

Minds and Machines

Disruptive Tech White Paper

NAVIGATING 2024'S AI REVOLUTION: TRIUMPHS, TRIALS, AND TRANSFORMATIONS

INTRODUCTION

As 2024 dawned, both hope and fear permeated global society in its perception of Artificial Intelligence (AI). There were those who hoped we would create Artificial General Intelligence (AGI), long the “Holy Grail” of the industry. When that moment occurred, they said we would have machines able to do anything a human could do, and even do it better. Other observers feared the same possibility, wondering what it would mean to be human if our planet suddenly harbored an intelligence equal to our own. The conversation was not binary, however. Other groups asserted that we would never achieve AGI, or that it would not happen in the near future, while their detractors claimed that existing AI systems already represented AGI. Almost everyone agreed that, whatever happened, AI would evolve rapidly, perhaps exponentially. Safety issues reached a fever pitch, in part because OpenAI employees charged with ensuring that the company produced safe AI “for the benefit of humanity” began leaving in droves. Prominent among them was Ilya Sutskever, who eventually surfaced after a long period in stealth mode. He had participated in the attempted coup to replace Sam Altman as CEO of OpenAI in mid-2023, and had more or less disappeared when that effort failed. Sutskever announced in mid-2024 that he had launched a startup called Safe Superintelligence, Inc and quickly raised several billion dollars in funding.

According to Sutskever, his new company would be unique in two ways: first, it would skip the AGI stage altogether and build “**superintelligence**,” a system that would not simply match human capabilities but would go far beyond us into a realm that is difficult to imagine. Second, he said that safety would be built into the AI from the beginning, rather than creating something very powerful and then going back to fix the safety issues its existence raised. Sutskever’s bold pronouncements are intriguing, but one wonders if “safety” is even a meaningful term when discussing a species with transcendent intelligence versus one with significant intelligence, but at a much lower level. If he succeeds, we will find out! **OpenAI, seemingly unperturbed by all the critical commentary, continued to enhance its offerings, turning out several new models, building on its flagship ChatGPT product.** Spurred on by the concerns about safety, regulation became an issue in 2024, with Europe enacting the “Artificial Intelligence Act” in March. The year also saw a passionate debate over “open” vs. “closed” AI systems. Despite its name, OpenAI led the fight against the open paradigm, calling it dangerous to let powerful AI code be accessible to anyone from school children to terrorists. Meta led the opposition, asserting that AI was too important to be overseen by a few Big Tech companies.

In the meantime, companies competing with OpenAI, like Microsoft, Google, Meta, Apple, and Anthropic brought out their own versions of AI. Anthropic’s Claude and Microsoft’s Copilot seemed especially to garner cadres of passionate followers. Apple build on a clever juxtaposition of letters to create their own brand of AI, “**Apple Intelligence**,” now available in iPhones, a huge global market. Becoming “multimodal” represented another significant evolution of the technology. As this report notes in subsequent pages, the industry moved from a “text-in, text-out” capability to an ability to integrate text, data, images, audio, and video. Another front was opened in the intelligence wars, as robots came into their own. For many years, robot manufacturers had been evolving their products’ physical capabilities. However, like the scarecrow in The Wizard of Oz, they could have done so much more “if they only had a brain.” In 2024, it became apparent that many of them did have a brain, as they became “AI-enabled.” Now, they are not just robots; they are “**humanoid robots**,” moving rapidly toward becoming the stuff of science fiction, androids. These are robots that not only have the intelligence of humans but also look like us, to the point that we will find it difficult to distinguish between them and ourselves. We aren’t there yet, but that day is likely on the horizon.

On the safety front, new concerns arose in 2024 about AI’s use in warfare. Weapons systems of all kinds are going to become more lethal if they are AI-enabled. From the drones of today to the robot soldiers of tomorrow, AI is going to be

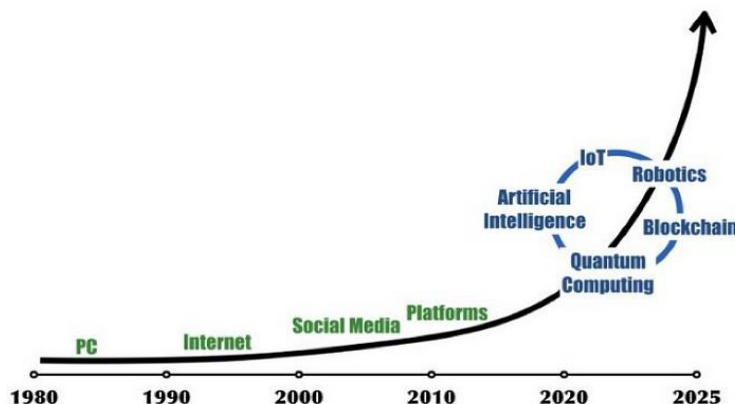
active on the battlefield, and the time has come to decide what limits should be placed on them. As the year roared toward its conclusion, the narrative around AI suddenly shifted. Some writers noted that huge investments had been made in the technology, but had not yielded much in the way of revenues. Others pointed out that **Large Language Models (LLMs)** might be reaching the limits of innovation. Since the advent of ChatGPT in 2022, the industry mantra had been “bigger is better,” with a voracious desire for more “compute.” However, this path led to more energy — intensive data centers, raising concerns about environmental damage. Moreover, the most recent offering of OpenAI disappointed many users, leading analysts to ponder whether LLMs, in general, might contain an inherent flaw that would require pivoting to some other design of AI systems across the board. So, there was great progress and innovation in 2024, along with concern and disappointment.

