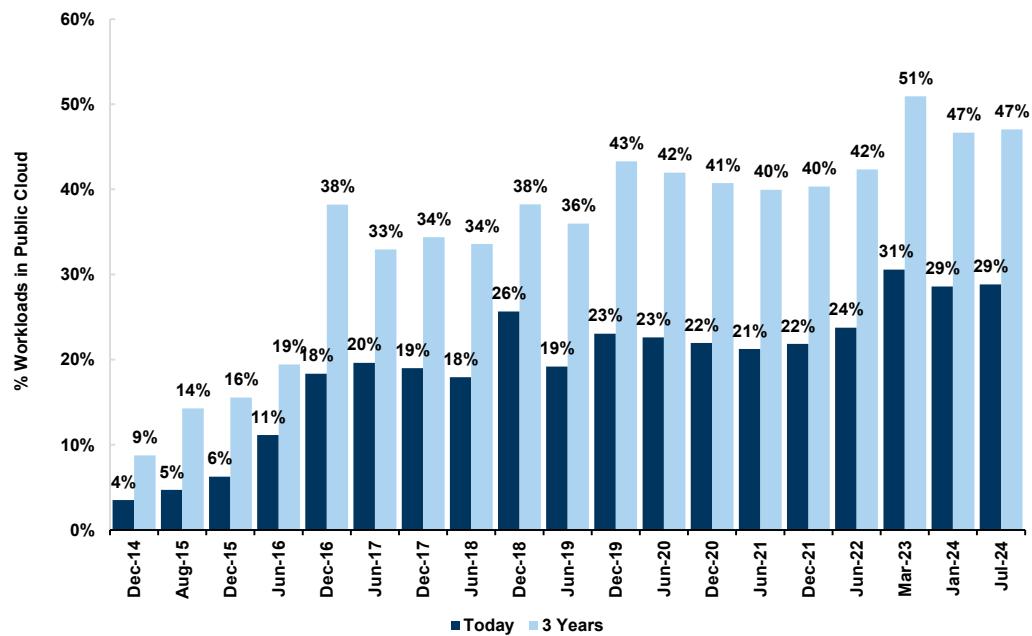
Gen-AI introduces complementary revenue opportunity, not cloud revenue

cannibalization. While investors may speculate that Gen-AI could cannibalize cloud budgets, we believe it should prove more complementary. As Gen-AI is inherently cloud native, it could help catalyze cloud migration as companies look to realize efficiencies by integrating it across various workflows. Indeed, we highlight that in our latest published [CIO survey](#), < 1/3 of workloads sit on public clouds ([Exhibit 62](#)), with this figure showing

relative consistency for the last three surveys (thereby implying minimal recent migration activity). However, should Gen-AI meet / exceed expectations, the 3-year forward expectations could prove conservative as companies look to leverage this technology to remain competitive. We also reference [Exhibit 10](#), [Exhibit 11](#), [Exhibit 12](#), as helping support the view that Gen-AI usage intent has remained consistent / trended positively year-to-date (implying that cloud migrations could accelerate). We believe that companies leveraging cloud infrastructure to utilize Gen-AI highlights the complementary (versus competing) nature of the two solutions.

Exhibit 62: Gen-AI could catalyze migration activity to the cloud

Question: "What percent of your applications have you moved to public cloud platforms today and what do you expect in 3 years from now (e.g. Amazon AWS, Microsoft Azure, Google Cloud Platform, etc.)?"



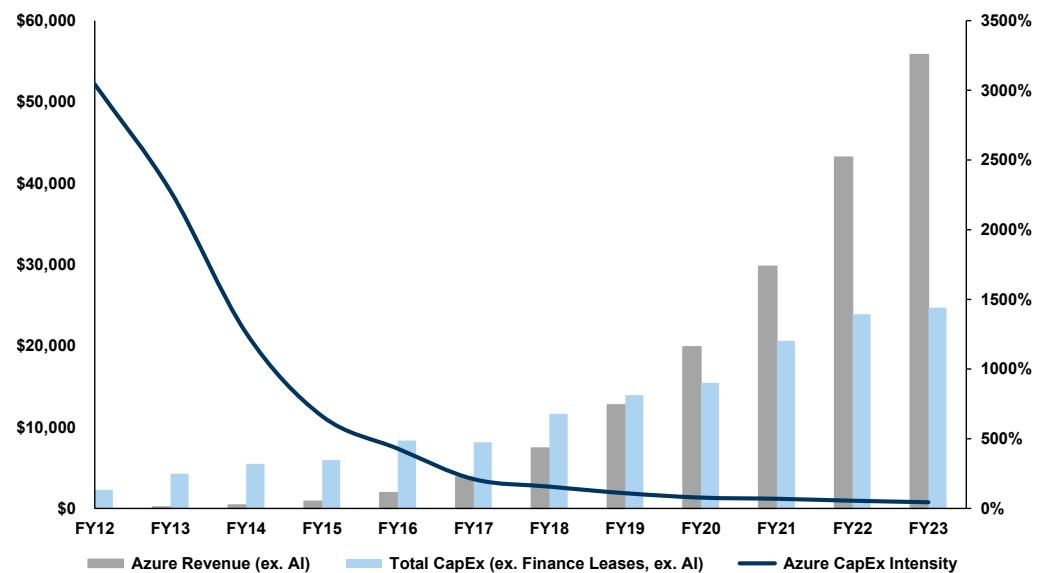
Source: Goldman Sachs Global Investment Research

AI CapEx intensity likely to fall below 300% within 2-3 years, move down

materially over time. As Azure has scaled, these investments have moderated out to a more normalized level at around 35% of Azure revenue (ex. AI) ([Exhibit 63](#)). Though we have parsed out numerous similarities to the cloud computing cycle in this report, we also note that in contrast to the Azure cycle, Gen-AI's build out benefits from: **1)** A significant amount of initial excitement in terms of pace of adoption and spend and, **2)** An investment philosophy that is catered around building agile, broad-based architecture that Microsoft can direct toward wherever tangible demand builds. For example, instead of Microsoft building out GitHub Copilot-specific infrastructure that can't be easily leveraged for another application, the company is structuring their investments around a common architecture. This can yield higher ROI and greater operating efficiency in this cycle as the company will be more aligned with end-market demand. Lastly, as companies start to iron out the right Gen-AI use cases in the platform and application layers, Microsoft's investment philosophy and presence across the IT stack position it well to continue to capture share of Gen-AI budgets. Therefore, we expect Microsoft to reach a level of critical scale (represented by CapEx / Revenue < 300%) at a faster clip than in Azure (where it took seven years). Given Microsoft is already operating at a

CapEx intensity of ~800% (where Azure was after ~4 / 5 years), we argue that we could see Gen-AI efficiency below 300% within 2-3 years.

Exhibit 63: The Azure CapEx buildout didn't see sub-300% CapEx intensity until FY17 (seven years into the Azure buildout)



Source: Goldman Sachs Global Investment Research, Company Data

Expect gross profit tailwinds despite margin pressure

Margin impacts from Gen-AI investments likely to prove palatable. Across the 10+ year cloud transformation, the hyperscalers showed a willingness to invest purposefully and aggressively ahead of the revenue opportunity they saw. We believe this investment cycle ([here](#), [here](#)) underpins the same level of conviction in the promise of Gen-AI.

Though investors are coming to the same conclusions, we see the market still finding its footing with these stocks as they wrestle with the gross margin impact on the back of the growing CapEx spend (and the flow through of D&A to the income statement).

This focus on margins discounts the profit dollars added to the business. The same dynamic was witnessed with cloud, which had ~70% GM initially versus 90% for on-premise solutions. In our view, this missed the mark as the revenue opportunity was growing at a faster clip. While history suggests that CapEx intensity has increased over a longer time horizon, the benefit has been the ability to accrue significant incremental dollars. Reiterating our belief that the long-term revenue opportunity is attractive, coupled with successful initial monetization, we believe this helps justifies the near-term CapEx.

- **Microsoft:** As initially discussed in our [Microsoft CapEx note](#), we currently expect Microsoft to encounter ~200 / 350bps of gross margin headwinds in F25 / F26, respectively ([Exhibit 64](#)). While investors may draw a pause at margin compression against a backdrop that remains keen on profitability, we remind investors that this drawdown is much more muted versus the 19pp drawdown witnessed with Azure's build out (80% → 61%). This is partially attributable to the associated Gen-AI dollars accumulating at a faster rate than was seen with Azure, thereby helping offset the

associated margin pressure. With our belief that Gen-AI gross profit dollars will be at comparable levels in F25 / F26 to that of Azure gross profit dollars eight years post-launch, investors are likely more willing to underwrite these investments, especially given the sizable medium-term revenue opportunity this technology presents. While expecting slight gross margins headwinds, Microsoft's track record of cost consciousness could support a continued top-tier operating margin profile (F24: 45%). Supporting this is the fact that Microsoft has continued to deliver an expanding operating margin profile since F16 (with F23 being flat). Furthermore, we highlight that over the LTM, Microsoft has contended with both Gen-AI investments (that were contributing lower revenue dollars versus current levels), and the integration of ATVI. Despite these factors, the company maintained operating margins >40%.

Exhibit 64: We estimate MSFT's AI CapEx has a ~200-350bps impact to GM in FY25 and FY26

	3Q23	4Q23	1Q24	2024	3Q24	4Q24	FY24	1Q25E	2Q25E	3Q25E	4Q25E	FY25E	1Q26E	2Q26E	3Q26E	4Q25E	FY26E	
Total CapEx (inc. Finance Leases)	\$7,800	\$10,700	\$11,200	\$11,500	\$14,000	\$19,000	\$55,700	\$19,390	\$18,872	\$18,270	\$18,453	\$74,984	\$17,537	\$17,025	\$17,150	\$17,641	\$69,353	
Non-AI CapEx (inc. Finance Leases) % yoy	\$6,988	\$6,988	\$6,988	\$6,988	\$6,988	\$6,988	\$27,950	\$6,988	\$6,988	\$6,988	\$6,988	\$27,950	\$6,988	\$6,988	\$6,988	\$6,988	\$27,950	
AI CapEx (inc. Finance Leases) % yoy	\$813	\$3,713	\$4,213	\$4,513	\$7,013	\$12,013	\$27,750	\$12,402	\$11,884	\$11,282	\$11,465	\$47,034	\$10,550	\$10,038	\$10,162	\$10,653	\$41,403	
Depreciation per Quarter (6-Year Useful Life)								175%	69%	-3%	-59%		69%	-11%	-11%	-77%	-12%	
3Q23	\$34	\$34	\$34	\$34	\$34	\$135	\$34	\$34	\$34	\$34	\$135	\$34	\$34	\$34	\$34	\$135		
4Q23		\$155	\$155	\$155	\$155	\$619	\$155	\$155	\$155	\$155	\$619	\$155	\$155	\$155	\$155	\$619		
1Q24			\$176	\$176	\$176	\$527	\$176	\$176	\$176	\$176	\$702	\$176	\$176	\$176	\$176	\$702		
2Q24				\$188	\$188	\$376	\$188	\$188	\$188	\$188	\$752	\$188	\$188	\$188	\$188	\$752		
3Q24					\$292	\$292	\$292	\$292	\$292	\$292	\$1,169	\$292	\$292	\$292	\$292	\$1,169		
4Q24						\$501	\$501	\$501	\$501	\$501	\$2,002	\$501	\$501	\$501	\$501	\$2,002		
1Q25E							\$517	\$517	\$517	\$517	\$517	\$517	\$517	\$517	\$517	\$2,067		
2Q25E								\$495	\$495	\$495	\$495	\$495	\$495	\$495	\$495	\$1,981		
3Q25E									\$495	\$495	\$495	\$495	\$495	\$495	\$495	\$495	\$1,880	
4Q25E										\$470	\$470	\$470	\$470	\$470	\$470	\$470	\$1,711	
1Q26E											\$478	\$478	\$478	\$478	\$478	\$478	\$1,519	
2Q26E												\$440	\$440	\$440	\$440	\$440	\$836	
3Q26E													\$418	\$418	\$418	\$423	\$423	
Total AI Depreciation		\$34	\$189	\$364	\$552	\$844	\$1,949	\$1,345	\$1,862	\$2,357	\$2,827	\$8,390	\$3,305	\$3,744	\$4,162	\$4,586	\$15,797	
Total AI Revenue % yoy % of Total Revenue		\$123	\$384	\$827	\$972	\$1,361	\$3,544	\$2,112	\$2,894	\$2,915	\$3,403	\$11,324	\$3,697	\$4,051	\$4,081	\$4,764	\$16,593	
Total Core Revenue % yoy		\$52,857	\$56,066	\$56,133	\$61,193	\$60,886	\$63,366	\$241,578	\$62,520	\$67,002	\$67,354	\$70,408	\$267,284	\$69,989	\$75,879	\$76,435	\$80,046	\$302,350
Total Revenue (Unchanged)		\$52,857	\$56,189	\$56,517	\$62,020	\$61,858	\$64,727	\$245,122	\$64,632	\$69,896	\$70,269	\$73,811	\$278,608	\$73,686	\$79,930	\$80,516	\$84,811	\$318,943
AI Gross Profit		\$46,109	\$48,871	\$40,215	\$42,397	\$43,353	\$45,043	\$171,008	\$44,722	\$47,573	\$48,093	\$50,587	\$190,976	\$51,086	\$55,155	\$55,875	\$58,937	\$221,053
Core Gross Profit		\$46,109	\$48,782	\$40,019	\$41,934	\$42,933	\$44,526	\$169,413	\$43,955	\$46,541	\$47,535	\$50,011	\$188,042	\$50,694	\$54,848	\$55,956	\$58,758	\$220,257
AI Impact to Gross Profit		\$0	\$89	\$196	\$463	\$420	\$517	\$1,595	\$768	\$1,032	\$558	\$576	\$2,934	\$392	\$307	(\$51)	\$178	\$796
AI Gross Margin Core Gross Margin		87.2%	87.0%	71.2%	68.4%	70.1%	69.6%	69.8%	69.2%	68.1%	68.4%	68.5%	68.5%	69.3%	69.0%	69.4%	69.5%	69.3%
AI Impact to Gross Margins		87.2%	87.0%	71.3%	68.5%	70.5%	70.3%	70.1%	70.3%	69.5%	70.5%	71.0%	70.4%	72.4%	73.2%	73.4%	72.8%	

Source: Goldman Sachs Global Investment Research, Company Data

- **Google (covered by Eric Sheridan):** We estimate that the current elevated CapEx cycle will result in depreciation expense growing at a ~19% 3-year CAGR (C23-C26E) and a ~16% 6-year CAGR (C23-C29E). Similar to our META analysis, we base our analysis on an assumed mix of future CapEx between Servers & Network Equipment and Other CapEx (based on historical categorization of gross additions to PP&E as disclosed by the company) and an assumed useful life applied to each category (informed by company disclosures).

Exhibit 65: GOOGL - CapEx and Depreciation analysis

Alphabet Capital Intensity Analysis	2018	2019	2020	2021	2022	2023	2024E	2025E	2026E	2027E	2028E	2029E
<i>Gross Additions to PP&E</i>												
Total Capex*	\$ 22,860	\$ 21,700	\$ 22,255	\$ 20,551	\$ 24,697	\$ 30,093	\$ 50,072	\$ 56,152	\$ 59,128	\$ 61,425	\$ 63,048	\$ 63,992
% y/y growth	89%	-5%	3%	-8%	20%	22%	66%	12%	5%	4%	3%	1%
Servers & Network Equipment	\$ 8,690	\$ 6,721	\$ 9,066	\$ 9,700	\$ 10,661	\$ 14,327	\$ 23,839	\$ 26,733	\$ 27,263	\$ 27,401	\$ 27,179	\$ 26,627
% y/y growth	63%	-23%	35%	7%	10%	34%	66%	12%	2%	1%	-1%	-2%
% of total capex	38%	31%	41%	47%	43%	48%	48%	48%	46%	45%	43%	42%
Other Capex	\$ 14,170	\$ 14,979	\$ 13,189	\$ 10,851	\$ 14,036	\$ 15,766	\$ 26,233	\$ 29,418	\$ 31,865	\$ 34,024	\$ 35,869	\$ 37,366
% y/y growth	109%	6%	-12%	-18%	29%	12%	66%	12%	8%	7%	5%	4%
% of total capex	62%	69%	59%	53%	57%	52%	52%	54%	55%	57%	57%	58%
<i>Useful Life Assumptions</i>												
Servers & Network Equipment	3.0	3.0	3.0	4.0	4.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
Other Capex:												
Land and Buildings	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0
Leasehold Improvements	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0
Furniture and Fixtures	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
Construction in Progress	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Total Annual Depreciation</i>												
Total Depreciation							\$ 13,614	\$ 18,744	\$ 23,404	\$ 27,643	\$ 31,572	\$ 34,710
% y/y growth	n/a						38%	25%	18%	14%	10%	
% of annual capex	27%						33%	40%	45%	50%	54%	

Source: Goldman Sachs Global Investment Research, Company Data

- **Meta (covered by Eric Sheridan):** We estimate that the current elevated CapEx cycle will result in depreciation expense growing at a ~28% 3-year CAGR (C23-C26E) and an ~18% 6-year CAGR (C23-C29E). We base our analysis on an assumed mix of future CapEx between Servers & Network Equipment and Other CapEx (based on historical categorization of gross additions to PP&E as disclosed by the company) and an assumed useful life applied to each category (informed by company disclosures). Along with our estimated higher CapEx on an absolute dollar basis going forward, we also assume Meta spends a higher percentage of CapEx on Servers & Network Equipment (mainly GPUs) than historical averages, which, given the shorter depreciable life vs. other CapEx, is driving the growth in estimated depreciation expense over the next ~5-6 years.

Exhibit 66: META - CapEx and Depreciation analysis

Meta Capital Intensity Analysis	2018	2019	2020	2021	2022	2023	2024E	2025E	2026E	2027E	2028E	2029E
<i>Gross Additions to PP&E</i>												
Total Capex*	\$ 13,236	\$ 14,413	\$ 15,065	\$ 16,838	\$ 26,604	\$ 25,228	\$ 39,479	\$ 49,599	\$ 54,639	\$ 56,571	\$ 58,106	\$ 59,394
% y/y growth	103%	9%	5%	12%	58%	-5%	56%	26%	10%	4%	3%	2%
Servers & Network Equipment	\$ 5,019	\$ 3,987	\$ 4,999	\$ 3,581	\$ 8,746	\$ 12,508	\$ 19,574	\$ 24,591	\$ 25,724	\$ 25,219	\$ 24,451	\$ 23,508
% of total capex	38%	28%	33%	21%	33%	50%	50%	50%	47%	45%	42%	40%
Other Capex	\$ 8,217	\$ 10,426	\$ 10,066	\$ 13,257	\$ 17,858	\$ 12,720	\$ 19,905	\$ 25,008	\$ 28,915	\$ 31,352	\$ 33,655	\$ 35,886
% of total capex	62%	72%	67%	79%	67%	50%	50%	50%	53%	55%	58%	60%
<i>Useful Life Assumptions</i>												
Servers & Network Assets	4.0	4.0	4.0	4.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Other Capex:												
Buildings	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Leasehold Improvements	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0
Equipment & other	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
Finance lease right-of-use assets	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0
Construction in progress	n/a											
Land	n/a											
<i>Total Annual Depreciation</i>												
Total Depreciation							\$ 14,186	\$ 19,721	\$ 25,885	\$ 30,739	\$ 35,073	\$ 37,789
YoY growth	n/a						39%	31%	19%	14%	8%	
% of annual capex	36%						40%	47%	54%	60%	64%	

Source: Goldman Sachs Global Investment Research, Company Data

Technological investments

The complexity of workloads has increased as the industry transitions from “retrieval-based AI” to “Generative AI”. Processing these complex workloads in a power efficient manner requires innovation across the entire Hardware and Software stack. Primary GPU designers like NVIDIA and AMD have moved to an annual cadence of product launches to cater to increased demand for Compute and related Networking. Consistent with the historical evolution of the semiconductor industry, we believe

providing customers with cheaper marginal cost of compute would eventually lead to them doing ‘more’. And, on that note, we are encouraged by the gen-to-gen uplift being provided by the merchant GPU designers, as well as hyperscalers upping their effort to design their own custom chips on more advanced nodes and including latest industry technologies like High Bandwidth Memory (HBM).

