

# Amazon's investment in Anthropic: who will win the race in the Generative Al space?

#### Introduction

On September 25, 2023, Amazon [NASDAQ: AMZN] announced its investment of up to \$4bn in Anthropic, a rival to the likes of OpenAI's ChatGPT and Google's Bard. The news follows a series of large investments by Microsoft [NASDAQ: MSFT], Google [NASDAQ: GOOGL], and Nvidia [NASDAQ: NVDA] into generative artificial intelligence (AI) based on large language models. While Amazon was falling behind rival providers of AI chip technology in terms of partnerships and deal-making, the company had reasons to remain positive, as experts believe the firm's custom silicon and specialized hardware is currently the best on the market. The deal is a statement from Amazon, signaling a long-awaited step-up in the race for provision of cloud computing services and the growing importance of generative AI.

### **Deal Structure**

The partnership between Amazon and Anthropic is designed to ensure a strong alignment of interests. Amazon's initial investment into the AI start-up has been reported to be \$1.25bn for a currently unknown minority stake in the firm, which was valued at \$5bn earlier this year. Amazon has committed to investing further, with the total value potentially reaching \$4bn in the event that either party triggers the additional funding. Amazon's market value soared \$21bn after the announcement. As part of the collaboration, Anthropic will use Amazon Web Services (AWS) as the "primary cloud provider for critical mission workloads" despite its earlier commitments to the same services offered by Google. In fact, Google has invested \$300mn into Anthropic prior to this deal under similar terms, that is, Anthropic using Google's cloud services.

#### **Deal Rationale**

In early 2023, Microsoft confirmed a multi-billion dollar investment in the overnight sensation ChatGPT owner OpenAI. This commitment marked a second deal between the two companies since 2019. In return, OpenAI agreed to exclusively use Microsoft's Azure – a cloud computing service just like AWS. Coupled by Google's \$300mn investment in Anthropic earlier this year, Amazon's competition in the cloud services industry has been ahead for some time. Additionally, Amazon has been rivaling Microsoft and Google in the demand for H100s – Nvidia's graphic processors units that power the AI race – ever since ChatGPT was released in November 2022. Even though the deal does not entail Anthropic will exclusively use AWS for computing power, it already has access to a significant quantity of Amazon's Trainium and Inferentia chips. These 2 types of AI chips are specifically built for AI and machine learning and can be adjusted to the needs of the client to unlock cost efficiency, performance, and energy efficiency for scaling. Furthermore, Amazon has already added Anthropic's AI chatbot Claude to the range of services in its AWS Bedrock – a service that allows its clients to leverage generative AI for building new applications in the cloud. The model within Amazon's service has already been used in creation of Bridgewater Associates' Investment Analyst Assistant, a large language model-powered tool for generation of elaborate charts, financial indicators, and summarization of results. Amazon and Anthropic believe that synergies between their business models will allow both firms to improve client service, cost efficiency, and safety of their products.



## Generative AI - value chain and industry dynamics

The generative AI industry, with its remarkable capacity to create new content, automate tasks, and enhance diverse sectors, operates within a complex and multifaceted value chain. At its core, generative AI harnesses a range of foundational technologies, including Machine Learning (ML) and Deep Learning (DL), which form the bedrock where algorithms learn from extensive datasets to make predictions or generate novel data; Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) each contribe unique capabilities to the generative AI landscape; Additionally, Natural Language Processing (NLP) and Transformative Neural Networks (TNNs) play pivotal roles in enabling the generation of coherent, contextually relevant content.

The generative AI value chain, exemplified by ChatGPT, represents a collaborative network that spans from the creation of specialized processors by chipmakers such as NVIDIA and AMD, to the deployment and interaction with the AI by end-users. These processors are the backbone of AI training, enabling the rapid computation needed for developing sophisticated AI models. Cloud platforms like AWS, GCP, and Azure provide the essential computational infrastructure, allowing AI researchers and developers to train and refine their algorithms on a massive scale without significant hardware investments. This cloud-based ecosystem facilitates the processing of large datasets, which is fundamental in teaching the AI to understand and generate human-like text. Once the AI models are perfected, they are deployed through these cloud services, making them widely accessible to users and developers across the globe. The AI becomes a service that can scale with demand, thanks to the cloud platforms' capability to distribute it efficiently. Ultimately, end-users reap the benefits of this intricate system through seamless interaction with AI applications, all made possible by the interconnected efforts of hardware manufacturers, cloud service providers, and AI innovators.

