AlphaSense

Generative AI in the Enterprise Software Sector: A Look At Adoption in 2024

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

February 2024

Generative AI In Enterprise Software

Executive Summary

2023 was the pinnacle of generative artificial intelligence (genAI) hype, with one company after another announcing breakthroughs and product launches. However, as we begin 2024, the reality of existing hurdles on the path to mass adoption of AI in enterprise software begins to set in. This report breaks down the current headwinds and tailwinds in the enterprise software subsector, provides perspective on the pace of maturation, and highlights potential unforeseen issues that might arise this year in 2024.

While recognizing the undeniable long-term secular impact of the AI wave, winners may emerge at a slower cadence than market exuberance currently reflects. At AlphaSense, we see early genAI in enterprise software success stories taking shape in two routes: companies benefiting directly through the application of genAI tasks and companies leveraging genAI to enhance an internal product offering.

In 2024, we'll look for signals that AI features are becoming more tightly integrated into software solutions and for potential vendor consolidation to achieve scale and innovation quickly. It's also likely that genAI-related products will displace current SaaS winners, a trend that has not yet surfaced but is worth monitoring. We will explore the current and developing backdrop utilizing the AlphaSense expert transcript library to better understand the maturation of genAI in the Enterprise Software space.

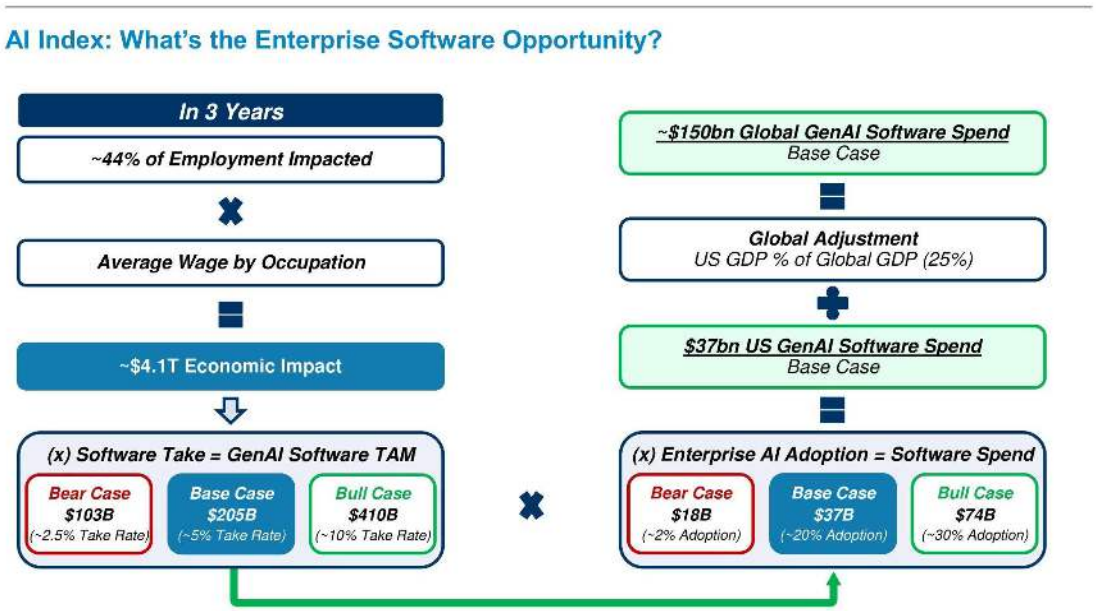
Generative AI Adoption in Enterprise Software

GenAI promises to be a disruptive force across several sectors over the coming years. Although specifically within the software sector, capital expenditure (capex) considerations may translate to revenue expanding at a separate pace than recent stock price appreciation for enterprise software companies. The sector is immature, builds are complex, and winners and losers will be defined by companies dedicated to solving open questions:

- Which opportunities offer the highest value? What is the return on investment (ROI), and in what timeframe?
- What are the tradeoffs between cost, benefits, and risks, and how can companies ensure that genAI is responsibly deployed while maximizing innovative potential?
- What preparation and infrastructure spending is required to ensure safety and compliance? Which partners can safely accelerate go-to-market timelines?

New technology adoption in enterprise requires meaningful investment in hardware and software frameworks that take time to plan, create, and implement. For genAI, obstacles include technological integration with existing systems, identification and implementation of security protocols, and obtaining and utilizing high-quality data. Once these obstacles are resolved, governance, intellectual property issues, and the potential for regulation remain headwinds. Beyond those considerations, pricing models and customer acceptance remain in the early stages.