Exhibit 67: Accelerating product releases and consistent gen-to-gen performance uplifts aimed to cater to increased demand for Compute and related Networking

AMD	Launch Date	Gen-to-Gen Performance Uplift
MI300	4Q23	6.8x better Training performance versus MI250
MI325	4Q24	288GB of HBM3e Memory versus 192GB HBM3 in MI300
MI350	2025	35x faster inference versus prior architecture
MI400	2026	
NVDA	Launch Date	Gen-to-Gen Performance Uplift
V100	2017	5x improvement versus Pascal architecture
A100	2020	20x higher performance versus Volta architecture
H100	2022	30x higher performance versus Ampere architecture
H200	2023	2x better inference performance versus H100
Blackwell Architecture	2024	25x less cost and energy consumption versus Hopper architecture
Blackwell Ultra-Architecture	2025	
Rubin Architecture	2026	

Source: Company data

Supply constraints likely to ease; may see industry diversification beyond GPUs

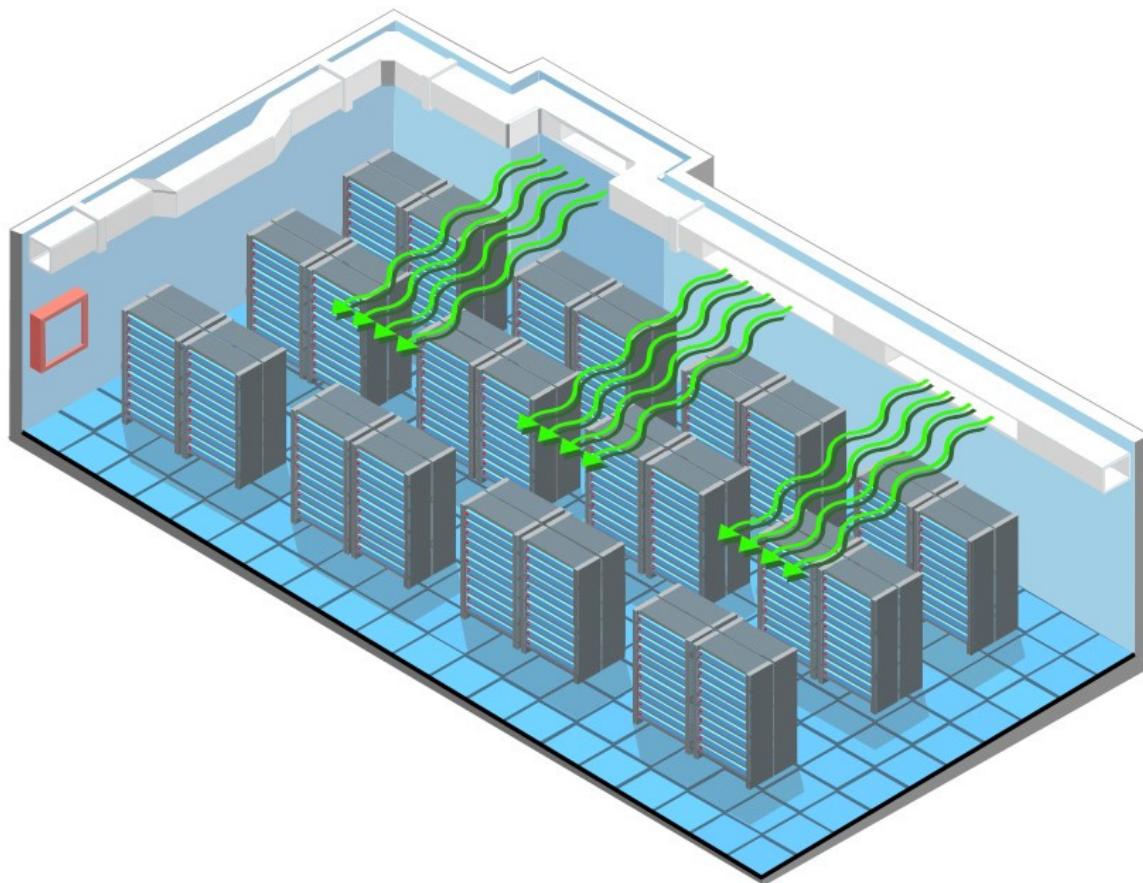
(per Toshiya Hari). Despite the scale of investments just discussed, GPU supply-demand remains imbalanced. For instance, we note that OpenAI recently announced a partnership with OCI to contract compute for inferencing due to Azure capacity constraints. Complementing the current supply strain, we highlight the notable cost of GPUs with ASP estimates (per chip) of \$15,000 (A100), \$25,000 (H100), and \$35,000 (B100 / 200). With tens of thousands of chips required to train newer LLMs, it can quickly prove to be a cost prohibitive endeavor. While difficult to time, we expect that supply-demand dynamics will normalize in the coming years. This is supported by NVIDIA's upcoming Blackwell release (early 2025), albeit prices could remain elevated against a robust near-term demand environment (somewhat evidenced by the estimated \$35,000 price tag on the Blackwell chip per GSe). We also think that hyperscalers' investments in custom chips to diversify exposure beyond NVIDIA (e.g., TPUs, ASICs) supports a normalization of supply-demand dynamics, too. As this plays out, we think two key trends will materialize: **1)** Consistency of increased supply will foster an eventual equilibrium in the demand environment, supporting the emergence of broader use cases. The evolution and optimization of these use cases can justify leveraging various hardware options that may be better suited for their specific needs (versus maintaining an increased reliance on GPUs). We've seen this dynamic somewhat materialize thus far with Groq's LPUs, which we think affirms the validity of our expectations for a diversification away from GPUs, and **2)** Investors' focus will shift even more so toward the actual ROI of surging CapEx, especially considering the scope of backlog increases in the LTM that have thus far kept the positive long-term AI narrative intact despite an air gap in spending further up the stack.

Implications of the new AI computing paradigm on data center operators (per Jim Schneider).

For data center operators, there are critical differences in the nature of conventional CPU-based computing relative to AI computing workloads, which are tied to GPUs. Importantly, AI- and other HPC-related workloads which rely on GPUs operate

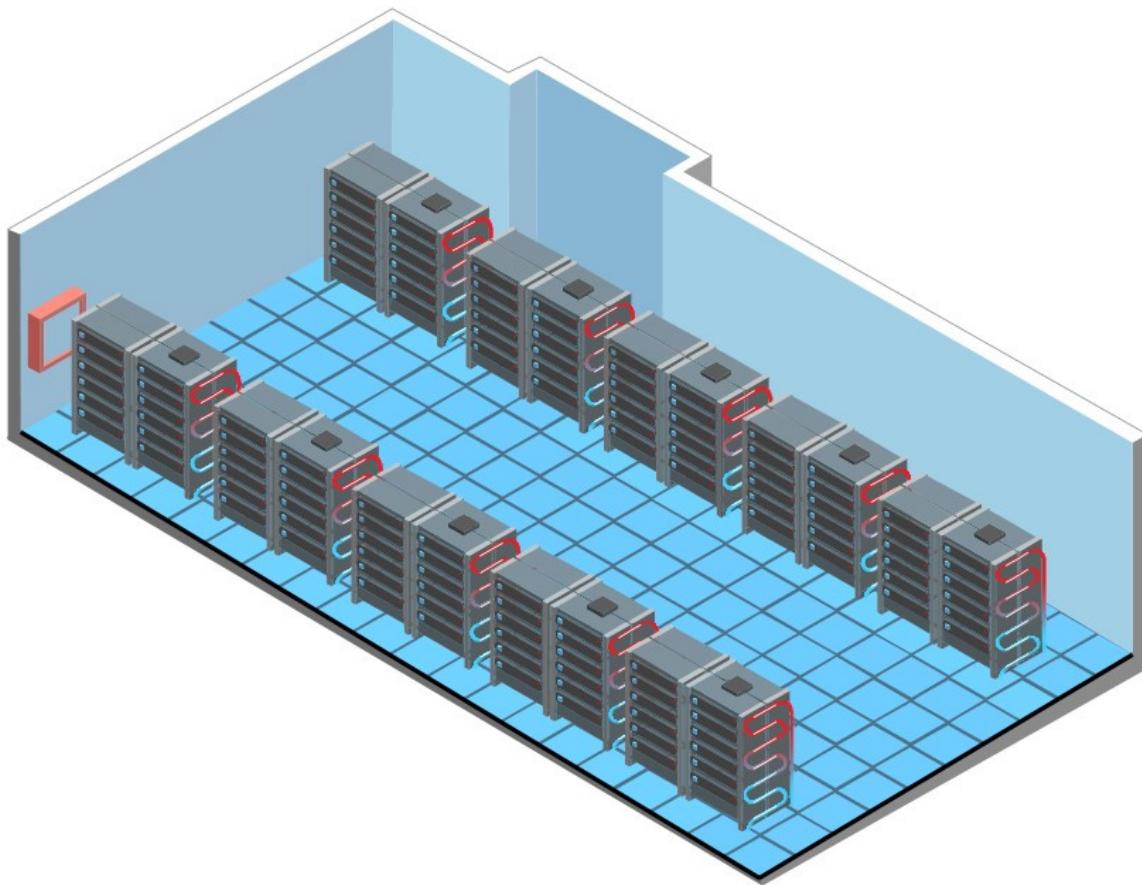
in configurations which requires much higher power densities relative to conventional workloads. General-purpose workloads could typically accommodate up to 42 conventional servers within a rack, consuming 4-10 kW of power per installed rack on average. However, for higher-performance GPU workloads targeting AI, the power density requirements can be significantly higher. Although a standard rack can only accommodate ~14 high-performance GPU servers for AI, the significantly higher power requirements translate to materially higher power of 30-50 kW per rack (inference) and 50-80 kW per rack (training). We believe this difference in power density is one of the key drivers to data center demand going forward, particularly in terms of power consumption.

This trend also has significant implications for data center layout and design. Most conventional data centers were designed and constructed for a conventional computing paradigm with a greater number of lower density racks and a tighter spacing between rows of racks. Data centers are typically designed for a certain number of computing racks in a certain areal density - and utilities which provide power to such facilities provision them for power delivery at a certain budgetary power level. It's also important to point out that most conventional data centers are cooled with standard forced-air HVAC systems - where the flow of ambient air is sufficient to relieve the heat emanating from the servers and other electronic systems ([Exhibit 68](#)).

Exhibit 68: Illustration of conventional datacenter footprint with densely-packed standard racks and air cooling

Source: Goldman Sachs Global Investment Research

In the emerging class of data centers being built to suit the high power densities required for Generative AI, a few important modifications are being made ([Exhibit 69](#)). First, data centers are being provisioned for significantly greater total power envelopes in order to accommodate the much higher power requirements of GPU-based computing systems. Second, rows of GPU-based server racks are often moved further apart in order to accommodate the increased amount of heat being generated from these cabinets. Third, in many cases forced air cooling is insufficient to relieve the heat generated by GPU-based servers - and a liquid cooling system must be installed where a chilled water loop brings cold water directly in contact with the GPUs or CPUs in each server in order to drain heat effectively. We highlight this as a potential risk for legacy data centers - especially those more than 5-10 years old, which could potentially be faced with significant retrofit costs in order to accommodate AI workloads. Although it is too early to know with certainty what fraction, if any, of legacy data centers might need to be retrofitted in this way, we highlight it as a potential obsolescence risk for legacy data center assets.

Exhibit 69: Illustration of Next-Gen-AI datacenter design with high power density, GPU-based liquid cooling

Source: Goldman Sachs Global Investment Research

- **Large Hyperscaler spending driving data center demand.** The AI imperative across hyperscalers is driving a rationalization of priorities within overall IT budgets as most corporate spenders must operate within a relatively narrow envelope of operating budgets. We believe there is a great disparity between customer types across different industries - with small-to-midsize data center customers likely still seeing some level of spending rationalization relative to larger enterprises. For third-party data center operators such as Equinix and Digital Realty, this is driving demand dispersion - with overall spending levels remaining healthy, but significantly more rationalization on the part of smaller retail data center customers (where Equinix is exposed) relative to the wholesale market which is dominated larger customers including hyperscalers (where Digital Realty is exposed) who are engaging in greater use of outsourced data centers relative to past spending cycles ([Exhibit 58](#)).

Data Center infrastructure investments to support Gen-AI workloads should be the primary driver of growth in the data center equipment market (per Mike Ng).

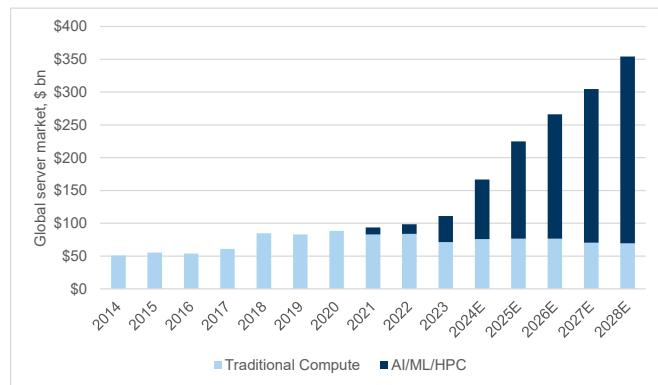
In particular, we expect **1)** AI Server market to grow from ~\$40bn in 2023 to ~\$285bn in 2028 at a 48% 5-year CAGR, and **2)** Total AI Data Center switching spend to grow from

\$3bn in 2023 to ~\$18bn in 2028 at a 43% 5-year CAGR.

First, companies are investing in AI infrastructure for training and inferencing. We forecast the AI server market to grow a 48% 5-year CAGR from \$40bn in 2023 to \$285bn in 2028 driven by data center buildouts for Gen-AI across hyperscalers, AI cloud service providers, tier 2 cloud, and enterprises. Since 2021, the AI server market has expanded ~375% (2021: \$10.6bn, 2023: \$39.9bn, according to 650 Group), while traditional server spend has declined 14% (2021: \$82.9bn, 2023: \$71.3bn, according to 650 Group). Currently, AI server spend is dominated by hyperscalers (72% share in 2023) and while we expect them to remain the dominant buyers of AI Servers through 2028, Enterprise and Tier 2 Cloud & SP customers should grow their share of AI server spend as companies discover suitable use cases. Currently the AI server market is led by whitebox players, followed by OEMs including SMCI, DELL, and HPE.

Exhibit 70: Growth in the server market should be driven by investment in AI-Servers (+48% 2023-2028 CAGR), while traditional compute server spend should be flat-to-down (-1% CAGR over the same period)

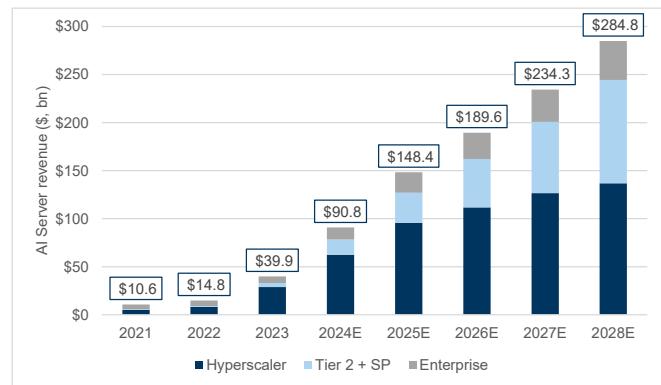
Global server market (\$, bn)



Source: 650 Group

Exhibit 71: AI server market spend should continue to be dominated by Hyperscalers, though Tier 2 + SP & Enterprise customers should gain spending share over time.

AI Server market revenue by customer vertical (\$, bn)



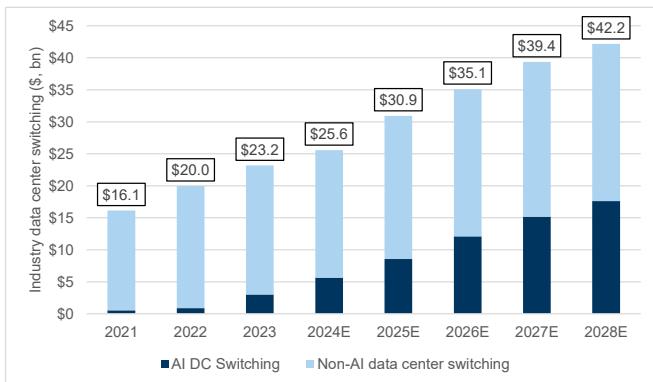
Source: 650 Group

Second, Gen-AI buildouts should benefit data center networking equipment spend.

Specifically, we expect the data center switching market to grow at a 13% 5-year CAGR from ~\$23 bn in 2023 to ~\$42 bn in 2028. As companies deploy AI workloads, they will invest in networking capacity within the data center to minimize latency, which should minimize job completion time and improve operational efficiency when building and running AI models. In doing this, companies will need to invest in improving and expanding data center infrastructure in order to support AI model training and inferencing. In particular, we expect AI data center switching spend to grow at a 43% 5-year 2023-28 CAGR to \$17.6 bn, while the non-AI data center switching market remains relatively muted growing at a 4% 23-28 CAGR. AI data center switching spend should largely be driven by Hyperscalers (CAGR: +36% C23-C28E), though Enterprise and Tier 2 Cloud and SP should be the fastest growing sources of spend (CAGR: +51% each, C23-C28E).

Exhibit 72: Growth in the data center switching market should be led by AI-data center switching spend (+43% 23-28 CAGR), while non-AI DC switching growth remains more muted (+4% 23-28 CAGR)

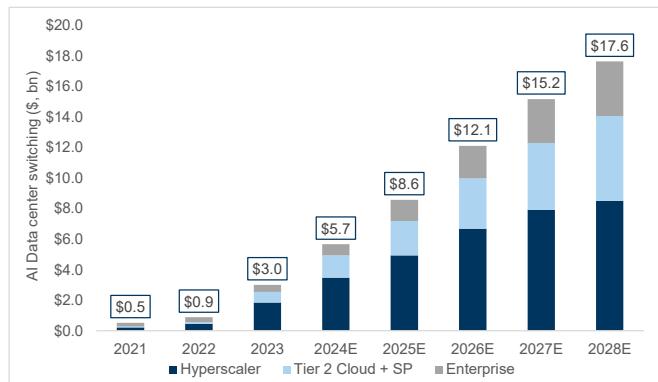
Ai vs. Non-AI DC switching market revenue (\$, bn)



Source: 650 Group

Exhibit 73: AI Data center switching spend should continue to be dominated by Hyperscalers, though Enterprise and Tier 2 Cloud & SP customers should gain share of spend over time.

AI data center switching market revenue (\$, bn)



Source: 650 Group

Fleshing out the foundation model layer

Fleshing out the foundational model layer. It is clear legacy players are determined to embrace this emerging technology and to not be outmaneuvered by upstarts. Beyond piling significant amounts of capital into internally developed AI projects, nearly all of the world's largest software companies are investing significant amounts of capital into partnerships and investments with AI-native startups. The most obvious example would be Microsoft's partnership with OpenAI, but other examples include Salesforce raising a \$500mn fund to invest in AI startups, Google establishing a similar fund and has done a large portion of its AI innovation through its acquisition of DeepMind, and Amazon and Google announcing investments of \$4bn and \$2bn in Anthropic respectively. Given the high hurdle rates to enter this ecosystem (discussed above) we largely see the infrastructure and foundational layers as crowded, with the key winners already in market.

