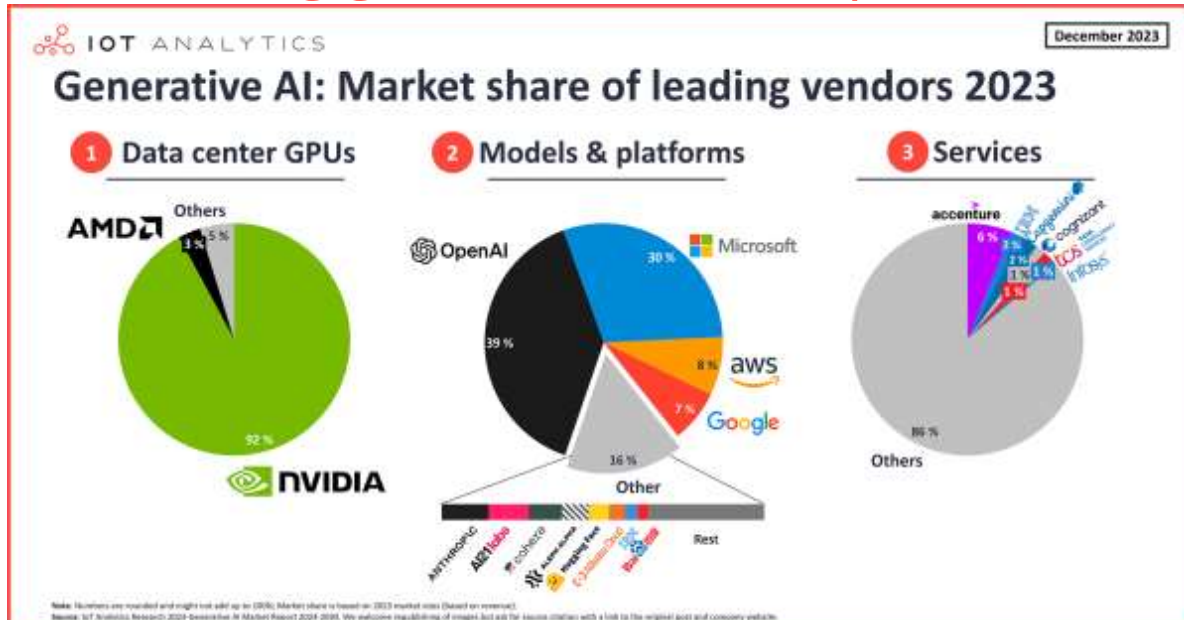FOR IMMEDIATE RELEASE

The leading generative AI companies

Hamburg, Germany, December 14, 2023: IoT Analytics, a leading provider of market insights and strategic business intelligence for the Internet of Things (IoT), AI, Cloud, Edge, and Industry 4.0, published an analysis based on the 154-page "Generative AI Market Report 2023–2030" report and highlights the landscape with its top players in the data center GPU, foundational model and platform, and generative AI services markets.

