

AlphaSense

Generative AI: The Road to Revolution

By Michelle Brophy, Director of Research, Technology, Media,
and Telecommunications

November 2023

Generative AI: The Road to Revolution

Executive Summary

Generative artificial intelligence (genAI) went prime time in 2023, as its vast potential to revolutionize technology applications captured headlines. Following months of market discovery, the conversation has turned to potential enterprise and consumer use case adoption, and more specifically, monetization opportunities and related timelines. However, questions around model foundations, cost, security and facilitation of genAI applications for internet and software companies are in the early stages of being addressed. By leveraging the power of the AlphaSense platform, we will explore the genAI landscape, potential near and long term hurdles, and separate the hype from the reality of company roadmaps.

Deployment of genAI will be driven by potential productivity gains, deeper user engagement, rising user proficiency and increased availability of infrastructure at a more reasonable cost. In the near term, the ability of genAI to deliver on expectations is more unclear than the buzz might suggest. Still, market estimates are staggering, with Markets and Markets [estimating the genAI market to grow](#) from USD 11.3 billion in 2023 to USD 51.8 billion by 2028, at a compound annual growth rate (CAGR) of 35.6% during the forecast period, and Microsoft recently commenting that technology would move from contributing 5% to closer to [10% of GDP](#), based on accelerated AI investments, as examples.

Unprecedented Demand

There's no question that longer term, the new generation of applications being developed is underway. The first step is having data center infrastructure in place to build and deploy applications. Starting with the infrastructure builds, Bank of America reports that [cloud-based capex](#) will come in +14% in 2023 at \$96bn, and is expected to grow another +22%/+7% in 2024 and 2025 to \$116bn/124bn respectively. On recent conference calls, the hyperscalers seem to have backed the enthusiasm:

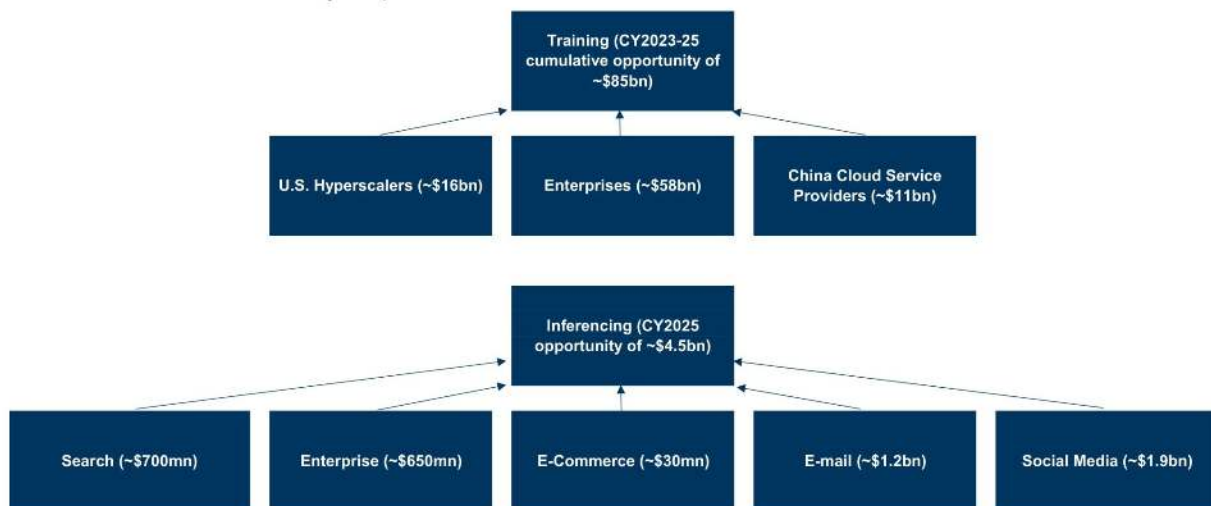
"My perspective is that the cloud is still in the early stages. If you think about 90% plus of the global IT spend being on premises, where I think that equation is going to flip in 10 years, I think cloud is early, so if you, with that lens on, I still think we're very early in generative AI."