Foundation Models illuminated the possibilities of Gen-AI; role to evolve as investments broaden beyond infrastructure layer. OpenAI's launch of ChatGPT in Nov. 2022 kicked off a wave of intrigue and discussion on the potential impact that Gen-AI could have on all sectors of the economy. Since then, we've seen a number of open and closed-source model providers emerge that focus either solely on image, video, text, or code-generation or a combination of several (e.g. multi-modal), with each iteration of these respective models leapfrogging the prior in terms of performance as GPU-powered compute and training datasets continue to scale. While not exhaustive, the primary FM providers today include Anthropic (Claude), OpenAI (ChatGPT, Sora, Dalle-3), Meta (Llama), Mistral, Midjourney, Stable Diffusion, Google (Gemini) and Databricks (DBRX). While we believe that the continued innovation spearheaded by these model providers will prove key to technology advancement with longer-term ambitions to unlock general artificial intelligence (GAI), we, like most, expect some level of commoditization across the model layer as incremental performance improvements prove to be less of a differentiating factor with most of the top models (Llama-70B,

DBRX, GPT-4) all reaching high levels of proficiency in core capabilities such as language understanding. As such, we expect the value-proposition of these base models to evolve beyond basic consumer services to powerful context engines underpinning AI-native applications.

- **Addressing the core competency of Foundation Models.** To understand how we expect the role of FMs to evolve, it's important to address their core competencies and the innovations we have observed to date. At a basic level, an FM (e.g., ChatGPT) is a deep learning algorithm that can perform a variety of natural language processing (NLP) tasks. These models are trained on trillions of words, images, and videos (depending on whether multi-modal or language-specific) across the public domain through a process called unsupervised learning. The models are then fine-tuned to be more performant in specific tasks, whether translation or image-generation. Today, most of the adoption and monetization tied to the FM providers has been on the consumer side, but we believe the real opportunity resides in Enterprise.
- **Expect commoditization at the model layer akin to renting compute from Hyperscalers.** FM providers have been investing heavily to increase the performance of their respective models, including trained knowledge, latency, throughput, and security, all of which we expect to be critical to enterprises' inevitable adoption of the technology. This can be validated by: **1)** The steady ramp in AI revenue across the Hyperscalers, predominantly tied to FM providers' training and inferencing workloads and **2)** The growth in model parameters from providers such as OpenAI, growing 2,000x from GPT-1 (100mn+) to GPT-4 (~2tn). As we observed during the cloud computing cycle, it became clear that the enabling factors to the Hyperscalers' growth over last decade centered on their ability to invest billions in data centers, servers, networking equipment and CPUs to provide the capacity pool for a growing cloud ecosystem (SaaS, PaaS). However, as the initial build-out subsided, we note that core computing became a commodity and the Hyperscalers differentiated by overlaying high-margin PaaS services (think data management services). Similarly, we anticipate that model providers are unlikely to compete solely on model performance, instead model specialization, AI enablement services, and security are likely to become key aspects of the value proposition.

Foundation Models to act as the context engines underpinning AI applications. As we highlight in the platform section of this report, we expect that highly performant base models will play an important role as an enabling technology for AI applications. At first, we expect Enterprises to tap into the base models themselves for simpler use cases such as customer support automation, code generation and data analysis. As the ecosystem surrounding FMs grows (PaaS), either through the providers themselves or external players, we expect that the role of FMs will center more so on providing the context engine. An example would be working in tandem with emerging technologies such as retrieval infrastructure (e.g. Retrieval Augmented Generation – RAG) to enhance the query output with proprietary and real-time information. The natural question is if and how FM providers can differentiate as commoditization around performance becomes more prevalent. We see multiple avenues, including: **1)** Developing

vertical-specific models trained on narrower datasets that can be sold into various industries as out-of-the-box offerings or used as a base model for further customization (fine-tuning or RAG), **2)** Providing additional tooling to customize the base model, either through partnerships or as a core service; OpenAI's Fine-Tuning service is a prescient example, **3)** Expanding the model library to include smaller language models that can dramatically reduce customers' inferencing costs and make AI application development more accessible to a broader customer demographic; ChatGPT-4o Mini, for instance, is 60% cheaper than ChatGPT-3.5 Turbo, **4)** Establishing robust compliance and governance frameworks, including tools to manage and audit AI systems, **5)** Building out a robust customer support organization and partner ecosystem, and **6)** Providing cost-effective deployment options (e.g. compute optimized for specific use-cases). As we can see, the FM providers will play a pivotal role as context engines, while also overlaying value-add services and solutions that will ultimately be crucial to driving broader AI adoption.

Training has proved instrumental to-date, mix-shift to eventually favor inferencing.

While focus and monetization has largely centered on model training (given LLMs' topicality), we believe that inferencing and fine-tuning will see durable long-term growth supported by continued Gen-AI usage and adoption. Typically, training is a more compute-intensive exercise as compared to inferencing. All else equal, we would expect Gen-AI compute costs to move lower directionally in the medium-to-long term due to **1)** Inference becoming a bigger sub-segment of Gen-AI in the medium-to-long term, in our opinion, **2)** continued innovation from the entire AI ecosystem across the entire Hardware and Software stack, and **3)** a natural progression towards better utilization of resources across the industry as we move up the learning curve of this new technology. However, we do note that we are still in early-days of proliferation of this new technology, and to that extent, it is difficult to quantify the step-down in the compute costs for inferencing. Key factors that could dictate the pace and shape of this step-down in the marginal cost of computing as the industry transitions from training to Inference would be **1)** the mix in the size of models that are being trained and would ultimately be inferenced (LLMs vs SLMs), **2)** pace of innovation and gen-to-gen performance uplifts going ahead as the industry continues to push the envelope on physics to pack as much performance as possible in smaller and smaller form factors to drive better power efficiency, and **3)** mix between GPUs and CPUs that would be used for inferencing depending on the workload.

- **Training** involves introducing data sets of various types (e.g., structured / unstructured, labeled / unlabeled) to AI models. Doing so helps refine the algorithms, with an eventual goal of generating highly accurate responses. Additionally, variations exist in the amount of training data utilized depending on if a model is an SLM or LLM. While no defined breakpoint exists to delineate "small" versus "large", those known to fall into the "large camp" are those which pull data across large datasets spanning the internet whereas smaller models tend to index toward curated and contained sources. In training LLMs, datasets are usually broad-based, like books, articles, web pages, images, and videos. Conversely, SLMs are likely to utilize more domain specific data sources. For instance, a healthcare focused SLM might be trained on various volumes and editions of medical journals,

resulting in domain expertise. Complementing the data volume intensity are the hardware requirements – specifically GPUs. Effective LLM training typically requires tens of thousands of GPUs (e.g., GPT-4: 25,000 A100s, Llama Family: >48,000 H100s), with model training times usually taking weeks to months. Accordingly, the crowding in the LLM arena, coupled with the capital intensity around producing larger models results in a natural competitive moat for the large hyperscalers and existing private model providers.

- **Inferencing** follows the training step and is focused on delivering outputs to entered prompts. For example, when a user enters a prompt, like “Write me a poem” or “Create a silly image”, the model takes that input (e.g., inference data) and delivers an output, with the end-result quality a direct result of the training data utilized. As Gen-AI usage and adoption continues growing, we believe that the long-term revenue mix-shift likely favors inferencing. While having limited historical context, we do highlight that NVIDIA estimated on their F1Q25 / F4Q24 earnings calls that AI Inferencing was responsible for ~40% of their TTM Data Center revenue.
- **Fine-Tuning** represents an optional step that involves taking a pre-trained model and introducing it to a more targeted data set. Doing so allows it to be used for specialized use cases like Customer Service, Healthcare, or Research. As companies look for ways to efficiently incorporate Gen-AI into throughout various workflows, fine tuning an existing LLM could prove attractive when considering the lower cost, relative compute efficiency (versus starting from scratch), and increased accuracy versus broader LLMs (high performance metrics for specialized use cases).

Various drivers can support a decrease in the intensity of Gen-AI compute costs.

Although hard to quantify and predict the exact timing, we expect a number of drivers to help reduce compute cost intensity associated with Gen-AI: **1)** Rising competition among LLMs to produce better models that extract more efficiencies from hardware, **2)** Emergence of small and medium language models where the scaling of compute costs can prove more muted relative to that seen with their larger model counterparts, **3)** Vertical domain-specific models, **4)** Price performance improvements expected in GPUs, **5)** New sources of competition in hardware (example being Groq's LPPUs), **6)** Shift in AI from learning and training to inferencing, the latter being several orders of magnitude less compute-intensive, and finally, **7)** Shift in the Gen-AI activity from infrastructure (training) to platforms and onto applications (inference). Something else could also come along that provides relief in addition to or substituting the above drivers.

Addressing the evolution of the hardware stack

Future hardware usage likely diverges from GPU dependency as application

development increases. While GPUs have proven fundamental to Gen-AI development to-date (via LLM training and inferencing), specific Gen-AI application requirements could result in a more diversified future hardware environment. Being in the nascent innings of this opportunity, the focus has understandably been more generalized, with much of the progress proving to be LLM-centric. This can be seen in the parameter count evolution of OpenAI’s GPT models. While exact figures remain unconfirmed, estimates pin GPT-1 (2018) at ~117mn parameters, which is dwarfed by the ~2tn

parameters of GPT-4 (2023). GPT-4 isn't alone, as other models like Olympus (Amazon) are rumored to have a similar number of parameters, too. Efficiently training these models requires vast amounts of efficient compute, which has been underpinned by GPUs thus far given their strong concurrent computing abilities. This concurrency allows models to be trained on an accelerated timeline versus their CPU counterparts.

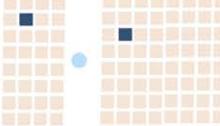
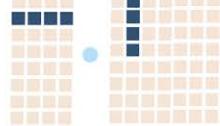
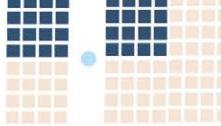
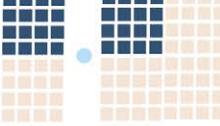
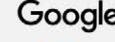
However, a broader ecosystem of more specialized Gen-AI applications (and subsequent adoption) could catalyze a shift away from the notable indexing to GPUs seen to-date.

Within the context of a broader ecosystem of Gen-AI applications, various use-case specific options are likely to emerge, such as applications that focus on CRM, HCM, and ERP. This dynamic can be somewhat seen with Adobe Firefly – a Gen-AI based application focused on content creation and editing (that has witnessed rapidly growing usage since its March 2023 release). As more niche applications are developed, there is likely to be a focus on ensuring the optimization of the underlying infrastructure – both software and hardware. From a software perspective, this could imply that the models underpinning these applications can be simpler and more specialized, potentially mirroring the more simplified yet domain-focused characteristics of SLMs. A simpler model could then result in utilizing more optimized hardware, like CPUs, TPUs, NPUs, or ASICs (application-specific integrated circuit) chips for example (Exhibit 74) – limiting the future reliance on GPUs (versus current levels).

We believe this dynamic has somewhat transpired to-date, evidence by Microsoft and OpenAI working to construct their own hardware options. Microsoft recently developed their own CPU – Azure Cobalt CPU, with a focus on optimizing performance, power, and cost. Initial testing on various workloads (e.g. Teams, SQL server) has shown promise, delivering 40% better performance versus the hardware they were currently utilizing in their data centers. This could set a strong foundation for future custom hardware development. For example, while Copilot for M365 could broadly rely on GPUs for general performance, specific tasks or operations with Copilot could prove more optimized by utilizing CPUs. In a similar vein, OpenAI has discussed developing custom AI chips with Broadcom. Doing so would result in a reduced dependency on NVIDIA GPUs (e.g. GPT-4 required 25,000 A100s to train), and allow OpenAI to better optimize the hardware for their current, and expected future, workloads / applications.

Highlighting key characteristics between various hardware options

Exhibit 74: Comparing the different processing units

	 CPU Central Processing Unit	 GPU Graphical Processing Unit	 TPU Tensor Processing Unit	 NPU Neural Processing Unit
Purpose, Specialization	Small dataset and single thread processing	Medium to large dataset processing, parallel processing capacities	Optimized for matrix calculations	Neural network processing on edge devices
Use-Cases	Diverse computing tasks	Image/video processing, data-center based AI model training	AI model training and inferencing	Image recognition, real-time AI inferencing
Architecture	Few strong cores, general purpose arithmetic logical units (ALUs)	Thousands of small cores, single instruction multiple data (SIMD) architecture	Multiple matrix multiply units (MXUs), custom application specific integrated circuits (ASICs) designed for specific operations	Custom low power (8,16-bit) cores, often integrated with system on chips (SoCs)
Scalability	Scalable within power constraints	Highly scalable to produce AI processing power across multiple GPUs in datacenters	Scalable across cloud environments	Scalable within the limits of edge devices (Phones, PCs)
Compute Base	 Scalar	 Vectors	 Tensors/Matrix	 Tensors/Matrix
Manufacturers	  NVIDIA    	 NVIDIA   	 HAILO 	  ARM   

For the exclusive use of CHRISTINA.LIU@GS.COM

Source: Goldman Sachs Global Investment Research, Analytics Vidhya

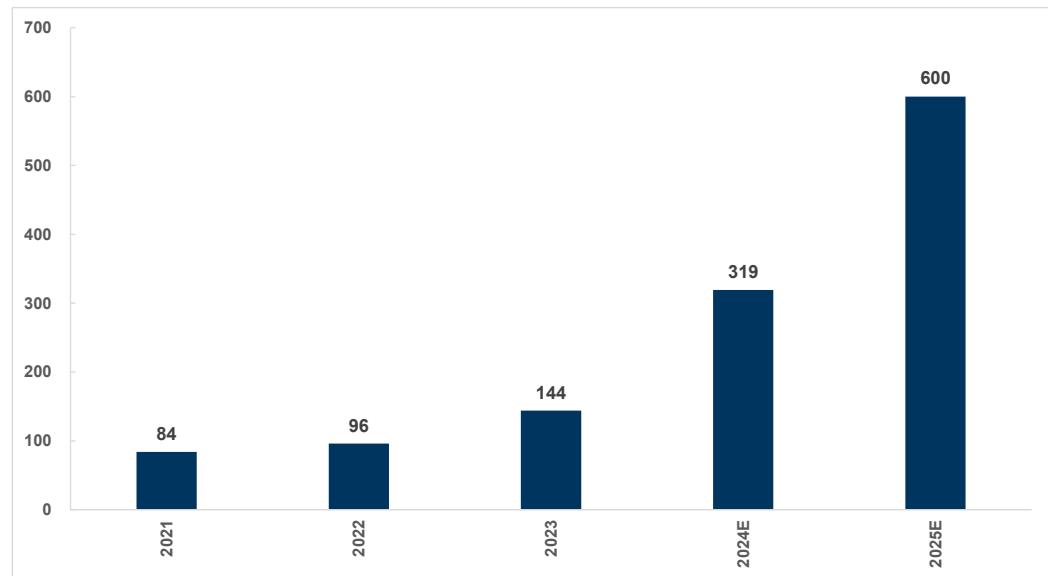
Hurdles to adoption

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

AI applications use two types of dynamic random-access memory (DRAM): HBM and DDR SDRAM. HBM is a revolutionary memory technology that stacks multiple DRAM dies—small blocks of semiconducting material on which integrated circuits are fabricated—on top of a base logic die, thereby enabling higher levels of performance through more bandwidth when interfacing with a GPU or AI chips more broadly. We expect the HBM market to grow at a ~100% compound annual growth rate (CAGR) over the next few years, from \$2.3bn in 2023 to \$30.2bn in 2026, as the three incumbent suppliers of DRAM (Samsung, SK Hynix, and Micron) allocate an increasing proportion of their total bit supply to meet the significant demand growth. Despite this ramp-up, HBM demand will likely outstrip supply over this period owing to growing HBM content requirements and major suppliers' supply discipline. We therefore forecast HBM undersupply of 3%/2%/1% in C24/C25/C26. Indeed as NVIDIA and AMD recently indicated, updated data center GPU product roadmaps suggest that the amount of HBM required per chip will grow on a sustained basis. And lower manufacturing yield rates in HBM than in traditional DRAM given the increased complexity of the stacking process constrains suppliers' ability to increase capacity.

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

Exhibit 75: TSMC's CoWoS capacity should grow >2x in 2024 and ~2x again in 2025, easing the packaging bottleneck
 TSMC's annual CoWoS capacity (k wafer per year)



Source: Goldman Sachs Global Investment Research

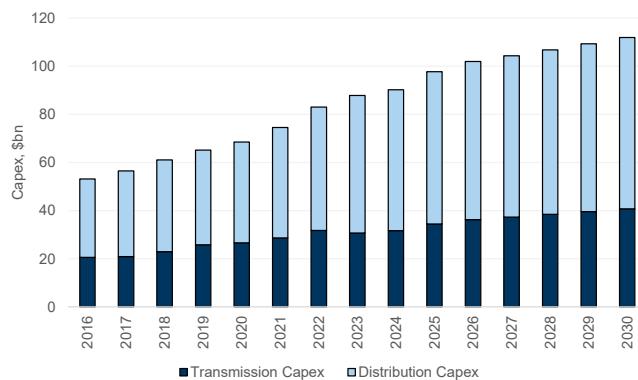
Energy infrastructure likely to require notable upgrade against increasing Gen-AI based demand (*covered by Carly Davenport*). The growth of data centers presents a unique challenge for the US power grid. Our current estimates show power demand in the US growing at a 2.4% CAGR through 2030, which is a significant change from the prior decade where power demand was flat due to energy efficiency efforts. Currently data centers make up just 3% of total power demand in the US, and we expect they will grow to be 8% of total demand by 2030. Data centers are increasingly power intensive and those building data centers, particularly hyperscalers, have demonstrated a desire to meet this increasing power demand with clean sources of power. This creates several challenges for local utilities given that building out transmission lines and generation assets takes time, and renewable generation is intermittent by nature once built. On the generation side, we believe that the US will need to build out 47 GWs of generation capacity to support this data center demand, which would be a significant undertaking for the utilities and renewable developers. Some larger renewable developers have stated that they believe that renewables and battery storage can economically meet roughly 80% of a data center's demand on a given day, with the rest likely coming from the grid. On transmission, we continue to view this as a key bottleneck given the time consuming permitting processes to build new lines and the lack of investment in transmission in recent years.

The data center theme will likely play out as a regional story, and several regions in the US have limited excess generation/transmission capacity to meet rising demand. We believe MISO in particular will face challenges as they have limited reserve margin, and that region will likely require a significant amount of CapEx to meet incremental demand. Utilities also have to navigate affordability concerns when requesting approval of CapEx to support data centers, as utility commissions will likely be hesitant to raise

residential customer bills for CapEx that supports a data center. We believe there is a path forward where utilities are able to manage all of these challenges without having to deny data center customer requests, but the companies will have to act in a proactive and prudent manner to achieve this.

Exhibit 76: We continue to view transmission as a bottleneck to accommodating load growth, and expect to see continued growth, with upside risk, in investment in this area. Overall we see \$720 bn of grid spend required through 2030, based on the CapEx growth outlook for our coverage

Grid CapEx, \$ bn



Source: EEI, Goldman Sachs Global Investment Research

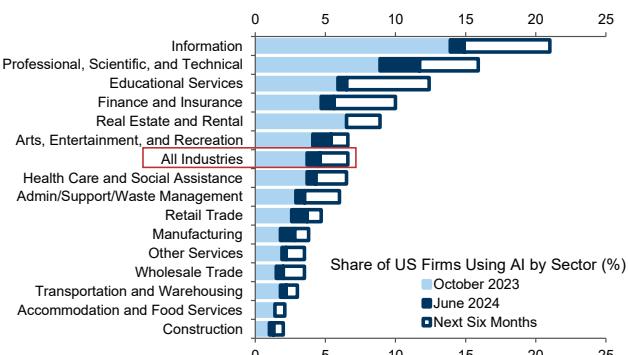
Formation of ecosystem to support broadening out of use cases

As the industry builds out the foundation of each layer in the IPA framework discussed above, it is likely to cultivate the groundwork for a wider ecosystem of Gen-AI applications. We look at different sectors from both a macro and micro lens to outline different viable use cases beyond the initial productivity boost evidenced in developer tools and digital content creation.