Predicting what will happen in 2025 is going to be challenging, because events are moving so rapidly. We should keep an eye on OpenAI, which will be shedding its nonprofit status completely while it is also being sued by Elon Musk. In addition, it makes sense to listen to those who claim to have a better path forward than LLMs. Finally, robotics will become more interesting and more important. In particular, it may begin to dawn on space companies that AI-enabled robots can be essential partners as humanity moves off-world. Please read on to gain additional insight into what 2025 might bring and where investors might look for opportunities.

LOOKING BACK AT 2024

The year 2024 was marked by significant transformations and pivotal innovations in the tech and IT sectors, particularly in AI. As we reflect on this year, it becomes evident that the industry has experienced significant achievements alongside challenging realities, navigated remarkable achievements alongside sobering problems, setting the stage for future technological advancements. **2024 began with unprecedented optimism surrounding AI, with GenAI dominating headlines and organizations racing to integrate these technologies into their workflows.** Promising applications spanned across industries from healthcare to finance, fueling a wave of investment and enthusiasm. However, midway through the year, a dose of reality emerged, as tech leaders like Elon Musk, Dario Amodei, Mark Zuckerberg, Sundar Pichai, and Sam Altman raised alarms about the reliability of AI models, ethical dilemmas, and economic sustainability.

Chart 1: Exponential Growth of Technology



Source: Intro-act, Intro-act, Medium

The integration of AI into real-world organizations often fell short of lofty expectations because of challenges such as algorithmic bias, hallucinations, and escalating costs of data and computational resources. Meanwhile, AI's dual-edged potential in cybersecurity became increasingly apparent. While it was celebrated for its speed in detecting and mitigating threats, malicious actors exploited AI for sophisticated attacks, automating phishing campaigns, and identifying vulnerabilities on an unprecedented scale. These developments highlighted the pressing need for strategic, regulated,

and balanced approaches to AI adoption. **As ChatGPT celebrates its second anniversary, we believe that responsible innovation will be key to harnessing AI's transformative potential while mitigating risks.** Cybersecurity reached a crisis point in 2024, with high-profile attacks eroding confidence in both public and private sectors. Incidents such as the cyberattack on Seattle-Tacoma International Airport and widespread outages stemming from a failed patch update at CrowdStrike underscored the urgent need for stronger defenses and robust disaster recovery planning. **Amid these challenges, compliance frameworks like Cybersecurity Maturity Model Certification (CMMC) 2.0 gained traction, becoming the de facto standard for organizations conducting business with the government, starting in December 2024.** This updated framework represents a shift from traditional checkbox compliance to proactive, threat-aligned defense strategies. We think that organizations will increasingly embrace such measures not only for regulatory adherence but also for safeguarding against increasingly sophisticated cyber threats.

As President-Elect Donald Trump prepares to take office in January 2025, speculation is mounting about significant regulatory shifts that could impact the tech industry. His incoming administration is expected to prioritize efforts to curtail Big Tech monopolies, signaling a renewed focus on fostering competition and addressing antitrust concerns. The Cybersecurity and Infrastructure Security Agency (CISA) is undergoing a leadership transition following Jen Easterly's departure, raising questions about the future direction of cybersecurity strategy in the public sector. Early indications suggest that the administration may streamline CISA's mission and relax AI regulations to encourage innovation while balancing national security concerns. Prominent industry figures like Elon Musk and Vivek Ramaswamy are advocating for government efficiency initiatives leveraging technologies such as automation and blockchain. We believe these efforts could push the boundaries of traditional strategies, potentially reshaping how technology is integrated into government functions.

NAVIGATING THE AI WAVE: A DEEP DIVE INTO TECHNOLOGY EQUITIES IN 2025

As we approach the midpoint of this decade, the technology sector's trajectory remains a focal point for investors. Portfolio Manager Alison Porter shares our team's bullish outlook on tech stocks in 2025, with GenAI acting as a powerful tailwind. However, she cautions that disruptive tech waves are typically lengthy and volatile, raising questions about the longevity of the current AI wave and whether we've reached peak hype. We believe it's crucial to differentiate between compute waves and themes like cybersecurity or electric vehicles. A compute wave necessitates broad investment across the entire technology stack, from silicon building blocks to user interfaces and applications.

Chart 2: AI-Focused ETFs Have Been Exhibiting Performance in Line with the Broader Market Trend



Source: Intro-act, Trading View, Data as on December 06, 2024

Historical parallels suggest that infrastructure must be built and scaled before innovative applications can flourish. The unprecedented capital expenditure in AI data centers and surging demand for accelerated computing power, exemplified by Nvidia's growth, indicate that we're still in the early stages of this GenAI wave. We anticipate significant developments across various sectors, including resource optimization, financial systems, autonomous capabilities, healthcare diagnostics, humanoid robots, and upgraded devices. As the GenAI wave matures, disruption across various sectors will accelerate, much like past tech waves. **Technology companies will continue to leverage their balance sheet strength to invest heavily in future research and development, supporting attractive returns for investors.** We remain excited about technology equities and continue to focus on identifying leaders across the sector by navigating the hype cycle and emphasizing stock fundamentals. Here are three among the many themes that we believe stood out in 2024:

MULTIMODAL AI: THE NEXT FRONTIER IN INDUSTRY'S AI TRANSFORMATION

In 2024, multimodal AI emerged as a powerful paradigm shift from previous unimodal AI systems. Its dual capability sets it apart from its predecessors, offering industries a transformative tool for enhanced decision-making and customer interaction. The multimodal models are designed to process and integrate information from multiple data types such as text, images, audio, video, and other sensory inputs. This capability enables these models to achieve a more comprehensive understanding of data and generate robust outputs. OpenAI's ChatGPT, launched in November 2022, initially demonstrated the potential of GenAI with its unimodal text input and output capabilities. However, it was OpenAI's subsequent multimodal implementations like GPT-4 that truly unlocked the power of GenAI by allowing multiple types of inputs and outputs. Imagine an AI system that can read, write, see, hear, and create across multiple modalities all at once — that's the essence of multimodal AI. At its core, multimodal AI's power lies in its ability to process and generate content across different forms such as text, images, audio, and video.

We believe that multimodal AI's ability to combine information across various data sources and media will revolutionize human-computer interaction. Imagine a virtual assistant that understands voice commands and visual cues simultaneously, or an app that identifies birds by analyzing both images and audio clips. These intuitive interfaces promise smoother, more efficient user experiences. The engineering challenge in multimodal AI lies in effectively integrating diverse types of data to create models that leverage the strengths of each modality while overcoming their limitations. Researchers are tackling this complexity through specialized neural networks for feature extraction, joint embedding spaces or attention mechanisms for representation learning, and advanced techniques like temporal and spatial alignment.

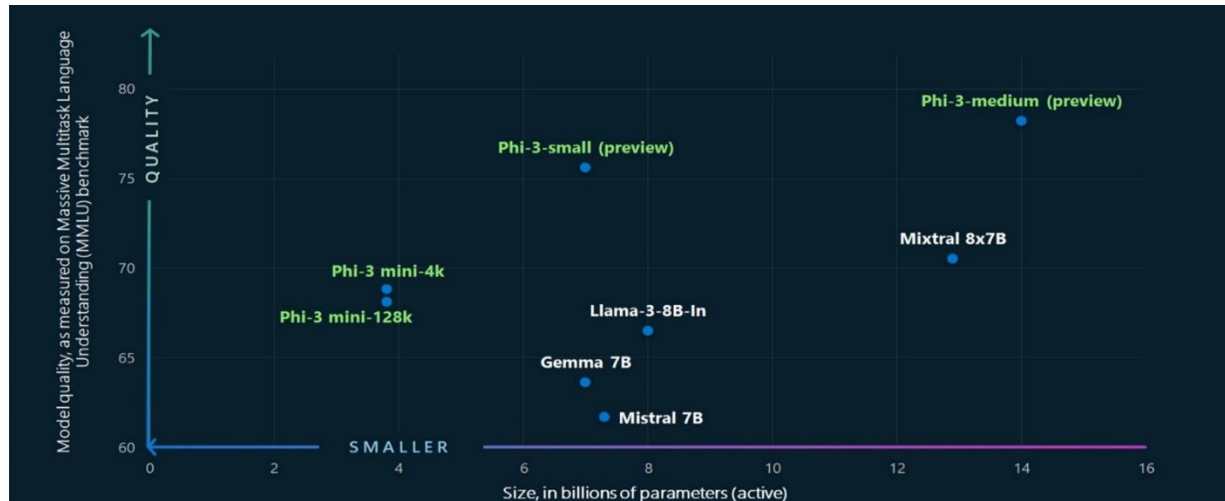
THE RISE OF SMALL LANGUAGE MODELS: A PARADIGM SHIFT IN AI EFFICIENCY

In the rapidly evolving landscape of AI and machine learning (ML), a compelling trend has emerged: the ascendancy of small language models (SLMs). While the tech industry has been captivated by the race to build and deploy large, complex models with billions and trillions of parameters, consumers have unwittingly become accomplices in this pursuit of "large." However, recent developments indicate a growing interest in smaller, more efficient models. This democratization of AI has spurred the development of "on-device AI," and has brought sophisticated natural language processing (NLP) into the mainstream. Within a week of its launch, ChatGPT garnered over a million users, demonstrating that AI is no longer confined to research and academic circles. This phenomenon has led to the integration of AI as copilots in various products, the rising interest behind SLMs, and the proliferation of startups focused on this space.

To understand the magnitude of this shift, let's examine the parameters of AI models. The number of parameters serves as a core measure of an AI's scale and complexity. GPT-2 had 1.5 billion parameters, but OpenAI's subsequent release of GPT-3 boasted a whopping 175 billion parameters—the largest neural network ever created at that time. Today, we're seeing trillion-parameter large language models (LLMs). SLMs, typically defined as models with approximately 10 billion parameters or fewer, are lightweight neural networks designed to process natural language with fewer computational resources than their larger counterparts. Unlike LLMs, which are generalized models, SLMs are usually purpose-driven

and tailored to specific tasks or applications. Additionally, **recent studies have shown that SLMs can be fine-tuned to achieve comparable or even superior performance compared to LLMs in specific tasks.** For instance, phi-3-mini—a 3.8 billion parameter SLM trained on 3.3 trillion tokens—rival's bigger models like Mixtral 8x7B and GPT-3.5 in terms of overall performance.