– Amazon | [Q3 2023 Earnings Call](#)

That's only one area underscoring demand in genAI across the broader technology ecosystem.

Sizing Up the GenAI Opportunity

Exhibit 4: Framework of our bottom-up analysis



Source: [Goldman Sachs, July 2023](#)

GenAI Document Trend at AlphaSense



Source: AlphaSense

Let's dig into the issues around genAI by exploring the AlphaSense platform, which has generated over 65,000 total documents on the topics around genAI since ChatGPT broke out in December 2022.

Infrastructure

Infrastructure to date has been a hurdle to many of the ambitions of cloud service providers and software companies. Large Language Models (LLMs) have been dependent on graphic processor units (GPUs) to power their data center productivity. This is because LLM's feed on data. Nvidia is currently the market leader in the GPU category, and is expected to keep that lead for the foreseeable future - giving it tremendous pricing power:

“You've heard of people making the snarky remarks of the inverse Moore's Law or a reverse Moore's Law where with Nvidia GPUs, you're finally starting to get less compute for your money as the years progress because the GPUs are getting so much more expensive.”

– Former Resident, Microsoft | [Expert Transcript](#)

Cost and availability have become stringent variables for build plan execution. This has led hyperscalers to seek alternative means to building their respective architectures. Google has led efforts to bring data center infrastructure partially in-house, with its homegrown tensor processing units (TPUs). Utilizing TPUs allows for Google to access required chips and bring costs down effectively, but could potentially expose them should technologies shift.

Longer term, it may be more efficient to outsource chip production. For now, it appears that Google, Meta, Microsoft and Amazon are all trying to reduce their reliance on Nvidia GPUs. The

market overall is looking to see Intel and AMD to compete with Nvidia, however, our experts view this as a ways out:

“The answer is yes, there's a possibility for Intel. If you can see what AMD is doing with its MI300, they've got the assets in GPU IP, in CPU IP, and even ASIC IP to go play in that market. The question is, how efficiently and effectively can you deploy it to capture some of those workloads? If you look at industry dynamics, for sure, the industry is looking for alternatives to Nvidia because, obviously, the classic Porter's Five Forces model suggests that the more players, the better bargaining power and cost will come down, etc.

The challenge is you've got the software moat that NVIDIA seems to have built with CUDA now, while some of that is also being eaten away by PyTorch and other deployments. The defendant's reliance on CUDA will likely devolve over time, whether it's coming from AMD's ROCm or other frameworks. The reality is if you want to do AI today, the fastest way to get there is Nvidia, which is why you've seen them signaling their business prospects, etc.”

– **Former Vice President, Intel** | [Expert Transcript](#)

Although companies are reluctant to rely solely on Nvidia systems, the company is also beginning to offer turnkey AI infrastructure solutions via its Nvidia DGX Platform. This has led Morgan Stanley to [introduce a potential TAM expansion](#) scenario for cloud service providers, and UBS sized it as a [\\$10bn opportunity for Nvidia](#) and ~\$1bn/yr in revenue for its base service alone.

Our expert is firmly in the camp that Nvidia has both the market lead and pricing power in the near term, but suggests we may see a meaningfully different landscape over the next five to ten years, with hyperscalers finding success with in-house silicon, AMD catching up, and start ups innovating to a viable point. This expert believes that over 5-10 years chip availability grows, but those who want to scale to capitalize on trends have no choice right now:

“If you are going to have to wait for the market leader [Nvidia] to get your parts, or you go with a start-up that's going to offer you an attractive price and customizations to prove it out because you want to [follow upfront] prove it out, proof of concept. You go into pilot mode, then you go to high-volume production.”

– **Former Unit Head, Global Foundries** | [Expert Transcript](#)

In the meantime, Nvidia will enjoy pricing power and command over which customers lead the way from an infrastructure perspective.

