MANHATTAN
VENTURE RESEARCH

Generative AI
Industry Report

Generative AI: Nearing the Ultimate Promise of AI

October 12, 2023
28 Pages

Summary

Generative AI is primed to drive a fundamental shift in the status quo while creating new business models and verticals. The technology has already begun to yield substantial benefits for businesses by effectively addressing the longstanding productivity paradox and raising the ROI for enterprises. For investors, opportunities abound in infrastructure, hardware, and application development. The report also highlights the role open-source generative AI projects could potentially play as this technology continues to evolve. Furthermore, the report addresses the critical ethical risks associated with the widespread use of generative AI such as bias in algorithms, data privacy, and potential misuse of AI-generated content. Additionally, the report delves into the growing importance of regulation to ensure the responsible and ethical deployment of generative AI. We round out the report by spotlighting eight emerging growth companies that we believe are in a strong competitive position to benefit from the explosive growth in the adoption of this technology.

Key Report Highlights

- Generative AI is primed to lead a fundamental shift in the workforce, eliminating few jobs while creating new ones.
- Generative AI provides businesses with tangible benefits like cost savings and improved efficiency; incorporating generative AI tools has the potential to increase productivity by 66%.
- As businesses increasingly acknowledge the importance of specialized AI solutions, the demand for domain-specific large language models (LLMs) will experience a remarkable upsurge.
- Companies need to prioritize investing in AI governance and focus on building trust, transparency, and accountability.
- While open-source generative AI models appear a tempting option due to their lower cost, there are risk factors associated such as trust boundary risk, data management risk, and inherent model risk.
- Generative AI offers a $1.3 trillion market opportunity with the hardware layer comprising nearly half of it.
- Key start-ups in generative AI include OpenAI, Anthropic, Jasper, Cohere, Coreweave and Contextual AI.

Author - Santosh Rao
Partner, Head of Research
santosh@mvp.vc
SantoshRao25
T: (212) 858-9908
M: (609) 240-1100

Manhattan Venture Research LLC
152 Madison Avenue, 7th Floor
New York, NY 10016
Main: (212) 858-9900
www.mvp.vc

Executive Summary

Generative AI is unquestionably a transformative technology in the same league as the Internet in the 1990s, cloud computing in the 2000s, and smartphones in the last decade. Generative AI is disrupting and reshaping the status quo across industries and institutions, with its applications expected to have broad socio-economic implications. Goldman Sachs estimates Generative AI to boost global GDP by 7% or nearly $7 trillion and enhance productivity by 1.5% over the next decade. It is fostering the development of an entire ecosystem that spans hardware providers to foundation model developers to application developers, with each playing a crucial role in realizing the technology's potential for businesses.

We believe the evolution of generative AI adoption will happen in two stages, each contributing to the technology's pervasive impact on various industries: stage one is currently underway with businesses using generative AI to enhance employee productivity. Verticals like sales and marketing, audio and video, product and service development, and copywriting are already witnessing transformative disruptions due to generative AI. The second stage we believe will unlock the true potential of generative AI when industries such as manufacturing, healthcare, and logistics, among others, start adopting it more pervasively. Businesses in these areas will embrace generative AI to drive automation and optimization of their complex processes.

As businesses across various verticals delve deeper into this transformative technology, they will uncover novel ways to leverage its potential, eventually reshaping the future of industries and paving the way for a new era of productivity and efficiency. Accordingly, the generative AI market is projected to reach $1.3 trillion by 2032 with plenty of opportunities for investors looking to invest in this dynamic space. While the attention often centers on software and applications layers, the hardware and infrastructure layer will hold a substantial 50% share of this thriving market. Emerging AI hardware start-ups will play a pivotal role in unlocking the true potential of this transformative technology.

Generative AI is at the cusp of driving a profound shift in the workforce, ushering in a process of creative destruction by eliminating some jobs but creating many new ones. Overall, businesses stand to gain tangible benefits from incorporating generative AI tools. As companies increasingly recognize the value of specialized AI solutions, the demand for domain-specific large language models (LLMs) is set to experience a remarkable upsurge. However, to fully leverage generative AI's potential, organizations must prioritize investments in AI governance. Building trust, transparency, and accountability are crucial to ensure responsible and ethical AI deployment.

Generative AI landscape is dynamic and evolving, with generalized solutions giving way to specialized tools and solutions tailored to address specific industry challenges and objectives. This strategic shift will drive increased adoption across enterprises, as they recognize the immense potential in custom-tailored solutions that align seamlessly with their unique requirements.

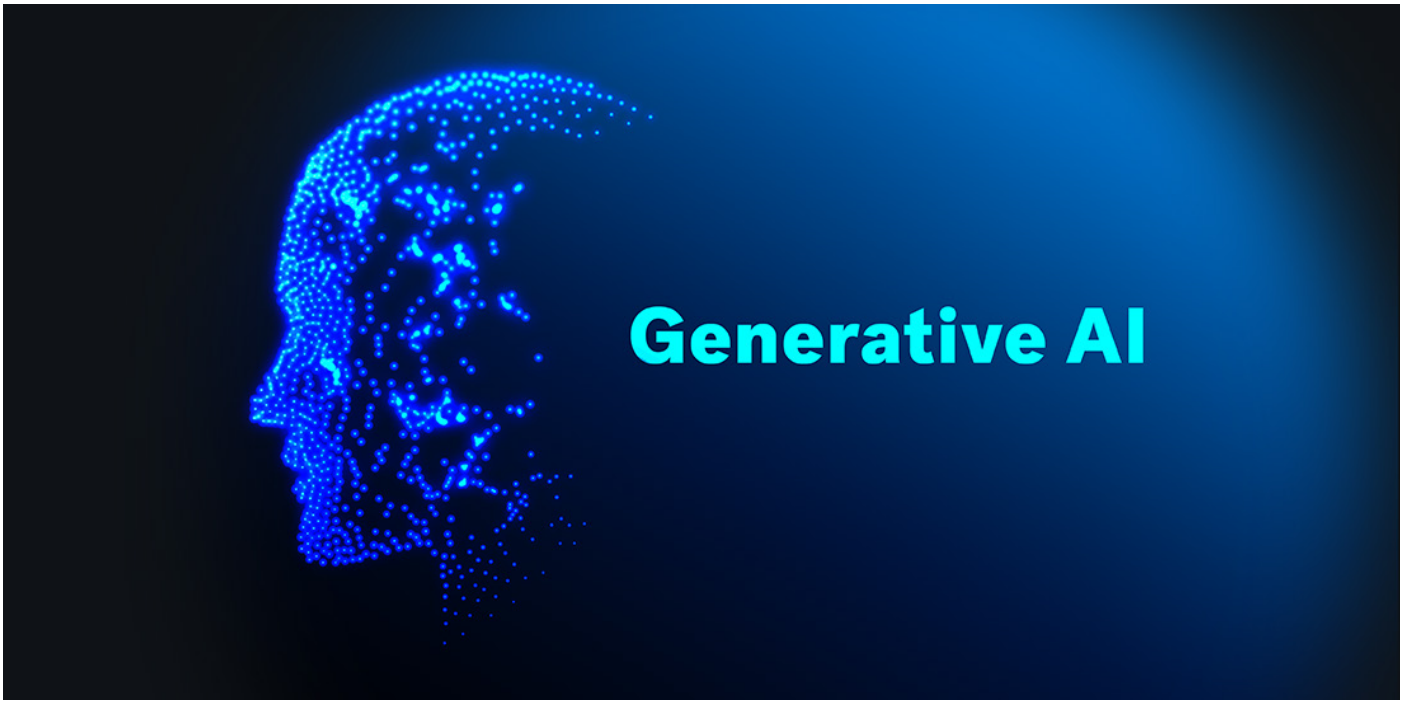
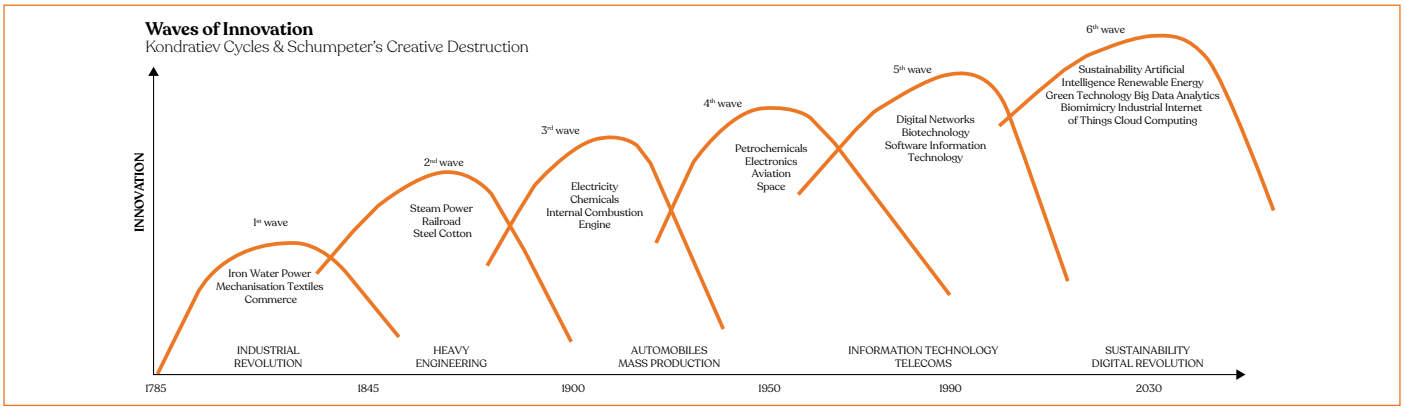


TABLE OF CONTENTS

Introduction

The history of human progress has been marked by waves of innovation, according to the theory of innovation cycles developed by Joseph Schumpeter, each triggering a process of creative destruction that reshapes industries and societies. Earlier waves have encompassed breakthroughs such as mechanization, which revolutionized manufacturing and set the stage for the Industrial Revolution. Further, the advent of steam power, railroads, and telegraphy, enabled unprecedented connectivity, transforming transportation and communication. In the third wave, electricity and chemicals emerged as game-changers, paving the way for mass production and the rise of modern chemical industries. The fourth wave introduced aviation and electronics. The fifth wave witnessed the rise of digital networks, biotechnology, and information technology, leading to the digital revolution and globalization of information. Generative AI is next such wave that will reshape industries across the board, from manufacturing and healthcare to finance and beyond, promising a future of unprecedented efficiency and transformative possibilities.

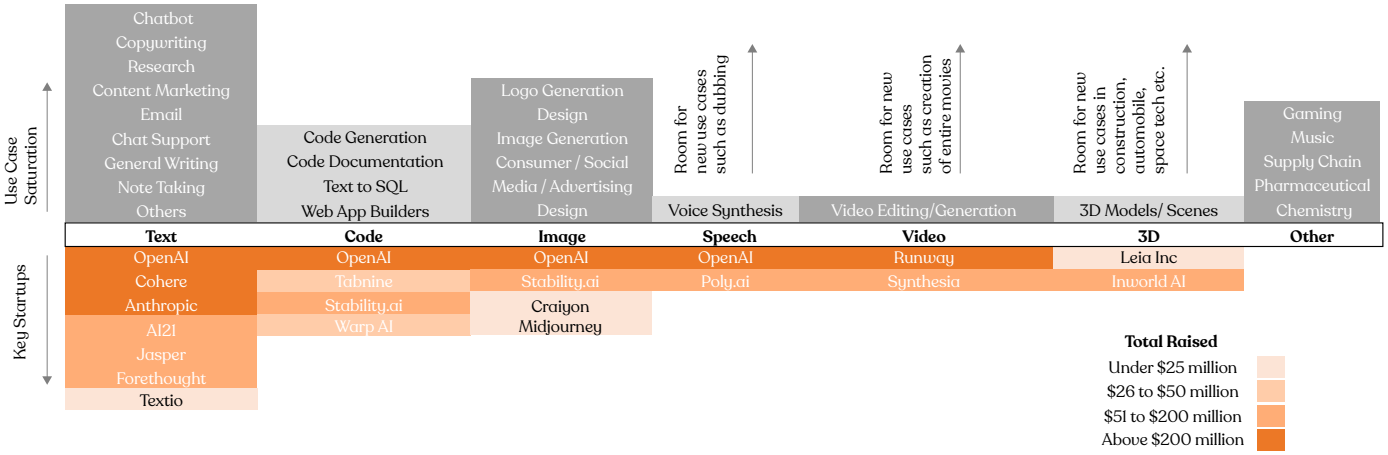
Figure 1: Waves of Innovation



Source: Oxford Research

The genesis of the generative AI revolution commenced with textual applications, achieving remarkable advancements. Pioneering companies such as OpenAI, Cohere, and Anthropic, among others, have harnessed substantial funding and established their credibility by leveraging their early-mover advantage. The next step in the generative AI proliferation was text-to-image applications in which startups such as OpenAI, Stability.ai, and Midjourney, among others, cemented their credibility. As we progress toward the subsequent wave of generative AI applications encompassing coding, video, gaming, music, supply chain, pharmaceuticals, chemistry, and beyond, there exists ample opportunity for emerging startups to establish differentiating value propositions. Consequently, startups pioneering novel use cases for generative AI are poised to evolve into highly valuable companies in the foreseeable future.