Balancing Demand and Market Expectations Against the Reality of Execution



Source: [Morgan Stanley, The Morgan Stanley AI Guidebook: Fourth Edition, 2024]

By all accounts, demand to deploy generative AI is present and accelerating. According to IDC, the global AI software revenue opportunity is projected to reach $944 billion by 2027, up 153% since 2022. AI Core software revenue is expected to reach $251 billion by 2027, expanding at a compound annual growth rate (CAGR) of 31.4%, and AI Feature software revenue is anticipated to reach $693 billion by 2027 at a CAGR of 17.6%.

After a year of non-stop genAI buzz, it seems counterintuitive to speak about demand trends. In reality, though, competition for budget dollars in 2024 will likely continue and be highly scrutinized by CIOs. Within their Q4'23 survey, Morgan Stanley notes:

> "Our CIOs survey suggests a broad interest in generative AI technologies, with 66% of CIOs reporting initial investments, the nature of these investments tend towards Evaluations (39%) and Pilot Projects (22%). The ramp in pushing these initiatives into production appears more weighted towards the back of 2024, meaning revenue ramps for many subscription-based application vendors likely do not become significant until CY2025."
>
> **– Morgan Stanley | 2023 Report**

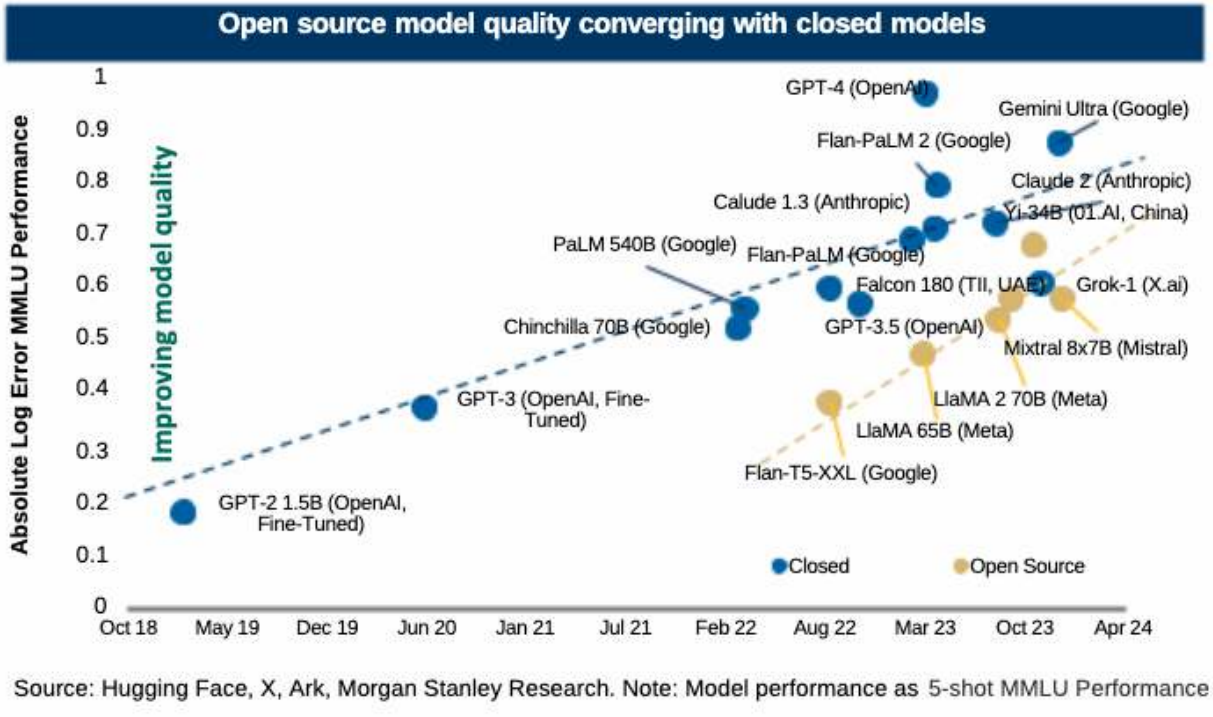
While enterprise software has big plans for genAI applications, there are several outstanding questions that management teams must examine before large-scale deployments can begin.