Macro Views

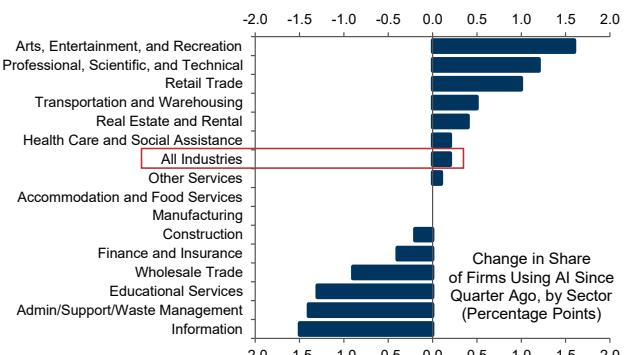
According to GS macro's quarterly AI adoption tracker, AI adoption by firms remains modest, with only 5% of US firms reporting use of AI to produce goods or services as of 2Q24 (Exhibit 77). While this share increased only marginally (+0.2pp) since 1Q24, arts and entertainment, professional, and retail trade firms reported higher adoption rates of at least 1pp while information, administrative, and educational firms reported lower adoption rates of at least 1pp (Exhibit 78). Cloud-related firms reported the highest expected increase in AI adoption over the next six months (Exhibit 79). Industry surveys released over the past quarter suggest that while a large share of small businesses are experimenting with AI, most have yet to define a clear business use case or employee training programs, and many still have concerns over data compatibility and privacy.

Exhibit 77: Approximately 5% of US Firms Are Using Generative AI to Produce Goods and Services, Although the Share Is Much Higher Among Information Service Companies



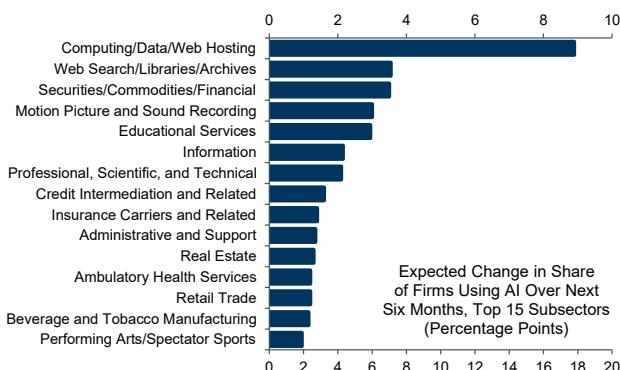
Source: Goldman Sachs Global Investment Research, Census Bureau

Exhibit 78: The Share of Firms Using AI for Production Ticked Up Just 0.2pp Over the Past Quarter, With Increases of at Least 1pp Among Arts and Entertainment, Professional, and Retail Trade Firms and Declines of at Least 1pp Among Information, Administrative, and Educational Firms



Source: Goldman Sachs Global Investment Research, Census Bureau

Exhibit 79: Across Subsectors, Cloud-Related Firms Most Expect to Adopt AI Within the Next Six Months



Source: Goldman Sachs Global Investment Research, Census Bureau

Sub-Sector Views

In our ['Gen-AI - Part I: Laying Out the Investment Framework'](#) note we outlined how the proliferation of this technology can enhance the workflows of a multitude of workforce personas. We update our thoughts with a more expansive list within TMT and in adjacent sectors. The evolution of the industry over the last year has given way to more concrete and profound use cases that can be enabled by the formation and proliferation of Gen-AI.

Hardware - Covered by Michael Ng

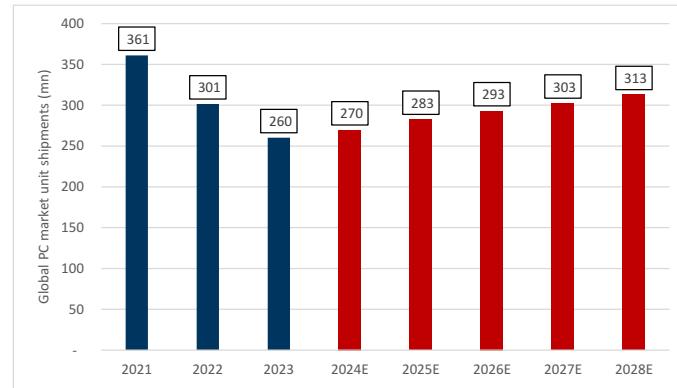
Apple Intelligence's Generative AI features to drive an uplift in iPhone demand. At [WWDC in June 2024](#), Apple announced Apple Intelligence, a personal intelligence system which includes features including **1)** improved Siri capabilities (deeper language understanding, text communication with Siri, tailored responses driven by user activity and information, etc), **2)** language features including writing tools that rewrite and

summarize text across apps including Mail, Notes, and Pages, and **3)** image features including Image Playground (image generation), new Genmojis, improved photo editing features, and more advanced search capabilities within a user's photo library. Apple Intelligence will only be available for the iPhone 15 Pro, iPhone 15 Pro Max, as well as future later models. Apple Intelligence should be released in the fall of 2024 and we believe these features should **1)** should drive product upgrades as customers refresh older iPhones to access AI capabilities; **2)** continue to drive a mix shift towards premium models, which should drive continued uplift in ASP; and **3)** could present an opportunity for an iPhone price increase. Accordingly, we forecast F2024/25/26 iPhone sell-in units of 232/241/257mn (+2/+4/+7% yoy).

AI PCs could drive additional upside to PC market recovery. PC industry unit shipments should return to growth in 2024 (GSe: 270mn units, +4% yoy growth) after two consecutive years of decline. Growth in PC shipments should be driven by **1)** the aging installed base of existing PCs (after outsized unit shipments during the work-from-home-era during the pandemic in 2021, ~354mn); **2)** Windows 10 refresh; and **3)** adoption of AI PCs. In particular, Canalys estimated that AI PCs made up 14% of unit shipments in 2Q24 and expects this share to continue to grow on a forward basis. Growing penetration of AI PCs should be driven by superior device capabilities including better power efficiency and the ability to run Generative AI workloads natively, and companies such as HPQ have expressed expectations for AI PCs to drive a 5-10% industry ASP uplift over time.

Exhibit 80: Global PC market unit shipments should return to growth in 2024 after two consecutive years of decline.

Global PC market unit shipments (mn)



Source: Data compiled by Goldman Sachs Global Investment Research, Company Data, HPQ, Dell, Apple, Asus, Lenovo, Taiwan ODM

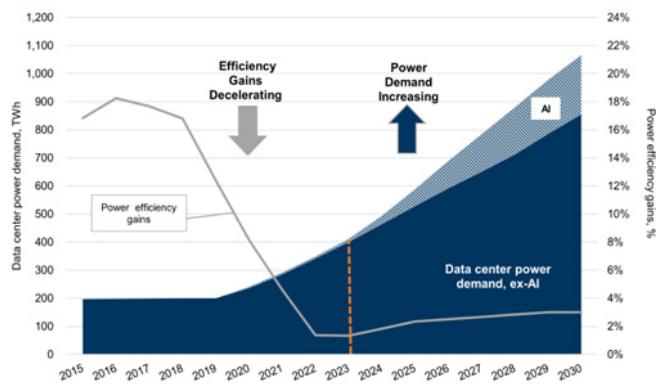
Data centers - Covered by Jim Schneider

The Long View - AI is driving an inflection in power demand, driving significant capacity growth

The combination of AI-driven demand and a material slowdown in power efficiency gains is making datacenters a critical driver of accelerating global and US electricity demand growth. In collaboration with our colleagues on the Goldman Sachs SUSTAIN, energy, utilities, and technology teams, we have developed a proprietary model of power

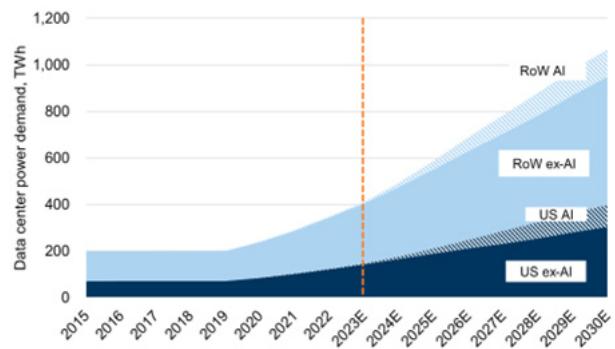
consumption based on our view of multiple supply and demand factors, including AI and conventional server shipments and efficiency gains over time - see AI/data centers' global power surge and the Sustainability impact. We model global datacenter power demand (ex crypto) will grow by 160% in 2030 vs. 2023 levels (bear/bull range of 80%-240% growth), representing an increase of about 650 TWh by 2030 in our base case (330/1,000 TWh in our bear/bull cases). Our US Utilities Research team estimates data centers alone will contribute a 0.9% CAGR to overall US power demand, bringing the total expected CAGR to 2.4% through 2030. We see data centers adding a 0.3% CAGR to overall global power demand. Our base case implies data center power demand moves from 1%-2% of overall global power demand to 3%-4% by 2030. The increase in the US is even greater — from 3% to 8%. Our estimates for overall data center power demand are above IEA forecasts (2026), and our outlook for AI to represent about 19% of data center power demand in 2028 is above recent corporate forecasts. On a medium-term basis, our model suggests that the growth is set to accelerate from 10%-11% growth in 2020-21 to 11%-12% in 2024-25 and further to 17%-18% by 2029-2030. Our view is that the amount of global datacenter demand dedicated to AI will move from 3%-4% in 2023-2024 to 17%-18% in 2029-2030.

Exhibit 81: As efficiency gains have decelerated, data demand growth is driving a surge in data center power use, with an AI kicker on the way



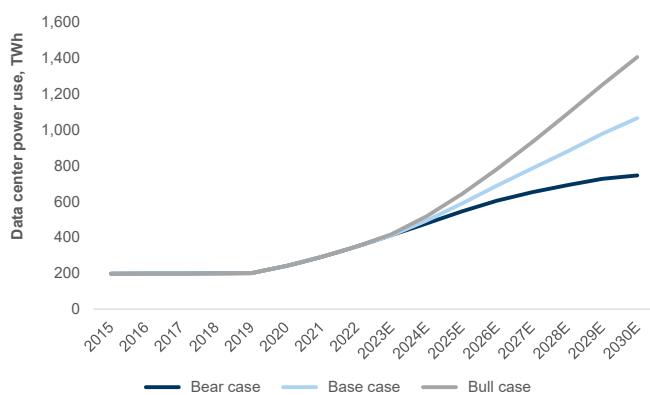
Source: Goldman Sachs Global Investment Research, Masanet et. al (2020), Cisco, IEA

Exhibit 82: After being flat for 2015-19, we have seen data center power demand accelerate in 2021-23 and expect a 160% increase through the rest of the decade



Source: Goldman Sachs Global Investment Research, Masanet et. al (2020), Cisco, IEA

Exhibit 83: We see 2030 power use from datacenters 1.8x-3.4x 2023 levels in our bear/bull case



Source: Goldman Sachs Global Investment Research, Masanet et. al (2020), Cisco, IEA

IT Services - Covered By Jim Schneider

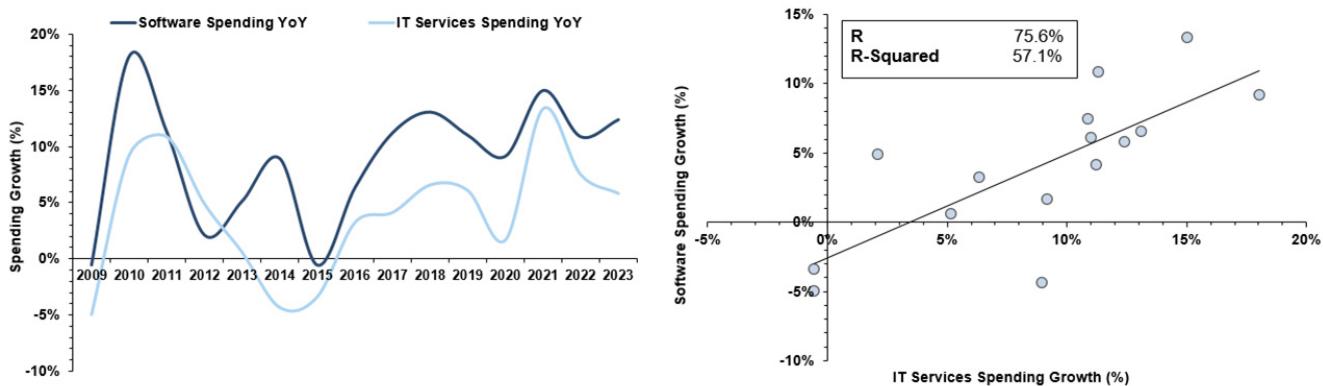
Is Gen-AI a Secular Threat or Net Beneficiary? - Considering the fundamental tailwinds and headwinds of AI.

At the heart of the debate around which sectors Gen-AI poses a threat or tailwind to is the IT Services sector, which comprises companies whose fundamental business model is based on the organization of labor into specific projects designed to deliver a new technology or capability to an organization. For corporations, the application of Generative AI comes down to how a company is able to gain insight from multiple proprietary data sources across its entire IT estate, which comprises multiple disparate applications, databases, and systems.

Generative AI - Net positive opportunity as more software translates to more services.

Ultimately, we believe that the long-term trend of more software spending as a percentage of overall corporate IT budgets will “pull along” additional IT Services work because of the high intensity of integration work tied to new applications and data sources. Simply put, more software applications require more integration work when installed in an enterprise environment. This has been proven out over the past 10-15 years, as Software and Services spending has been highly correlated with an R-squared of 0.57 ([Exhibit 84](#)).

Exhibit 84: Software spending tends to pull along IT services spending historically



Source: Gartner, Goldman Sachs Global Investment Research

But the risks are real - AI could cannibalize certain classes of IT Services. While IT Services companies can capitalize on many of the emerging transformational data trends tied to Generative AI, there are also portions of the industry that are under significant threat. Typically, types of work tied to the low-cost processing of records-outsourcing of business processes or repetitive coding tasks are well served by both traditional and Generative AI. We see the following segments of the industry as being under the greatest secular pressure from Generative AI:

- **Classic Business Process Services:** Some of the most traditional IT Services work is tied to the offshoring of manual and semi-automated corporate functions such as processing insurance claims, employee travel expenses, and the like. Over time, these services have increasingly moved offshore and have seen a greater portion of their content either partially or fully automated. We believe Generative AI may accelerate these declines - and while certain modern service providers could benefit, others are likely to see significant declines. Based on Gartner estimates, we

believe this category of services is about \$208bn in size (2023), and we believe this is likely to decline at a rate of 1% - 3%.

- **Legacy Applications Maintenance Services:** Many older applications, particularly those on legacy mainframe platforms (written in COBOL, for example) continue to operate, either in their original environment (such as mainframes) or after having been containerized and moved to the cloud. We believe that over time, there is an opportunity to reconstitute many of these workloads, based entirely on an English-language specification of the algorithm, which is translated into code by an LLM. Based on Gartner estimates, we estimate this category of services is \$101bn in size (2023), and we believe this segment is likely to decline at a rate of 1-3%.

Security Software - Covered by Gabriela Borges

We expect Gen-AI to provide meaningful tailwinds to Security Software both in the near term with product cycles to secure AI use cases and long term with an opportunity to automate the Security Operations Center. We highlight PANW, ZS, and CRWD from our coverage as the stocks where Gen-AI impact is most meaningful:

- **Palo Alto Networks** - The company offers an AI security suite to protect enterprises from employees using AI within firm systems including AI Access Security, AI SPM and AI Runtime. Furthermore, the company announced its Precision AI bundle combining current ML techniques with Gen-AI to predict evolutions for zero day attacks. In its most recent quarter, Palo Alto Networks disclosed \$200mn in ARR from its AI product integrations across XSIAM and newer products launched earlier this year.
- **Zscaler**: We see Zscaler's infrastructure as well positioned to ramp on its DLP functionality to protect against data leakage as developers look to leverage third-party AI tools. We believe the company is close to parity with best in class DLP peers by year-end as there is the potential to consolidate policy into a centralized DLP engine vs. the fragmented ecosystem across network, email, endpoint & cloud.
- **CrowdStrike** - We expect CrowdStrike to be a key player in using endpoint telemetry, security analytics and log aggregation to improve automation in the security operations center. Its acquisition of LogScale in early 2021 allows it to process large amounts of data real-time at low cost, with ~85% of this data being contextually-rich data collected from the Endpoint agent. Coupled with its ongoing investments in AI we believe CrowdStrike is well-positioned to address the single biggest pain point in Security today which is finding talent in the SOC.

Vertical Software - Covered by Gabriela Borges

Vertical software companies are typically slower to follow technology trends vs. horizontal providers. We expect that the same is proving to be true with more sophisticated AI/Gen-AI use cases, except note that these companies have been fairly quick to leverage Gen-AI models by adding industry specific data. Because of this data, we expect that Vertical Software companies will be able to come up with more sophisticated AI use cases over time – particularly companies that focus on verticals that are less adept at internal development (ex. government, insurance, construction).

However, given that we have yet to see killer applications by the faster based horizontal application companies, we view this as one of the last phases of the Gen-AI cycle (albeit with the potential to be one of the most impactful).

We highlight specific companies in this theme below:

- **CCC Intelligent Solutions (CCCS):** In the insurance end market, CCC has made significant investments in AI over the past decade: the company has ~\$1tn in historical claims data and collects ~500bn car photos a year. CCC is focused on incorporating AI into existing customer workflows in order to maximize the ROI. Key AI driven products include Estimate-STP and Subrogation. Estimate STP is an AI solution that enables low-touch claims processing, this product can turn photos into damage cost estimates in seconds. Subrogation uses an AI driven analysis to determine that maximum dollars have been recovered and has been shown to reduce arbitration by 25-50%.
- **Clearwater Analytics (CWAN):** In the insurance and asset management market, Clearwater has been investing in Gen-AI and has talked about having multiple clients signed up for new products driven by Gen-AI. Where Clearwater has seen more immediate benefits from Gen-AI is in internal efficiencies which we have seen in gross margin improvements.
- **Veeva (VEEV):** In the life sciences space, Veeva is adding an AI powered CRM Bot as part of their new Vault CRM. The company has also announced an AI Partner Program that includes API training & support and an Application Sandbox in order to help partners integrate Gen-AI solutions into Veeva Vault applications.
- **Tyler Technologies (TYL):** In the government software space, Tyler is doing multiple AI driven acquisitions. Most notably CSI (Computing System Innovations) which uses AI to provide automated redaction and indexing for courts.

Infrastructure Software – Covered by Gabriela Borges

As companies look to implement Gen-AI, we view Palantir as uniquely positioned to leverage its core data stitching competencies into enterprises that are looking to deploy AI use cases.

- **Palantir:** Palantir's Artificial Intelligence Platform, AIP, was launched in 2023 and enables customers across commercial and government to connect LLMs and other AI with their data and operations. The platform provides unified access to open-source, self-hosted, and commercial LLMs that can transform structured and unstructured data into LLM-understandable objects. The platform is model agnostic and can leverage all publicly available LLMs, and is designed to be bundled with existing Palantir offerings such as the Foundry, Gotham, and Apollo platforms. At Palantir's AIPCon, customers showcased how they have gone from prototype to production-grade AI via AIP bootcamps – one example is an insurance company that leveraged AIP in its underwriting application to provide guidance for complex policy requests by putting together the information at the same time, identifying all relevant underwriting standards set by the company, and searching conversations between the sales representative and the customer - based on this information, the

program determines the suggested action plan. We view Palantir as uniquely positioned to benefit from medium-term enterprise AI engagements given its ability to structure data, upgrade IT infrastructure and custom build AI applications - and this is translating into a much stronger demand backdrop for Palantir exhibited in its US Commercial revenue growth (up 55% yoy in 2Q24).