Foundational Models

Beyond the GPU dilemma, LLM's thrive on high-quality data for both training and inference. These models learn patterns and structures from extensive datasets and generate new content

by extrapolating from the learned information. There are debates around the size of parameters, pros and cons of open versus closed, and the ability to innovate, however quality of the data is of paramount importance.

There are two types of sourced LLMs, open source and closed, which are defined by the availability of the underlying code being publicly available. The advantage of an open source model is firstly transparency. An open source model promotes transparency, allowing developers, researchers and communities access to the underlying code.

This transparency fosters innovation and collaboration. With a community of developers scrutinizing the code, it also fosters discovery of security issues. On the flip side, closed source models offer monetization through licensing through more stable software offerings. The market may head to a point that one size does not fit all applications:

“To me, it's a lot more interesting because they're actually open-sourcing them more frequently and inviting more developers and machine learning scientists' community conversation into it. That seems to be probably one of the more fruitful areas as well, for sure. The open-source models, Mistral and LLaMA, are really promising because they're the most economical, most configurable, most open book model that we've had since GPT-J, about a year and a half, two years ago.”

– **Engineer, Latitude** | [Expert Transcript](#)

A current competitor of OpenAI believes that consumers will encounter challenges with large open-source models:

“I think it'll be hard for consumers to serve increasingly large open-source models but I imagine that they'll keep creeping ahead and have their performance approach the state-of-the-art proprietary model. I think the goalpost keeps shifting through. A year ago, if Llama 70b was released, it probably would have been the state-of-the-art model on the market but because GPT-4 is out, that's not the case anymore.”

– **Unit Lead, Cohere** | [Expert Transcript](#)

While the jury is out on the ultimate winner, the field is beginning to evolve in a way that categorizes LLMs based on task and context, potentially helping to define the advantage of using one over the other depending on the specific use case. Over time, the market will move to multi-modal LLMs, to support the combination of text with other types of information, such as images, videos, audio and sensory data.

Enterprise GenAI

One of the biggest opportunities in genAI is in enterprise applications. Morgan Stanley [sizes up the enterprise software TAM](#) to be in the range of ~\$205bn in the US and \$820bn globally, based upon estimates of 25% of the US labor market potentially impacted today, and rising to 44% within three years. ServiceNow commented in their latest earnings report:

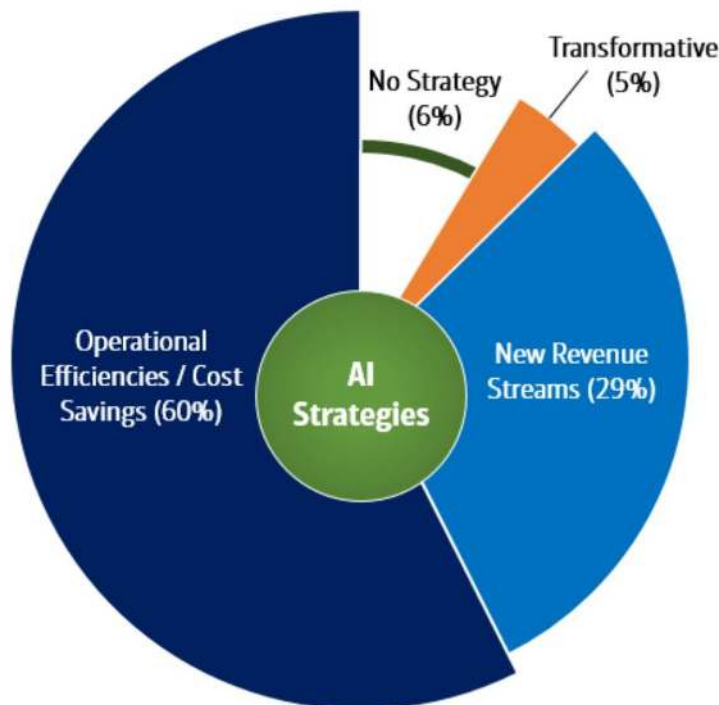
“AI has strengthened the market dynamics for enterprise software. This is a unique, highly differentiated company that is rewriting the benchmarks to be best-in-class in the SaaS industry. Looking beyond the quarterly results, while the world's challenges are sobering, the digitization imperative is stronger than ever.