GenAI Infrastructure: Build, Buy, or Partner?

One of the biggest impediments to AI adoption is whether companies buy, build, or partner with a provider. Interest in genAI has driven a massive investment cycle in GPUs, servers, and LLM creation—all of which are incredibly expensive. The costs include training and inferencing an LLM and establishing infrastructure to catalyze the application, store the data, and produce the end product. An example of how to think about these costs comes from a recent D.A. Davidson report on Microsoft:

> "Ultimately, companies such as Microsoft, who are spending large amounts of Capex on hardware to train or inference large language models will require a reasonable return on investment to justify the Capex. We show that for every $1 of GPU spend, Microsoft will have to generate $0.13 of incremental recurring revenue in order to justify their investment."
>
> **– D.A. Davidson | 2024 Report**

Growing Number of Models Coming to Market



Source: Hugging Face, X, Ark, Morgan Stanley Research. Note: Model performance as 5-shot MMLU Performance

Source: [Morgan Stanley, The Morgan Stanley AI Guidebook: Fourth Edition, 2024]

While early adopters' genAI rollouts have favored transformer technology in the form of LLMs, particularly OpenAI's GPT-4, considerable competition among LLMs has emerged in recent months. The introduction of smaller open-sourced models has debunked OpenAI's assertion that bigger is better for model performance. Customers deploying smaller models can optimize the balance between model complexity and computational efficiency. They require less memory for storage and operation, directly improving the training and inferencing costs.

> "Many of the almost half a million models listed on the Hugging Face platform are based on Meta's open-source Llama, which leaked within days of its launch in late February, or its successor Llama 2, launched in July in lightweight 7 billion-,13 billion- and 70 billion-parameter versions. New techniques such as Low-Rank Adaptation (LoRA) are also making it cheap and easy to churn out new fine-tuned versions of large-scale pre-trained models."
>
> **– Deutsche Bank | 2024 Report**

Introducing open-source models has begun to diminish LLM's commanding lead, with experts believing that lower price points are on the horizon. Enterprise software companies such as Oracle and Salesforce have moved to an LLM-agnostic model, signaling a potential shift in strategy in the future.

> "I think an early adopter phase [of LLM's] is coming to an end. That push is going to take all of the industry leaders into another direction, whether they go in the merger and acquisition or they start investing heavily into their own operation. Of course, they're not going to die out, but they're going to face huge competition."
>
> **– Former Manager, Cube | Expert Transcript**

System architects are beginning to layer innovations into the AI-tech stack, which could reduce complexities, enabling enterprise applications to run more efficiently. One such innovation is the deployment of a vector database with an LLM in the tech stack, enabling a developer to create vector search-powered experiences:

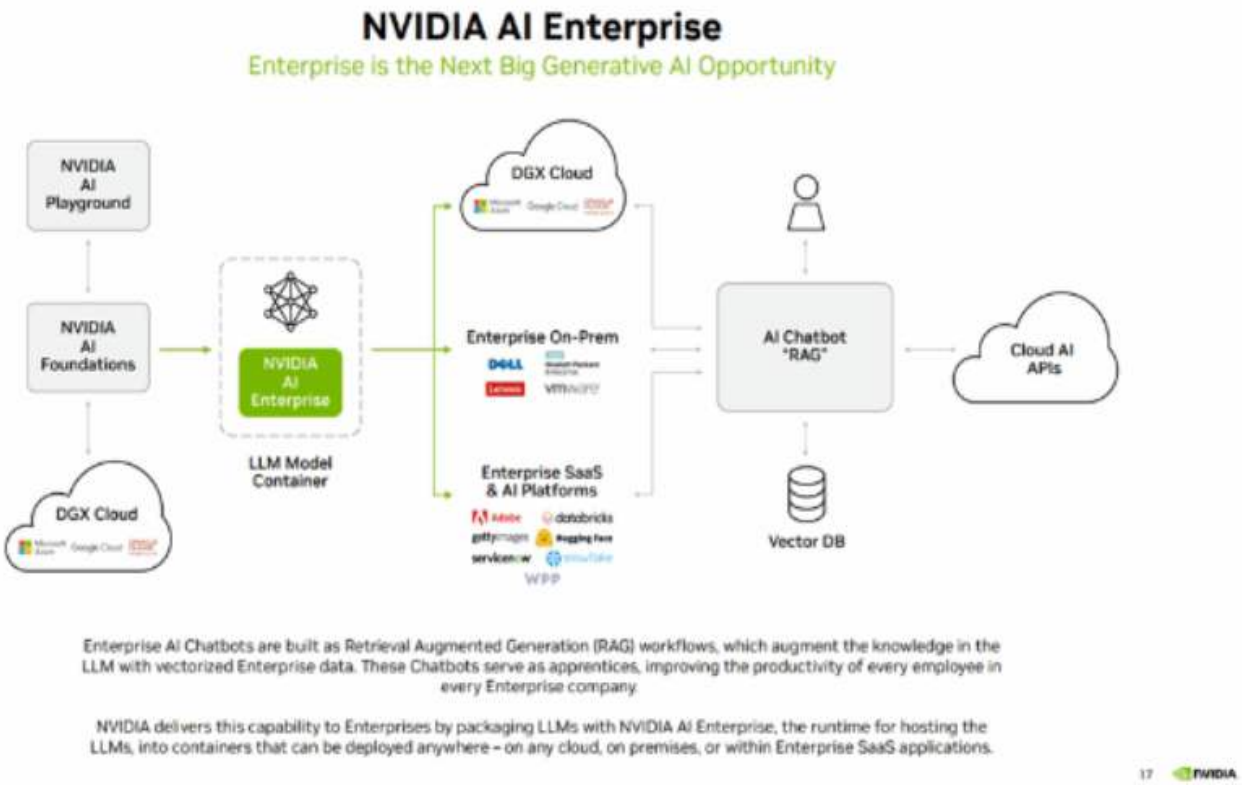
> "It's really important that this kind of engine, like Elasticsearch with a strong technology, to process distributed data and load amounts of data into the cluster have the features to understand and ingest natural process language with some of the additional features, to convert all of these different contexts with the potential of AI and put into the search engine is really powerful because you can boost your use cases. For example, to incorporate natural process language coming from any other different sources that in the past were not possible."
>
> **– Customer of Elastic | Expert Transcript**

Alternatives to Building

Given the high costs and limited availability of GPUs, servers, memory, and other hardware, the "build" option is limited to big tech players who have the purchasing power, engineers, and roadmaps to secure profitable market growth. As an alternative to building or buying, Nvidia has introduced AI Enterprise, a full-stack platform designed to optimize AI adoption for enterprise-based customers.

> "NVIDIA currently notes that AI Enterprise offers 100+ frameworks, pre-trained models, and development tools to accelerate data sciences and streamline the development + deployment of production AI (genAI, computer vision, and speech AI). As noted above, NVIDIA's AI Enterprise is supported via cloud partners - AWS EC2 instances, Google Cloud Platform (GCP), Microsoft Azure, and Oracle Cloud Infrastructure (OCI). Additionally, AI Enterprise is supported on +400 NVIDIA-certified servers and workstations (e.g., Dell, HP Enterprise, Supermicro, Lenovo, Inspur, Cisco, etc.)"
>
> **– Wells Fargo | 2023 Report**

Nvidia Moving Up The Tech Stack By Offering Enterprise-Ready Solutions



Source: [Nvidia, Nvidia Investor Presentation October 2023]

> "[AI has] real solutions, but people also see what are the exit opportunities for start-ups. [There are unknowns]. First, we don't have that long-term ROI, especially with compute costs. The second one, the things that generally keep me up at night, is the trust and safety parts. The third one is, what is this doing for the environment? I've read studies where running a lot of these models is pretty expensive, but very costly. It requires a lot because these are data centers."
>
> **– Former Product Leader, Google | Expert Transcript**

While a clear winning infrastructure strategy has yet to be solidified, it is evident from our research that enterprise customers will be looking at new ways of increasing ROI while innovating business models with generative AI applications. Options include enhancement of LLM performance, cost optimization, deploying smaller models, and potentially new architectures displacing transistor-based technology altogether. Currently, larger enterprise software companies such as Salesforce and Oracle are pivoting towards partnerships to enhance capabilities and go-to-market timelines.

Data Security and Governance

As foundational models mature, a significant concern remains in the existence of hallucinations. Hallucinations in LLMs refer to the generation of inaccurate or nonsensical information. This problem can be mitigated by training the model on domain-specific knowledge and adding additional guardrails. The reliability of generated data is crucial to enterprise software customers as adoption increases.