Figure 2: Generative AI Application Snapshot

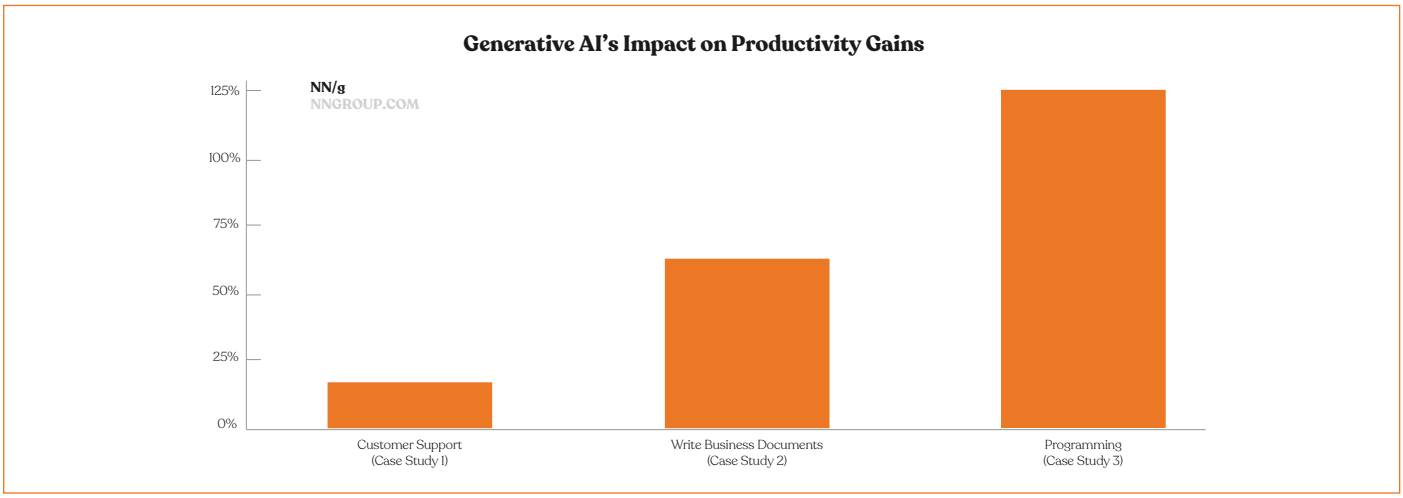


Source: Manhattan Venture Research

Solving the Productivity Paradox

Businesses are embracing the power of generative AI to revolutionize and optimize their work processes, yielding remarkable outcomes. By deploying generative AI technologies, businesses are unlocking the potential for automation, continuous improvement, and heightened performance across a diverse range of tasks and processes. The findings of studies by Nielsen Norman Group underscore the remarkable efficiency gains experienced by AI users across various domains.

Figure 3: Generative AI's Impact on Productivity Gains

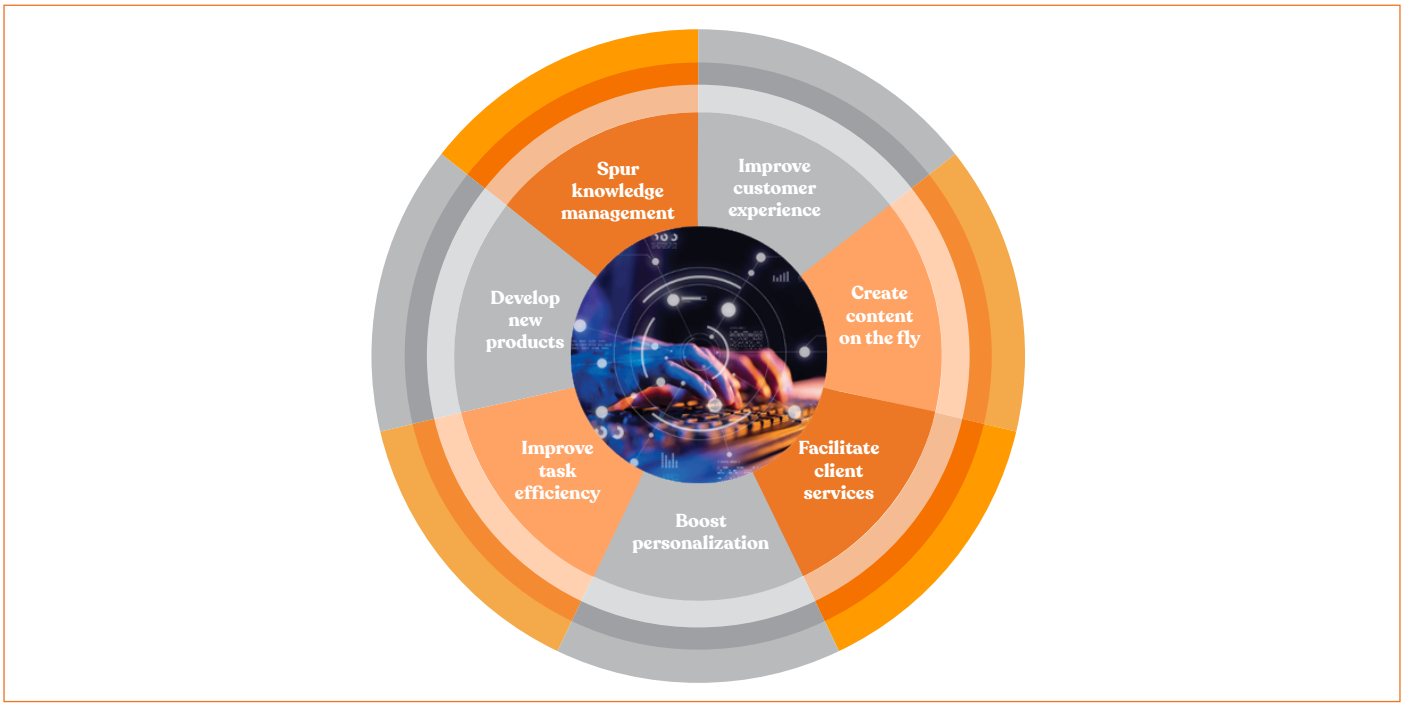


Source: Nielsen Norman Group

- Study 1 reveals that support agents, bolstered by AI assistance, are able to adeptly handle 13.8% more customer inquiries per hour, thus substantially enhancing customer service capabilities.
- Study 2 demonstrates a significant 59% increase in business professionals' output, as they harness AI to expedite the creation of critical business documents.
- Study 3 provides remarkable insights into the coding domain, showcasing a staggering 126% improvement in programmers' ability to deliver projects within compressed timelines when they employ AI support.

On average, across all three studies, generative AI tools have bolstered business users' throughput by an impressive 66% when engaged in realistic tasks. Comparing these numbers with labor productivity growth rates in the US offers a clearer picture of the impact of AI on productivity gains. During the 12-year period leading up to the pandemic (2007–2019), average labor productivity growth in the United States stood at a modest 1.4% per year, per data from the Bureau of Labor Statistics. This number reflects the average value created by a worker per hour worked, irrespective of the total economic output or the number of workers. Assuming there is no AI, it would take 47 years to reach this level of productivity gains.

Figure 4: Generative AI Benefits for Businesses



Source: Tech Target

Further research highlights the impact of generative AI on productivity gains. According to a study by the National Bureau of Economic Research, generative AI tools like ChatGPT have the potential to increase workforce productivity for customer support agents by an average of 13.8%. The study further suggests that these tools will have the greatest impact on novice and low-skilled workers. Additionally, data from Artisana, a platform focused on generative AI, reveals that Generative AI tools are leading to a 40x increase in productivity specifically in the Chinese video game art industry. By automating repetitive and mundane activities, generative AI liberates human resources to focus on higher-value and strategic initiatives.

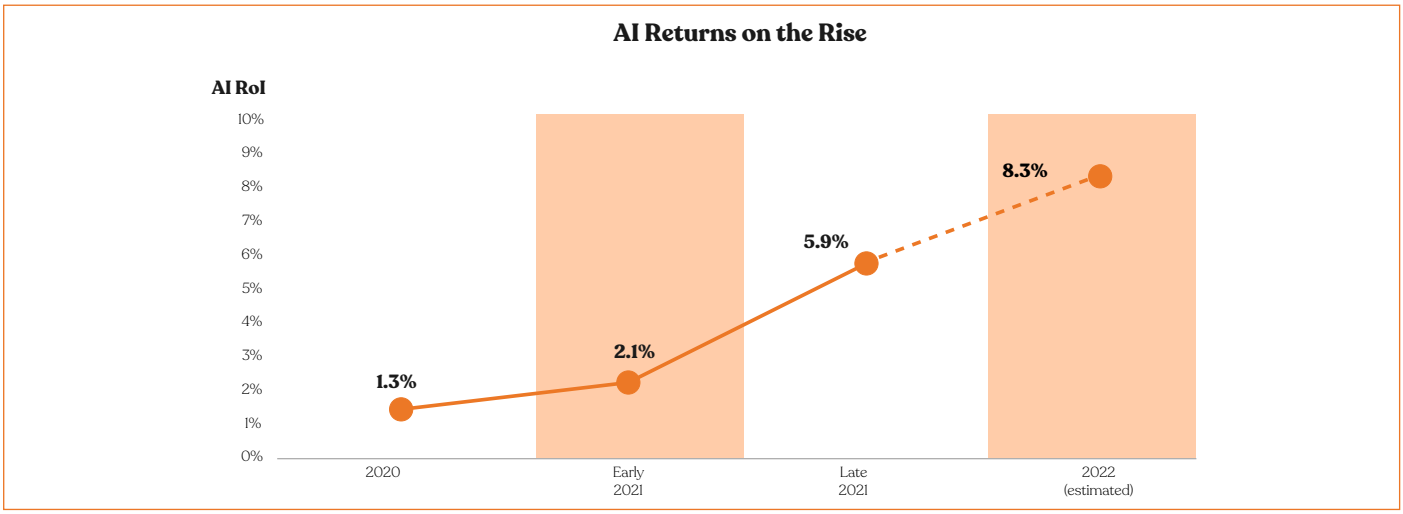
As generative AI continuously evolves and adapts, it sets the stage for a future where intelligent machines collaborate seamlessly with humans, amplifying our collective potential and propelling us towards unprecedented levels of creativity and progress.

Generating ROI with Generative AI

The most compelling promise of generative AI lies in its potential to yield substantial monetary benefits for businesses. As this transformative technology is still in its early stages of adoption, accurately gauging its projected impacts on revenue and economic growth remains a challenge. However, insightful data from an IBM report sheds light on the remarkable progress achieved. It reveals that the average Return on Investment (ROI) for enterprise-wide generative AI initiatives grew from just over 1% in early 2020 to a significant 6% by the end of 2021.

This data signifies the rising recognition of generative AI's immense value in driving business outcomes, including improved efficiency, enhanced customer experiences, and cost savings. As businesses continue to embrace and integrate generative AI into their operations, the expected growth trajectory of AI ROI has not only been sustained but has further surged. A continuation survey conducted in April and May 2023 revealed that the estimated AI ROI reached an impressive 8.3% in 2022.

Figure 5: AI Returns on the Rise



Source: 2020: Deloitte/ESL AI survey; Early 2021: IBV AI ethics survey; Late 2021: IBV AI capability survey; 2022: IBV generative AI pulse survey, April-May 2023.

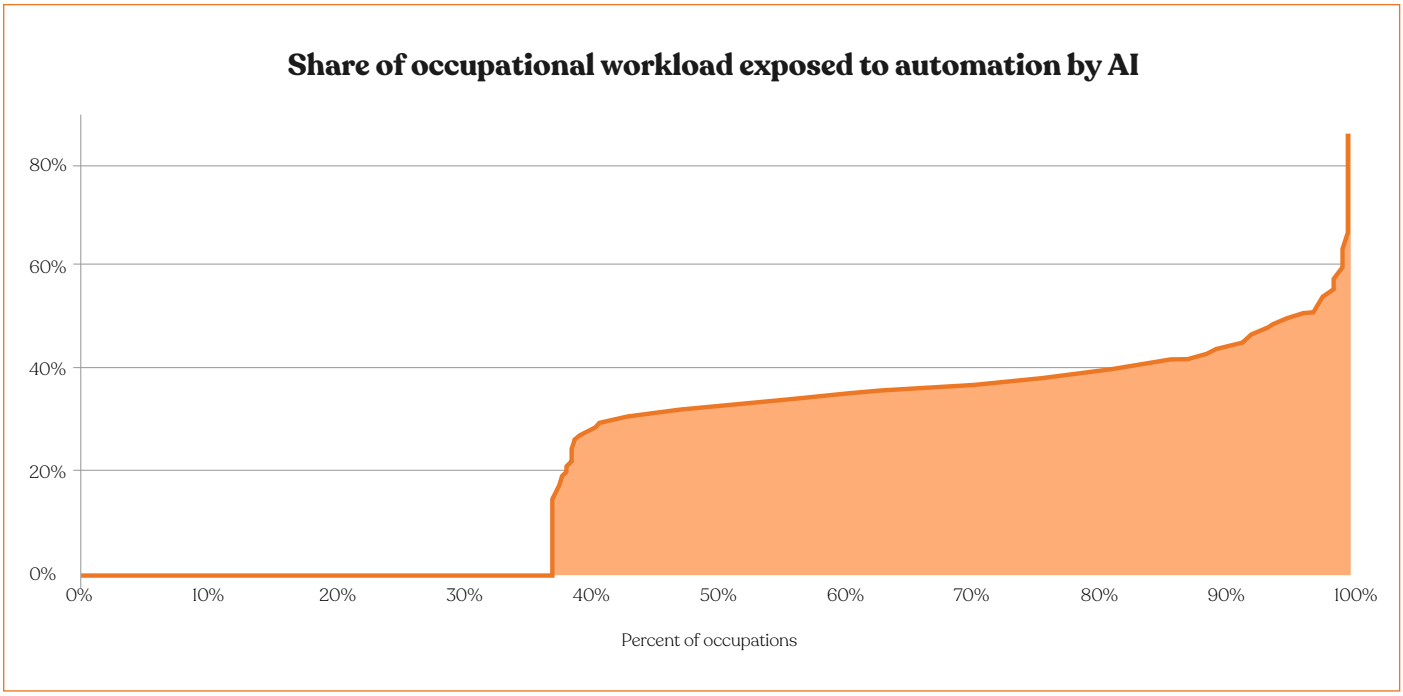
This sustained upward trajectory of AI ROI underscores the technology's transformative potential in generating substantial value across various industries. The data serves as a promising indicator that businesses investing in generative AI are reaping significant rewards, with increased profitability and competitive advantages becoming increasingly evident.

Moreover, the increasing adoption and success of generative AI have far-reaching implications for the workforce and the economy at large. As businesses achieve higher levels of efficiency and productivity through AI integration, they are well-positioned to invest in other strategic initiatives and drive further economic growth. This cascading effect is likely to ripple through the entire business ecosystem, leading to a virtuous cycle of progress and prosperity

Generative AI Applications across Industries

Generative AI, fuelled by machine learning and deep neural networks, is poised to transform numerous industries in the future. Its applications will revolutionize content creation, product design, virtual reality experiences, natural language processing, data augmentation, personalized marketing, and risk analysis. By leveraging generative AI, businesses can expect increased efficiency, accelerated innovation, and heightened competitiveness, paving the way for remarkable advancements in sectors such as entertainment, healthcare, finance, and more.