Key quotes

Knud Lasse Lueth, CEO at IoT Analytics, remarks: "*The speed of generative AI innovation with new offerings coming on the market on a weekly basis is fascinating monitor. Nvidia with 92% market share for data center GPUs as well as Microsoft and OpenAI with a combined 69% market share in the models and platforms segment are firmly in the lead in their respective market segments. With hyperscalers developing their own data center chips, with the availability of powerful open-source models and with giants like AWS and Google looking to differentiate with their new offerings, it will be interesting to watch how much the early lead is worth for the current market leaders. I personally do expect both Microsoft and Nvidia to maintain their strong positions in the coming years but the gap to the competition will likely close a bit*."
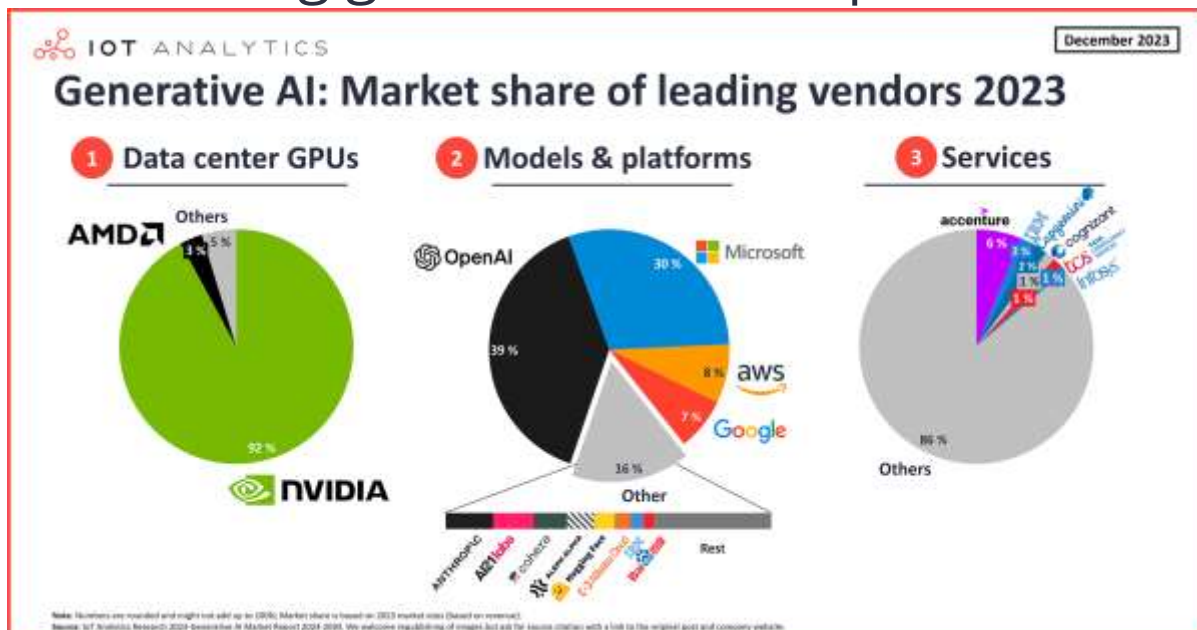
Philipp Wegner, Principal Analyst at IoT Analytics, adds that "*The Generative AI market is rapidly evolving, with established leaders and a growing number of startups. In 2024, it's a make-or-break year for Gen AI vendors, as they navigate a crowded field of competitors.*"

Key insights

- The generative AI market went from nearly nothing to a hot market within a year, as shown by IoT Analytics' latest <u>research report</u>.
- IoT Analytics analyzed 3 interconnected markets for generative AI: 1) data center GPUs, 2) foundational models and platforms, and 3) generative AI services. Each has distinct aspects and market players.
- NVIDIA leads the data center GPU segment with a 92% market share, while OpenAI and Microsoft have a combined share of 69% in the foundational models and platforms market. The services market is more fragmented, with Accenture currently seen as the leader with a 6% market share.

[The full research article is attached below]

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

The leading generative AI companies

The rise of generative AI

Following its release of ChatGPT in 2022, **OpenAI** experienced an impressive one-year, zero-to-$1 billion revenue bump—surpassed only by US-based chipmaker **NVIDIA**, which managed to increase its data center GPU sales from $3.6 billion in Q4 2022 to an expected $16 billion in Q4 2023. When it comes to generative AI companies, these two stand out.