Our experts have recently detected that new attack points come with expanded enterprise genAI usage, creating an increased security vulnerability. Additionally, enterprises utilizing LLMs face competitive hiring issues, and software engineers may not be equipped to identify insecure code. Our experts believe that this insecurity also poses a tailwind for cybersecurity vendors. Given the risks associated with enterprise genAI deployments, our experts suggest that functions with less sensitive business information, like Sales and Marketing, are more likely to see early adoption of AI solutions.

> "After the whole drama that happened with OpenAI and the fear-mongering around AGI and things of that nature, enterprises are going to take a little bit of extra time to go full on. Right now, what I'm seeing is that, R&D departments or let's say some non-critical functions like marketing or sales are seeing a lot of these genAI use cases, but the finance and all of those functions are going to be slower to pick up the pace on that."
>
> **– Industry Expert | Expert Transcript**

Data and governance are almost sure to get more attention as genAI applications begin scaling more meaningfully, especially if damaging incidents arise over time.

When Will We See the Benefits of AI Applications?

Monetization

While revenue for most companies is unlikely material for 2024, early identification of spending patterns will be top of mind for investors. Finding areas of value in features or products in early production may be leading indicators of successful monetization strategies. Direct monetization will likely come through strategies similar to Microsoft's, where customers pay directly for usage. Although it's unclear if Microsoft's AI offerings will remain profitable with general availability.

Beyond direct payments, features gated by a paywall or premium tier could serve as upgrade catalysts, while other companies may charge for using genAI-powered features. Additionally, indirect monetization opportunities, such as cloud providers, drive usage and/or retention on the platform. AI has the opportunity to create tremendous value and efficiency for enterprises, but

it's still an open question what commercial models providers will use to translate that productivity gain into revenue.

> "70% of the Copilot users from our early access, were more productive, and 68% of them said it improves the quality of their work.. Fortune 100 customers have their employees say 68%. 29% said it helped them do multiple tasks much faster, searching, writing, and summarizing."
>
> **– Former Director, Microsoft | Expert Transcript**

In the near to medium term, we see two areas in which enterprise software companies will monetize on genAI: companies that benefit directly from workloads and companies that leverage genAI to enhance a current offering or a combination of the two. Microsoft Copilots and GitHub Copilot are clear examples of the first way to monetize. However, it's not clear if their ROI is sufficient. When asked if the company would need to raise prices to be profitable at scale, our expert offered:

> "It's expensive on a training part as well as an inference and maintenance part. What was causing most of that was because the chips were so expensive. Early adopters get access to it, and there's a lot of testing taking place and feedback changes and refinements. If general availability was made to every single enterprise customer and small business and everyone who has Office 365 subscriptions, you are right [on compute usage]."
>
> **– Former Director | Expert Transcript**

Emerging Enterprise Software GenAI Trends

Prospects for Open-Source and Closed-Source Models

Are open-source or closed-source models more appropriate for the enterprise market? Smaller models introduced by the likes of Meta Platforms show chops, particularly for enterprises, where cost optimization, productivity, and security may be better managed. This could lead to enterprises shying away from Open's app store, as data security concerns persist.

While GPT-4 and similar models trained on public data owned the headlines in 2023, smaller customized open-source models trained on private data may cross a threshold whereby they are on par with GPT-4 performance in 2024. As such, there could be a significant shift towards the change in architecture for enterprise-based applications.

> "That's the fork in the road right now, whether you take Llama or Mistral or one of the many other smaller models that are easily available on-prem or put your trust in one of these big boys and use their enterprise solution. There's still not enough trust in these bigger companies to do better data stewardship."
>
> **– Industry Expert | Expert Transcript**

Foundational models represent an essential component of cost, and some organizations are already bucking the "bigger is better" trend that OpenAI introduced to the market with ChatGPT. A new set of complexities will likely be introduced as we move into a realm where LLMs evolve into multi-modal capabilities.