Figure 6: AI to Partially Automate Two-Third Occupations



Source: Goldman Sachs

Mobility Tech: In the realm of Mobility Tech, Generative AI is already making waves with its ability to improve the conversational interface to cabin controls, optimizing user experiences within vehicles. This technology is also enhancing public transit scheduling and routing, enabling efficient and streamlined transportation solutions. Furthermore, ride-hailing apps are benefitting from Generative AI's capacity to deliver personalized and efficient services, enhancing customer satisfaction and loyalty. As technology advances, we can expect to witness accelerated design and development processes, leading to more agile and cost-effective mobility solutions. Additionally, the optimization of production lines through Generative AI will foster increased efficiency and productivity, revolutionizing the manufacturing and automotive industries.

Supply Chain Tech: In Supply Chain Tech, Generative AI is revolutionizing order tracking through the integration of a conversational tone, simplifying and streamlining interactions between suppliers and customers. Data analytics and actionable insights are empowering businesses to make informed decisions, optimize inventory management, and reduce costs. This technology's potential in optimizing schedules, prices, and routes will have a profound impact on supply chain efficiency and responsiveness. In the long term, Generative AI holds the potential to turbocharge productivity across various supply chain operations, from booking and pricing to processing and tracking. As a result, businesses can expect to witness significant cost savings, reduced lead times, and increased customer satisfaction.

E-commerce: In the realm of e-commerce, the integration of Generative AI with search bars is already improving product discovery and recommendations, elevating the shopping experience for consumers. Customer support is becoming more efficient and personalized, enhancing customer loyalty and retention. As Generative AI evolves, it will play a vital role in enabling deeper personalization of products and services, ensuring that customers receive tailored experiences. Moreover, the application of Generative AI in B2B email outreach and support will drive more effective marketing strategies and foster stronger business relationships. The rise of startups like Anyword, Treat, and Copysmith in this space indicates the growing recognition of Generative AI's potential in transforming the e-commerce landscape.

FinTech: The FinTech sector is experiencing a significant transformation with the implementation of Generative AI. Operational efficiency is being optimized through automation and AI-driven processes, streamlining financial operations and reducing costs. Enhanced marketing strategies, powered by Generative AI, are enabling financial institutions to target their offerings with greater precision, driving customer engagement and satisfaction. In the long run, Generative AI's potential in budgeting and forecasting will have a profound impact on financial planning and risk management. Additionally, regulatory compliance monitoring will be significantly strengthened through AI-powered tools, ensuring adherence to complex financial regulations. The notable valuations of industry leaders like Stripe and Complyadvantage underscore the immense market potential for Generative AI in the FinTech landscape.

Health Tech: The integration of Generative AI in Health Tech has immense implications for patient care and medical practices. Patient history summaries generated by AI will streamline healthcare workflows, allowing medical professionals to make informed decisions swiftly. Symptom checkers are empowering individuals to access initial health assessments conveniently. In the long term, Generative AI's role in insurance claims and personalized medicine will have far-reaching consequences, enabling personalized treatment plans and more efficient claim processing. The rise of startups like AKASA, Woebot, and Infinitus showcases the increasing focus on harnessing Generative AI's potential to revolutionize the healthcare industry.

Gaming: Generative AI applications in the gaming industry are redefining the development process. Code assistance is streamlining game creation, reducing development time, and enhancing the overall quality of gaming experiences. As the technology evolves, we can expect entire game development to be impacted significantly, opening new avenues for creativity and immersive gameplay. Startups like Inworld AI, Promethean AI, and Wonder Dynamics are leading the charge in the gaming domain, bringing innovative AI-driven solutions to the market.

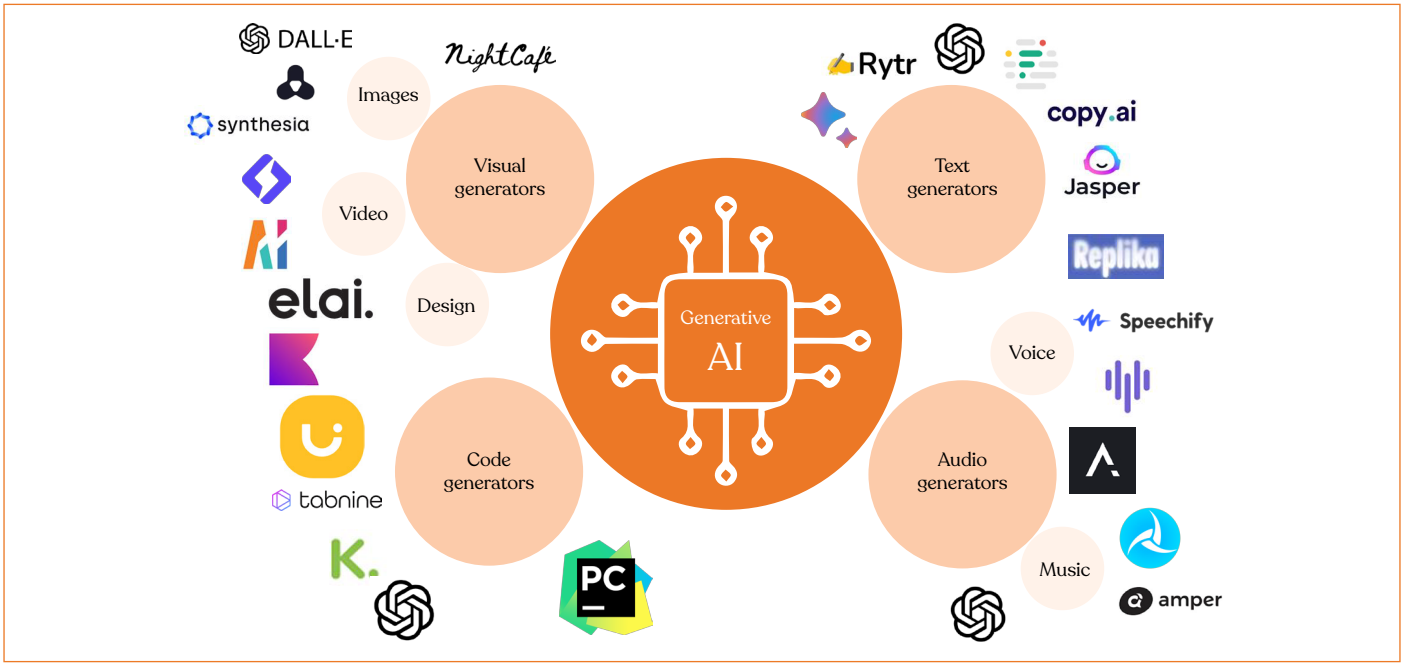
AgTech: In the agricultural sector, Generative AI is making significant strides, aiding farmers in yield prediction and livestock monitoring. In the long run, this technology will play a crucial role in rural transformation, driving efficiency and sustainability in farming practices. The environmental implications are particularly noteworthy, as Generative AI will contribute to promoting greener agricultural practices. Notable startups such as Gamaya, Hortau, and Cropin are at the forefront of leveraging Generative AI to revolutionize the agricultural industry.

Food Tech: Generative AI is revolutionizing the Food Tech sector by optimizing menu creation and personalized diet planning, catering to individual preferences and dietary requirements. As food waste continues to be a global concern, Generative AI's role in reducing food waste is of utmost significance, promoting more sustainable practices in the food industry. Leading startups like Instacart and Uptake are redefining food delivery and waste reduction, showcasing Generative AI's potential to shape the future of food-related services.

Climate Tech: In Climate Tech, Generative AI is being applied to critical areas such as carbon accounting and offset evaluation, assisting in environmental sustainability efforts. The integration of Generative AI in the development of greener AI hardware and computationally lightweight models is vital for promoting sustainable technology adoption. With startups like Carboncrop, Brainbox AI, and Climate AI leading the charge, Generative AI is playing a pivotal role in driving innovations to tackle climate change and foster a greener future.

Enterprise SaaS: Generative AI is disrupting the Enterprise SaaS landscape with its ability to facilitate first-draft content creation and enhance enterprise search capabilities. In the long term, the emergence of virtual avatars for marketing, sales, and customer service will revolutionize customer interactions, personalizing user experiences and driving customer engagement. Startups like Jasper, Glean, and Super Normal are pioneers in deploying Generative AI to empower businesses and streamline enterprise workflows.

Figure 7: Key Generative AI SaaS Platforms



Source: Canalys

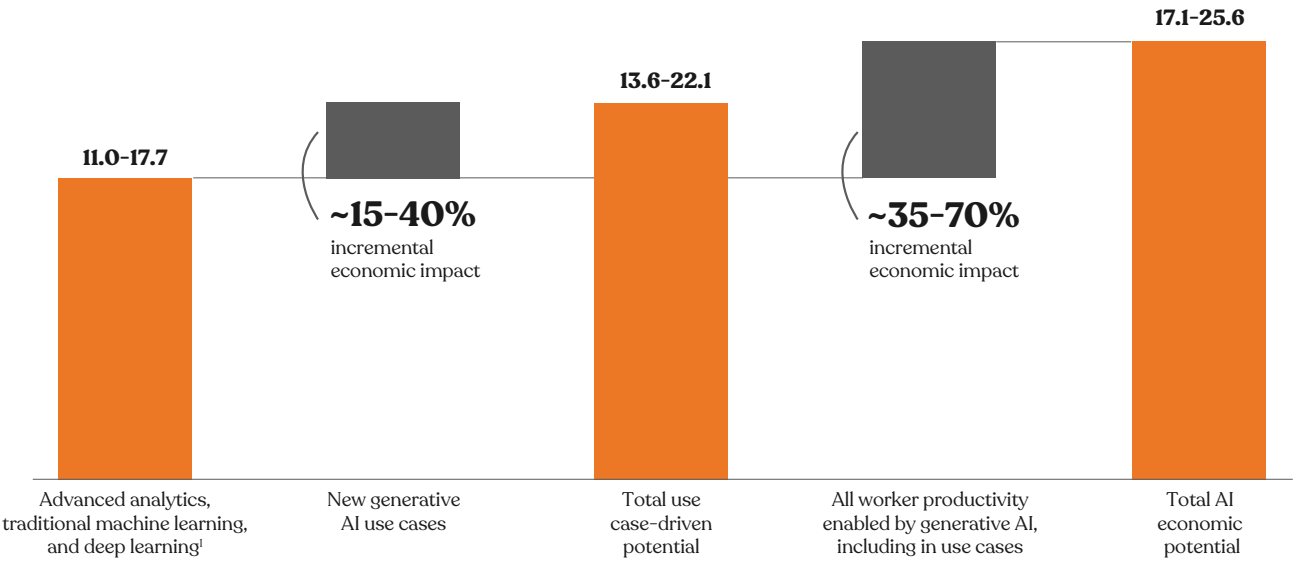
Infosec: In the realm of Infosec, Generative AI is revolutionizing data analysis automation and secure coding practices. The advent of advanced Generative AI tools may render traditional SIEM (Security Information and Event Management) solutions obsolete, paving the way for more sophisticated security frameworks. Noteworthy startups like Ocra Security, Torq, and Log Point are shaping the future of cybersecurity, harnessing Generative AI to fortify digital defenses.

IoT: In the Internet of Things (IoT) domain, Generative AI is driving connected car enhancement and video surveillance analytics, optimizing efficiency and safety. The application of Generative AI in industrial designing and the creation of digital twins for industrial systems will drive transformative changes in manufacturing and automation. Startups like Ground Light and Diabatix are leveraging Generative AI to redefine the IoT landscape, propelling the world.

The Economic Potential of Generative AI

Generative AI possesses immense economic potential, catalyzing transformative growth across various industries. Sequoia Capital estimates the potential economic value that generative AI can generate could reach trillions of dollars, by empowering businesses to unlock new opportunities, streamline operations, and enhance creativity and innovation. Additionally, research from McKinsey estimates that generative AI could add the equivalent of $2.6 trillion to $4.4 trillion annually to the global economy across the 63 use cases analyzed by the consulting firm. The research further states that 75% of the value that generative AI use cases could deliver falls under 16 business functions under four areas – customer operations, marketing and sales, software engineering, and R&D.

Figure 8: Generative AI's Potential Impact on Global Economy ($T)

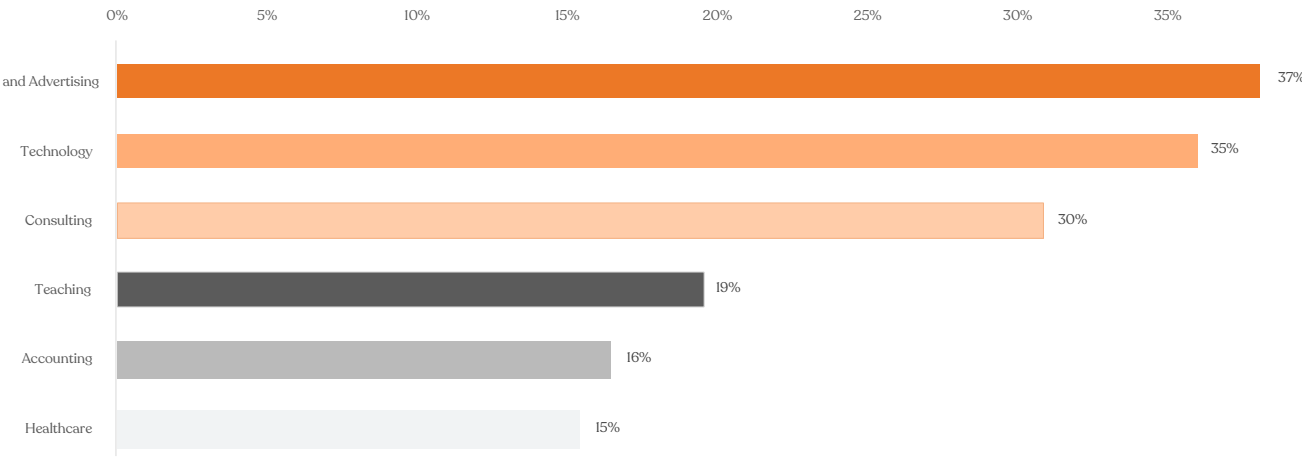


Updated use case estimates from "Notes from the AI frontier: Applications and value of deep learning," McKinsey Global Institute, April 17, 2018

Source: McKinsey

The pervasive adoption of generative AI in various industries underscores its profound impact on the modern workplace. The marketing and advertising sector, in particular, has an impressive 37% adoption rate, embracing generative AI's capacity to craft personalized content, optimize campaigns, and foster exceptional customer engagement. The technology sector (35%) has also embraced generative AI to fuel innovation and automate processes. For comparison, it took 10 years for cloud adoption to reach 31% adoption. Gartner predicts that generative AI will replace 20% of procedural code professionals' work by 2025 and will be used to create 30% of outbound marketing messages in large organizations. By then, 50% of drug development will be made by generative AI.