Emerging Application Software – Covered by Adam Hotchkiss

Generative AI-driven productivity improvements are front-and-center for smaller application software businesses, though uniform adoption and monetization is further down the curve. Across our emerging back-office, front-office, and vertical software coverage, nearly all of our companies have introduced some level of Generative AI functionality into their products or organizations. We believe the most low-hanging fruit opportunities exist in replacing highly manual, time-intensive processes often associated with the vast data sets and repetitive processes that software applications are designed to address; and ***most of the companies under our coverage today are framing these use cases for Generative AI as customer retention & acquisition opportunities rather than meaningful near-term monetization opportunities.*** We think the next leg of innovation, providing more value through original insight and content generation, is likely to come to fruition over time given the proprietary, often niche data that sits within these companies. That said, given the output accuracy required for these tools to be widely usable (particularly in vertical-specific use cases and back-office departments like Office of the CFO), we think that more resources and refinement of the technology, which requires investment and time, are needed for applications to be broadly viewed as value-add with incremental monetization potential. We distill key themes across subsegments of emerging software with company-specific examples below:

- **Back Office: Companies within our back-office coverage generally provide software to address workflows in the Office of the CFO, which contain numerous repetitive, time-consuming, and manual processes.** Generative AI has so far given software companies the ability to reduce the need for overhead and quicken workflow through automation of tasks, which gives products incremental value for customers. A few examples from our coverage include: **1) BlackLine's (BL)** ability to automatically detect anomalies in accounting entries within the financial close process, **2) Vertex's (VERX)** recently acquired AI tax-specific capabilities that reduce the time required to accurately categorize products and services across millions of tax rules and jurisdictions, and **3) Workiva's (WK)** embedded, off-the-shelf LLMs that have been trained to generate targeted responses for a wide range of user queries (how to use the platform, clarifying and explaining regulatory statutes, workflows, and more). Given the generally risk-averse nature of the office of the CFO, partners are critical to the process of getting customers comfortable with using these capabilities, as they can often be a core part of the heavy lifting in the validation, accuracy, and security process, which is happening today. That said, these enhancements will likely require continued human oversight, at least in the near-term, as accuracy at scale will need to be proven out over time. Given this, we expect a longer timeframe of adoption and broader-based monetization.

- **Front Office: Companies within our front-office software coverage possess the greatest near-term opportunity to monetize Gen-AI with products that leverage unique datasets to generate value-add content for customers.** Within our coverage, Semrush (**SEMR**) has been the most aggressive in its Gen-AI monetization strategy with several products including 1) ContentShake AI, which utilizes data to create unique content designed to appear in the company's tone and writing style while being optimized to increase traffic and engagement through pointed customer targeting (priced at \$60 per month), 2) AI Writing Assistant, which generates long-form content requiring only a few manual inputs (priced at \$25 per month), and 3) AI Social Content Generator, that creates ads, videos, and marketing posts for individual brands (priced at \$35 per month). Given these products are in the early innings of adoption, we expect to see continued momentum in the monetization of Gen-AI across other front-office software companies within our coverage once these features are advanced enough to provide an ROI than justifies the added cost.
- **Vertical Software: We believe companies within our vertical software coverage have a meaningful opportunity to leverage their long-term customer relationships and often large, proprietary data moats to add to competitive advantages in verticalized markets.** Although these products will likely require longer build times given the specialization required, companies like Waystar (**WAY**) in the healthcare services space are already executing on this opportunity through their Hubble AI, that 1) assists in the reduction of manual tasks required for hospital systems to receive revenue from patient claims, and 2) leverages existing data to provide customers with pricing accuracy in advance of patient visits (rather than post-visit), improving point-of-sale conversion. Similarly, nCino's (**NCNO**) newer AI-powered product suite (nIQ) includes Gen-AI Banking Advisor, which reduces manual workloads in the loan origination process, and Commercial Pricing & Profitability, which provides insights allowing banks to better understand pricing and profitability within their loan portfolios. NCNO has noted ~39% of existing customers utilize at least one nIQ product. ***Over time, we believe Gen-AI has the potential to provide additional pricing power for vertical software companies as they extend competitive advantages in niche software categories, something that is likely to play out over a multi-year period.***

Consumer Internet - Covered by Eric Sheridan

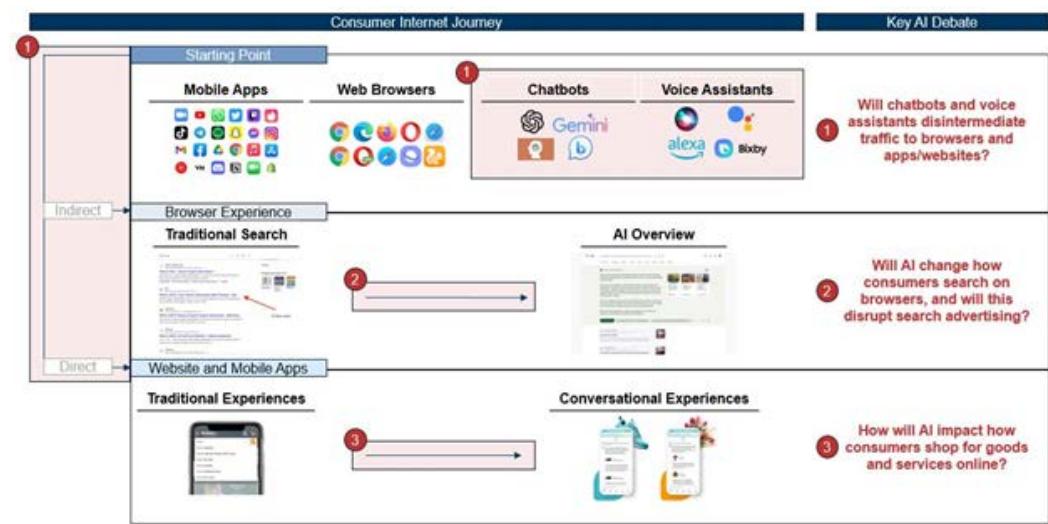
Since the public launch of ChatGPT in November 2022, the rise of Generative AI has been the dominant theme across the technology industry (inclusive of our Consumer Internet and Interactive Entertainment coverage) with debates revolving around companies being impacted by a mixture of rising capital expenditures, increased computing costs and the potential for business model innovation/disruption if AI leads to a shift towards a 3rd wave of computing in the coming 5-10 years (following the introduction of desktop and mobile). ***Similar to prior consumer computing cycles, we expect that any meaningful change to consumer Internet habits will play out over several years and that the disruptive impact of AI on existing business models and industry operating trends will likely be more incremental in the next 12-18***

months.

We see three potential sources of disruption from AI as it relates to consumer internet habits: 1) AI powers new interfaces that can cause shifts in how consumers start their online journey (chatbots, voice assistants); 2) Generative AI offers a more streamlined search experience on traditional browsers which can disrupt the current digital advertising landscape; 3) Generative AI can be deployed inside of mobile applications and website to offer more conversational experiences to consumers across a wide range of use cases (shopping, media, education). In addition, we categorize three key layers by which Generative AI is currently being developed and implemented: **1) the infrastructure layer** - fulfillment of the Cloud Computing needs associated with this technology; **2) the AI model/platform layer** - dynamics around the development and ownership of large-language models (LLMs), where we view that only a select few have both the scale of capital and technical resources to compete effectively (GOOGL, AMZN, META and MSFT [covered by Kash Rangan], either developing internally or investing in partnering with 3rd party AI companies); & **3) the application layer** - potential shifts in market share, unit economics and monetization opportunities, resulting from changes in behaviors and the introduction of new ways to interact with mobile and desktop products.

In our view, the next iteration of the Internet (the Interactive Web) will have a series of characteristics: **1) Device / User Interface** – With the primary user interface/internet access point still being defined by existing devices (probably mobile), the next evolution of the smartphone (elements of increased personalization and aggregated assistant driven behaviors) could be driven by increasing elements of AI through mobile apps (ChatGPT, Perplexity) or revamped on-device assistants (Google's Gemini), leading to increased efficiency and operating processes with less friction; **2) Platform Ecosystem (“Open Web” vs. “Closed/Walled Gardens”)** – still a debate with low visibility on how the mixture of walled garden behavior (a hallmark of mobile computing) will persist or be disintermediated by more Open Web practices; **3) New Monetization Models** – in addition to the forms of monetization (advertising, commerce, subscription) we see the rise of sustained momentum on elements of the Creator Economy, Data Licensing (given the importance of data to evolving/iterating AI) & Paid Consumer AI models as rising forms of industry revenue sources; & **4) Product / Content Discovery** – rise of automated and curated recommendations which accelerate potential outcomes with purchase pathways with less friction.

Exhibit 85: Consumer Internet Journey and Key AI Debates



Source: Company data, Goldman Sachs Global Investment Research

Business Services - Covered by George Tong

Financial Services: Gen-AI is changing the way financial professionals interact with data by enabling contextual search with natural language queries, improving the efficiency of information retrieval and eliminating the need for complex query languages. Before Gen-AI, pulling information from data service terminals such as Refinitiv, FactSet, CapIQ or Bloomberg was made possible with specialized add-ins and Excel functions, which required users' familiarity with specific formulas and syntax to query the companies' database. For other website-based data platforms, users needed to navigate the website, locate specific data links and download the file before manually analyzing the data in Excel. In contrast, Gen-AI-powered search enables users to query databases using natural language to automatically download data and/or generate charts, graphs and summaries based on the retrieved data without manual intervention or technical expertise, which we believe improves the user experience and increases product usage. Information services companies have widely implemented Gen-AI-powered conversational data query capabilities as the first phase of Gen-AI adoption, as shown in [Exhibit 86](#).

Exhibit 86: Gen AI/Gen-AI powered data search functionality has been widely adopted by information services companies

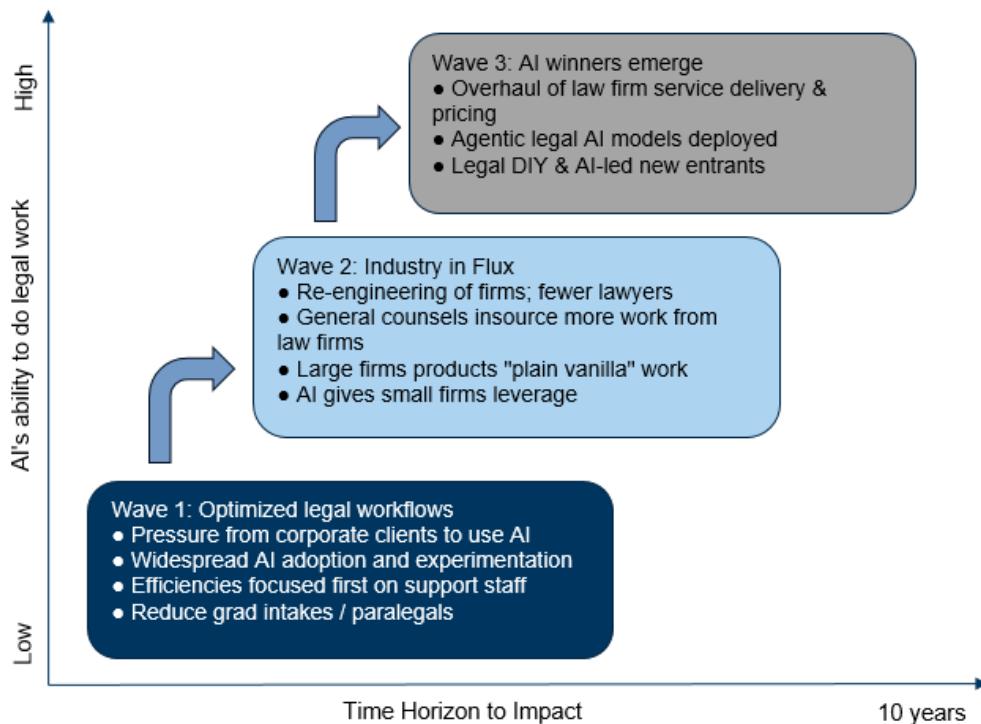
Existing Gen-AI powered data search and analytics products offered by financial services companies

Companies Covered by Info & Business Services	Gen AI Powered Products	Other Companies	Gen AI Powered Products
FactSet (FDS)	<ul style="list-style-type: none"> • FactSet Mercury • Portfolio Assistant 	AlphaSense	<ul style="list-style-type: none"> • AlphaSense Assistant - Gen AI chat extracts insights from different documents real-time • Enterprise Intelligence - market research
Moody's (MCO)	<ul style="list-style-type: none"> • Moody's Research Assistant 	Bloomberg	<ul style="list-style-type: none"> • Virtual assistant powered by BloombergGPT - providing Gen AI services for finance professionals
MSCI Inc. (MSCI)	<ul style="list-style-type: none"> • Conversational AI for portfolio and risk management solutions • Risk Signals 	Hebbia AI	<ul style="list-style-type: none"> • Matrix - portfolio of efficiency tools for unstructured data analysis
S&P Global (SPGI)	<ul style="list-style-type: none"> • ChatIQ • Generative AI search in the S&P Global Marketplace 	Nasdaq (NDAQ)	<ul style="list-style-type: none"> • Global market surveillance offerings with Gen AI
Verisk (VRSK)	<ul style="list-style-type: none"> • Discovery Navigator for P&C insurance claims 	News Corp Dow Jones	<ul style="list-style-type: none"> • Dow Jones Risk & Compliance - Dow Jones Integrity Check for financial crime investigation

Source: Goldman Sachs Global Investment Research

Legal Services: Gen-AI is changing the workflows of corporate legal departments and law firms. Thomson Reuters indicated that 70% of legal professionals believe AI and Gen-AI will have a transformational impact on the legal profession within the next five years, and the legal industry should expect wide-scale business model changes within the next 10 years, as shown in [Exhibit 87](#). The acquisition of Casetext by Thomson Reuters underscores how the combination of extensive legal data sets with Gen-AI technology can help determine future market leadership. Gen-AI-enabled Westlaw and Practical Law solutions at Thomson Reuters streamline the litigation preparation process and potentially improves the likelihood of success in court. Before Gen-AI, legal search technology was limited to keyword searches, which could be inadequate given language used in court rulings often varies and relevant cases may appear to address different subjects. With semantic search capabilities, machines can understand the intent and context behind search queries, which allows for a more comprehensive retrieval of pertinent case precedents. Gen-AI is also changing the deposition preparation process for litigators. Given the ability to analyze a large amount of court records, case files and documents, Gen-AI-powered legal tools can reduce the time required to assemble a case from days to minutes, and assist with strategizing the litigation process, including generating realistic mock deposition scenarios, anticipating opposing counsel's strategies and refining the litigator's inquiry techniques. Additionally, Gen-AI-powered legal tools enhance E-discovery workflows by instantly processing large amounts of electronic documents, organizing unstructured data, identifying relevant information and summarizing findings, which significantly reduces the labor and time required. In transaction law and patent law practices, LLMs trained on legal knowledge can handle contract drafting and forms completion with minimal manual intervention, allowing legal professionals to focus on more complex aspects of legal work.

Exhibit 87: Gen-AI will likely transform legal working practice and law firm cost bases in near term and lead to business model transformation across the industry in about ten years
 Three overlapping waves of Gen-AI transformation in legal industry projected by Thomson Reuters



Source: Company data, Goldman Sachs Global Investment Research

Tax Services: Gen-AI has improved the personal tax filing workflow, reducing tax preparation time while enhancing filing accuracy. Gen-AI-powered assisted products, such as HRB's AI Tax Assist, provide personalized answers to users' tax questions based on uploaded documents and historical data. LLMs are able to pull real-time data and discover recent changes to the tax code, providing dynamic responses that are up-to-date with the latest tax laws. LLMs also perform comprehensive accuracy checks, flagging errors such as missing fields or incorrect entries. Additionally, the integration of Gen-AI into corporate tax workflows improves tax professionals' efficiency, allowing companies to redirect their workforce toward more strategic areas. Gen-AI powered corporate tax products pull real-time data updates on evolving tax regulations and facilitate corporate tax code research that supports natural language queries, reducing the risk of tax penalties and reputation damage. For example, Thomson Reuters is in the process of integrating CoCounsel's Gen-AI capability into its Tax & Accounting solutions, improving the efficiency of tax research workflow for corporate clients. A research survey of tax professionals from the Thomson Reuters Institute identified most common use cases cited by at least half of respondents who are using or planning to use Gen-AI tools in their tax practice, which includes tax research, tax return preparation, tax advisory, document review and correspondence drafting.

IRM Data Centers Business: We believe IRM is a key beneficiary of rising data center demand brought on by Gen-AI. IRM expanded into the data center business in 2017 beginning with the Northern Virginia campus, which consists of nine operational or planned data center buildings. The company now has data centers in major markets

including Northern Virginia, Phoenix, London, Frankfurt and Amsterdam. Across the globe, IRM has approximately 500 MW of operational or in-development data center capacity. IRM's data center business began primarily with co-location clients, which currently still represent the majority of leased MW. The company experienced a shift to more hyperscale contracts beginning in 2020, when it won a contract from a reputable hyperscaler that legitimized its data center capabilities. The rise of Gen-AI has served to intensify hyperscale demand. IRM has 861 MW of potential total data center capacity based on its signed power and land commitments, up from ~750 MW one year ago, creating a healthy runway for growth from the current 500 MW of operational or in-development data center capacity. In its data center business, IRM signed 124 MW in 2023 and 30 MW in 1Q 2024. IRM's data center revenue is growing 20%+ y/y organically.

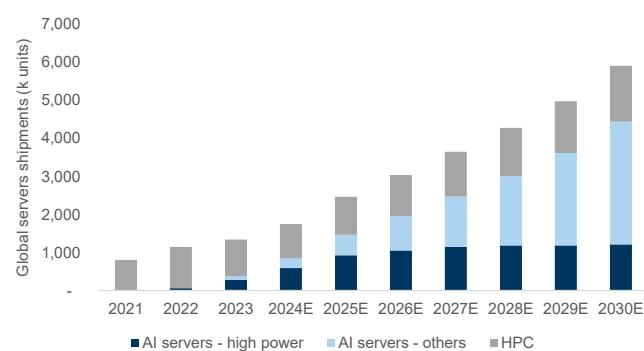
Industrial Tech - Covered by Mark Delaney

AI infrastructure is most meaningful for VRT, FLEX, JBL, APH, and TEL in our Industrial Tech coverage.

Industrial Tech coverage. There are several companies in our industrial tech coverage that we believe are well-positioned to support the infrastructure build-out tied to AI. While the pace of deployments and shipments to support the AI build-out could be somewhat volatile depending on supply chain availability, the growth of applications that utilize Generative AI (as discussed throughout this report), and the project nature of investments, we see this as a multi-year opportunity for several companies in our coverage.

Importantly, with our GS global tech colleagues expecting AI servers to comprise just 6% of server mix in 2024 and for AI servers to grow at least 20% for each of the next five years ([Exhibit 88](#)), we believe this will be a multi-year tailwind for the industry.

Exhibit 88: The GS global tech team expects AI server shipments to grow at a double digit CAGR for the next several years



Source: Company data, Goldman Sachs Global Investment Research

We highlight **VRT, APH, TEL, FLEX, and JBL** from our coverage as the stocks where the AI infrastructure market is most meaningful:

- **Vertiv** - We believe Vertiv is one of the key beneficiaries in our coverage from growing hyperscale investment including for AI, with 35-40% of Vertiv's total revenue tied to hyperscale/co-location datacenter customers. Vertiv's key products

include air handlers, in-room cooling, liquid cooling (including CDUs), chillers, heat exchangers, UPS systems, switchgear, and busbars. Vertiv historically estimated that it had mid teens market share in the power market, and >30% in thermal/cooling, although Vertiv has said that it's too early to define market share in liquid cooling.