New Techniques: Retrieval-Augmented Generation and Vector Databases

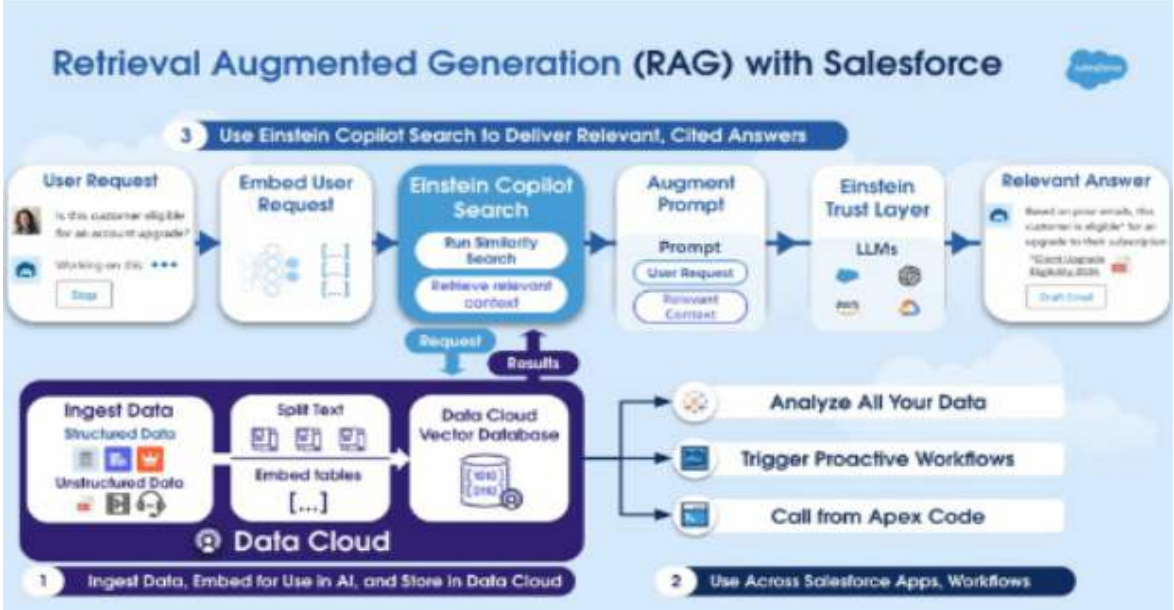
LLMs approximate patterns in text rather than duplicate knowledge. Developers can utilize retrieval-augmented generation (RAG) to combine search with LLMs to generate more accurate answers. RAG reduces hallucinations and enables verification through citations embedded in the model's answers. Our expert believes that Databricks is following this path:

> "Databricks offers the path wherein you put in a query, the query goes into a database, the database matches with specific context, that context gets added on to a query, and then goes into an LLM. Sometimes [Lang] implementation is very unique and available only on Databricks, and that makes it a big advantage for deployment, especially on productivity-based applications, which is heavily reliant on a lot of documents."
>
> **– Senior Director | Expert Transcript**

Similarly, an expert discusses the growing commoditization of LLMs but the potential differentiation of RAG and Langchain:

> "A lot of folks in AI, what they're doing is, and I mentioned this before, is making these different AI systems like OpenAI or Google's system or Facebook's system, they're making a commodity to choose one of these and making it interchangeable. Being able to choose different models at will and having the LangChain or the RAG supplementary training to still work with these commodity LLMs is pretty powerful."
>
> **– Industry Expert | Expert Transcript**

Salesforce Deploying RAG and Vector Database for Einstein Copilot



Source: Salesforce, Solid Risk/Reward Makes CRM Our Top Pick – Feedback from Our Upgrade

> "Salesforce also recently announced an important update to Data Cloud and its Einstein 1 Platform, adding the Data Cloud Vector Database and Einstein Copilot Search; both will be in pilot next month, with Einstein Copilot generally available by then. This marks a critical development in Salesforce's ability to bring quality genAI capabilities to customers, as the addition of a vector database within Data Cloud allows for new unstructured data pipelines to be tapped, utilizing semantic search and Einstein
>
> Copilot prompts to perform Retrieval Augmented Generation (RAG) activity directly within the Salesforce ecosystem, thus delivering higher-fidelity LLM-generated outputs for users of the Einstein 1 Platform and Customer 360 application suite."
>
> **– Morgan Stanley | 2024 Report**

As AI architectures evolve, other infrastructure elements potentially change specific to the enterprise. Companies like Vast Data and Databricks are redefining hardware within enterprise to handle AI workloads innovatively. Our expert thinks Vast Data's innovation is something to watch in the storage arena:

> "What Jeff and Renen have signaled out to the market with the VAST DataEngine, the VAST DataBase, the VAST DataSpace, I think it's revolutionary. The concept of building thinking machines and putting all the thinking on the storage layer so that AI could leverage it, from a vision perspective, sounds like a standalone company to me."
>
> **– Former Executive | Expert Transcript**

Addressing New Cybersecurity Threats

The dots between cybersecurity and enterprise genAI applications have yet to be fully drawn, but the risks are clear and on the minds of CIOs looking to innovate securely.