Chart 3: Quality v/s Size in Small Language Models (SLMs)



Source: Intro-act, Capgemini, Microsoft

It's not about choosing one over the other—LLMs are generalists with extensive knowledge across various subjects but come at a high operational cost. SLMs, on the other hand, are specialized and powerful but struggle with complex context if not used within their specific use case. The choice between SLM and LLM depends on the need and availability of resources. Nevertheless, **SLMs are undoubtedly a game-changer in the AI era, offering efficiency, sustainability, economic viability, and accessibility that will shape the future of AI.**

HARNESSING GENERATIVE AI AGENTS: THE NEXT WAVE OF PRODUCTIVITY AND INNOVATION

The world has been captivated by the capabilities of GenAI, with LLMs performing impressive feats across various media. However, we are on the cusp of an even more transformative stage—one that moves from knowledge-based tools to GenAI-enabled agents capable of executing complex workflows independently. These agentic systems, powered by foundation models like LLMs, promise a new era of productivity and collaboration. Traditional agentic systems have been challenging to implement due to their reliance on laborious rule-based programming or highly specific ML training. However, the advent of GenAI has changed this landscape. Foundation models, trained on vast amounts of unstructured data, enable agents to adapt to different scenarios much like LLMs respond intelligibly to prompts they haven't been explicitly trained on. Moreover, using natural language instead of code allows users to direct agent systems more intuitively, making complex workflows accessible even to non-technical employees.

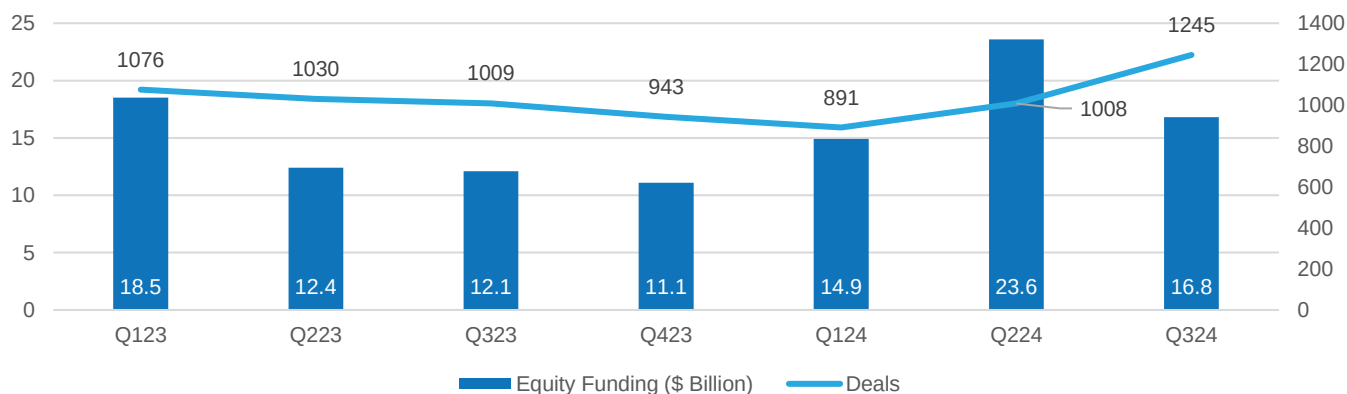
Major tech players such as Google, Microsoft, and OpenAI have already begun investing in software libraries and frameworks supporting agentic functionality. Applications like Microsoft Copilot, Amazon Q, and Google's upcoming Project Astra are shifting toward action-based models, while companies like Adept, crewAI, and Imbue are developing agent-based models and multiagent systems. Given the rapid pace of GenAI development, agents could soon become as ubiquitous as chatbots are today.

THE AI INVESTMENT LANDSCAPE: DOES IT REMIND YOU OF THE DOTCOM BOOM?

The tech investing landscape has undergone a seismic shift with the advent of AI, compelling investors to adopt a more macroeconomic perspective. Historically, tech investments have been driven by product features and growth strategies, with little regard for aggregate growth or inflation. However, AI's capital-intensive nature is forcing investors to reassess this approach. **The projected spending on AI infrastructure by big tech firms in 2025 alone—over \$200 billion—is almost double their 2021 expenditure, and nearly entirely dedicated to developing GenAI capabilities.** This shift toward hardware-intensive investments highlights a key difference from the prior high-tech boom: while software was nimble and scalable with low fixed costs, AI requires substantial capital outlay and time to build infrastructure. If economic conditions deteriorate, tech executives may reconsider these ambitious spending plans, complicating investors' return calculations.

The rise of GenAI has also changed the startup narrative. Unlike their software counterparts, AI startups are highly capital-intensive and sensitive to market conditions. **In the first half of 2024, investments in AI and ML accounted for nearly half of all VC funding in the U.S. Notably, OpenAI raising a staggering \$10.6 billion in October 2024 alone.** While these investments have been made during favorable economic conditions, a downturn could make it challenging for startups to fund their ambitious visions, potentially slowing growth and innovation in the broader AI ecosystem and reducing demand for big tech's AI infrastructure. Moreover, hardware businesses exhibit more cyclical characteristics than software due to their reliance on substantial inputs and manpower. For instance, Nvidia has doubled its product release cadence, subjecting it to traditional inventory cycles influenced by economic conditions. Semiconductor sales have historically been correlated with manufacturing PMIs, but this relationship began breaking down in 2022 as AI demand surged. If historical patterns hold, the global semiconductor boom may be due for a correction.