Gartner forecasts that \$3 trillion will be spent on AI and gen AI between 2023 and 2027. Gen AI represents 36% of AI spending overall. We believe every dollar of global GDP will be impacted by AI over the next several years. This isn't a hype cycle. It is a generational movement.”

– Service Now | [Q3 2023 Earnings Call](#)

Where is Enterprise in Defining Their Future AI Strategy?

■ Only 6% of covered companies do not have an AI strategy



Source: BofA Global Research. Results are based on survey responses.

BofA GLOBAL RESEARCH

Source: [Bank of America, October 2023](#)

Despite the excitement over future applications, near term deployments are another story. Arete Research says they see [investment in this early phase tied to headcount](#) as enterprises prepare the ground (data, security, etc) before more significant direct spend on AI services.

CIO conversations suggest spending remains in the \$100Ks run rate (on direct services) into CY24 even as headcount costs mean total investment is running in the millions of dollars. Slower capex investments may be partially explained by lack of hardware availability, but

enterprise use cases will require a level of complexity and security requirements that will command gradual scaling.

There are a number of considerations around privacy and security that need to be addressed for wide scale adoption:

“There's also some danger in there about the privacy that can be involved. I know that there is actually some sector of research to understand how to protect people's privacy in the large language model training. I agree that if you can scrape as much different data as possible from different domains, you can improve the performance or you can improve the data coverage inside the large language model. I personally am against using whatever data that is out there, regardless of what they entail because there could be some intrusion of privacy in the processing here.”

– Research Lead, Iris.AI | [Expert Transcript](#)

Successful early adopters are driving use cases by automating and simplifying processes. Use cases such as summarization, simplification and code generation are easier. Smaller LLMs have the capacity and knowledge to accomplish specific tasks while operating within reasonable operational constraints. Legal firms, tech solutions, financial services platforms, sales teams with access to Gong and Salesforce Einstein will also benefit from AI deployments in the shorter term.

Productivity Scaling Across Enterprise

While Still Early, Productivity Uplifts Are Becoming Clearer...

Customer Support	National Bureau of Economic Research	<ul style="list-style-type: none"> Access to AI-based conversational assistants increases productivity, as measured by issues resolved per hour, by 14% on average, with the greatest impact on novice and low-skilled workers, and minimal impact on experienced and highly skilled workers.
AI-Assisted Coding	GitHub Copilot	<ul style="list-style-type: none"> "We're now seeing that the developers using GitHub Copilot are 55% more productive on tasks." – Scott Guthrie, Microsoft EVP of Cloud & AI
	AMZN Code Whisperer	<ul style="list-style-type: none"> "Internal test showed 57% faster task completion and 27% higher likelihood of success." – Adam Selipski, AWS CEO
	PayPal	<ul style="list-style-type: none"> "AI is already making the coding development team 30% more productive and expects similar improvement across front office and back office." – Dan Schulman, PayPal CEO
	Airbnb	<ul style="list-style-type: none"> "I think our employees could easily be, especially our developers, 30% more productive in the short to medium term and this will allow significantly greater throughput through tools like GitHub's Copilot." – Brian Chesky, Airbnb CEO
Writing Productivity	MIT Study	<ul style="list-style-type: none"> Study of 453 college-educated professions show that ChatGPT raises the average productivity of those who used the technology and increased job satisfaction. Average time taken decreased by 40% and output quality rose by 18%.
General Productivity	Adobe	<ul style="list-style-type: none"> Early customer feedback suggests 20% productivity gains from Firefly, saving ~8hrs per week and freeing up time for more high value-add work.
	HBS Study	<ul style="list-style-type: none"> Study of 758 BGC consultants highlights GPT4 productivity gains... consultants using AI completed 12.2% more tasks on average, and 25.1% more quickly, and produced significantly higher quality results (more than 40% higher quality compared to a control group).
	Unity	<ul style="list-style-type: none"> "Now, something that might have taken a group of artists 3 months can now take a server run about 10 minutes. And the interesting point about the 3 months to 10 minutes statistic is that's not just a productivity uplift per se... originally most people would have looked at the challenge to create new 3D content taking 3 months and go, "we can't afford that" and so they didn't do it. Generative AI will allow all those who never started to get there in 10 minutes instead; market expansion." – Marc Whitten, Unity Create President