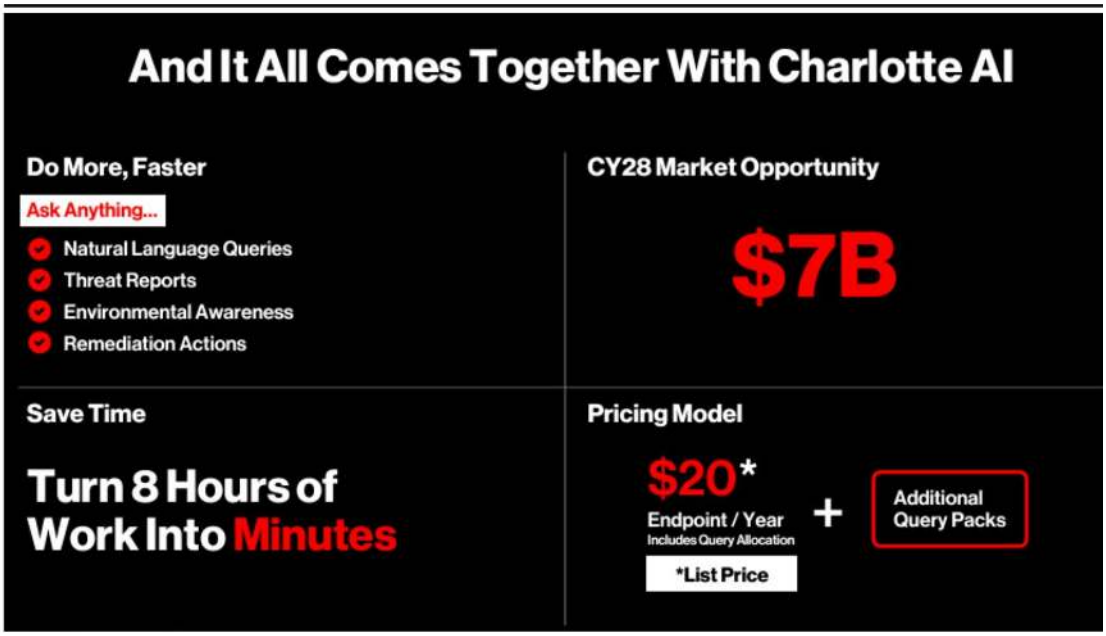
> "I would say 99% of enterprises have not yet truly integrated generative AI into their IT and business logic. Many more sophisticated companies have already declared a moratorium on generative AI use cases until they go through these kinds of security reviews because of the concern that once you attach these generative AI models to your corporate data."
>
> **–Director, Intel | Expert Transcript**

AI model security, zero-trust identity, and cloud security are all top of mind as AI applications expand. UBS expects future plans to include specialized tools to improve the security and rigor of proprietary AI app models. Our expert also highlights the increased risk in less trained staff, as advancements in genAI may be ahead of technical staff in some instances:

> "You've got less trained software engineers using code from AI systems, trained on insecure examples, that's going to create a significant amount of more risk. You're generating more insecure code from people with less training in the systems. That can cause a much greater risk for organizations that leverage AI to build their technologies and products."
>
> **– Former VP Engineering, Rapid7 | Expert Transcript**

At AlphaSense, we're monitoring potential cybersecurity intersection genAI enterprise applications in 2024, as leading platforms will likely be required to increase innovation. Category leaders, such as Crowdstrike, have defined the market opportunity, while other newcomers are on the horizon.