Vertiv estimates that its revenue opportunity increases to \$3-\$3.5 mn per MW for an AI workload, up from \$2.5-\$3mn for a traditional compute application. We believe that demand for liquid cooling is one of the drivers of this increase in its addressable market for AI, and Vertiv is on track to increase the capacity of CoolTera (a key historical partner for Vertiv in CDUs that Vertiv acquired in December 2023) by 45X over the course of 2024. Vertiv's orders grew roughly 60% yoy in both 1Q24 and 2Q24, and its backlog was about \$7 bn at the end of 2Q (with some of the backlog for 2026 delivery). Vertiv guided its 3Q24 orders to grow 10-15% yoy, citing order lumpiness as a key factor in the slower 3Q order outlook vs. 2Q. We believe some investors had viewed this guidance as conservative (especially given that Vertiv has generally reported results above guidance in recent quarters). However, with the mix of Blackwell servers (that would have higher power/cooling needs than the current Hopper products) potentially lower in 2H24 than we previously expected due to supply chain constraints, this could limit the degree of 3Q24 order/2024 EPS upside vs. company guidance compared to the last few quarters.

- **Amphenol** - Amphenol is a provider of high-speed, low-latency connectivity products for AI datacenter applications, and we believe the company is benefiting in particular with cable assembly and connector products that are utilized in Generative AI deployments and the fabric approach to processing. We estimate that Amphenol generated ~\$250-\$350mn of revenue from AI applications in 2023, and that quarterly revenue is currently \$200-\$250mn (or \$800mn - \$1bn annualized), with the company noting on its 2Q24 earnings call that the vast majority of its yoy growth in IT Datacom came from AI related interconnect. When we consider the potential for the technology needs for more advanced chips and the GS hardware team's server forecast, we believe Amphenol's AI related revenue could be ~\$2bn in 2025.
- Although **TE Connectivity's** AI revenue exposure is somewhat smaller, the company expects its revenue to grow from about >\$250mn in FY24 to >\$500mn in FY25 (and reach ~\$1bn a few years after that), and we note the company increased its outlook to these levels on its 3QFY24 (2QCY24) earnings report as we wrote in our wrap note on 7/24.
- We also continue to believe that **Flex** and **Jabil** are well positioned to benefit from AI/datacenter builds, and we believe that each company has >10% of revenue tied to the datacenter end market (including power products for Flex and high-speed networking for Jabil). For Flex, the datacenter business across both assembly and power grew about 60% in 2QCY24 (1QFY25), and the company expects a 20% revenue CAGR through FY2029.

Healthcare - Covered by Salveen Richter and Chris Shibutani

Viewing the Generative AI revolution through a healthcare lens. Overall, we see value in Generative AI's potential to create high-quality content at a faster rate, which

can increase labor productivity and drive new insights while providing cost savings. As it relates to potential use cases and advantages in healthcare, we highlight:

- **Synthetic data generation for drug development and diagnostics.** AI/ML is highly dependent on the availability/quality of training data, and Generative AI could help mitigate the negative impact of inadequate/low-quality datasets by producing artificial (synthetic) data that augments areas where collected data may be scarce or expensive. For drug development, synthetic data could virtually augment chemical libraries, providing the data on which to train ML models for drug design/optimization purposes, at a lower cost than physically synthesizing such molecules in the lab. Synthetic data can also aid in database diversification and anonymity for diagnostic purposes when data collection is unfeasible.
- **Generating designs for novel drugs.** Traditional de novo discovery (the design of novel chemical entities that fit a set of constraints) typically involves the time- and resource-intensive process of screening large libraries of variants against a target to generate binding sites and may be limited by design constraints, chemical accessibility and the quality and diversity of the underlying data. Generative AI could be used to expand the chemical library and take a more flexible, targeted approach to designing compounds with specific properties, including factors such as stability and reactivity that impact the practical feasibility of generated designs.
- **Personalized medicine.** Generative AI can analyze large datasets and include multiple patient-centric factors such as genetics, medical history, and lifestyle as well as considerations such as current practice guidelines to assist providers in generating a personalized, informed treatment plan for patients - thereby potentially improving adherence and treatment outcomes. In addition, applications from large language models could help patients better understand their diagnosis and treatment by summarizing and simplifying the language (which can be quite complicated, with medical jargon) for different reading abilities and languages. We also see the potential for Generative AI to reduce the burden of clinical documentation for physicians so that they can spend more time with patients.
- **Diversity and equity in healthcare.** Generally, the real-world use of therapeutics is informed by the results of controlled clinical trials, and as such, it is important that clinical trials are well-designed, avoid bias and include participants that are representative of the full spectrum of patients that the drug is meant to treat after approval. While racial and genetic factors can significantly impact a patient's response to a therapeutic, minority groups are historically underrepresented in clinical trials, resulting in insufficient data to evaluate the efficacy and safety of new drugs in certain patients. While this problem persists in traditional AI/ML given the analysis is performed on existing training data that may contain implicit bias and skew towards favoring certain groups over others, Generative AI may offer the potential to improve imbalanced/unrepresentative datasets by selectively augmenting the representation of minority datapoints (via synthetic minority oversampling), thereby mitigating bias in ML uses for healthcare. In addition, Generative AI could be used to produce simulations and text for educating potential

study participants (e.g., on the treatment, benefits and risks, protocol, etc.) in a personalized, easily understandable fashion to improve enrollment and adherence across patients of different backgrounds and languages in clinical trials.

- **Manufacturing and supply chain efficiency.** While global supply chain continues to be tenuous across the pharmaceutical landscape, notably with disease areas such as obesity, many companies look to leverage Gen-AI to better predict and scale manufacturing infrastructure. Using Generative AI, companies can detect anomalies and product real time recommendations to operators in hopes of improving product yield and decreasing product cycle time. However, not only can these technologies help with internal efficiency, but also with predicting product level supply and demand from current market trends.
- **Approval and launch materials.** In the lifecycle of drug development, the process of filing for approval is often arduous, noting that individual applications must be sent to the governing regulatory body for each individual territory (i.e. FDA, EMA, etc.). Gen-AI offers the potential to accelerate regulatory data submissions through crafting the necessary documentation, thus potentially decreasing the amount of time needed until drug approval. Upon drug approval, the company typically deploys digital content to promote the drug and cultivate brand awareness; On the forward, we expect pharmaceutical companies to improve their direct-to-consumer digital content strategy using Generative AI insights on regional end-market trends.

We expect that Big Tech will continue to look to healthcare for optionality as a growth vertical and use their expertise and resources to support innovation from healthcare companies in the space. Historically, companies such **AMZN** (AWS), **MSFT** (Azure), and **GOOGL** (Google Cloud), have stepped in as cloud service providers, essentially helping healthcare companies with productivity gains relative to using on-premise infrastructure for data analytics, managing unstructured data, while saving on CapEx by using computing resources as needed. In therapeutics, **NVDA** is involved both with software and hardware, providing supercomputers/chips that are crucial for tasks such as AI-powered imaging and analysis and foundation models. For providers, Generative AI can help automate tasks such as drafting clinical notes from patient conversations, integrating these in electronic health records and drafting insurance authorizations for productivity gains - one player here is Microsoft's Nuance for automated clinical documentation, which incorporates GPT-4.

Regarding the risks, much has been written about the risk of Generative AI models providing inaccurate information without realizing it (called "hallucinating"), even citing plausible-looking sources that do not exist to support its conclusions. **Such hallucinations could hinder uptake in life sciences given the impact on human health (e.g., making diagnoses) and, even if used as a tool, clinicians may require a clear trail of source material that the model used to arrive at its decision.** Thus, humans may need to develop efficient methods on how to reliably detect misinformation in AI, as manual verification would undermine AI's convenience/speed advantage.

European Financials Group - Chris Hallam

The use of AI across financial services continues to be a focus in investor and corporate discussions. We estimate the potential AI adoption uplift to ROE at c.200bp (excluding the upfront investment required). We highlight: **1)** The AI impact is most clearly driven by numerous modest cost gains, with the cost opportunity c.3x the revenue opportunity. **2)** Gains are far from evenly distributed, as differences in bank cost and revenue make-up lead to a c.180bp spread within our coverage between those seemingly better placed to capitalise on the AI opportunity vs. those less well positioned. **3)** Almost every bank is already undertaking AI initiatives, and for many this builds on long-standing endeavours in big data analytics and machine learning. **4)** Given the early stage of many of these technologies, pricing mechanisms are still in their infancy and some may prove to be more or less economic over time. Investment flexibility (and budget) will therefore be key for banks to pursue the right initiatives for their business. **5)** The latest AI product offerings from leading vendors focus on customer service, risk management and fraud detection. We see the greatest AI opportunity among relatively capital-light and cost-intensive business models where greater process and cost efficiencies can provide a meaningful uplift to returns. These capital-light business models are typically higher ROTE, with greater through-cycle organic capital generation and larger excess capital positions. These businesses should therefore also find it easier to self-fund AI investment; the adoption of AI across European Banks should accordingly serve to reinforce the returns of industry leaders relative to peers.

Across European Banks we currently see five broad use cases for Gen-AI.

- ***the below bullets are all expounded on than above version*** **1) Enhanced coding efficiency:** Gen-AI powered copiloting tools such as Github Copilot suggest code to programmers (similar to predictive text), leading to faster and more accurate coding. For Banks, not only is there a clear efficiency opportunity given the large tech efforts, but the technology can be particularly helpful on maintenance of older tech stacks where programming languages are less widely used today.
- **2) Data extraction:** Gen-AI can extract and synthesise data and content from large datasets or documents such as annual reports or legislative and regulatory documentation far faster than the traditional manual method. This frees up more time for analysis and decision-making — this is particularly powerful when it is typically the same team who would handle data extraction and data analysis (i.e. where there is an opportunity cost between the breadth of data extracted and the depth of the data analysis).
- **3) Chatbots:** While chatbots aren't a new technology to European Banks, Gen-AI offers a step change in interface performance (the relevance of the questions and answers) as well as widening the scope for products which can be proposed to the customer. Overall, this more personalised and efficient setup offers banks the chance to reduce operating costs associated with client queries as well as improve revenue generation through higher product sales. From the customer perspective, the chatbot evolves into a digital assistant. These improved chatbots and digital assistants can also be used by call centre staff promptly to provide better proposals and solutions to the client on the call.

- **4) Automation:** Gen-AI presents numerous opportunities to automate tasks across a bank. In client onboarding, taking client provided data and inputting it into firm systems can be partly or fully automated. In client communication, the outcome of automated decision models (e.g. credit card applications) can be automatically transformed into client-friendly communication. Overall, many repetitive manual tasks have the potential to be partly or fully automated.
- **5) Human/AI collaboration:** Similar to the potential in coding, Gen-AI technologies offer copiloting opportunities across a range of programmes and applications used day-to-day in banking. Microsoft 365 Copilot enables Office users to turn to Copilot to generate emails, analyse data, prepare presentations and much more. In our view, the human/AI collaboration has the potential to deliver significant time savings and output improvements across a very broad base of daily tasks.

AI in Financial Services: The focus from large technology incumbents

- **Customer conversations & recommendations:** businesses across the globe are leveraging solutions offered by artificial intelligence firms in order to improve their customer service. For the financial services industry, the range of solutions includes personalisation, robo-advisory, and recommendations. For instance, *Google Cloud* is partnering with **KeyBank** to enable the bank to leverage *Google's* virtual agents and contact centre AI solutions to offer new tools for client interaction. Moreover, **Scotiabank** is also partnering with *Google Cloud* to provide its clients with predictive offers as well as to leverage NLP, voice, and vision capabilities to improve customer interactions with the bank. Another example is **Banco de Crédito e Inversiones'** move to *Microsoft Azure* to improve performance and reliability of the bank's system while making better use of customer information to support decision-making for product development and improvement.
- **Process enhancement & modelling:** the big-three vendors also offer solutions for the financial services industry to improve internal processes, streamlining actions that were previously more personnel-intensive. One example of this is *Mphasis* who run their expenses forecasting software on *Amazon's AWS* to offer companies 30 weeks' forward forecasts of operating expenses. Moreover, *Google Cloud* has enabled **HSBC** to launch a more efficient scenario risk-modelling tool that, for example, runs simulations 16x faster than their previous solution — enhancing visibility of risks. An example of an application of AI to improve internal processes while also enhancing the customer journey is **Mr. Cooper's** usage of *Google Cloud AI and ML* to analyse and process lending documents, increasing efficiency and hence improving customer experience.
- **Controls and anomaly detection:** For example, **BNY Mellon** partnered with *Google Cloud* not only to use AI and ML to improve settlement and clearance processes but also to develop AI-powered solutions for anomaly detection in transactions. Moreover, **HSBC** leveraged *Microsoft Azure* to build its PayMe for Business app where the bank can, through the use of machine learning, observe patterns and relationships as well as spotting anomalies in customer activity.

Banks working with Newer AI players

- **Risk management:** Financial firms, and banks in particular, are focused on leveraging AI in order to improve their risk management processes. As a result, AI unicorns globally have started developing advanced machine learning algorithms and models to offer to these institutions. *OpenAI*, for instance, uses its Machine Learning (ML) algorithms to assist financial firms in processing large datasets, which in turn enables better analysis and risk assessment. Other than this, *Databricks* has offered new AI-driven risk models to both **HSBC** and **ABN Amro**, whereas *Tanium* helped **Barclays** improve their data protection capabilities.
- **Fraud detection:** Financial fraud has become increasingly prevalent and costly for global financial institutions; artificial intelligence companies are accordingly enhancing capabilities to deal with this issue. Among these players, both *Databricks* and *OpenAI* offer AI-enabled solutions to enhance fraud prevention and detection systems for financial firms.
- **Improved services:** For the financial services industry, the range of solutions includes personalisation, robo-advisory, cloud security services and compliance reporting. For instance, *Databricks* provides use cases around modern cloud data infrastructure, hyper-personalisation (with **HSBC**), compliance and regulatory reporting (with **FINRA**). *OpenAI*, on the other hand, has developed advanced ML models to i) improve financial analysis for these firms, and ii) develop robo-advisory platforms to improve investment services to the clients. Finally, *Lacework's* cloud security solutions for fintech accelerates product development and *Tanium's* platform helps in speeding up processes.

Spotlight on Generative AI Use Cases

Within the broader field of AI technologies, the past year has undoubtedly seen a step change in Generative AI, particularly the range of use cases and potential beneficiaries. While there are certainly a large number of potential use cases for 'traditional AI' (Machine Learning, Big Data Analytics etc), we are also beginning to see some Gen-AI use cases emerging. Overall, how does this frontier technology suit Banks? On the positive side, generative large language models are currently well suited to: i) textual analysis and creating summaries, ii) content creation, and iii) knowledge mining (e.g. finding common patterns or information across diverse media). However, they are not yet well suited to: i) numerical calculation and representation, such as creating financial tables or complex risk models; ii) real-time information, given large language models are limited by content accessible up to their training date; and iii) direct interaction with customers, where new content or advice is developed by AI without any human oversight, as generative AI has the potential to hallucinate. Therefore, at this stage, we expect banks to focus in on a select number of Gen-AI use cases, rather than aiming to apply the technology as broadly as possible across the bank.

Consumer Staples - Covered by Bonnie Herzog

Consumer Staples companies are actively exploring how Gen-AI can help tailor customer solutions, sharpen innovation, increase speed-to-market, and ultimately drive incremental volume & revenue – While early days, many consumer staples companies have begun to explore potential use cases for Gen-AI. This includes

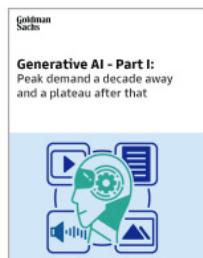
companies such as KO that are piloting AI-driven initiatives to improve their offerings and speed-to-market, including an AI-based price pack channel optimization tool across several markets and a separate AI-driven initiative to push personalized SKU recommendations to retailers. KO is encouraged by the YTD results and believes it is just scratching the surface of what is possible with AI. PEP has also announced plans to execute an aggressive data and AI strategy against its core capabilities, with the goal of improving their consumer insights, strengthening their forecasting capabilities, and ultimately delivering better execution/performance from the plant to the shelf – which ultimately should drive faster growth and productivity. In HPC, EL's collaboration with MSFT has allowed it to embed AI into its system to drive faster speed-to-market and local relevance, as well as improve the customization of its messaging at scale. CL expects Generative AI to be significantly disruptive in areas like innovation and content creation. To that end, CL is funding various test case initiatives to see which have the potential to be scaled, with the goal of integrating AI into CL's broader 2030 strategy. For its part, CLX recently deployed an AI-enabled digital core into all of its businesses, which uses a signal sensing technology to sense trends after combing through vast amounts of social media and marketplace data, then leverages Generative AI to create digital prototypes of innovative product ideas that get immediate consumer feedback in a virtual marketplace. While still early, CLX already finds AI has helped reduce the cycle time on innovation ideas by 50%. Overall, consumer staples companies are broadly optimistic about the long-term potential of Generative AI and are committed to being at the forefront of AI exploration, learning and implementation.

Retail and Hardlines - Covered by Kate McShane

Key anticipated benefits and risks of deploying Gen-AI. Companies are targeting Improving employee productivity and cost savings in the near term and improved customer service longer term. Some companies plan to use tools for generating marketing copy for articles and website. All such generated material have manager oversight reviewed for accuracy to reduce risk around biases and errors. As companies get advanced and put up the necessary guardrails, the anticipated benefits would move to faster delivery of business insights. Companies are leveraging tech partners like Snowflake, Salesforce, Bloomreach, Microsoft, OpenAI, Google Gemini and Cordial who are enhancing their tools to allow clients to take advantage of GenAI and reduce risk exposure by limiting the dataset the AI can train from. They noted a heavy lean towards inference (pre-built model) in their Gen-AI deployments versus training (building).

- Key risks are **1)** Proprietary information being leaked – Mitigation: Guard rails of what can be uploaded, structured data classification preventing shareability outside the organization; **2)** Incorrect information potentially discrediting reputation– Mitigation: Human monitoring; **3)** Fraudsters leveraging GenAI for more sophisticated phishing campaigns – Mitigation – Continuous training. For financial authorizations, leveraging Wet Signatures is one way to mitigate risks; **4)** The degree of data quality required is more critical in this space; and **5)** The need to be purposeful in investments with an eye on ROI.

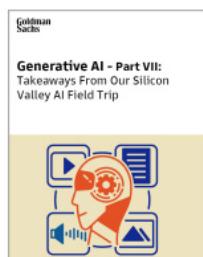
Related Research



Americas Technology: Generative AI - Part I: Laying Out the Investment Framework

26 March 2023

Cloud computing created new investment opportunities by enabling the delivery of software as a utility. Generative AI further unlocks value as it extends this utility and provides new tools for enhancing end-user productivity. While traditional AI has been helpful in making predictions of outcomes, Generative AI is about generating content such as text, video, images, or computer code, which was previously not possible. Large Language Models (LLMs) are a key enabler of GAI, with a profound level of proficiency and intelligence. AI has the potential to establish new companies, while providing incumbents with new growth avenues by turbo-charging end-user productivity.



Americas Technology: Gen AI Part VII: Takeaways From Our Silicon Valley AI Field Trip

21 Aug 2024

We hosted our Silicon Valley AI field trip focused on exploring the use cases and application layer of Gen AI. This has been a key area of investor interest given the significant investments made into the infrastructure layer of Gen AI that are prompting the question of when companies will begin to see real world use cases and returns on their investment. Our lineup for the field trip included presentations and product demos from leading Gen AI private companies, including Datastax, Glean, Hebbia AI, Tera AI and Enquire AI, followed by visits to top VC investors in Gen AI, including Kleiner Perkins, Google Ventures and Lightspeed Venture Partners. The day concluded with a visit to a Stanford University Professor of Computer Science who is a thought-leader on Gen AI.



Top of Mind: Gen AI: too much spend, too little benefit?

25 June 2024

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



Generational Growth: AI, data centers and the coming US power demand surge

28 April 2024

We forecast a 15% CAGR in data center power demand from 2023-2030, driving data centers to make up 8% of total US power demand by 2030 from about 3% currently. We now see a 2.4% CAGR in US power demand growth through 2030 from 2022 levels vs. ~0% over the last decade. Of the 2.4%, about 90 bps of that is tied to data centers.