Chart 4: Global AI Equity Funding and Deal Size



Source: Intro-act, CB Insights

OPENAI'S RAPID ITERATION IN GENERATIVE AI: A YEAR IN REVIEW

OpenAI has continued its meteoric rise in the world of AI, solidifying its position as one of *Fast Company's* Most Innovative Companies of 2024. Ranking ninth on the list of the World's 50 Most Innovative Companies, OpenAI has shown no signs of slowing down, especially with its groundbreaking work in LLMs. We believe that their relentless pursuit of innovation is not only pushing the boundaries of what's possible but also setting new standards for the industry. The release of GPT-4 last March was a testament to OpenAI's commitment to advancing AI capabilities. Ten times more powerful than its predecessor, GPT-4 demonstrated improved context discernment and nuance understanding, delivering more accurate and coherent responses. Its multimodal nature allowed users to interact with it using both text

and images, further enhancing its practicality. We think that topping the Massive Multitask Language Understanding (MMLU) benchmark across 57 subject areas and maintaining its State of the Art (SOTA) designation for most of the year is a remarkable achievement.

However, OpenAI didn't stop at GPT-4. In November, they unveiled GPT-4 Turbo, which can process about 300 pages of text in a single prompt and has knowledge of more recent information than its predecessor. This update not only expands the model's capabilities but also makes it significantly less expensive for developers to access, potentially democratizing AI innovation. We believe this strategic pricing could lead to wider adoption and more diverse use cases. Further, OpenAI's impact extends beyond just model improvements. **With over 2 million developers building on their API and more than 90% of Fortune 500 companies using their products, OpenAI is becoming the go-to platform for AI integration.** Moreover, their launch of new products like ChatGPT Plus, GPT 4o, ChatGPT Enterprise, and customizable GPTs in the GPT Store shows a clear understanding of different user needs. We think this tailored approach will help OpenAI maintain its competitive edge. Yet, such rapid iteration comes with its challenges. The destabilizing effects of GenAI across various sectors, including business and government, are evident. The surprise firing (and reinstatement) of OpenAI CEO Sam Altman late in 2023 is a stark reminder of the turmoil that can arise from the fast-paced development and deployment of such powerful technologies.

AI INVESTMENT LANDSCAPE IN Q3'24: SOARING DEAL COUNT AMIDST FUNDING FLUCTUATIONS

In Q3 2024, the global AI investment landscape witnessed a significant surge in deal activity, bucking the broader venture market trend. We believe this stark contrast indicates investors' confidence in AI's long-term potential. **The global AI deal count skyrocketed by 24% QoQ, reaching an impressive 1,245 deals—the highest level since Q1 2022.** This robust growth occurred despite a 10% QoQ decline in overall venture capital deals, which hit their lowest point since 2016/2017. However, while deal activity flourished, global AI funding value experienced a notable downturn. Total funding dropped by 29% QoQ to \$16.8B, primarily driven by a substantial 77% decline in funding from \$1B+ AI rounds QoQ. This discrepancy between soaring deals and decreasing funding suggests that investors are increasingly favoring strategic investments over large, risky bets. We think this shift could indicate a more cautious approach to AI investing amid economic uncertainties.

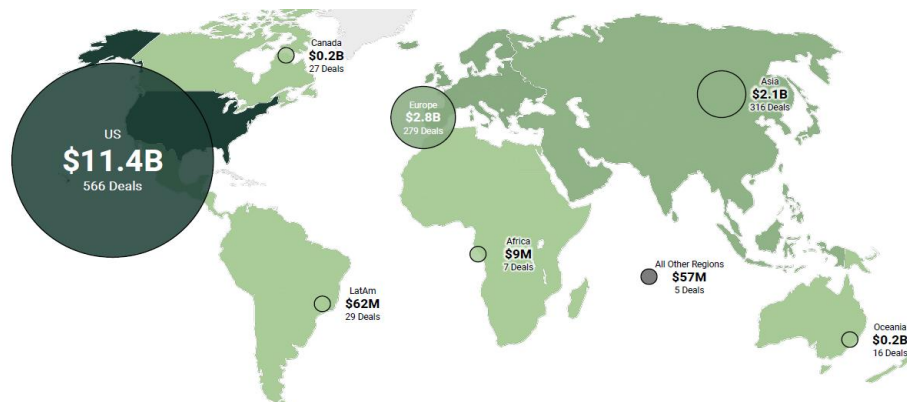
Chart 5: The Top 10 Companies by Funding in 2024 to Date

Company Name	Amount (\$ Billion)	Latest Valuation (in \$ Billion)
Anthropic	7.7	40
xAI	6.6	24
OpenAI	6	157
G42	1.5	-
Anduril	1.5	14
Moonshot AI	1.3	3.3
CoreWeave	1.1	19
Scale	1	13.8
Wayve	1	-
Safe Superintelligence	1	5

Source: Intro-act, CB Insights

The average AI deal size has been on an upward trajectory in 2024, with a YTD figure of \$23.5M—up 28% compared to the full-year 2023 average of \$18.4M. This trend is partly attributable to the rise in massive \$1B+ deals, with nine such deals closed in 2024 YTD compared to four in full-year 2023. Notable rounds include xAI's \$6B Series B at a \$24B valuation and Anthropic's \$2.8B Series D at an \$18.4B valuation. We believe this trend reflects investors' willingness to back established AI startups with proven track records. However, it is essential to note that the median deal size has also increased by 9% in 2024 YTD, indicating a broader shift in investment patterns.