Crowdstrike's Charlotte AI is a $7 Billion Opportunity

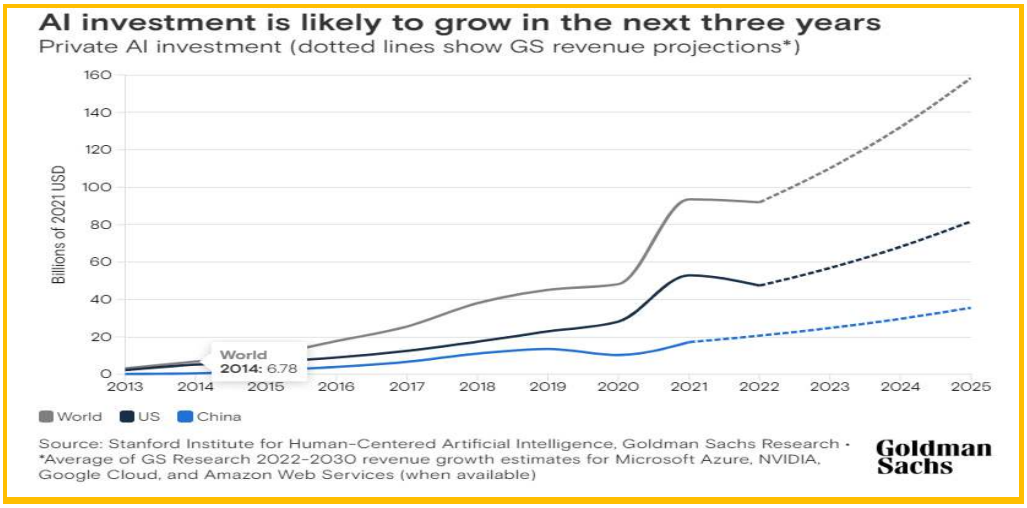


Source: Crowdstrike, Investor Briefing 2023

Private Investment Fueling Innovation

According to Goldman Sachs Research, firms are prioritizing private AI investment globally; the market is forecasted to increase to more than $160 billion by 2025. The generative AI ecosystem will empower incumbent enterprises across the spectrum. Last year, corporate venture arms accounted for 90% of capital raised for genAI privates. Microsoft, Nvidia, Amazon, and Google have all heavily invested in the genAI ecosystem, as funding into foundational model vendors has fostered cloud computing revenue as those vendors rent compute space.

Goldman Sachs Predicts Continued Private AI Investment in Next Three Years



Source: Goldman Sachs, How to Unlock an AI-drive M&A Supercycle

Source: Goldman Sachs, *How to unlock an AI-drive M&A supercycle* Within AI private funding, foundational models received the lion's share of funding in 2023. Newer areas of the tech stack for AI applications are garnering interest beyond foundational models, including in cloud security. Experts suggest that this may be an area where traditional cyber vendors may be heading towards:

> "Palo Alto and CrowdStrike, two players are in inorganic or acquisition mode. There is a high probability that when a niche player becomes successful, they will go ahead and pick that niche player and integrate them in their solution suit and offer that to the customers. Today, it is difficult to say who will have that advantage because they are not investing in-house on AI and ML solutions."
>
> **– Former VP, Qualys | Expert Transcript**

AI models present new risk areas, like model poisoning or hijacking, genAI detection and response, and data governance. Conversely, companies like Zscaler are leveraging genAI to predict breaches and recommend policies for threat detection and response.

Introduction of Standard Safety Protocols and Regulation

Since its inception, LLMs have been prone to error and hallucination. While foundational model companies such as Anthropic focus specifically on safety, there is no current industry standard, which is a significant risk at the enterprise level. We recently saw the New York Times sue OpenAI, which may bring safety issues into the spotlight. An AlphaSense expert reviewed the case and referred to ongoing hallucinations, which would lead to a destructive outcome for the New York Times trademark.

> “I believe they called those AI hallucinations, at least that's what OpenAI referred it to where the AI just believes that something is true and keeps repeating that information, but that is destructive to The Times trademark to be publishing false information and attributing it to them. Trademark law is a lot more sound than copyright law, especially when it comes to litigating about it.”
>
> **– Legal Industry Expert | Expert Transcript**

2024: Year of Maturation for Enterprise GenAI

It doesn’t appear that 2024 will bring seismic change to the enterprise landscape. However, there are significant changes on the horizon as enterprise companies move through the discovery phases of genAI and towards implementation in specific sub-verticals. ROI on investments made over the last 12-28 months may or may not come to fruition in upcoming earnings reports, and those results will likely guide further spending on genAI-related applications.

Major topics such as security and governance, open-source models versus closed-source models, and data integration into existing systems still need to be answered. The foundation is being laid, and enterprise software companies are leading explorations of genAI while enhancing top and bottom-line profiles. Our experts at AlphaSense are making calls in these essential focus areas; our expert transcripts are critical pieces to develop a 360-degree view as the industry gains clarity on how genAI deployments unfold.

Ready to dive deep into the expert transcript library? Start your free trial of AlphaSense today.

AlphaSense

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