More related reads:

- Americas Technology: Software: Cyclical or Structural: Don't Count Software Out of Gen-AI (3 June 2024)
- Americas Technology: Software: Cusp of a Platform Shift from Compute to Cognition, Catalyzed by Gen-AI (14 June 2024)
- Americas Technology: IT Spending Survey: 2024 spending expectations inflect positively, Gen-AI sees strong future budget intentions (18 March 2024)

Valuation and Key Risks

Amazon (AMZN): We are Buy rated. Our \$230, 12-month price target is based on an equal blend of: (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/Sales applied to our NTM+1 estimates for 1P, 3P, Subscription, AWS and Other segments; (3) SOTP of EV/EBIT applied to North America and AWS segments, and EV/Sales applied to the International segment, also based on our NTM+1 estimates.

Risks to our Buy rating include: Any impact to eCommerce or Cloud growth from competition; lack of success in scaling high margin businesses including Advertising, Cloud, third-party selling and the subscription business; investments across any array of initiatives creating a headwind to gross or operating margin; any product or platform changes necessary to comply with changes to the global regulatory environment; exposure to the volatility caused by the global macroeconomic environment and investor risk appetite for growth stocks.

Apple (AAPL): We are Buy rated on Apple with a 12-month target price of \$275 reflecting 33X our NTM+1Y EPS. Key risks include:

- Weakening consumer demand for products and services. Apple's products and services are typically sold to consumers and any weakness in the macroeconomic environment could reduce demand for Apple products and services. Apple generated over 50% of its revenue from iPhones (F2022), which is highly dependent on purchases driven by upgrades. Lengthening replacement cycles due to macroeconomic headwinds, improved product durability, or lackluster product innovation could all negatively impact upgrade demand.
- Supply chain disruption. Although Apple's suppliers have a global footprint, the majority of final assembly occurs in China. Increased geopolitical tension may result in disruptions to global trade including through tariffs. While Apple has a robust supply chain network, it may rely on one or a few key suppliers for unique or hard-to-manufacture at scale parts.
- Intensifying competition. Apple competes across personal devices (e.g., handsets, tablets, PCs, headphones) and a variety of services (e.g., video streaming, app distribution, advertising, music streaming, online fitness, cloud storage, product warranties). Although Apple is the largest company and the most well resourced among its competitors, it is not the market leader in every single business line. For instance, in video streaming, it faces several key competitors that invest more heavily than Apple in content.
- Regulatory risks. Apple is subject to intense regulatory scrutiny in all the major markets that it operates in. Regulatory intervention could result in weakening Apple's competitive advantages if it is forced to make its proprietary products or services available for competitors to use.
- Capital allocation execution. Apple has a history of M&A, which has no guarantee of

success. AAPL also has a history of share repurchases, which could prove to offer lower ROI or come under deeper regulatory scrutiny.

Adobe (ADBE): We are Buy rated and reiterate our 12-month price target of \$640 on ADBE. Our price target is based on a three-pronged valuation framework based on equal weights to a DCF, EV/Sales multiple, and a P/E multiple. Our DCF assumes a 4% perpetual growth rate (unchanged). We use a 12x Q5-Q8 EV/Sales (unchanged) and a 28x Q5-Q8 P/E multiple (unchanged).

Key risks include: 1) prolonged and worse-than-expected COVID-19 impact causing slower net new business, deal delays, and longer sales cycles, 2) slower and more volatile Digital Experience growth, 3) slower net new subscriber additions, 4) higher expense growth limiting margin expansion, and 5) increased competition.

Arista Networks (ANET): We are Buy rated on ANET with a 12-month target price of \$390 reflecting 37X our NTM+1Y EPS.

Key downside risks include slower cloud CapEx spending; customer concentration, with META and MSFT representing 26% and 16% of total revenue in 2022, respectively; competition, including from Cisco, major Chinese providers Huawei & H3C and whitebox switching solutions; margin degradation from investment in enterprise sales force, elevated costs associated with supply chain headwinds; broader pricing pressure from commoditization of networking hardware; inability to execute on enterprise/campus strategy & development of new SD-WAN platform.

Amphenol Corporation (APH): We are Buy-rated on APH shares. Our 12-month price target is \$75, which is based on 35X multiple applied to our normalized EPS estimate of \$2.15. Key downside risks to our view relate to revenue growth, margins, the trade situation with China (given Amphenol's presence in the country with several manufacturing sites), and Amphenol's ability to continue to execute well on M&A.

Braze (BRZE): We are Buy rated. Our 12-month price target of \$62 is based a combination of our fundamental value (85%) and M&A value (15%). Our fundamental value of \$60 is based on 9x our Q5-Q8 revenue estimates and our M&A theoretical value of \$72 is based on 11x our Q5-Q8 revenue estimates. Key risks include: reduced marketing spend due to macro, gross margin erosion due to mix shift, normalization of MAUs, intensification in the competitive landscape, and the TAM becoming limited to only very sophisticated customers.

CCC Intelligent Solutions Holdings Inc (CCCS): We are Buy rated. Our 12-month price target of \$13 is based on 32x EV/FCF on our Q5-Q8 FCF estimates. Key risks include: 1) unexpected impacts of inflation on the auto insurance ecosystem, 2) changes in the competitive landscape e.g. from disruptive insurance tech companies, 3) ongoing litigation (lawsuits allege that the total loss valuation from CCC's solution is lower than the actual total loss incurred by the insured party).

Clearwater Analytics Holdings Inc. (CWAN): We are Sell rated. Our 12-month price target of \$20 is based on 36x EV/FCF on our Q5-Q8 FCF estimates. Key risks include: stability in financial markets driving stability in the asset manager and corporate pipeline,

company-specific momentum offsetting a broader slowdown in technology investment at customers, company-specific drivers of NRR such as the new pricing model offsetting the impact of rates on AUM, and the potential for Street estimates to be reset to reflect potential for a slowing pipeline and the impact of rates on NRR.

CrowdStrike (CRWD): We are Buy-rated. Our 12-month price target of \$295 is based on 50x FY27 EPS of \$6.25, discounted by half a year to align with our broader price target methodology for Q5-Q8. Key risks include enterprise TAM saturation, competition from Microsoft, Palo Alto and SentinelOne, and a slowdown in endpoint demand environment.

Datadog (DDOG): We are Buy rated and maintain our 12-month price target of \$150. Our price target is based on a 50%/50% weighting of EV/Sales and DCF. We use an EV/NTM Sales multiple of 16x (unchanged). Our 10-year DCF assumes a perpetuity growth rate of ~3% (unchanged).

Key downside risks include: 1) Decline in sales growth from lower SMB contribution resulting from COVID-19, 2) Declining sales productivity, 3) Increased competition from both observability and hyperscaler vendors, 4) Lack of traction of newer products, and 5) optimization headwinds persist.

Dell Technologies (DELL): We are Buy-rated on DELL with a 12-month target price of \$160 reflecting 16.5X NTM+1Y EPS.

Key risks include: Weaker than expected consumer and commercial PC market demand; Worse-than-expected enterprise IT spending; headwinds from hybrid work could result in reduced spending in enterprise IT; Macroeconomic weakness could result in lowered consumer demand; Pricing pressure could occur as competitors discount due to excess channel inventory and reduced demand; Competition from white box manufacturers could occur, particularly in servers and storage.

DigitalOcean Holdings (DOCN): We are Buy rated. Our 12m price target is \$42 based on 24x Q5-8 uFCF. Key downside risks include: further degradation in macro environment, company specific actions not contributing to improved retention, growth comes below expectations when lapping pricing and M&A, and graduation to hyperscalers.

Digital Realty Trust (DLR): We are Buy rated. Our 12-month price target of \$175 is based on an equal weighting of P/AFFO (25X) and DCF. Downside risks: (1) Excess supply-side dynamics in the datacenter market; (2) weaker-than-expected demand dynamics from key datacenter customers, including hyperscalers; (3) the impact of higher rates on core returns; (4) pricing pressure.

Equinix (EQIX): We are Buy rated. Our 12-month target price of \$870 reflects a 50/50 blend of our Price/AFFO (23X) and DCF methodology. Key downside risks: (1) excess supply in the datacenter market; (2) weaker-than-expected demand dynamics from key datacenter customers; (3) obsolescence risk to existing facilities; (4) more intense price competition.

Fair Isaac (FICO): We are Buy rated. Our 12-month price target of \$1,879 is based on

60x our NTM + 1 YR EPS estimate of \$31.32. Our target multiple comes at a premium to the Info Services peer group median of 29.9x due to FICO's strong pricing power in the Scores business and attractive long-term earnings potential from its transformation into a decision management software provider. Key risks include a lack of traction with FICO's decision management software transformation, an inability to sustain elevated pricing increases in the Scores business, a macro and credit market downturn, and competitive threats.

Flex (FLEX): We are Buy-rated on FLEX shares. Our 12-month price target is \$39, which is based on 15X applied to our Q5-Q8 EPS estimate inc. SBC. Key risks to our view relate to macroeconomic demand, supply/demand dynamics, Flex's ability to expand margins, geopolitical uncertainty, and FCF.

Alphabet (GOOGL): We are Buy rated and our 12-month Price Target is \$217. Our price target is based on an equal blend of (1) EV/GAAP EBITDA 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. Specifically:

- 16.5x EV/GAAP EBITDA or ~1.0x EV/GAAP EBITDA-to-growth applied to our NTM + 1 year estimates. We apply a 16.5x multiple to maintain our prior ~1.0x multiple-to-growth.
- 23.0x EV/FCF-SBC applied to our NTM + 4 years estimates discounted back 3 years at 12%. The discount rate represents CAPM using the blended average of companies within our coverage universe consisting of: (1) 3% risk free rate (based on the normalized 10-year rate); (2) average beta of ~1.3; (3) equity risk premium of 7%.

Risks to our Buy rating include: a) competition of product utility levels and advertising dollars; b) headwinds to monetizable (product) search from industry disruption; c) shifting media consumption habits; d) heavy investments depress operating margins for longer than our forecasts; e) no/low levels of incremental shareholder returns going forward; & f) regulatory scrutiny and industry practices altering the business model's prospects. In addition, Alphabet is exposed to the volatility caused by the global macroeconomic environment and investor risk appetite for growth stocks.

HubSpot (HUBS): We are Buy rated. Our 12-month price target of \$626 is based on 10x EV/Sales on our Q5-Q8 FCF estimates. Key risks include: tighter lending standards could have an outsized negative impact on access to funding for SMBs, increased competition, inability to expand beyond core hubs, and weaker-than-expected momentum upmarket.

IBM (IBM): We are Buy rated. Our 12-month price target of \$220 is based on 19X our one-year forward earnings estimates. Downside risks: 1) cyclical headwinds in software, 2) further slowdown in consulting, 3) AI bookings deceleration, 4) dilutive M&A.

Intuit (INTU): We are Buy rated. We reiterate our 12-month price target of \$765. Our price target is based on a three-pronged valuation framework based on equal weights to a DCF, EV/FCF multiple, and a P/E multiple. Our EV/FCF target multiple is 32x our Q5-Q8

FCF estimates (unchanged) and our P/E target multiple is 33x Q5-Q8 EPS (unchanged). Our 10-year DCF assumes a perpetual growth rate of ~3% (unchanged).

Key downside risks include increased SMB failures and churn related to extended COVID headwinds, higher than expected competition in both SMB and consumer businesses limiting Intuit share gains, decelerating customer growth and higher than expected attrition, slower than expected adoption of QuickBooks Online, QuickBooks Advanced, and TurboTax Live, slower adoption of Intuit's higher priced assisted offerings, thereby limiting ARPC growth, online ecosystem growth slowing and not reaching long-term targets for 30% growth, synergies between core Intuit, Credit Karma, and Mailchimp taking longer than expected to materialize, incremental competition from Microsoft, Salesforce, Hubspot in the SMB segment, and slower than expected margin expansion.

Iron Mountain (IRM): We are Buy rated. Our 12-month price target of \$120 based on 23.0x our NTM + 1YR AFFO/shr estimate of \$5.21. Our target multiple is in-line with the median multiple of 23.0x for self-storage and industrial REIT peers reflecting attractive organic growth trends as the company's growth portfolio continues to expand, mitigated by IRM's transformation from a legacy records management company into a data and technology company that comes with execution risk and medium-term financial uncertainty. Risks to the downside for IRM shares include competing technologies in records management, acquisition integration risk, financial leverage and FX risk.

Jabil (JBL): We are Buy-rated on JBL shares. Our 12-month price target is \$132, which is based on 15X applied to our Q5-Q8 EPS estimate. Key risks to our view relate to revenue growth (including due to macro factors, market share, and program mix), its ability to continue to expand margins vs. history, and FCF.

Meta (META): We are Buy rated and reiterate our 12-month price target of \$555. Our price target is based on an equal blend of (1) EV/GAAP EBITDA applied to our NTM + 1 year estimates and (2) a modified DCF using EV/FCF-SBC multiple applied to our NTM + 4 years estimates. Risks include: a) competition for user growth, user engagement & advertising dollars across an array of incumbent and emerging Internet, media and commerce companies; b) large investments in long-tailed initiatives depress operating margins for longer than our forecasts; c) no/low levels of incremental shareholder returns going forward; d) regulatory scrutiny and industry practices altering the business model's prospects; e) potential antitrust scrutiny could prove to be a headwind to M&A aspiration (to add talent and product innovation inorganically) and/or result in a break-up or dis-synergies of prior period M&A activity; & f) inability to monetize upside optionality opportunities (messaging layer, social commerce/shopping, Oculus/Metaverse). In addition, Meta is exposed to the volatility caused by the global macroeconomic environment and investor risk appetite for growth stocks.

MongoDB (MDB): We are Buy rated and reiterate our 12-month Price Target of \$325. Our Price Target is derived from an equal weighting of a DCF using an ~3.5% perpetuity growth rate (unchanged) and 12x Q5-Q8 EV/Sales (vs. unchanged).

Key downside risks include: 1) Adverse changes in the IT spending environment and

spending priorities, 2) Competition in the non-relational database market, particularly from proprietary offerings from cloud service providers (CSPs), 3) The ramp to profitability as the company continues to invest ahead of a large and growing market opportunity, 4) Slowing pace in public cloud adoption, and 5) Potential for rising interest rates.

Monday.com (MNDY): We are Buy rated and reiterate our 12-month price target of \$340. Our price target is based on an equal weighted DCF analysis and EV/Sales multiple. Our DCF assumes a 3% perpetual growth rate (unchanged). We use an 11x Q5-Q8 EV/Sales (unchanged).

Key downside risks to our thesis include: 1) incremental competition as Monday.com faces competition from a number of pure-play project management vendors such as Asana, Smartsheet, Atlassian, and Microsoft, 2) slower-than-expected adoption of adjacent use-cases, 3) slower-than-expected traction with the upmarket motion and enterprise customer growth, 4) higher-than-expected customer attrition, and 5) longer-than-expected free cash flow burn.

Microsoft (MSFT): We are Buy rated and reiterate our 12-month price target of \$515. Our price target is derived from an equal weighting of a DCF (~3% perpetuity growth rate, unchanged), 35x Q5-Q8 EV/FCF (unchanged), and 35x Q5-Q8 P/E (unchanged).

Key downside risks include: 1) Slower than anticipated public cloud adoption, 2) Overall slowdown in IT spending, 3) Slower pace of margin expansion, and 4) Adverse competitive landscape.

Nvidia (NVDA): We are Buy rated. Our 12-month price target of \$135 is based on 50x our normalized EPS estimate of \$2.70. Key downside risks to our estimates and price target include: 1) a sudden decline in Gen AI infrastructure spend by the major CSPs and/or enterprises; 2) further restrictions on GPU exports; 3) delays in new product introductions; 4) weaker-than-expected demand for Gaming GPUs; and 5) supply chain issues.

Palo Alto Networks (PANW): We are Buy rated and reiterate our 12-month Price Target of \$376. Our Price Target is derived from a 32x (unchanged) multiple on our Q5 - Q8 FCF estimates.

Key risks include: firewall cyclical, competition in cloud, competition in SASE, and obsolescence in next gen technology.

Palantir (PLTR): We are Neutral rated and reiterate our 12-month price target of \$16. Our Price Target is derived from a 30x (unchanged) multiple on our Q5 - Q8 FCF estimates.

Key risks include: (-) momentum in the commercial segment slows due to weaker macro, increase in the competitive environment, additional alternative/SPAC investments. (+) tailwinds in the government segment drive revenue acceleration, strong hiring trends in commercial sales reps lead to faster-than-expected growth.

SAP (SAP): We are Buy-rated. Our 12m PT of €240 and our ADR PT of US\$261 are

based on c.35x 3Q25-2Q26E PF EPS (including SBC).

Key risks to our view and price target are as follows: (1) macro risks; (2) cloud and subscription risks; (3) traction in S/4 HANA/cloud; (4) opex spending and (5) further management changes.

S&P Global (SPGI): We are Buy rated. Our 12-month price target of \$561 is based on 32.0x our NTM + 1YR EPS estimate of \$17.53. Our target multiple comes above the Info Services peer group median of 27.8x due to S&P's strong track record of execution, idiosyncratic opportunity for revenue synergy realization from the IHS Markit merger, wide competitive moat due to its proprietary and essential data, and higher EBITDA margins relative to peers. Risks to our rating include global debt issuance volume volatility, financial services budget and headcount pressures that could weigh on Market Intelligence growth, a market pull-back that could put downward pressure on Indices growth, failure to achieve remaining IHS Markit revenue synergies and FX risk.

ServiceNow (NOW): We are Buy rated and reiterate our 12-month Price Target of \$940. Our Price Target is derived from an equal weighting of a DCF (perpetuity growth rate of 3%, unchanged), 41x Q5-Q8 EV/FCF (unchanged), and 13x Q5-Q8 EV/Sales (unchanged).

Key downside risks include: 1) Execution risk limiting growth in new markets, 2) Slower new business growth and longer sales cycles, and 3) Higher expense growth limiting margin expansion. We also note a potential slowdown in net new ACV or customer expansion would likely limit revenue growth in the near term. Furthermore, investors may view the potential of a large M&A transaction as a risk to margin expansion.

Snowflake (SNOW): We are Buy rated and reiterate our 12-month Price Target of \$220. Our price target is derived from an equal weighting of a DCF analysis and EV/Sales multiple. Our DCF analysis is based on a terminal growth rate of ~4% (unchanged) and our relative valuation is based on a 17x Q5-Q8 EV/Sales multiple (unchanged).

Key downside risks include: 1) Adverse changes in the IT spending environment, 2) Competition - particularly from cloud service providers (CSPs) and Databricks, and 3) Outages from reliance on CSPs.

SentinelOne (S): We are Neutral rated with a 12-month price target of \$23, based on 7x Q5-8 EV/sales.

Key risks include competition from CrowdStrike and Microsoft; and balancing between revenue growth and profitability.

Salesforce (CRM): We are Buy rated and reiterate our 12-month price target of \$315. Our price target is derived from an equal weighting of a DCF (~2% perpetuity growth rate, unchanged), 24x Q5-Q8 EV/FCF (unchanged), and 7x Q5-Q8 EV/Sales (unchanged).

Key downside risks include: 1) Sales execution, 2) Macroeconomic slowdown, 3) Unsustainable pace of acquisitions, 4) Slower than expected operating margin expansion or higher than expected expense growth, and 5) Adverse changes in the IT spending environment.

TE Connectivity Ltd (TEL): We are Buy rated on TEL shares. Our 12-month price target

is \$189, which is based on 21X our normalized EPS estimate of \$9.00. Key downside risks to our Buy thesis relate to automotive end demand, TE executing on its margin expansion plan, the impact of EVs and new car architectures on automotive content for TE, unfavorable commodity pricing, macroeconomic demand trends, and TE successfully expanding into the sensor market.

Thomson Reuters (TRI): We are Neutral rated. Our Price Target of \$167 is based on 38x (unchanged) our NTM + 1YR EPS estimate of \$4.40.