The generative AI foundational models and platforms market is expected to reach nearly 5% of global software spending by 2030
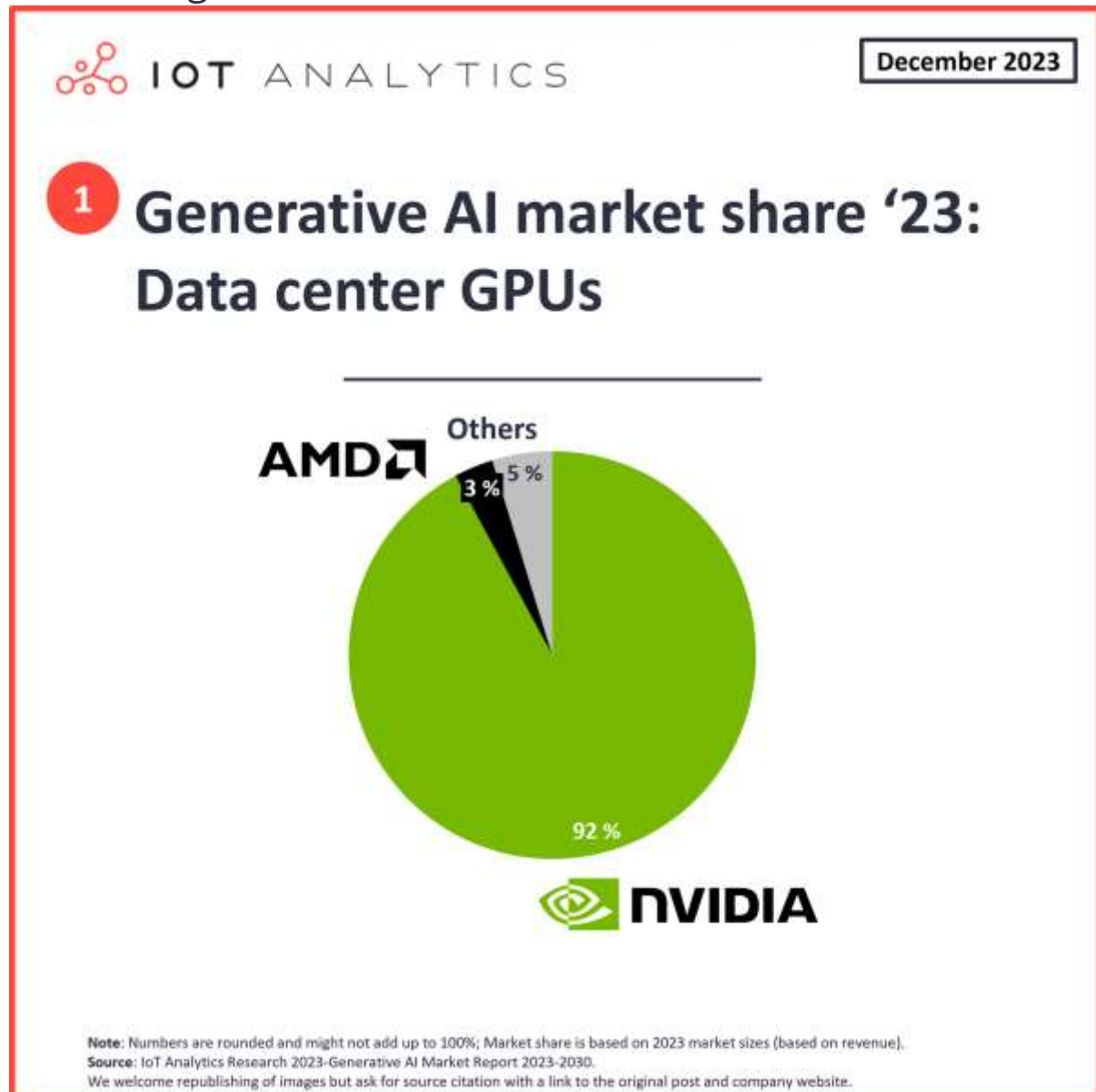
According to IoT Analytics' 154-page Generative AI Market Report 2023–2030 (published December 2023), the generative AI software and services market reached $6.2 billion in 2023. Although it is still very early to forecast where things are going from here, the IoT Analytics research team expects the generative AI foundational models and platforms market to make up nearly 5% of global software spending by 2030 due to its disruptive nature and tremendous value potential.

However, this does not include the market for individual generative AI solutions. The team believes generative AI will become standard within most software in the near future. This also does not include the hardware market, such as for

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

data center GPUs, since this market is looked at separately from software but is discussed below.

In this article, we dive into the data center GPU, generative AI foundational model and platform, and generative AI services markets, discussing what aspects of the generative AI field make up each market and highlighting the leading generative AI companies within them.