Figure 9: Generative AI Adoption in Workplace by Industry (US)

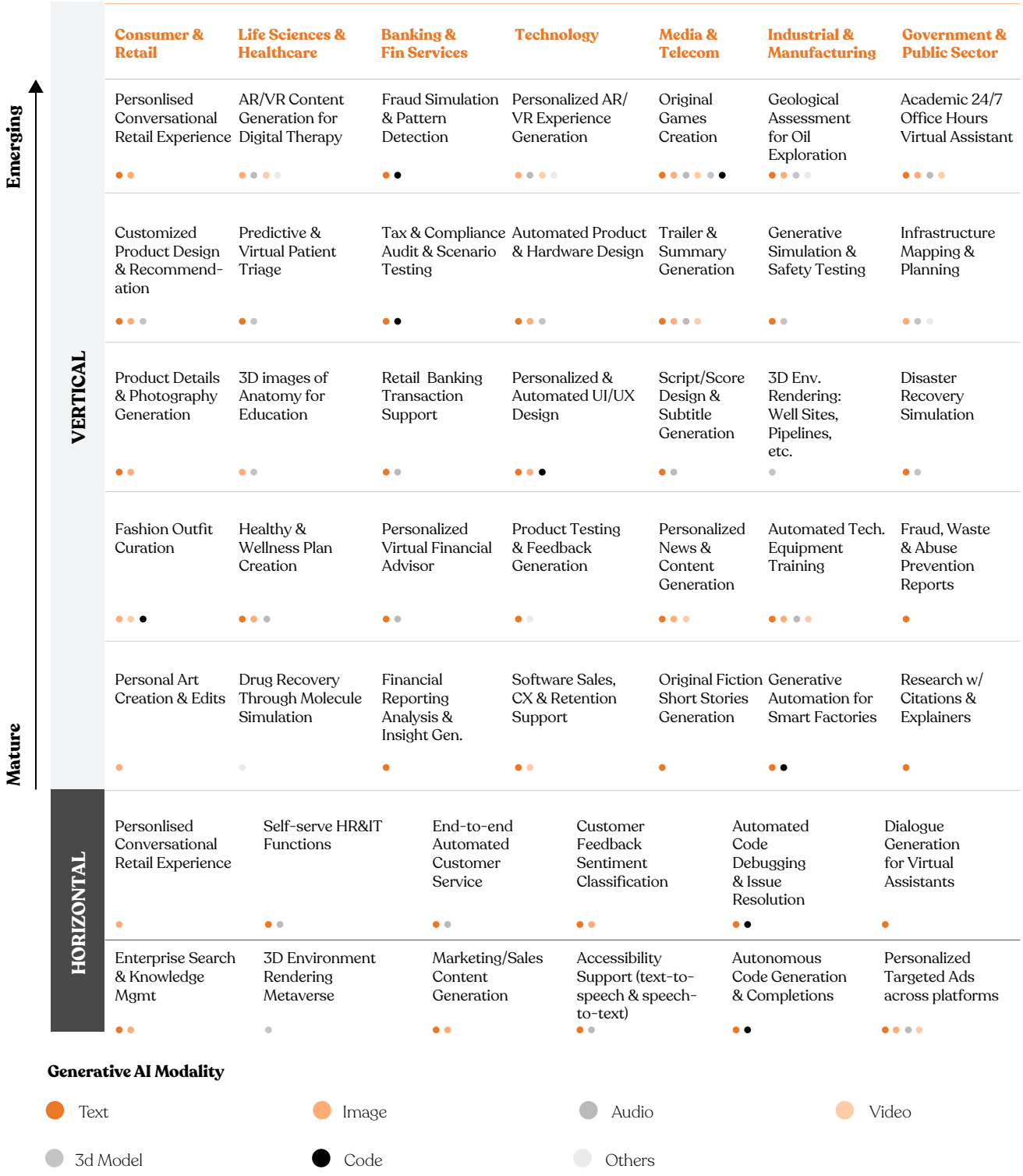


Source: Enterprise Apps Today

Moving from Consumer to Enterprise Use-Cases

The potential for generative AI in enterprise use cases is immense, and it is poised to drive widespread adoption of this technology in various industries. By leveraging generative AI, enterprises can unlock unprecedented opportunities for innovation, efficiency, and competitive advantage. From automated content generation and product design to personalized customer experiences and data-driven decision-making, generative AI empowers businesses to create, iterate, and optimize at scale.

Figure 10: Potential Enterprise Applications Across Industries



Source: Deloitte

Generative AI Market Opportunity

As the generative AI market evolves, there are immense opportunities across the generative AI value chain. Edge AI hardware presents the biggest opportunity, representing a $54.9 billion market by 2032. The highest growth is likely to come from the generative AI market in manufacturing and DevOps with a CAGR (2024-2032) of 39.6% and 36.8% respectively. Generative AI in healthcare and marketing is also projected to grow at an impressive CAGR (2024-2032) of 35.1% and 31.3% respectively.

Figure 11: Generative AI Market Opportunity

Market	2024 ($M)	2032 ($M)	CAGR
Generative AI in Financial Services	1,390.2	9,475.2	27.1%
Generative AI in Manufacturing	444.5	6,398.8	39.6%
Generative AI in HR	550.0	1,660.0	14.8%
Generative AI in Legal	89.8	781.6	31.1%
Generative AI in Marketing	35.9	316.7	31.3%
Generative AI in DevOps	1,800.1	22,100.0	36.8%
Edge AI Hardware Market	13,146.5	54,858.9	19.6%
Generative AI in Healthcare	1,954.1	21,740.0	35.1%
Generative AI in Media and Entertainment	1,929.6	12,077.0	25.8%
Generative AI in Cybersecurity	740.9	2,654.0	17.3%

Source: Manhattan Venture Research, MarketResearch.biz, Allied Market Research, Precedence Research

Exploring the Investment Potential of AI Hardware

The explosive growth of generative AI has triggered a sharp increase in demand for graphics processing units (GPUs). AI could allow semiconductor companies to capture 40-50 % of total value from the technology stack, per McKinsey and Company. This opportunity will be spread across computer vision AI products, conversational AI products, AI servers, AI storage, and Generative AI Infrastructure-as-a-service. While Nvidia continues to run the generative AI chip show, there are immense opportunities for VC-backed start-ups as well.

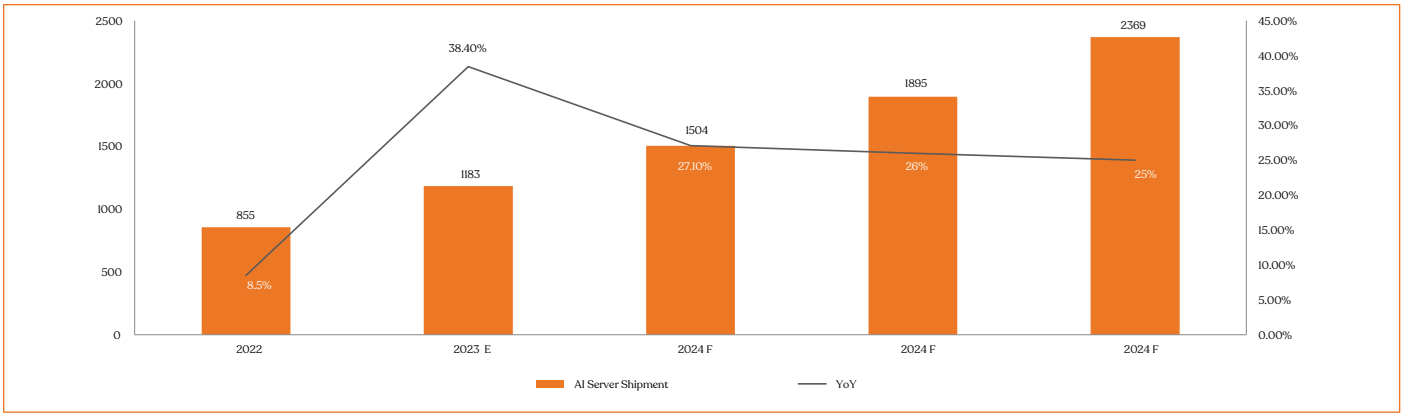
According to projections by market intelligence provider Trendforce, a GPT model required approximately 20,000 GPUs (estimations use NVIDIA's A100) to process training data in 2020. That estimate has now risen to over 30,000 GPUs. Should Google plan to integrate ChatGPT-like features into its search, it will require 512,820 A100 HGX servers with a total of 4,102,568 A100 GPUs, representing $100 billion in servers and networking capex, according to Semi Analysis. These numbers are indicative of the sheer scale of computational resources that are necessary to support the development and deployment of advanced AI models like GPT. Meta, another AI heavyweight, announced a shift from CPU-based to GPU-based models in its February 2023 earnings call, highlighting the increasing importance of GPUs in the AI ecosystem.

One company that has partnered with Nvidia and looks to benefit from the shift to GPUs is CoreWeave. Founded in 2017, CoreWeave is a specialized cloud IaaS provider that offers cloud infrastructure to train, serve, infer, and fine-tune models using GPUs. They differentiate themselves with their fast and flexible infrastructure that allows it to scale up and down in seconds. CoreWeave caters to enterprises of all sizes including VC-backed startups, SMEs, and large enterprises. They have raised a $371 million series B valuing CoreWeave at $2 billion. In 2022, they generated revenue of around $13.1 million by providing solutions 35x faster than legacy cloud providers like Google, AWS, and Azure and look to continue building on the momentum of AI.

The share of specialized chips such as GPUs used in data centers is expected to exceed 15% by 2026, up from less than 3% in 2020 (Gartner). Nvidia estimates that the need for processing power for most machine-learning or AI tasks has increased by 25 times every two years. With the growing dominance of generative AI, the demand for GPUs and specialized chips is expected to rise significantly, benefiting all participants in the value chain.

Furthermore, Trendforce estimates that around 1.2 million servers, equipped with GPUs, FPGAs, and ASICs, will be deployed across global markets in 2023, signifying a year-on-year growth rate of 38.4%, reflecting the escalating demand for AI servers and chips. The implications of this growth are significant as AI servers are expected to make up approximately 9% of the total server shipments, a figure that is projected to rise to 15% by 2026.

Figure 12: Global AI Server Shipments Estimates (2022-2026) (Unit=1,000 units)



Note: Estimates based on AI servers equipped with accelerators such as GPUs, FPGAs, and ASICs, which are used for AI training and inference
Source: TrendForce

McKinsey further estimates the demand for AI-related semiconductors to grow at a rate of 18% over the next few years, five times higher than the rate projected for non-AI applications. By 2025, AI-related semiconductors could constitute around 20% of all demand, generating an estimated revenue of $67 billion. This rapid growth presents opportunities for semiconductor companies at data centers and the edge. If the growth rate remains consistent, semiconductor companies could capture up to 40-50% of the total value generated by the AI technology stack.

Figure 13: AI Chip Startups to Watch

Company	Product	Founded Year	Total Raised ($M)	Valuation ($M)	Last Round
SambaNova Systems	SN10 Processor Chip	2017	1,132	5,000	$676M Series D
Cerebras Systems	Cerebras WSE-2	2016	718	4,000	$250M Series F
Graphcore	IPU-POD256	2016	682	2,770	$222M Series E
Groq	GroqChip Processor, GroqCard Accelerator, etc.	2016	362	1,000	$300M Series C-II
Mythic	M1076 AMP, MM1076 key card, etc.	2012	172	406	$13M Series C-II

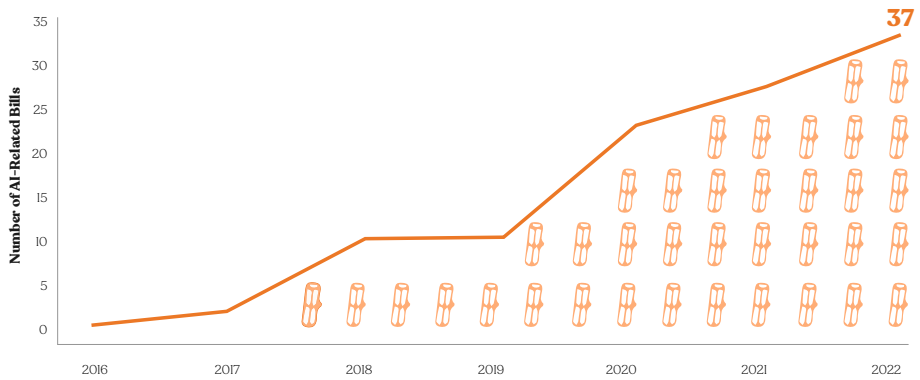
Source: Manhattan Venture Research, CB Insights

Regulation Around AI

The rapid progress of AI has not only prompted regulatory scrutiny but also a comprehensive evaluation of its potential impact. Regulators worldwide are actively working to regulate generative AI. Notably, OpenAI CEO Sam Altman is urging members of Congress in the US to implement AI regulations. In a significant development, the European Parliament's committee of lawmakers approved the EU's AI Act in May, bringing it closer to becoming a law. This regulation adopts a risk-based approach and includes specific requirements for developers of "foundation models" like ChatGPT, ensuring compliance with copyright laws in their training data. China is also poised to present a draft of its AI law to the country's lawmakers for review later this year. The Canadian parliament is also considering its own hotly contested Artificial Intelligence and Data Act.