Risks to the downside include competition in the legal and tax & accounting markets, the inability to drive new customer growth among small and medium-sized businesses, the inability to adequately upsell enterprise customers, and market share loss from a fragmented product set. Risks to the upside include market share gains from product rationalization, increased cross-selling and EBITDA margin expansion from cost management actions and fixed cost leverage.

Tyler Technologies (TYL): We are Buy rated. Our 12-month price target of \$627 is based on 52x EV/FCF on our Q5-Q8 FCF estimates. Key risks include: 1) Tyler could face execution issues on cross-sell with its broader suite and especially with NIC, 2) there could be changes in legislation that impact budgets, 3) changes in transaction volumes, 4) increased competition, and 5) Tyler could see slower organic growth if government technology adoption slows post a COVID-catalyzed cycle of adoption.

Veeva Systems (VEEV): We are Buy rated. Our 12-month price target of \$235 is based on 28x EV/FCF on our Q5-Q8 FCF estimates. Key risks include: further headwinds in Veeva's CRM business, the timing of new product cycles in R&D, execution on M&A, macro factors that impact life science companies, impacts from the Inflation Reduction Act, and competitive disruption as a result of the CRM transition.

Vertiv Holdings Co (VRT): We are Buy rated on VRT shares. Our 12-month price target is \$104, which is based on a 20X multiple applied to our Q5-Q8 EBITDA estimate. Key downside risks to our view include: 1) Revenue growth – Vertiv has a portion of its datacenter revenue still tied to traditional on-premise datacenters, which is a part of the market that could be impacted by either enterprise spending weakness or by the shift to the cloud. In addition, Vertiv's ability to continue to improve its cloud/hyperscale revenue and grow in the channel will be important factors in how fast its revenue will increase. Also, market share will be a factor in revenue trends. 2) Margins – There are risks to our expectation for margin expansion. This could come from increased competitive pressure, mix, or execution.

Zscaler (ZS): We are Neutral rated. Our 12-month price target of \$189 based on 43x (unchanged) Q5-Q8 FCF.

Key risks: faster adoption of next-generation products (ZPA, ZS for workloads, CWP); saturation in the enterprise, and a change in competitive landscape.

Disclosure Appendix

Reg AC

We, Kash Rangan, Gili Naftalovich, Matthew Martino, Toshiya Hari, Eric Sheridan, Gabriela Borges, CFA, Mohammed Moawalla, Adam Hotchkiss, Michael Ng, CFA, James Schneider, Ph.D., Mark Delaney, CFA, George K. Tong, CFA, Chris Hallam, Salveen Richter, CFA, Chris Shubhani, M.D., Bonnie Herzog, Kate McShane, CFA, Carly Davenport, Jacob Staffel, Henry Dane, Nishad Patwardhan, Ben Miller, Alex Vegliante, CFA, Joshua M. Frantz, CFA, Patty Kanada, CFA, Pierre Riopel, Anmol Makkar, Max Gamperl, Carolyn Valenti, Anna Wu, Greyson Skiba, Morgan Leung, Will Bryant, Tommie Reerink, Michael Smith, Zorayda Montemayor, Karishma Raghuram and Uzair Merchant, 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.

GS Factor Profile

The Goldman Sachs Factor Profile provides investment context for a stock by comparing key attributes to the market (i.e. our coverage universe) 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 (DACP) (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

Disclosure

The spouse of a GIR analyst, who is not on the covered company's coverage team but is a member of the same industry business unit, is the Chief Executive Officer and shareholder of HubSpot.

Logo disclosure

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

Rating and pricing information

Adobe Inc. (Buy, \$558.30), Alphabet Inc. (Buy, \$165.62), Altair Engineering Inc. (Sell, \$90.41), Amazon.com Inc. (Buy, \$177.04), Amphenol Corp. (Buy, \$67.35), Apple Inc. (Buy, \$226.84), Arista Networks Inc. (Buy, \$355.13), Atlassian Corp. (Buy, \$160.00), Autodesk Inc. (Sell, \$255.26), Bentley Systems Inc. (Neutral, \$50.22), Braze Inc. (Buy, \$44.86), CCC Intelligent Solutions Holding (Buy, \$10.94), Clearwater Analytics Holdings (Sell, \$24.63), Cloudflare (Sell, \$82.39), Confluent Inc. (Neutral, \$21.36), Couchbase Inc. (Sell, \$18.27), CrowdStrike (Buy, \$271.54), Datadog Inc. (Buy, \$117.23), Definitive Healthcare Corp. (Neutral, \$4.53), Dell Technologies Inc. (Buy, \$112.01), Digital Realty Trust Inc. (Buy, \$152.18), DigitalOcean Holdings (Buy, \$38.00), Dropbox Inc. (Sell, \$24.33), Dynatrace Inc. (Buy, \$50.24), Elastic NV (Neutral, \$105.66), Equinix Inc. (Buy, \$827.96), Fair Isaac Corp. (Buy, \$1,745.38), Flex (Buy, \$32.93), GitLab Inc. (Buy, \$46.98), HashiCorp Inc. (Not Rated, \$33.90), HubSpot Inc. (Buy, \$503.07), Informatica Inc. (Buy, \$24.56), International Business Machines Corp. (Buy, \$196.10), Intuit Inc. (Buy, \$619.85), Iron Mountain Inc. (Buy, \$113.35), Jabil Circuit Inc. (Buy, \$108.51), Meta Platforms Inc. (Buy, \$528.00), Microsoft Corp. (Buy, \$416.79), Monday.Com Ltd. (Buy, \$263.42), MongoDB Inc. (Buy, \$247.61), Nvidia Corp. (Buy, \$129.37), Oracle Corp. (Neutral, \$139.19), PagerDuty (Neutral, \$20.17), Palantir Technologies (Neutral, \$31.78), Palo Alto Networks (Buy, \$350.75), Procure Technologies Inc. (Buy, \$61.40), RingCentral (Neutral, \$33.81), Rubrik Inc. (Buy, \$34.95), S&P Global Inc. (Buy, \$500.59), SAP (Buy, \$219.15), SAP (Buy, €195.62), Salesforce Inc. (Buy, \$264.00), Samsara Inc. (Buy, \$41.69), SentinelOne (Neutral, \$25.26), ServiceNow Inc. (Buy, \$828.27), Snowflake Inc. (Buy, \$115.86), SolarWinds Corp. (Sell, \$12.53), TE Connectivity Ltd. (Buy, \$153.08), TechTarget (Not Rated, \$26.95), Thomson Reuters Corp. (Neutral, \$166.24), Twilio (Neutral, \$60.76), Tyler Technologies Inc. (Buy, \$571.73), Unity Software Inc. (Neutral, \$17.88), Veeva Systems Inc. (Buy, \$199.25), Vertiv Holdings (Buy, \$78.55), Weave Communications Inc. (Neutral, \$10.91), Workday Inc. (Buy, \$259.95), Zoom Video Communications Inc. (Neutral, \$70.14), ZoomInfo Technologies Inc. (Sell, \$9.72) and Zscaler (Neutral, \$199.43).

The rating(s) for Amphenol Corp., Flex, Jabil Circuit Inc., TE Connectivity Ltd. and Vertiv Holdings is/are relative to the other companies in its/their coverage universe: Amphenol Corp., Aptiv Plc, Aurora Innovation Inc., Belden Inc., BorgWarner Inc., Cerence Inc., ChargePoint Holdings, Flex, Ford Motor Co., General Motors Co., Gentex Corp., Innoviz Technologies, Jabil Circuit Inc., Keysight Technologies Inc., Lear Corp., Luminar Technologies Inc., Magna International Inc., Mobileye Global Inc., QuantumScape Corp., Rivian Automotive Inc., Sensata Technologies Holding,

Symbotic Inc., TE Connectivity Ltd., Tesla Inc., Vertiv Holdings, Visteon Corp.

The rating(s) for Apple Inc., Arista Networks Inc. and Dell Technologies Inc. is/are relative to the other companies in its/their coverage universe: Apple Inc., Arista Networks Inc., Axon Enterprise Inc., Blend Labs, Cisco Systems Inc., Comcast Corp., Compass Inc., Dell Technologies Inc., F5 Inc., Fox Corp., HP Inc., Offerpad Solutions Inc., Opendoor Technologies Inc., Redfin Corp., SoFi Technologies Inc., Stagwell Inc., Super Micro Computer Inc., TD SYNNEX Corp., Upstart Holdings, Walt Disney Co., Warner Bros Discovery Inc., Zillow Group

The rating(s) for Nvidia Corp. is/are relative to the other companies in its/their coverage universe: ARM Holdings, Advanced Micro Devices Inc., Amkor Technology Inc., Analog Devices Inc., Applied Materials Inc., Arrow Electronics Inc., Avnet Inc., Broadcom Inc., Cohu Inc., Credo Technology Group, Entegris Inc., Impinj Inc., Intel Corp., IonIQ Inc., KLA Corp., Lam Research Corp., Marvell Technology Inc., Microchip Technology Inc., Micron Technology Inc., NXP Semiconductors NV, Nvidia Corp., ON Semiconductor Corp., Qorvo Inc., Seagate Technology, Skyworks Solutions Inc., Teradyne Inc., Texas Instruments Inc., Western Digital Corp.

The rating(s) for Adobe Inc., Datadog Inc., Intuit Inc., Microsoft Corp., Monday.Com Ltd., MongoDB Inc., Salesforce Inc., ServiceNow Inc. and Snowflake Inc. is/are relative to the other companies in its/their coverage universe: Adobe Inc., Altair Engineering Inc., Atlassian Corp., Autodesk Inc., Bentley Systems Inc., Confluent Inc., Couchbase Inc., Datadog Inc., Definitive Healthcare Corp., Dropbox Inc., Dynatrace Inc., Elastic NV, GitLab Inc., Informatica Inc., Intuit Inc., Microsoft Corp., Monday.Com Ltd., MongoDB Inc., Oracle Corp., PagerDuty, Procore Technologies Inc., RingCentral, Rubrik Inc., Salesforce Inc., Samsara Inc., ServiceNow Inc., Snowflake Inc., SolarWinds Corp., Twilio, Unity Software Inc., Weave Communications Inc., Workday Inc., Zoom Video Communications Inc., ZoomInfo Technologies Inc.

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., Chegg Inc., Chewy Inc., Clear Secure Inc., Corsair Gaming Inc., Coursera Inc., Cricut Inc., DoorDash Inc., DoubleVerify Inc., DraftKings Inc., Duolingo Inc., Electronic Arts Inc., Etsy Inc., Expedia Group, Fiverr International Ltd., Frontdoor Inc., Genius Sports Ltd., GoodRx Holdings Inc., Groupon Inc., IAC Inc., Ibotta Inc., Instacart, Lyft Inc., Match Group, MediaAlpha Inc., Meta Platforms Inc., Nerdy Inc., Netflix Inc., Nextdoor Holdings, Opera Ltd., Peloton Interactive Inc., Pinterest Inc., Playtika, Reddit Inc., Roblox, Snap Inc., Spotify Technology S.A., Take-Two Interactive Software Inc., TripAdvisor Inc., Uber Technologies Inc., Ubisoft Entertainment SA, Upwork Inc., Vacasa Inc., Wayfair Inc., Webtoon Entertainment Inc., Xometry Inc., Yelp Inc., ZipRecruiter Inc., eBay Inc.

The rating(s) for Digital Realty Trust Inc., Equinix Inc. and International Business Machines Corp. is/are relative to the other companies in its/their coverage universe: AT&T Inc., Accenture Plc, Altice USA Inc., American Tower Corp., Charter Communications Inc., Cogent Communications Holdings, Cognizant Technology Solutions, Crown Castle International Corp., Digital Realty Trust Inc., EPAM Systems Inc., Equinix Inc., Frontier Communications Parent Inc., Globant SA, IHS Holding, International Business Machines Corp., Lumen Technologies Inc., SBA Communications Corp., Softchoice Corp., T-Mobile US Inc., TaskUs Inc., Verizon Communications

The rating(s) for SAP is/are relative to the other companies in its/their coverage universe: Adyen NV, Capgemini, Dassault Systemes, Hexagon AB, Indra, Nemetschek, Nexi, Ovh Groupe, SAP SECO, Sage Group, Sinch AB, TeamViewer, Temenos, TietoEVRY, Wise PLC, Worldline

The rating(s) for Fair Isaac Corp., Iron Mountain Inc., S&P Global Inc. and Thomson Reuters Corp. is/are relative to the other companies in its/their coverage universe: ADT Inc., Bright Horizons Family, BrightView Holdings, Cintas Corp., Clarivate, CoStar Group, Dun & Bradstreet, Equifax Inc., FactSet Research Systems Inc., Fair Isaac Corp., Gartner Inc., H&R Block Inc., Iron Mountain Inc., Korn Ferry, MSCI Inc., Manpower Group, Moody's Corp., NCR Atleos, Robert Half International Inc., Rollins Inc., S&P Global Inc., The Brink's Co., Thomson Reuters Corp., TransUnion, UL Solutions, Verisk Analytics Inc., Vestis Corp.

The rating(s) for Braze Inc., CCC Intelligent Solutions Holding, Clearwater Analytics Holdings, CrowdStrike, DigitalOcean Holdings, HubSpot Inc., Palantir Technologies, Palo Alto Networks, SentinelOne, Tyler Technologies Inc., Veeva Systems Inc. and Zscaler is/are relative to the other companies in its/their coverage universe: Appian Corp., AvePoint, BigCommerce Holdings, Braze Inc., CCC Intelligent Solutions Holding, Check Point, Clearwater Analytics Holdings, Cloudflare, CrowdStrike, Dayforce Inc., DigitalOcean Holdings, Docebo Inc., Enfusion Inc., EverCommerce Inc., Fortinet, Guidewire Software Inc., HubSpot Inc., Kaltura Inc., Klaviyo, Okta, Olo Inc., Palantir Technologies, Palo Alto Networks, Paycor HCM Inc., Pegasystems Inc., Sapiens International Corp., SentinelOne, Shopify Inc., Tyler Technologies Inc., Veeva Systems Inc., 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	48%	35%	17%	64%	57%	41%

As of July 1, 2024, Goldman Sachs Global Investment Research had investment ratings on 2,925 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

Regulatory disclosures

Disclosures required by United States laws and regulations

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

The following are additional required disclosures: **Ownership and material conflicts of interest:** Goldman Sachs policy prohibits its analysts,

professionals reporting to analysts and members of their households from owning securities of any company in the analyst's area of coverage.

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

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, 951-A, Rational House, Appasaheb Marathe Marg, Prabhadevi, 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 Investor Grievance E-mail: india-client-support@gs.com. Compliance Officer: Anil Rajput |Tel: + 91 22 6616 9000 | Email: anil.m.rajput@gs.com. **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, The Investment Trusts Association, Japan, and Japan Investment Advisers Association. 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, 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 not provided or have been suspended 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, when the company is an early-stage biotechnology company, and in certain other circumstances. **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. **Not Available or Not Applicable (NA).** The information is not available for display or is not applicable. **Not Meaningful (NM).** The information is not meaningful and is therefore excluded.

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 contato@goldmansbrasil@gs.com. Available Weekdays (except holidays), from 9am to 6pm. Canal de Comunicação com o Públíco Goldman Sachs Brasil: 0800 727 5764 e/ou contato@goldmansbrasil@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: GSI, authorised by the PRA and regulated by the FCA and the PRA, disseminates research in the following jurisdictions within the European Economic Area: the Grand Duchy of Luxembourg, Italy, the Kingdom of Belgium, the Kingdom of Denmark, the Kingdom of Norway, the Republic of Finland and the Republic of Ireland; GSI - Succursale de Paris (Paris branch) which is authorised by the French Autorité de contrôle prudentiel et de resolution ("ACPR") and regulated by the Autorité de contrôle prudentiel et de resolution and the Autorité des marchés financiers ("AMF") disseminates research in France; GSI - Sucursal en España (Madrid branch) authorized in Spain by the Comisión Nacional del Mercado de Valores disseminates research in the Kingdom of Spain; GSI - Sweden Bankfilial (Stockholm branch) is authorized by the SFSA as a "third country branch" in accordance with Chapter 4, Section 4 of the Swedish Securities and Market Act (Sw. lag (2007:528) om värdepappersmarknaden) disseminates research in the Kingdom of Sweden; Goldman Sachs Bank Europe SE ("GSBE") is a credit institution incorporated in Germany and, within the Single Supervisory Mechanism, subject to direct prudential supervision by the European Central Bank and in other respects supervised by German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin) and Deutsche Bundesbank and disseminates research in the Federal Republic of Germany and those jurisdictions within the European Economic Area where GSI is not authorised to disseminate research and additionally, GSBE, Copenhagen Branch filial af GSBE, Tyskland, supervised by the Danish Financial Authority disseminates research in the Kingdom of Denmark; GSBE - Sucursal en España (Madrid branch) subject (to a limited extent) to local supervision by the Bank of Spain disseminates research in the Kingdom of Spain; GSBE - Succursale Italia (Milan branch) to the relevant applicable extent, subject to local supervision by the Bank of Italy (Banca d'Italia) and the Italian Companies and Exchange Commission (Commissione Nazionale per le Società e la Borsa "Consob") disseminates research in Italy; GSBE - Succursale de Paris (Paris branch), supervised by the AMF and by the ACPR disseminates research in France; and GSBE - Sweden Bankfilial (Stockholm branch), to a limited extent, subject to local supervision by the Swedish Financial Supervisory Authority (Finansinpektionen) disseminates research in the Kingdom of Sweden.

General disclosures

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

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

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

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.fiadocumentation.org/fia/regulatory-disclosures_1/fia-uniform-futures-and-options-on-futures-risk-disclosures-booklet-pdf-version-2018.

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.

© 2024 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 third-party 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.

MINDCRAFT: OUR THEMATIC DEEP DIVES

For the exclusive use of CHRISTINA.LU@GS.COM

The Future of Batteries



Carbonomics



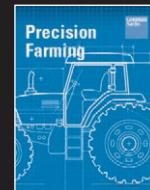
Europe's Energy Crisis



China Agriculture



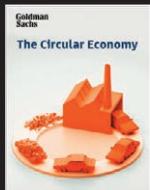
Precision Farming



Green Capex



The Circular Economy



Byte-ology



Gene Editing



The Metaverse



Cloud Computing



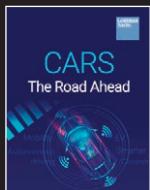
5G



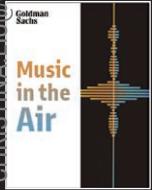
Blockchain



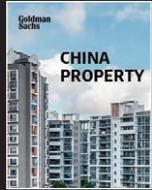
Cars: The Road Ahead



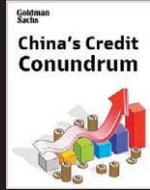
Music in the Air



China Property



China's Credit Conundrum



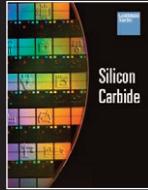
Age of Automation



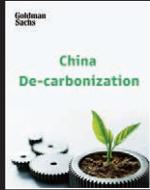
China Post-95s



Silicon Carbide



China Decarbonization



The Survivor's Guide to Disruption



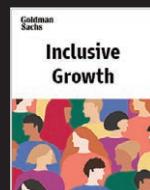
Sustainable ESG Investing



Black Womenomics



Inclusive Growth



Market Cycles



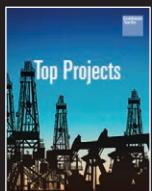
Top of Mind



What Matters for IPOs



Top Projects



Tracking the Consumer



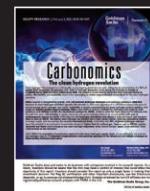
EU Taxonomy



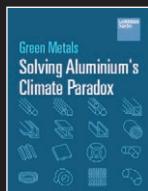
Balanced Bear



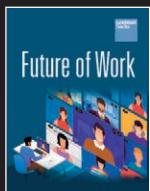
Clean Hydrogen



Green Metals



The Future of Work



What the Market Pays For



The Great Reset



The Competitive Value of Data



MINDCRAFT
GS Research's Theme Bookshelf