Market segment 1: Data center GPU market

IoT ANALYTICS — December 2023

1 Generative AI market share '23: Data center GPUs

AMD 3 %   Others 5 %

NVIDIA 92 %

Note: Numbers are rounded and might not add up to 100%; Market share is based on 2023 market sizes (based on revenue).
Source: IoT Analytics Research 2023-Generative AI Market Report 2023-2030.
We welcome republishing of images but ask for source citation with a link to the original post and company website.

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

a.) Market overview

The data center GPUs market refers to specialized GPUs designed to handle the extensive computation demands of modern data centers, which are the backbone of generative AI. Originally designed for rendering graphics, GPUs excel at parallel processing, which is fundamental for deep learning computations used in generative AI.

Note: This market does not include CPUs, consumer GPUs, or TPUs, but it does include GPU systems intended for data center use.

The report shows the data center GPUs market reached $49 billion in 2023—a booming increase from 2022 (+182%), mostly driven by one company alone: **NVIDIA**. Although the market for data center GPUs has seen steep price increases and is undergoing severe supply constraints, there is currently no reason to believe demand will decline in the next two years.

b.) Leading data center GPU companies

The data center GPU market at this point has one very clear leader. However, the market report shows that there are other promising startups and other established companies trying to make inroads.

The data center GPU market at this point has one very clear leader. However, the market report shows that there are other promising startups and other established companies trying to make inroads.

1. NVIDIA

NVIDIA leads the data center GPU market by a long shot, owning 92% of the market share. In 2023, the company's quarterly revenue jumped 272%, from $4.3 billion in Q1 to a forecasted $16 billion in Q4.

The NVIDIA A100 Tensor Core GPU is the de facto standard for data center GPUs. However, as discussed in the report, hardware is not the only differentiator for NVIDIA. Some consider their developer ecosystem, CUDA, as NVIDIA's biggest moat, and it is often cited as the key reason why NVIDIA is not set to lose its dominant position anytime soon.