The generative AI landscape is rapidly evolving, making it challenging to craft specific and comprehensive regulations that can keep pace with the technology's rapid advancements. Instead of rigid regulations, a more suitable approach would be to establish guardrails that foster responsible development and deployment of generative AI. Also, it is important to recognize that the legal framework surrounding AI already exists, albeit in a broader context. Laws such as intellectual property rights, privacy regulations, and anti-discrimination laws can effectively address potential issues related to generative AI without stifling its progress.

Figure 14: Global AI-Related Bills



Source: AI Index, 2022

AI is already subject to significant regulation, including data protection laws like the EU's General Data Protection Regulation and the US's Health Insurance Portability and Accountability Act. These regulations play a crucial role in safeguarding personal information and ensuring accountability. Additionally, generative AI is already being utilized in heavily regulated sectors such as healthcare and finance, highlighting the existing framework of regulations in these critical domains. Moving forward, the regulation of specific sectors within the AI landscape will remain vital. As AI continues to advance and permeate various industries, targeted regulations tailored to each sector's unique challenges and requirements will play a crucial role in ensuring ethical and responsible AI deployment. Such sector-specific regulations can address concerns related to privacy, security, transparency, bias, and accountability, among other important considerations.

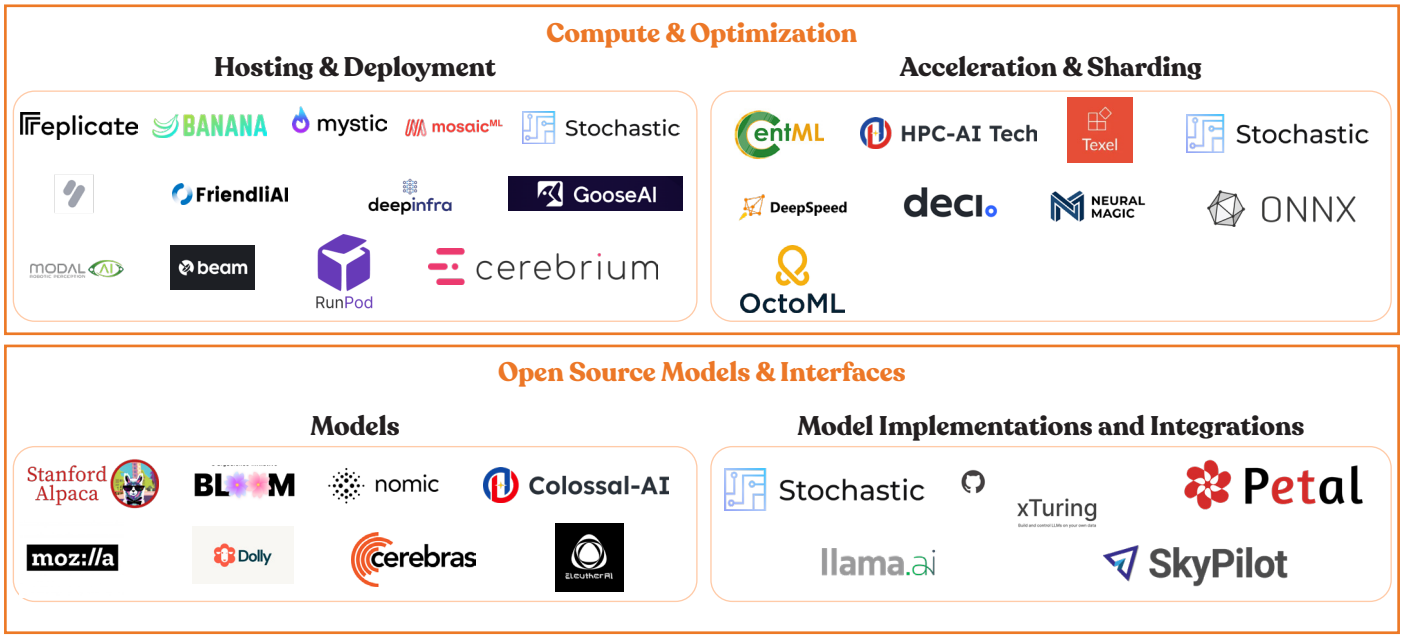
However, lawmakers must remain wary of the lobbying efforts exerted by influential Big Tech companies such as Alphabet, Microsoft, and OpenAI. These companies possess significant resources and political influence, which can sway policy and regulatory decisions in their favor. While their contributions to innovation and technological advancement are undeniable, it is essential for lawmakers to maintain an independent and objective perspective. By critically evaluating the lobbying activities of these companies, lawmakers can ensure that regulations are formulated in the best interest of the public and not unduly influenced by private agendas.

At the heart of the ongoing debate surrounding the regulation of AI lies the fundamental challenge of striking the right balance between mitigating risks and fostering its continued growth. This intricate task demands an approach that involves comprehensive deliberation and engagement from various stakeholders, rather than solely relying on technologists to determine the course of action.

Open-Source AI Vs Close-Source AI

Open-source generative AI platforms embrace collaboration and decentralization, empowering developers, researchers, and enthusiasts to access, modify, and distribute the underlying code and models. This collaborative approach taps into the collective intelligence of the community, fostering creativity, knowledge sharing, and rapid innovation as well as driving forward the democratization of AI. The open-source community operates without institutional constraints, enabling agile experimentation and diverse perspectives.

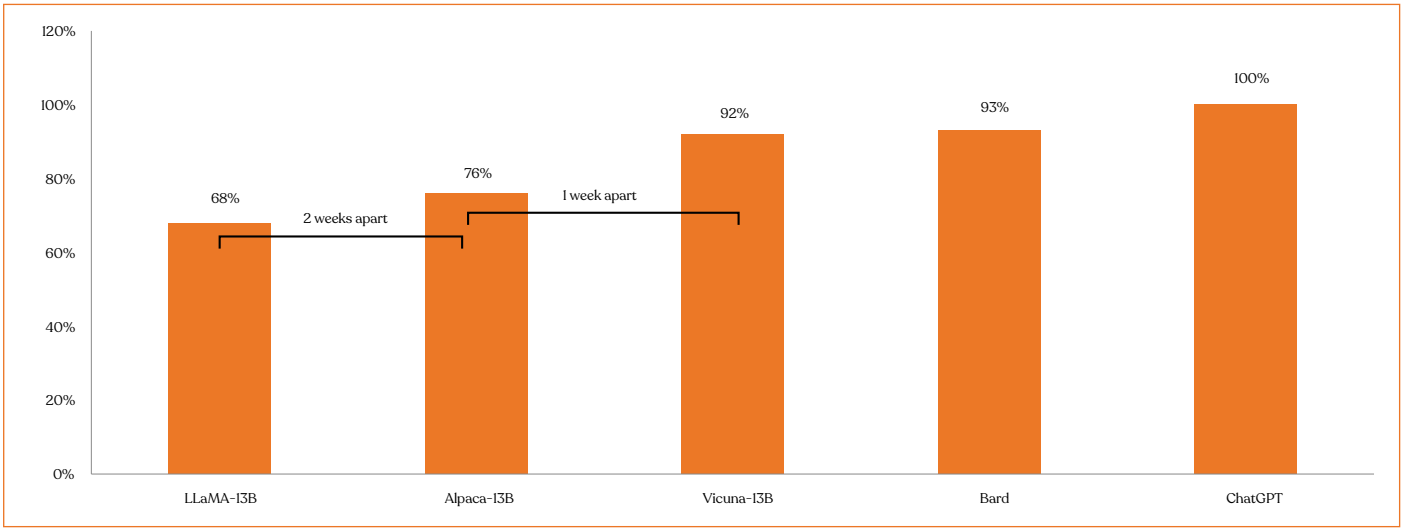
Figure 15: Open Source LLM and Infrastructure Ecosystem



Source: Scale Venture Partners

One of the major advantages of open-source AI models lies in their significantly lower training costs. This cost efficiency gives them a distinctive edge over proprietary models. To illustrate this point, consider an open-source model priced at $100, equipped with 13 billion parameters. Surprisingly, such a model has the potential to rival a high-end Google model costing $10 million, boasting a staggering 540 billion parameters. The open-source community has made remarkable strides in addressing various scaling challenges by implementing a wide array of optimizations. MosaicML showcases the community's ability to train a Stable Diffusion model with remarkable cost-effectiveness. They replicated Stable Diffusion 2 for less than $50k, an 8x reduction from the original Stable Diffusion 2.

Figure 16: Open Source LLMs Move Fast



Source: LMSYS

However, there are some downsides to open source LLMs. We firmly believe that in order for open-source projects to transition into successful ventures, they must address key factors such as sustainability, risk-to-success ratio, and profitability while maintaining a high standard of quality and accuracy to deliver tangible value. It is premature to ascertain whether open-source models trained at a cost of $100 or $300 will truly fulfill their intended purpose. While these models may serve well for experimental purposes, attaining profitability and ensuring long-term viability for open-source AI systems necessitates substantial investment in the range of tens of millions of dollars to properly operate and manage them.

Additionally, in a recent research study by Rezilion, an automated software supply chain security platform, experts have discovered several issues related to open-source LLMs. These issues include trust boundary risk, data management risk, and inherent model risk.

Trust boundary risk: Open-source LLMs depend on external data sources and pre-trained models. This raises concerns about the security and reliability of the data used for training. If these external elements have vulnerabilities, it could compromise the entire language model's security.

Data management risk: Handling large amounts of sensitive data during the training process poses risks. Ensuring the privacy and proper management of this data is crucial. Any data security lapses could result in data breaches and unauthorized access to sensitive information.

Inherent model risk: Open-source LLMs may inadvertently amplify biases present in the training data. While these models excel at generating text and understanding language, they are not immune to inherent biases, which could have ethical and societal implications.

While the inclusivity of open source projects fosters a wider range of contributors in AI research and development, it also escalates the potential for misuse and bias. Conversely, closed-source systems offer protection against external entities, shielding them from unauthorized interference. Companies like OpenAI and Google also benefit from massive pools of user data, enabling them to deliver quality information. Access to more data and feedback will help Google and OpenAI in the accuracy and relevancy of responses.

Additionally, the impending regulations surrounding generative AI will impose high entry barriers for new open-source entrants, putting closed-source AI companies at an advantage as they already adhere to the required compliance measures and possess the resources to navigate the evolving regulatory landscape

Emerging Trends

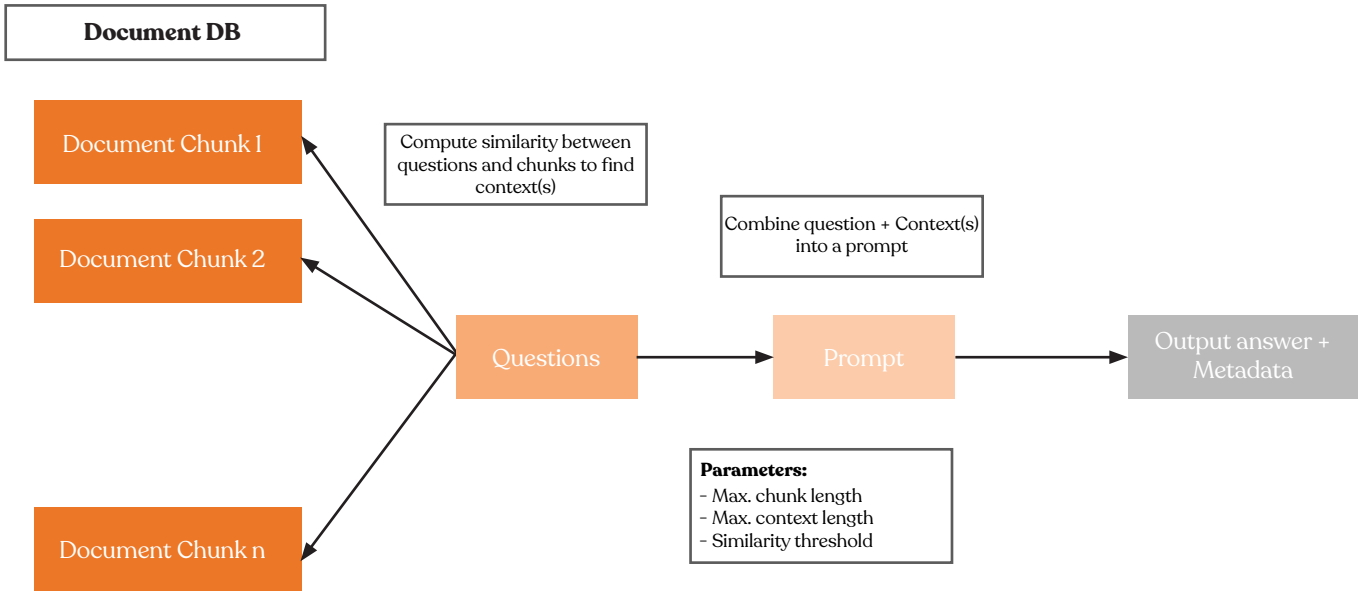
Enterprise-Grade LLMs

Generative AI presents a world of possibilities for organizations, offering a powerful means to expedite high-impact work, unleash creativity, and drive substantial cost and time savings. However, harnessing these benefits requires a solution that meets rigorous enterprise-grade standards. To fully realize the potential of Generative AI, organizations must prioritize the adoption of enterprise-grade solutions. These solutions offer the necessary reliability, scalability, and security required for large-scale deployment. Enterprise-grade technology ensures that Generative AI systems can seamlessly integrate into existing infrastructure, adhere to compliance standards, and protect sensitive data, instilling trust and confidence in their application.

While LLMs like OpenAI's GPT-4 are powerful tools, their training on general internet data often leads to inaccuracies when confronted with specific fact-based questions, making them less attractive to enterprises, particularly the ones with strict compliance and governance requirements. However, companies like Contextual AI offer a solution by building platforms that enable businesses to develop, train, and deploy their own customized LLMs, addressing compliance and privacy concerns while delivering significantly improved performance.

Contextual AI's platform focuses on grounding LLMs in underlying data sources, allowing them to generate factual responses more reliably. By incorporating external sources such as files and webpages, Contextual AI's retrieval augmented generation (RAG) technique enhances the performance of LLMs, providing context-aware responses.

Figure 17: RAG Models for Industry-Specific Q&A Models



Source: Towards Data Science

The RAG model (as depicted above) involves splitting documents into manageable chunks, concatenating relevant chunks within the maximum context length, and comparing questions with document chunks using similarity metrics to identify the most likely answer-containing chunks.