October 2023

Source: [Morgan Stanley](#)

Certainly, some large companies have begun to find their paths to monetization. Bernstein believes that Adobe is one of the cleanest examples, showing both productivity gains and clear ROI:

“Adobe has shown that its generative AI capabilities are differentiated and provide some serious ROI for the end user — the two factors that create a winner in the AI race. We believe Adobe will be the first of the SaaS vendors to generate meaningful revenue and profit from AI starting next year.”

– **Sanford Bernstein** | [Report](#)

At AlphaSense, we have also taken deep dives into startups around the genAI landscape which may emerge as future leaders:

“I would say probably I think 6sense is doing a lot. I would say that the other competitors will probably need to step it up in order for them to achieve where they're at today. I think they would look at all the different aspects of AI.

AI that's powered by the platform at 6sense is more accurate and actionable. 6sense probably focuses more on revenue than Demandbase does. I would say that the difference in integration platforms is probably more geared towards 6sense where they have marketing and field integration.”

– **Former Senior Engineer, LeadSpace** | [Expert Transcript](#)

Consumer and Internet GenAI

While the headlines have been vast since news of ChatGPT broke in December 2022, we remain in early innings for consumer and internet applications featuring genAI. As Barclays notes:

“Early data around AI services like ChatGPT and others suggests a few major consumer use cases: 1) work productivity, 2) information retrieval, and 3) services that resemble companionship. Character.ai is seeing two hours of daily engagement with some of its companionship applications, and ChatGPT is seeing significant time spent in the same kind of context, these are massive per-user engagement figures along the lines of what Tik Tok was seeing in its early adoption stage. So the initial engagement appears to be strong, but it is not yet clear what the 6-month and 12-month retention rates look like.”

– **Barclays** | [Report](#)

Despite claims that ChatGPT would immediately unseat Google Search, market share shifts around search have not yet surfaced.

Global Search Market Share

All Search	Jul-22	Aug-22	Sep-22	Oct-22	Nov-22	Dec-22	Jan-23	Feb-23	Mar-23	Apr-23	May-23	Jun-23	Jul-23
Google	91.5%	92.0%	92.4%	92.3%	92.2%	92.6%	92.9%	93.4%	93.2%	92.8%	93.1%	92.6%	92.1%
Bing	3.3%	3.3%	3.5%	3.6%	3.4%	3.0%	3.0%	2.8%	2.9%	2.8%	2.8%	2.8%	3.0%
Yahoo	1.3%	1.3%	1.3%	1.3%	1.2%	1.2%	1.2%	1.1%	1.1%	1.1%	1.1%	1.1%	1.2%
Other	3.9%	3.3%	2.8%	2.8%	3.2%	3.2%	2.9%	2.7%	2.8%	3.3%	3.0%	3.5%	3.8%

Source: [Bank of America](#)

Advancements in technology have typically been seen via increases in consumer engagement and adoption (mobile internet, streaming as examples). As consumer driven genAI becomes more mainstream, it is unclear if the value chain of consumer internet applications will shift meaningfully.

ChatGPT and Google have gained mindshare with genAI search, and deliver a new more detailed experience via chatbots. It's equally unclear if we'll see a new type of search emerge: Apple and Google have yet to unveil mobile offerings incorporating genAI, and Samsung just introduced its Samsung Gauss model, designed for AI applications on devices.

Our expert commented on the topic:

“Models like ChatGPT where they can be specific on specific topics. I'm really interested in how that will come together and how someone like Apple will bring that into the Siri ecosystem or how they'll leverage those individual models to make certain workflows more efficient and build moats in that regard.”