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

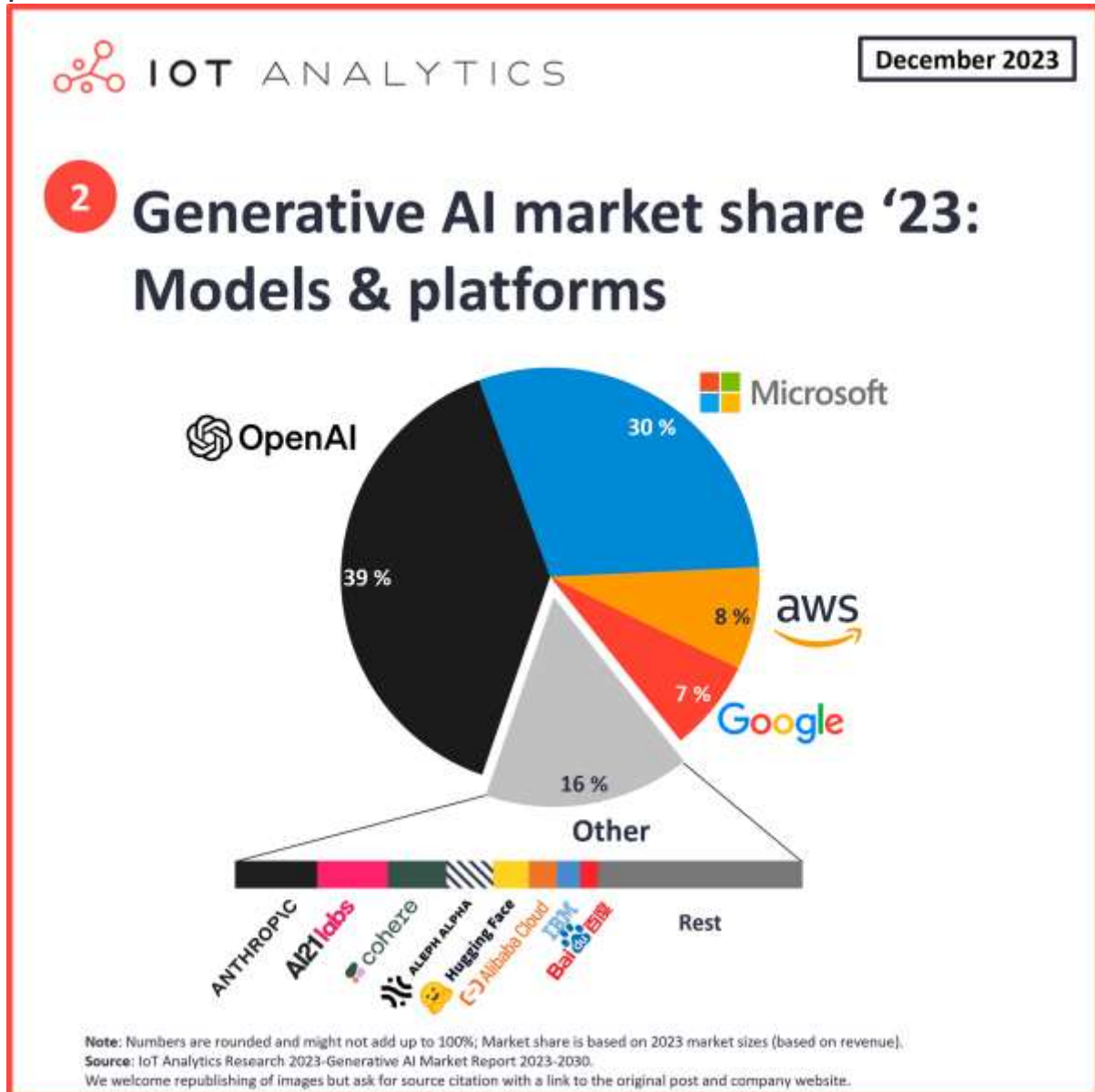**NVIDIA A100, the company's flagship GPU for data centers (source: NVIDIA)**

2. AMD

The Data Center segment of US-based semiconductor **AMD player**, **NVIDIA's** first real GPU challenger, grew by 21% from Q2 2023 to Q3 2023 and shared 3% of the market. However, AMD has big ambitions in 2024 to eat into NVIDIA's market share. In early December 2023, it announced the release of its Instinct MI300 Series accelerators, which are cheaper than NVIDIA's comparable accelerators and, as AMD claims, faster. AMD's CEO, Dr. Lisa Su, forecasted at least $1 billion in revenue in 2024 through this chip alone, and **Microsoft**, **Meta**, and **OpenAI** stated they would use the Instinct MI300X in their data centers. AMD also recently launched ROCm 6.0 to provide developers with an ecosystem that is equally attractive to CUDA.

3. Intel and others

US-based chipmaker **Intel**, the traditional competitor to **NVIDIA** and **AMD**, has lagged behind on the data center GPU front. In May 2022, Intel's Habana Labs released its second generation of AI processors, Gaudi 2, for training and inferencing. Though not as fast as NVIDIA's popular H100 GPU, it is considered a viable alternative when considering price to performance.

Meanwhile, in July 2023, startup chipmaker **Cerebras** announced it had built its first of nine AI supercomputers in an effort to provide alternatives to systems using NVIDIA technology. Cerebras built the system, Condor Galaxy 1, in partnership with the UAE, which has invested in AI research in recent years.

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

Market segment 2: Generative AI foundational models and platforms market

a) Market overview

The foundational models and platforms market comprises two related areas. Foundational models are large-scale, pre-trained models that can be adapted to various tasks without the need for training from scratch, such as language processing, image recognition, and decision-making algorithms.

Generative AI platforms, in turn, refer to software that enables the management of generative AI-related activities outside of foundational models. Notably, IoT Analytics identified six platform types: 1) development, 2) data

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

management/databases, 3) AI IaaS/GPU as-a-service, 4) middleware & integration, 5) MLOps, and 6) user interface and experience (UI/UX).