This approach not only ensures better accuracy but also eliminates the need for extensive retraining or fine-tuning when adding data sources—a significant advantage over conventional LLMs. The typical enterprise environment deals with vast amounts of unstructured data generated through various channels like text, images, and video. While conventional LLMs excel in comprehending and extracting insights from this unstructured data, their size, and their resource requirements pose implementation challenges for many enterprises. California-based generative AI start-up Contextual AI's platforms offer a solution by enabling businesses to build and deploy smaller, yet equally performant, LLMs. These models not only provide faster processing and lower latency but also alleviate the burden of infrastructure costs.

Responsible AI

In the rapidly advancing landscape of generative AI, the concept of responsible AI emerges as a compelling path toward a brighter future for humanity. As AI continues to permeate every aspect of our lives, it has become increasingly vital to acknowledge and address its potential impact on society, the environment, and future generations. A whitepaper from IDC, backed by Credo AI, underscore the significant opportunities awaiting B2B enterprises in 2023 through the thoughtful and ethical deployment of generative AI. Organizations that embrace an "AI-first, ethics-forward" approach are projected to experience remarkable business improvements, with an estimated year-on-year growth of 22-29% across various critical metrics, including revenue, customer satisfaction, sustainable operations, profits, and reduced business risks. The survey reveals that business executives eagerly anticipate the adoption of responsible AI, recognizing its potential to foster customer satisfaction, drive sustainability, and enhance profitability.

Major tech giants including Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI, recently made commitments aimed at promoting the responsible AI. Central to these commitments are the principles of safety, security, and trust, which bind the companies to a series of ethical obligations. The tech firms have unanimously pledged to prioritize the safety of AI products, conducting rigorous testing before their public release. Moreover, they have committed to actively guard against potential threats, misuse, and risks associated with AI applications, ensuring the technology is harnessed responsibly. Earning the public's trust is another crucial aspect of their commitment, emphasizing transparency and accountability in their AI endeavors.

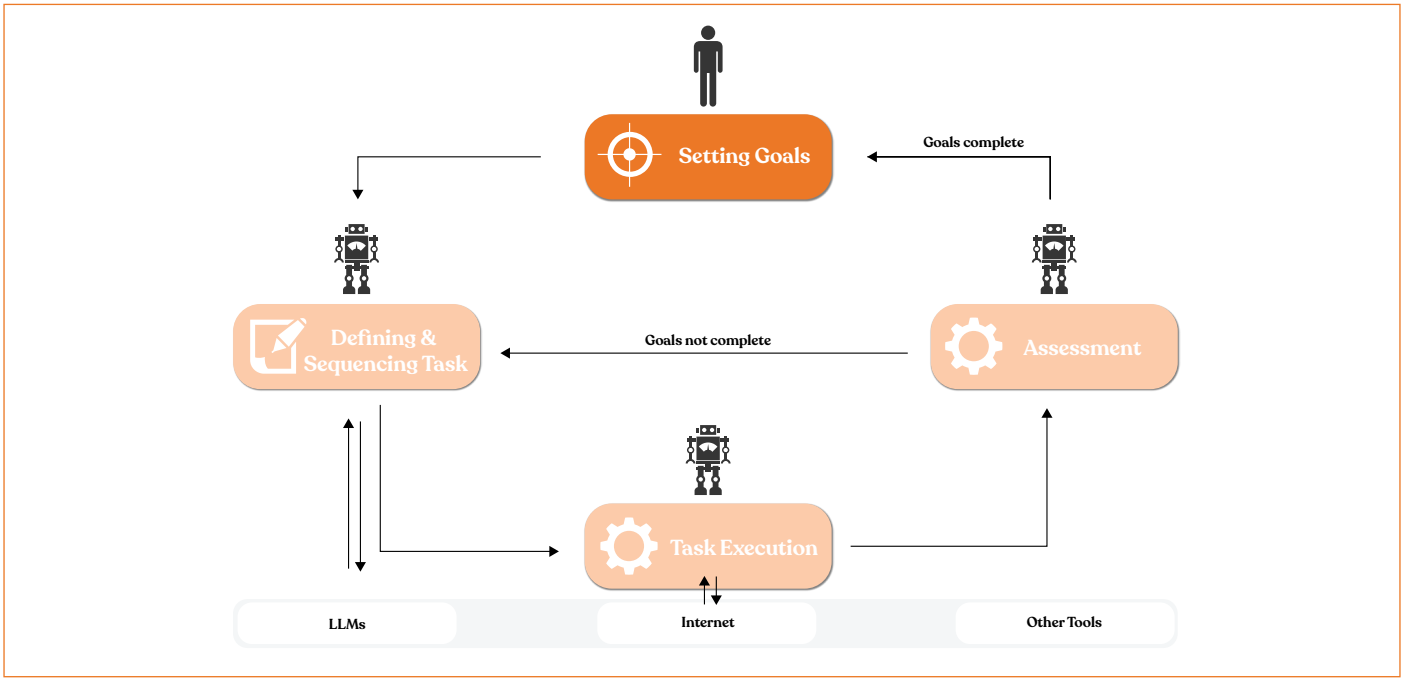
Artificial General Intelligence

Agentic AIs represent the future of AI automation and play a crucial role in advancing artificial general intelligence (AGI). These cutting-edge systems are specifically designed to automate GPT-4 tasks, enabling the development of autonomous agents capable of completing tasks without any human intervention. By leveraging the self-promoting mechanism, agentic AIs effectively break down complex tasks into smaller sub-tasks that can be assigned and executed independently.

The integration of a self-prompting system empowers agentic AIs to plan, justify decisions, and generate detailed task execution. This remarkable capability signifies a significant step towards achieving AGI, where AI-driven systems possess the intelligence to operate autonomously, without relying on human involvement. Agentic AIs utilize various APIs to seamlessly integrate GPT-4 with LangChain, an open-source software tool that facilitates the linking of multiple prompts. Additionally, Pinecone, a vector database, serves as a memory for GPT-4, allowing it to refer back to external documents or its own prior responses to prompts.

The open-source nature of agentic AIs such as AutoGPT and Agent GPT raises concerns for certain well-funded startups dedicated to developing commercial AI assistants. These startups may face a potential threat as AutoGPTs provide a powerful alternative with their advanced capabilities and accessibility

Figure 18: Autonomous AI Mechanism



Source: Manhattan Venture Research

Key Features of Agentic AIs

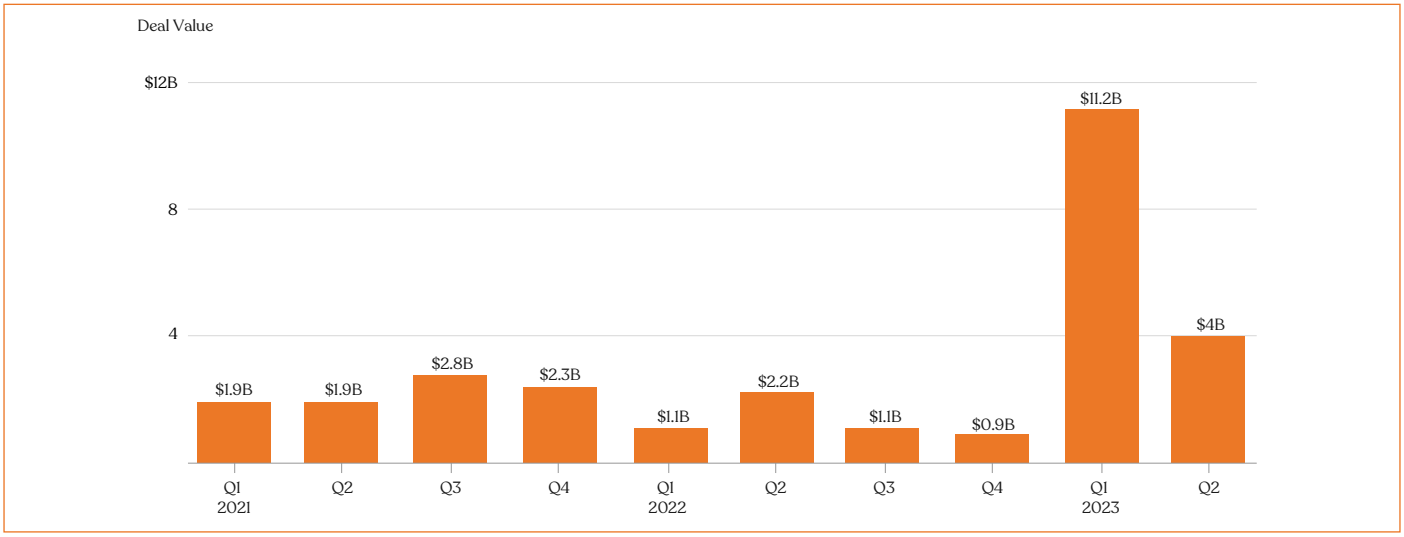
- Assign tasks/goals to be worked on automatically until completed
- Chain together multiple GPT-4s to collaborate on the tasks
- Internet access and ability to read/write files
- Memory to know what has been done

Investment and Funding Landscape

VC Investments in Generative AI

VC investments in generative AI have been on a steady rise from $408 million in 2018 to $4.8 billion in 2021, per Pitchbook. While the funding fell slightly in 2022 to $4.5 billion, funding in 2023 has improved considerably. VCs poured nearly five times more investment into generative AI firms during 1H23 when compared to 1H22. This surge in funding highlights the growing recognition of generative AI's potential to revolutionize various industries and address complex challenges through its wide capabilities. Various companies across the generative AI value chain such as OpenAI, Anthropic, Jasper, Inflection AI, and Cohere benefitted from the overall positive trend around generative AI even while the overall VC funding continued to struggle to gain momentum.

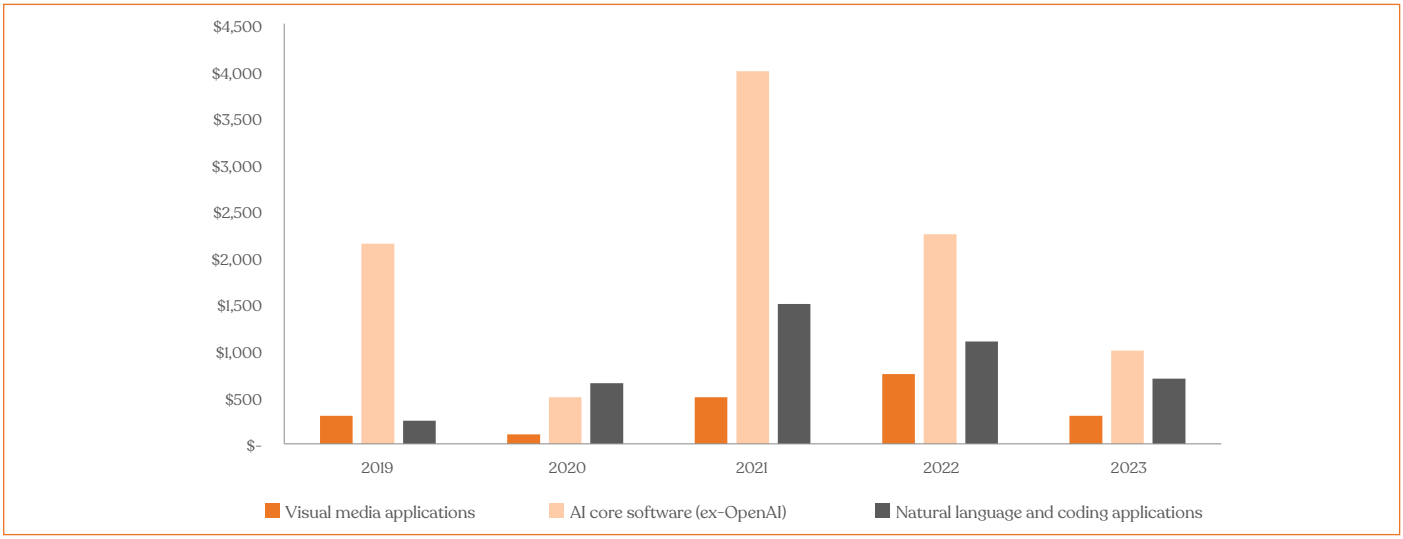
Figure 19: VC Funding Stays Strong in Generative AI



Source: Pitchbook

A notable trend in the global funding landscape since 2019 is the dominance of AI core software start-ups in securing the largest funding share, per Pitchbook. This underscores the growing recognition of the immense potential and transformative power of artificial intelligence across various industries. Start-ups focusing on natural language and coding applications, as well as visual media applications have also garnered significant investor interest. The funding trend also shows a transition from generative applications to infrastructure, which can be attributed to the rising emphasis on building internal tools within enterprises.

Figure 20: Global Generative AI VC funding by Segment ($M)



Source: Pitchbook

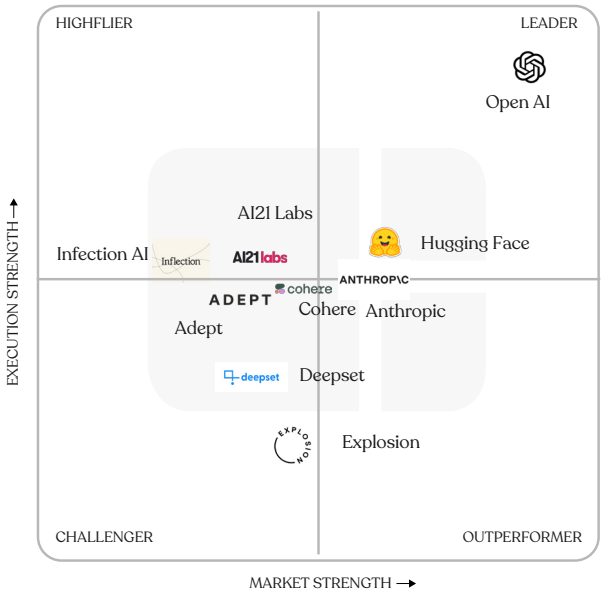
Company Spotlight

In the thriving landscape of the generative AI where innovation and creativity converge, several key companies have emerged as trailblazers, which are shaping the future of this transformative field. While the generative AI space is vast, dynamic, and growing, we have identified eight best-of-breed companies that are in a strong competitive position to benefit from the rapid adoption of the technology:

- OpenAI
- Anthropic
- Hugging Face
- Jasper
- Cohere
- Adept
- Sambanova Systems
- Contextual AI

Each of these startups brings a unique perspective, expertise, and innovative solutions to the table. From state-of-the-art language models and advanced research to democratizing access to powerful AI tools, these companies showcase the dynamic and ever-evolving nature of the generative AI landscape.