Lastly, to properly understand why prominent companies such as Nvidia, along with Big Tech giants like Google and Facebook, as well as innovative companies like OpenAI and Anthropic, are actively investing in and participating in the generative AI race, it is crucial to grasp the underlying dynamics of this industry. Several key industry drivers are steering its growth and development. These include the availability of vast and diverse datasets, advances in deep learning, and the increased computational power that has significantly bolstered generative AI capabilities. Additionally, continuous efforts in research further enhance the quality, efficiency, and versatility of generative models, allowing industry-specific solutions that can be used to optimize processes, enhance user experiences, and drive innovation in several industries, such as healthcare, marketing, finance, and education.

## The race to AI dominance

The generative AI arena is witnessing an intense race to dominance, with tech giants deploying strategic investments and forging alliances. Specifically, Amazon, Google, and Microsoft have spent the past year investing billions of dollars in artificial-intelligence startups, while also charging those fledgling companies a similar amount to use their cloud platforms.



At the forefront of the pivotal developments that are shaping the sector, Microsoft has made a resounding statement with its colossal investment in OpenAI, rumored to be around \$13 billion. This partnership has not only amplified Microsoft's AI prowess but has also cemented its status as an AI powerhouse, particularly in enhancing its cloud offerings. This move is a testament to Microsoft's dedication to weaving generative AI into its product ecosystem, potentially revolutionizing business and consumer interactions with technology.

Google, in a bid to keep pace, has been executing its own strategic plays. A notable \$300 million investment in AI startup Anthropic underscores Google's commitment since the beginning of 2023, followed by a recent agreement to pour up to \$2 billion more into the company. With Google Cloud's emphasis on AI platforms, especially in the realm of enterprise search, Google is tailoring its AI solutions to deliver enterprise-grade capabilities. The launch of specialized models and the Gen App Builder indicates Google's focus on AI's transformative potential for future business infrastructures.

Meanwhile, Apple is reportedly preparing to infuse generative AI features into its flagship iPhone and iPad devices by late 2024. With an investment that's said to be in the ballpark of \$4.75 billion by the end of next year, Apple appears poised to leap into the AI fray, potentially altering the landscape of mobile and personal technology. While some of these funds are earmarked for the procurement of state-of-the-art AI servers, like Nvidia's HGX H100 8-GPU, the company's plans don't stop there. Future server acquisitions may even transition to Nvidia's next-gen B100 solution if speculations hold. Despite this considerable investment, Apple's purchasing strategy, with an estimated 20,000-unit acquisition, lags behind that of competitors.

Lastly, Meta Platforms has notably shifted gears from its metaverse aspirations to a concentrated focus on generative AI, marked by the rumor of expected approximate purchases of 40,000 AI servers in 2024. Establishing a dedicated top-level product group for generative AI, Meta is signaling its intent to become a key player in this burgeoning field. The launch of their AI tool, LLaMA, further indicates Meta's ambition to stake a claim in the generative AI space.

Amidst this backdrop, the industry is rife with rumors and speculation. There's chatter about Nvidia's potential role, given its trillion-dollar valuation and the pivotal nature of its chips in AI development, hinting that it might emerge as a formidable contender in the AI race. Whispers of potential collaborations between the tech titans and specialized AI startups also hint at a dynamic and disruptive future for the sector.

#### The race to AI dominance

So far, the world has experienced a major transformation brought by generative AI and the race of tech giants for market share and consumers. The public perceives Amazon's investment into the technology and its derivatives as a major positive, recognized by the rally in the company's shares of 5.58% since the announcement of partnership. Investors' heavy interest in the sector has also notably been displayed by Nvidia's performance. Its stock increased in value by 214% year-to-date. We can expect to see strong progression in the industry, with likely larger investments that will secure exclusivity for tech giants like the agreement between Microsoft and OpenAI. Quite possibly, the AI chips used for cloud computing services of Microsoft, Amazon and Google will be more diversified, for a wider range of applications. So far, we have seen AI being used for just about anything – from market research and information technologies to marketing and sales and customer service. What will be



vital for the chip producers, however, is the ability to achieve improved customization for delivery of excellent service, efficiency, and scaling. The fine-tuning developed by tech giants will certainly ensure more action and deal-making in the competition for generative AI.

TAGS: Deal, Investment, Amazon, AWS, Anthropic, Generative AI