Chart 6: Global AI Funding & Deals by Region in Q3'24



Source: Intro-act, CB Insights

The AI unicorn births more than doubled QoQ in Q3 2024, reaching 13—a staggering 54% of the broader venture total for the quarter. GenAI continues to be a prominent theme among new unicorns, with over half working across various sectors such as 3D environments, code generation, and legal workflow automation. Safe Superintelligence, co-founded by OpenAI co-founder Ilya Sutskever, led the pack with a \$5B valuation after raising a \$1B Series A round in September 2024. We think this strong showing of GenAI unicorns underscores investors' enthusiasm for cutting-edge AI applications. Meanwhile, AI mergers and acquisitions (M&A) exits fell by 48% QoQ to hit 62 in Q3 2024. Despite the decrease, strategic acquisitions remained prevalent, with AMD's acquisition of Silo AI and Salesforce's pickup of Zoomin highlighting enterprises' continued pursuit of competitive advantages through AI integration. Regionally, the United States maintained its dominant position in AI funding and deals, accounting for over two-thirds of global AI funding (\$11.4B) and 45% of global AI deals (566) in Q3 2024. While Silicon Valley remains the epicenter of U.S. AI investment, other metros like Los Angeles and New York are gaining traction, with both regions experiencing QoQ deal count increases while Silicon Valley saw a consecutive quarterly decline.

AI is driving significant M&A activity across the U.S. tech sector. Notable deals include HPE's acquisition of Juniper Networks, Nvidia's strategic purchases, Amazon and Microsoft's investments in Anthropic and OpenAI respectively, and potential IPOs from Cerebras Systems and CoreWeave. We anticipate that this momentum will continue into 2025 as companies across the tech landscape look to capitalize on the rapidly evolving AI space. The new U.S. administration is expected to remain friendly to M&A, further fueling deal-making activity.

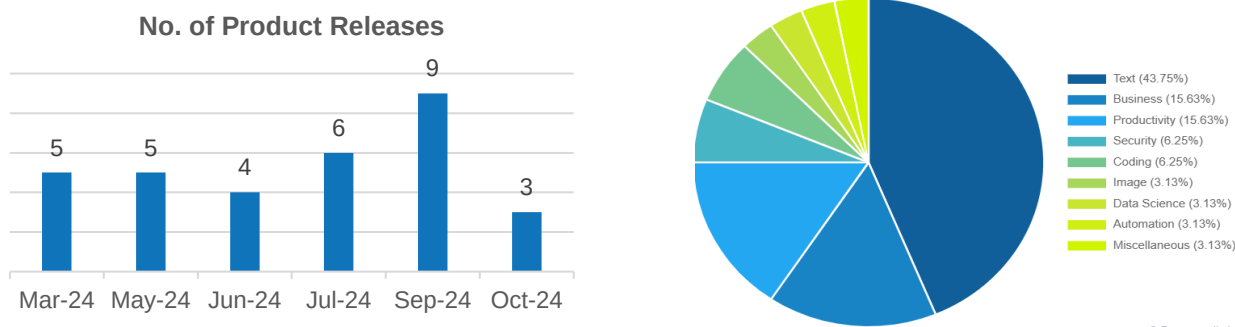
HIGHLIGHTS OF AI DEVELOPMENTS BY SOME TOP TECH PLAYERS

NVIDIA - In 2024, NVIDIA made a major announcement, the introduction of Neural Interface Models (NIM) at the AWS re: Invent event. It focused on enhancing AI inference in cloud environments by optimizing model processing with reduced computational overhead. NVIDIA also revealed the upcoming release of the H200 NVL PCIe GPU in December, offering substantial upgrades, including a 1.5x increase in memory and 1.2x increase in bandwidth, making it ideal for data centers focused on efficient AI model fine-tuning and lower power consumption. Additionally, NVIDIA launched the Omniverse Blueprint for Real-Time CAE Digital Twins, allowing simulations that typically take weeks to be completed in real time, improving capabilities across industries like fluid dynamics.

Anthropic - In 2024, Anthropic made significant strides in AI development, particularly with its Claude model family. A key innovation was the introduction of the "Computer Use" feature in October, enabling the AI to autonomously perform complex tasks like moving the mouse, clicking, and typing, streamlining work for software developers. Over the past six months, the company launched 22 new AI solutions, focusing on text generation, productivity, and business processes. Additionally, Anthropic secured a \$4 billion investment from Amazon in November, fostering deeper collaboration to

integrate their AI models into AWS services. The company also funded initiatives to create independent AI performance benchmarks, focusing on safety and ethics. **The Claude 3.5 model demonstrated notable improvements in coding tasks and performance metrics, outperforming competitors like OpenAI's GPT-4 in areas such as quantitative analysis and object identification.**

Chart 7: Anthropic's AI Innovations by Number and Type of Product Releases Between Dec 09, 2023 - Dec 09, 2024



Source: Intro-act, Futurepedia

Meta - Meta made significant strides in AI, in 2024 focusing on open-source projects, creative tools, and increasing user accessibility. **A key development was the release of Llama 3.1 in July, featuring the 405B model, which is the largest open-source AI model to date.** Designed to compete with top closed-source models, Llama excels in reasoning and coding tasks, while promoting innovation by enabling smaller organizations and researchers to use advanced AI. Meta also introduced creative tools, such as GenAI Ad Tools and the Advantage+ Creative Portal, which automate routine tasks and provide real-time feedback to improve content quality, making powerful tools more accessible to creators. Furthermore, Meta has expanded its AI services globally, supporting multiple languages and localizing tools for a diverse user base. The company is also integrating AI across platforms like WhatsApp and Meta Quest, enhancing user experience with features like AI-assisted messaging, customer support chatbots, and personalized recommendations.