Figure 21: Key Generative AI Start-ups

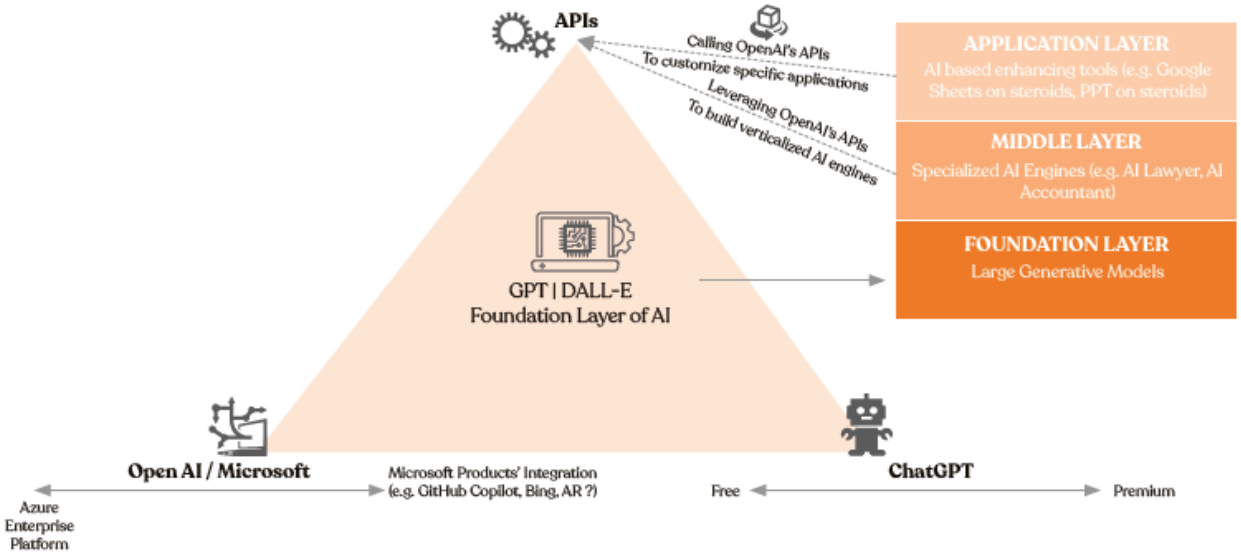


Source: CB Insights

OpenAI

Founded in 2015, OpenAI is a trailblazer in the field of generative AI, standing tall as a beacon of innovation. With the highest total raise in the generative AI space, amounting to a staggering $4.3 billion, this Microsoft-backed startup possesses immense potential to spearhead a paradigm shift across various industries. The company offers a cost-efficient alternative to various white-collar tasks. Moreover, according to Reuters, OpenAI's revenue is projected to hit $200 million this year before climbing to $1 billion in 2024. The viral adoption of OpenAI among users establishes the credibility of the platform.

Figure 22: OpenAI Business Model



Source: FourWeekMBA

Key Positives

Technological Leadership: OpenAI has established itself as a leader in the development of generative models. OpenAI's GPT-4 model can solve difficult problems with greater accuracy, thanks to its broader general knowledge and problem-solving abilities. The precision and accuracy exhibited by this model surpasses that of its peers, placing it in a league of its own within this domain. Its models also have broad applicability across various industries and use cases. From natural language understanding and dialogue systems to computer vision and content generation, OpenAI's technologies can be leveraged in sectors such as healthcare, gaming, entertainment, marketing, and more. The chart below outlines a few parameters for GPT-4 model.

Figure 23: GPT-4: Performance Benchmarking

Benchmark	GPT-4 Evaluate few-shot	GPT-3.5 Evaluate few-shot	LM SOTA Best external LM evaluated few-shot	SOTA Best external model (includes benchmark-specific training)
MMLU Multiple-choice questions in 57 subjects (professional & academic)	86.4% 5-shot	70.0% 5-shot	70.7% 5-shot U-PaLM	75.2% 5-shot Flan-PaLM
HellaSwag Commonsense reasoning around everyday events	95.3% 10-shot	85.5% 10-shot	84.2% LLAMA (validation set)	85.6% ALUM
AI2 Reasoning Challenge (ARC) Grade-school multiple choice science questions, challenge-set	96.3% 25-shot	85.2% 25-shot	84.2% 8-shot PaLM	85.6% ST-MOE
WinoGrande Multiple-choice questions in 57 subjects (professional & academic)	87.5% 5-shot	81.6% 5-shot	84.2% 5-shot PaLM	85.6% 5-shot PALM
HumanEval Python coding tasks	67.0% 0-shot	48.1% 0-shot	26.2% 0-shot PaLM	65.8% CodeT + GPT-3.5
Drop (f1 score) Reading comprehension and arithmetic	80.9% 3-shot	64.1% 3-shot	70.8% 1-shot PaLM	88.4% QDGAT

Source: OpenAI

Competent Leadership Team: OpenAI's illustrious trajectory towards a highly valuable future is underpinned by a leadership team with rich and diversified expertise. Noteworthy figures such as Sam Altman, the former president of Y Combinator, bring invaluable strategic insights and a deep understanding of the startup ecosystem. Greg Brockman, the former CTO of Stripe, adds unparalleled technical prowess and a keen eye for innovation. Ilya Sutskever, previously employed by Google, contributes a wealth of expertise in machine learning and artificial intelligence. Mira Murti, a former employee of Tesla, brings to the table invaluable insights into cutting-edge technologies. Lastly, Brad Lightcap, formerly of J.P. Morgan Chase, contributes his astute financial acumen and business insights. With this accomplished leadership team at the helm, OpenAI is poised for remarkable success in the future as well.

Strategic Partnerships and Collaborations: OpenAI has established strategic partnerships and collaborations with leading organizations including Microsoft, Salesforce, Shutterstock, BuzzFeed, and Bain & Company. These partnerships not only provide financial support and resources but also foster knowledge exchange and accelerate research and development efforts. Collaborations with companies like Microsoft and access to the Azure cloud platform have enabled OpenAI to scale its offerings and reach a broader customer base. Strategic partnerships enhance OpenAI's credibility, distribution channels, and market reach, contributing to potential revenue growth.

Robust Customer Base: – OpenAI's platform is robust and scalable, while their user-friendly interface renders them a leader to seize the lion's share of the generative AI market. The model has utility in various industries. For instance, according to HG Insights, 1,327 companies are using OpenAI with a 23.89% growth in 12 months.

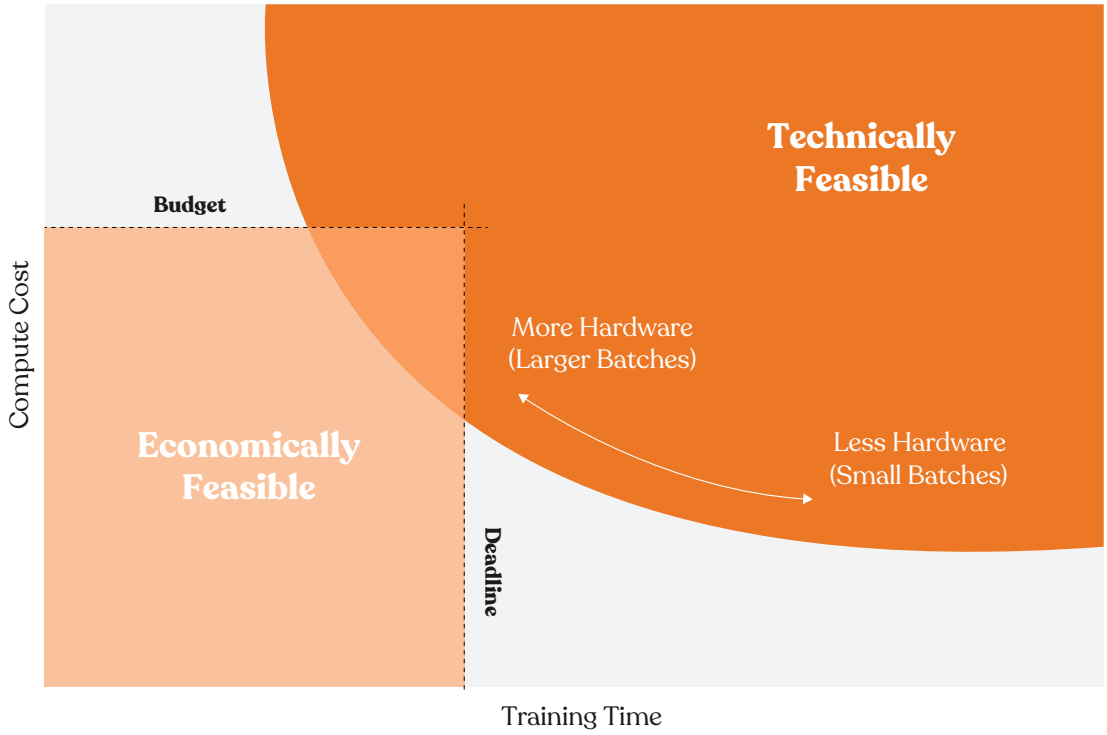
Figure 24: Some Major Companies Leveraging the Power of GPT-4 include

Company	Description
Microsoft	GPT-4 powered chat option is now live on the Bing search engine
Duolingo	GPT 4-powered features are added to this language-learning app
Be My Eyes	GPT-4 model is helping the visually impaired get details about images through this app
Khan Academy	GPT-4 powered chatbot is integrated with this EdTech platform
Stripe	GPT-4 is making digital payments safer for this FinTech startup

Source: Manhattan Venture Research

Declining Model Training Costs : – OpenAI stands to enhance its profit margins in the coming years through a remarkable reduction in training expenditure. For instance, according to the ARK Invest Big Ideas 2023 report, the costs associated with training a language model akin to GPT-3 have descended from a staggering $4.6 million in the year 2020 to a mere $450,000 in 2022 – an annual decline of 70%.

Figure 25: Open AI Model: Compute Cost vs Training Time



Source: OpenAI

Key Concerns

Knowledge Cutoff: The OpenAI model lacks access to the most recent database. For instance, the current source database of ChatGPT is updated until September 2021. Though, the new chat feature by Microsoft Bing, powered by OpenAI's GPT-4, has access to the internet and thereby the latest databases.

OpenAI SWOT Analysis

Strengths
- Largest financial backing in the industry
- Strategic support from Microsoft
- Most precise model in the market

Weaknesses
- Lacks access to the latest data

Opportunities
- Integration with LinkedIn, Facebook, Twitter, etc

Threats
- Authorship concerns on new content
- Cybersecurity risks
- Open source platforms

Source: Manhattan Venture Research

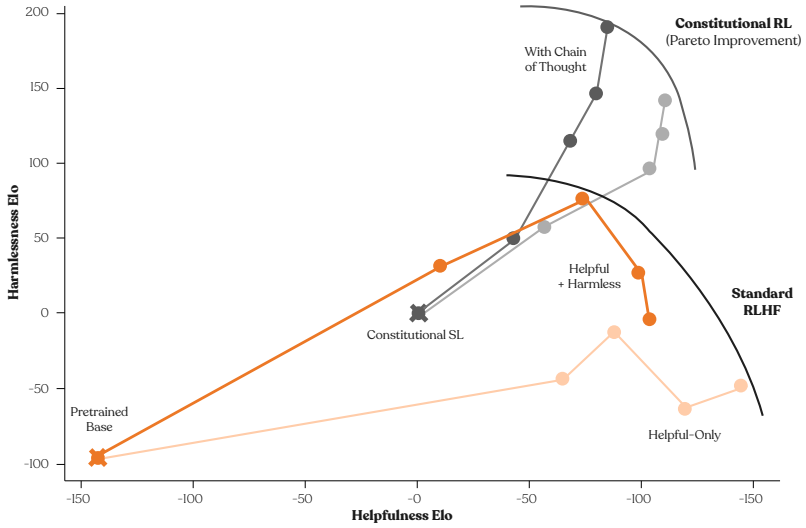
Anthropic

ANTHROP\C

Founded in 2021, Anthropic is an AI safety and research company focusing on a new approach to AI safety that shapes the outputs of AI systems according to a set of principles. Called Constitutional AI, this approach combines both supervised learning and reinforcement learning (RL) to achieve its objectives. In the supervised phase, the initial model is sampled, self-critiques and revisions are generated, and the original model is fine-tuned based on the revised responses. In the RL phase, the finetuned model is sampled, and a separate model evaluates the quality of the two samples to determine which one is better. A preference model is then trained using the dataset of AI preferences. RL is subsequently employed using the preference model as the reward signal, utilizing 'RL from AI Feedback' (RLAIF). This methodology enables the training of an AI assistant that is harmless yet capable of addressing harmful queries by providing explanations for its objections to such queries.

Anthropic's unwavering commitment to responsible AI positions it as a beacon of ethical and moral integrity in an AI landscape fraught with potential challenges such as bias and factual inaccuracies. Anthropic's focus on building reliable, interpretable, and steerable AI systems represents a quantum leap in the AI industry. In a world increasingly reliant on AI, the responsible deployment of this technology is paramount. Anthropic leads the charge in ensuring that AI is harnessed for the betterment of humanity, rather than its detriment.

Figure 26: Anthropic's Constitutional RL Approach



Note: This graph shows harmlessness versus helpfulness Elo scores (higher is better) computed from crowdworkers' model comparisons. It displays a Pareto improvement (i.e., win-win situation) where Constitutional RL is both more helpful and more harmless than standard RLHF

Source: Anthropic

With an impressive fundraising of $2.9 billion and backing from Google and Amazon, Anthropic aims to create AI systems that are not only technically advanced but also ethically and socially responsible. As the demand for ethical and value-aligned AI continues to grow, Anthropic's commitment to bridging the gap between cutting-edge technology and human values sets it apart and positions it for long-term success in the rapidly evolving AI landscape.