– Former Manager, Apple | [Expert Transcript](#)

Identifying new services with an existing distribution mechanism might be the winning combination for meaningful adoption of consumer internet products. In the near to medium term, we believe there will be a focus on iterations of ChatGPT, Meta's AI Agent Launch, Google's Bard, and what comes from Apple's yet to be announced consumer facing AI strategy.

Looking Ahead

A key to success will be market differentiation and defensibility of elements of the genAI tech stack, which will evolve as the space matures. The LLM ecosystem for text generation is likely to become more commoditized over time, for instance, and as it does, costs will likely come down allowing for demand elasticity to take shape in a more linear way.

At some point soon, we will begin to see a shift towards genAI at the edge – which will enable data to be processed where the data is collected (a mobile phone for example) versus in the cloud. This is an area AlphaSense is paying close attention to, as the opportunity to compute, create, and establish workflows with AI on a mobile device is another breakthrough that doesn't seem all that far away:

“We have seen the increasing popularity of generative AI since late-2022, yet only performed by cloud computing so far. During recent earnings calls, many key semis firms (TSMC, MTK, QCOM, Samsung, etc.) talked about the next wave of development.”

– **Bank of America** | [Report](#)

We recently had a semiconductor expert weigh in on the progress being made at the edge:

“What they [Qualcomm] have mostly been focused on is their AI-focused Snapdragon processors. Again, this is used in mobile devices. This hardware is really designed for that edge computing. We haven't talked about it yet, but IoT applications, the Internet of things. This is an interconnected world.

That also is on the edge, so you're doing this computational thing that's there, you're doing inferences, and then you're connecting to these other devices for ideally more simplicity for human life. These guys are also going to be very relevant because they are going to be the leader in these mobile applications.”

– **Former Director, Veeco** | [Expert Transcript](#)

Hype Versus Reality

The hype around AI is real, but it will take time for the reality to live up to elevated expectations. In the next couple of years, roadmaps depend on infrastructure availability, complexity and cost, and innovation as the sector matures. Andreessen Horowitz discusses the hype versus reality argument, saying the generative AI boom has been accompanied by [real gains in real markets](#), and real traction from real companies. Despite the clear use cases forming, and infrastructure strains easing into next year, Q2 2023 earnings season has proven that the conversation is has turned towards future profitability:

“What I predict is that the fundamental technology around generative AI is going to transform meaningfully how people use each of the different apps that we build. I think for the Feed apps, over time, more of the content that people consume is going to be either generated or edited by AI.

Some of it will be creators who will now have all these tools to make content more easily and more fun. And I think over time, maybe we'll even get to the point where we can just generate content directly for people based on what they might be interested in.”

– **Meta Platforms** | [Q3 2023 Earnings Call](#)

AI chatbots, like the ones from Meta and Snap, or AI infused into Google Search, offer much sleeker advancements in technology. However, if profitability doesn't meaningfully set in, the offerings may become less enticing as time goes by. On the flip side, AI is showing more progress on productivity and efficiency for enterprise applications. Customer service, creative

advertising and e-commerce have shown promise in earnings already. Regulatory risks on both genAI usage, and AI chips and fears around privacy will likely persist in the short term.

At AlphaSense, we'll continue to leverage our platform and first party expert insights to monitor key trends related to genAI as they emerge.

Ready to dive deep into the expert transcript library? [Start your free trial of AlphaSense today.](#)

About AlphaSense

AlphaSense is a market intelligence and search platform used by the world's leading companies and financial institutions. Since 2011, our AI-based technology has helped professionals make smarter business decisions by delivering insights from an extensive universe of public and private content—including company filings, event transcripts, expert calls, news, trade journals, and equity research. Our platform is trusted by over 3,500 enterprise customers, including a majority of the S&P 500.

Headquartered in New York City, AlphaSense employs over 1,000 people across offices in the U.S., U.K., Germany, Finland, and India.

[Learn More at Alpha-Sense.com](https://www.alpha-sense.com)