Microsoft - In 2024, Microsoft made great progress in AI development, focusing on productivity, new models, and industry integration. **The company introduced two SLMs, Phi and Orca, which deliver performance comparable to larger models but are more accessible and efficient.** Microsoft also enhanced its multimodal AI capabilities, integrating tools like Microsoft Copilot, which analyzes both images and text, and applications like Microsoft Designer for graphic design and custom neural voices for accessibility. The launch of Copilot wave 2 improved collaboration and business automation, while new features in Copilot Studio allowed users to create tailored autonomous agents for specific industries. Microsoft also expanded AI models for sectors like manufacturing and healthcare, including the Factory Namespace Manager to optimize production data.

Apple - In 2024, Apple made notable advancements in AI with the introduction of **Apple Intelligence, a generative AI system that enhances user experience across its devices.** Unveiled at the Worldwide Developers Conference (WWDC) on June 10, Apple Intelligence integrates with iOS 18, iPadOS 18, and macOS Sequoia, using generative AI to improve user interactions by understanding and creating language and images. Key features include AI writing tools for refining text and summarizing messages, enhanced Siri capabilities that allow for context retention and app control, and visual search tools for locating moments in videos and removing unwanted objects from photos through a "Clean Up" feature.

THE EXPANDING AI ECOSYSTEM: OPPORTUNITIES AND TRENDS IN 2025

The AI landscape is evolving rapidly, creating a multitude of opportunities across the tech ecosystem. As we approach 2025, the landscape has shifted significantly. Major tech companies like Microsoft, Amazon, etc. are now firmly entrenched in the AI revolution, dominating the data centers that fuel AI, holding substantial stakes in leading AI model companies, and heavily investing in emerging AI startups. With this increased confidence, 2025 could be viewed as a year of stabilization in AI capital expenditures.

Chart 8: Increasing Capex Allocation Among Big Tech Players (Quarterly Capex in \$ billion)

	Q22023	Q22024	Q32024
Microsoft	\$11	\$19	\$20
Amazon	\$10	\$17	\$21
Google	\$7	\$13	\$13
Meta	\$6	\$9	\$9

Source: Intro-act, Sequoia, Earnings Transcripts, Public Filings

As enterprises increasingly monetize their AI investments, we see the following key trends emerging that could shape the U.S. AI market in 2025.

- Monetization drives industry expansion:** The commercialization of enterprise AI is fueling growth across various sectors. **Pure-play AI firms like OpenAI and Anthropic are experiencing staggering revenue increases, with OpenAI projected to reach \$5 billion in net revenues by the end of 2024, up 225% YoY.** This trend extends beyond developers of LLMs to consulting firms like Accenture and IBM, cloud computing companies such as ServiceNow and Oracle, data centers like Equinix, and even advertising platforms like Meta. We believe that companies with unique market positioning, robust distribution, and access to privileged data will be primary beneficiaries of this expanding AI services demand in 2025.
- Infrastructure development will be essential:** The rapid evolution of AI infrastructure is underscored by growing investments in data centers and GPU clusters. XAI’s “Colossus” and Meta’s 24,000 GPU data center-scale cluster are just two examples of the unprecedented level of investment in AI hardware. Hyperscalers are projected to invest over \$250 billion in capital expenditures (CapEx) by 2025, with a significant portion allocated to AI infrastructure. We anticipate that this growth will drive demand for ancillary products such as low-power CPUs, memory, storage systems, and cooling equipment. Additionally, **the energy intensity of data centers is prompting investments in alternative power sources and management solutions.**
- Proprietary models unlock AI applications:** Enterprises are shifting toward developing proprietary LLMs to create highly customized AI solutions tailored to their specific business needs. Financial services firms like JPMorgan Chase and Bloomberg, cybersecurity companies such as Palo Alto Networks, and defense contractors like Palantir are investing in proprietary LLMs. As the cost of AI infrastructure continues to decline, **we expect robust demand for supporting technologies like data warehousing and advanced analytics platforms.**
- AI moves to most everyday devices:** Smartphones and personal computers are at the forefront of the AI device revolution. **Apple’s integration of Apple Intelligence and Samsung’s Galaxy AI features are expected to spark significant upgrade cycles in 2025, benefiting component suppliers and vendors.** The impact extends beyond smartphones to industries such as wearables, medical devices, industrial equipment, and home automation. We believe these upgrades will be critical for optimizing manufacturing processes, improving automation, and reducing inefficiencies.

IMPORTANT DISCLOSURES

This third-party research report has been prepared by **Intro-act**, and PartnerCap Securities, LLC (henceforth "**PartnerCap**") has reviewed the said research report for compliance with the applicable provisions of the relevant regulations. Based on representations of the research provider, PartnerCap has a reasonable belief such research is objective and reliable and is not in possession of information that would prove otherwise. PartnerCap Securities, LLC and Intro-act have common ownership.