Backing from giants such as Google and Amazon on Anthropic's mission and technology speaks volumes about its credibility and potential. These companies recognize the significance of Anthropic's work in shaping the future of AI. Their strategic backing not only provides financial support but also opens doors to synergistic collaborations and partnerships.

Anthropic SWOT Analysis

Strengths

- Sophisticated constitutional model – human oversight is provided through a list of rules
- Strategic support from Google and Amazon
- Most natural conversational experience – well suited for customer support

Weaknesses

- Needs improvement in analysis and reasoning

Opportunities

- Social media integrations
- Growing demand for responsible AI

Threats

- Low entry barrier for new entrants

Source: Manhattan Venture Research

Hugging Face

Established in 2016, Hugging Face is emerging as a well-recognized start-up, offering an exceptional open-source library, which empowers users to construct, train, an d deploy artificial intelligence chat models. Moreover, Hugging Face offers a multitude of datasets, including the popular IMDb movie scores, designed to empower developers in their model training endeavors.

Hugging Face's prominence in the generative AI space is established through the success of its frameworks, including PyTorch and TensorFlow, among others. These tools have ascended to the upper strata of the GitHub ratings, revered by machine learning professionals. Although Hugging Face's financials remain veiled from the public, Forbes estimated that the company generated a remarkable $10 million in revenue in 2021. The predominant source of Hugging Face's revenue stems from licensing fees for the utilization of its models and software. The diversified use-case of its models, ease of testing the model on the website itself for accuracy, and a library of algorithms for all the prediction problems will augment the adoption in the coming years.

Key investment positives of hugging face include partnerships with eminent cloud providers, including Amazon's AWS, and a customer roster ranging from startups like Grammarly to technological behemoths like Google and Salesforce. However, the prerequisite for a high computational infrastructure, such as GPU, to implement Hugging face will hinder the adoption.

Hugging Face SWOT Analysis

Strengths

- A history of successful AI tools such as PyTorch, TensorFlow, etc
- Cash flow positive
- Strategic partnership with cloud providers such as Amazon's AWS
- Diversified customer roster – Grammarly, Google, Salesforce, etc

Weaknesses

- High cost for deployment

Opportunities

- Offering industry-specific solutions

Threats

- Illegal use of the tool

Source: Manhattan Venture Research

Cohere

Founded in 2019, Cohere represents a pioneering force in the generative AI space, with a strategic focus on serving enterprise customers. In a landscape filled with generative AI startups, Cohere's laser focus on catering to enterprise use cases, while providing the freedom to choose preferred cloud providers, positions it as a standout player. This unique selling proposition enhances data privacy and simplifies implementation for discerning enterprise clients.

Figure 27: Benefit from Cohere Integration

Company	Benefit from Cohere Integration
Loom	The support automation rate increased to 51%
OpenPhone	Achieved over 54% self-serve rates with a 4X return on investment
Ramp	Over 60% of in-app support inquiries got automated
Canix	Saving 15 hours a week using Cohere
GrowSurf	Reduced support resolution time by 41%

Source: Cohere

Cohere's strategic use of cutting-edge LLMs built on the Transformer architecture and trained on supercomputers, sets it apart in the fast-growing generative AI industry. Cohere's strategic partnerships with industry giants like NVIDIA and Oracle validate its technology and market potential. These partnerships not only provide financial support but also open doors to integration opportunities within large enterprises. Cohere's product portfolio includes a range of generative AI models accessible through cloud-hosted APIs. These models, akin to ChatGPT, offer versatility for developers and businesses seeking to leverage AI for various applications, from content generation to text summarization and website search.

Cohere's association with reputable clients such as Jasper, LivePerson, and HyperWrite underscores its capability and reliability. These partnerships are a testament to Cohere's capacity to deliver world-leading AI solutions tailored to meet the unique needs of diverse enterprises. Additionally, the rapid increase in users also establishes the credibility of the platform. For instance, Cohere witnessed an 800% increase in users between its Series A round (September 2021) and Series B round (February 2022).

Cohere SWOT Analysis

Strengths

- Strategic investment from tech giants
- Track record with quantifiable benefits to customers
- Experienced leadership team from Coinbase, Ramp, Opendoor, etc
- Partnership with Google's Cloud TPU
- Rapid increase in users

Weaknesses

- High upfront cost for deployment coupled with ongoing costs for inference
- Less mature product than competitors such as Open AI and AI21 Labs

Opportunities

- Understanding unstructured data in various industries such as finance, shipping, legal, healthcare, etc

Threats

- Illegal use of the tool
- Free open-source platforms

Source: Manhattan Venture Research

Jasper

Jasper

Founded in 2021, Jasper stands at the confluence of AI and creative content generation. By offering unparalleled assistance in crafting compelling content for social media, advertising, articles, emails, websites, and art, Jasper has garnered the attention of around 100,000 paying customers by 2022. This remarkable growth reflects not only the market demand but also Jasper's efficacy in delivering value to its customers.

Building upon this remarkable achievement, Jasper concluded the year 2022 with a remarkable 100% increase in annual recurring revenue, showcasing the inherent strength of its business model. Moreover, the introduction of Jasper Art, a cutting-edge text-to-image generation feature, has unveiled exciting creative possibilities for users. Furthermore, the launch of a browser extension enabled seamless integration with various websites and content platforms. Through such initiatives, Jasper has positioned itself to unlock new avenues for revenue growth. Moreover, Jasper acquired Outwrite, an Australia-based AI content startup, utilized by over a million users worldwide. This strategic integration of expertise expands Jasper's offerings and bolsters its customer base.

The stamp of approval from enterprise clients such as HubSpot, Zillow, and Zoom underscores Jasper's suitability for professional and large-scale content generation needs. These high-profile partnerships not only boost Jasper's credibility but also open doors to even larger market segments.

The burgeoning demand for swift and cost-effective content marketing positions Jasper at the forefront of this rapidly evolving industry. For reference, industry projections indicate that the revenue generated by the content marketing sector is poised to exceed a staggering $135 billion by 2026, as reported by AI Multiple. However, it is important to note that while Jasper excels at producing concise content pieces, its application for longer projects has demonstrated some limitations. As a result, its adoption has primarily been focused on swiftly generating social media blurbs or meta descriptions for blog posts.

Jasper generated an estimated $75 million in 2022, a 1.75x growth over 2021 – as per Sacra. A strong year-over-year revenue growth cements the company as a compelling investment opportunity.

Jasper Swot Analysis

Strengths

- Growing paid subscriber base• Cash flow positive
- Rapidly growing revenue

Weaknesses

- Strong competition from Copysmith, Grammarly, Canva, Notion AI, Copy.ai, Dall-e, etc.

Opportunities

- Solutions for the entertainment industry such as comics, cartoons, etc

Threats

- Intellectual property issues

Source: Manhattan Venture Research

CoreWeave

Founded in 2017, CoreWeave has solidified its position in the field of cloud infrastructure with its fine-tuned models. The company claims 35x faster and 80% less expensive solutions than legacy cloud providers such as AWS, Google, and Microsoft. The company also boasts over 1000 customers across four key verticals as of January 2023, establishing the credibility of the business model.

CoreWeave has raised $576.5 million through 6 funding rounds at a valuation of $2 billion post its $371 million Series B. Moreover, the introduction of its accelerator program to support startups with compute credits, discounts, and other resources on the CoreWeave cloud has positioned it to unlock new avenues for revenue growth.

Key investment positives of CoreWeave include a strong roster of partners and investors including Nvidia, former GitHub CEO, Nat Friedman, and ex-Apple exec, Daniel Gross, establishing the credibility of the technology. Moreover, a reasonable pricing strategy, allowing customers to pay only for the resources they use and while they are using them, creates a competitive edge for Coreweave.

CoreWeave SWOT Analysis

Strengths

- Strong backing by key strategic partners
- Reasonable pricing
- Offers solutions that are 35x faster and 80% less expensive than legacy cloud providers

Opportunities

- Solutions for AI start-ups looking for cheaper and more efficient cloud providers

Threats & Weaknesses

- Strong competition from AWS, Google, and Microsoft.

Source: Manhattan Venture Research

Sambanova Systems

Sambanova Systems, an enterprise-ready AI platform, and AI chips provider has solidified its position since its establishment in 2017. Financial backing from investors such as Softbank, Intel, Micron, and Samsung, establishes the credibility of the platform. Moreover, the introduction of Datascale, a platform for generative AI development and innovation, has unveiled exciting creative possibilities for users. Furthermore, the launch of SambaNova Suite has enabled seamless integration either on the ground or through the cloud for accurate generative AI models with enterprise data for greater accuracy. Through such initiatives, Sambanova has positioned itself to unlock new avenues for revenue growth. This extensive product offering expands and bolsters Sambanova's customer base.

The burgeoning demand for AI chips positions Sambanova at the forefront of this rapidly evolving industry. According to Allied Market Research, the global market for AI chips will expand from $11.2 billion in 2021 to $263.6 billion by 2031, at a CAGR of 37.1% from 2022 to 2031. Moreover, US-based AI chip startups are set to benefit from the US government's focus on reducing reliance on China, exemplified by the $280 billion CHIPS Act that aims to revive domestic chip manufacturing. For reference, ChatGPT was trained using 10,000 of Nvidia's graphics processing units (GPUs) clustered together.

Key investment positives of Sambanova include technologically advanced products. For instance, the Cardinal SN30 accelerator by Sambanova claims better performance than NVIDIA's A100 while SN30-based DataScale systems are claimed to be six times faster than Nvidia's DGX A100 servers. Moreover, recent hires of former AWS Managing Director and Google Cloud VP also fuel the bright future of the company.

Sambanova SWOT Analysis

Strengths

- Extensive product offerings
- Technologically superior products

Opportunities

- Rapidly growing demand for AI chips

Threats & Weaknesses

- Intensifying competition from NVIDIA, AMD, Groq, etc

Source: Manhattan Venture Research

Contextual AI

Contextual.ai, an AI platform specifically for enterprises, offers an enterprise-ready AI platform that augments LLMs with external sources, like files and webpages, to improve their performance. They have garnered the attention of investors like Bain Ventures, Lightspeed, Greycroft, and SV Angel thereby establishing the credibility of the business model. Contextual AI has raised $20 million in seed funding, showcasing the inherent need for enterprise-friendly LLMs. Moreover, their foundational models are customizable, less hallucination-prone, and privacy-aware.

Key investment positives of Contextual AI include a robust founding team with experience from Facebook AI Research and Hugging Face. Moreover, the rapidly growing generative AI marketing creates ample opportunities for startups in this space.

Contextual AI Swot Analysis

Strengths

- Renowned pool of investors• Technologically superior products
- Privacy-aware AI model
- Founding team of ex-Facebook, ex-Hugging Face, etc

Opportunities

- Solutions for the enterprise industry scared of data leakage

Threats & Weaknesses

- Intensifying competition from Anthropic, Cohere, and OpenAI

Source: Manhattan Venture Research

About Manhattan Venture Partners

Our Research Methodology

Manhattan Venture Research provides clients with accurate, timely, and innovative research into the companies and sectors we cover. To that end, we have established an experienced team of analysts, researchers, economists, and industry veterans that focus exclusively on private companies with a proven track record of success. Producing quality research on a private company is uniquely challenging. Our analysts communicate with ex-employees, early investors, VCs, competitors, suppliers, and others to gather valuable information about the company under coverage. This information enables us to create unique financial models that value the underlying company and provide insight to our clients and industry experts, leveraging years of experience working for bulge bracket firms. Manhattan Venture Research reports include business and financial aspects of late-stage companies. These reports include but are not limited to industry overviews, competitor analysis, SWOT analysis, products (existing and in development), management and key directors, risks and concerns, other proprietary channels, historical financials, revenue projections, valuations (using various matrices and valuation recommendation), waterfall analysis, and a capitalization table.

Information Access Level Classification System (IALCS)

Manhattan Venture Research uses an Information Access Level Classification System (IALCS) to make clear the degree of access offered by the company(s) covered in all research reports.

Each research report is classified into one of three categories depending on its classification. The categories are:

I++: The company covered by the research report provided substantial disclosures to Manhattan Venture Partners.

I+: The report was prepared following partial disclosure by the company, including publicly available financial statements, and/or is based on conversations with past or present company employees.

I: All reports are prepared using a mosaic research approach. Not all companies are willing and able to provide substantive access to management and information. In I reports no direct access was granted.

About the Analyst

Santosh Rao

Santosh Rao has over 25 years of experience in equity research with a primary focus on the technology and telecom sectors. He started his equity research career at Prudential Securities and later moved to Dresdner Kleinwort Wasserstein, Gleacher & Co, and Evercore Partners, where he followed Telecom and Data Services. Prior to joining Manhattan Venture Partners, he was the Managing Director and Head of Research at Greencrest Capital, focusing on private market TMT research. Santosh has an undergraduate degree in Accounting and Economics, and an MBA in Finance from Rutgers Graduate Business School. While at Gleacher & Co he was ranked leading telecom equipment analyst by Starmine/Financial Times

Disclaimer

I, Santosh Rao, Head of Research, certify that the views expressed in this report accurately reflect my personal views about the subject, securities, instruments, or issuers, and that no part of my compensation was, is, or will be directly or indirectly related to the specific views or recommendations contained herein.

Manhattan Venture Research is a wholly-owned subsidiary of Manhattan Venture Holdings LLC ("MVP"). MVP may currently and/or seek to do business with companies covered in its research report. 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. This document does not contain all the information needed to make an investment decision, including but not limited to, the risks and costs.

Additional information is available upon request. Information has been obtained from sources believed to be reliable but Manhattan Venture Research or its affiliates and/or subsidiaries do not warrant its completeness or accuracy. All pricing information for the securities discussed is derived from public information unless otherwise stated. Opinions and estimates constitute our judgment as of the date of this material and are subject to change without notice. Past performance is not indicative of future results. MVP does not engage in any proprietary trading or act as a market maker for securities. The user is responsible for verifying the accuracy of the data received. This material is not intended as an offer or solicitation for the purchase or sale of any financial instrument. The opinions and recommendations herein do not take into account individual client circumstances, objectives, or needs and are not intended as recommendations of particular securities, financial instruments, or strategies to particular clients. The recipient of this report must make its own independent decisions regarding any securities or financial instruments mentioned herein. Periodic updates may be provided on companies/industries based on company-specific developments or announcements, market conditions, or any other publicly available information.