PartnerCap shall not be considered to have distributed a third-party research report where the research is an independent third-party research report and is made available (a) upon request; (b) through a member-maintained website; or (c) to a customer in connection with a solicited order in which the registered representative has informed the customer, during the solicitation, of the availability of independent research on the solicited equity security and the customer requests such independent research.

In addition, PartnerCap assumes no warranty, liability or guarantee for the current relevance, correctness or completeness of any information provided. Furthermore, PartnerCap assumes no liability for any direct or indirect loss or damage or, in particular, for lost profit, which you may incur as a result of the use and existence of the information, provided within the report. This electronic message and any attachments has been provided only to the person or entity to which it is addressed and has been provided for informational purposes only and should not be used or construed as an offer to sell, a solicitation, an offer to buy, or a recommendation for any security and is not purported to be tailored to any particular investor and may not be re distributed.

Other than disclosures relating to PartnerCap and its affiliates, the information herein is based on sources we believe to be reliable but has not been independently verified by us and does not purport to be a complete statement or summary of the available data. No representation is made that such information is accurate in all material respects, complete or up to date. PartnerCap is under no obligation to update, modify or amend the information and it does not accept any liability for any loss or damage, however caused, arising from any errors, omissions, or reliance on any information or views presented.

Information and securities mentioned may reflect a third party's independent opinions and are not recommendations of PartnerCap. The commentary is a general discussion based on short term market factors or current market trends, and may differ from, or be inconsistent with the views of other sales and trading personnel within the Firm or the long-term views of research analysts employed by PartnerCap.

The third-party research sourced information is the exclusive property of Intro-act; and without its permission, this information and any other intellectual property may not be reproduced, re-disseminated or used to create any financial products, including any indices. This information is provided on an "as is" basis. The user assumes the entire risk of any use made of this information. The third-party research analyst, its affiliates and any third party involved in, or related to, computing or compiling the information hereby expressly disclaim all warranties of originality, accuracy, completeness, merchantability or fitness for a particular purpose with respect to any of this information. Neither PartnerCap or Intro-act beneficially own 1% or more of any class of common equity securities of the subject company covered in the said research report.

Intro-act Disclosures

Analyst Certification: I, Frank White, certify that the views expressed in the research report accurately reflect my professional views about the subject securities or issues. Intro-act issued this report and may seek fees for the assistance with investor targeting, access, and further investor preparation services. Intro-act will not seek remuneration for any investment banking service or investment advice.

Accuracy of content: All information used in the publication of this report has been compiled from publicly available sources who are believed to be reliable. However, the issuer and related parties, as well as Intro-act, do not guarantee

the accuracy or completeness of this report, and have not sought for this information to be independently verified. Opinions contained in this report represent those of the Intro-act analysts at the time of publication. Forward-looking information or statements in this report contain information that is based on assumptions, forecasts of future results, and estimates of amounts not yet determinable, and therefore involve known and unknown risks, uncertainties, and other factors that may cause the actual results, performance, or achievements of their subject matter to be materially different from current expectations.

AI-generated content: Content in this research report is AI-generated. Intro-act employs its proprietary AI research system for content generation, based on a universe of credible and authentic source documents. The system functions under an overall manual supervision, and thus the creative control over the content generated by the system lies with its human authors. The AI content is not mere “mechanical reproduction” and instead is the result of the human author’s “own original mental conception, to which [the author] gave visible form.” The AI-generated content from the system is sufficiently modified so as to meet the standards for copyright protection. Intro-act thus claims that the work as a whole constitutes an original work of human authorship, and the content in this research report has sufficient human authorship to support a copyright claim, according to the Copyright Registration Guidance, published by the U.S. Copyright Office on 03/16/2023.

Exclusion of Liability: To the fullest extent allowed by law, Intro-act, Inc. shall not be liable for any direct, indirect, or consequential losses, loss of profits, damages, or costs or expenses incurred or suffered by you arising out of or in connection with the access to, use of, or reliance on any information contained in this note.

No personalized advice: The information that we provide should not be construed in any manner whatsoever as personalized advice. Also, the information provided by us should not be construed by any subscriber or prospective subscriber as Intro-act’s solicitation to affect, or attempt to affect, any transaction in a security. The securities described in the report may not be eligible for sale in all jurisdictions or to certain categories of investors.

Investment in securities mentioned: Intro-act has a restrictive policy relating to personal dealing and conflicts of interest. Intro-act, Inc. does not conduct any investment business and, accordingly, does not itself hold any positions in the securities mentioned in this report. However, the respective directors, officers, employees, and contractors of Intro-act may have a position in any or related securities mentioned in this report, subject to Intro-act’s policies on personal dealing and conflicts of interest.

Copyright: Copyright 2024 Intro-act, Inc. (Intro-act).

Intro-act is not registered as an investment adviser with the Securities and Exchange Commission. Intro-act relies upon the “publishers’ exclusion” from the definition of investment adviser under Section 202(a) (11) of the Investment Advisers Act of 1940 and corresponding state securities laws. This report is a bona fide publication of general and regular circulation offering impersonal investment-related advice, not tailored to a specific investment portfolio or the needs of current and/or prospective subscribers. As such, Intro-act does not offer or provide personal advice, and the research provided is for informational purposes only. No mention of a particular security in this report constitutes a recommendation to buy, sell, or hold that or any security, or that any particular security, portfolio of securities, transaction, or investment strategy is suitable for any specific person.