The foundational models and platforms market exploded with the public release of ChatGPT in late 2022, reaching $3.0 billion in 2023. This is substantial growth over 2022, which saw next to nil in terms of revenue. IoT Analytics' analysis projects strong market growth in the coming years as enterprises invest billions in—and report real value from—generative AI implementations and continuous improvements.

b.) Leading generative AI foundational model and platform companies

Unsurprisingly, the foundational model and platform market are currently led by OpenAI, with several well-known technology companies trying to catch up.

1. OpenAI

With the November 2022 launch and subsequent success of ChatGPT, **OpenAI** leads in the share of the foundational model and platform vendors market with 39%. Since the release of ChatGPT, OpenAI's generative pre-trained transformer (GPT) models went from GPT-3.5 to GPT-4 to GPT-4 Turbo, showcasing the continued development of the model. OpenAI's models continue to impress in independent model assessments and rankings—often coming out in the top three of all tested models. Although many experts expect the foundational model space to become a commodity over time, at this point, OpenAI's flagship models remain the top foundational model on the most common benchmarks.

According to IoT Analytics' What CEOs Talked About series, in 2023, ChatGPT skyrocketed in boardroom discussions in Q1, but as other foundational models and generative AI applications became available, mentions of ChatGPT steadily declined as "generative AI" separated and continued to rise. (The What CEOs Talked About in Q4 2023 report and blog is expected to be released mid-December 2023.)

2. Microsoft

On OpenAI's heels at 30% market share is **Microsoft**, its largest shareholder. Microsoft's platform, Azure AI, offers Azure OpenAI, which uses OpenAI's LLMs but goes beyond the public ChatGPT offering by promising greater data security and custom AI apps. This is suited for enterprises who want to secure their

IoT Analytics GmbH
Zirkusweg 2
D-20359 Hamburg

T: +49 (0) 40- 63911891
M: press(at)iot-analytics.com
www.iot-analytics.com

proprietary data when leveraging the benefits of generative AI since ChatGPT's terms of use state that they can store and use content (both input and output) to improve their services. In November 2023, Microsoft reported over 20,000 active paying customers for its Azure AI platform, adding that 85% of Fortune 100 companies used it in the past year.

Despite Microsoft's strong partnership with OpenAI, Microsoft also heavily promotes the usage of other models, such as Llama 2, via its platform, thereby enabling customers to freely choose and test different models and providers. Another key priority for Microsoft is integrating AI capabilities into its existing product portfolio, such as Azure, Microsoft/Office 365, and Bing.

3. AWS

AWS has an 8% share of this market. Its Bedrock service, publicly released in September 2023, provides access to models from several AI companies, such as **Anthropic**, **AI21 labs**, and **Cohere** (each with a 2% share of this market), and combines them with developer toolsets to help customers build and scale generative AI applications.

AWS has quickly claimed the third spot in this market because the company is the market leader in public cloud services and quickly got its existing customer base excited about its differentiated approach to Generative AI. In contrast to Google and Microsoft, AWS Bedrock focuses on providing a platform service that gives users access to a number of both general and domain-specific foundational models from a variety of vendors—providing choice, flexibility, and independence.

4. Google

In 2022, most experts credited Google as being the one tech company at the forefront of AI. Many experts interviewed by the IoT Analytics team continuously praised Google for its AI and its data products and innovations. In 2023, the picture is different, and Google is fighting to defend its position as an AI leader.

Vertex AI is **Google** Cloud'splatform focused on machine learning (ML) ops. It is integrated with other Google Cloud services, such as BigQuery and Dataproc, and offers a Jupyter-based environment for ML tasks. In early December 2023, Google released a preview version of its new multi-modal flagship model, Gemini. The related technical report states that the largest of the Gemini family outperformed other existing models in 30 out of 32 common ML benchmarks.

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

Initially, the announcement of Gemini was widely received as positive, but a popular demo video released by Google later turned out to be <u>staged</u>.

Market segment 3: Generative AI services market
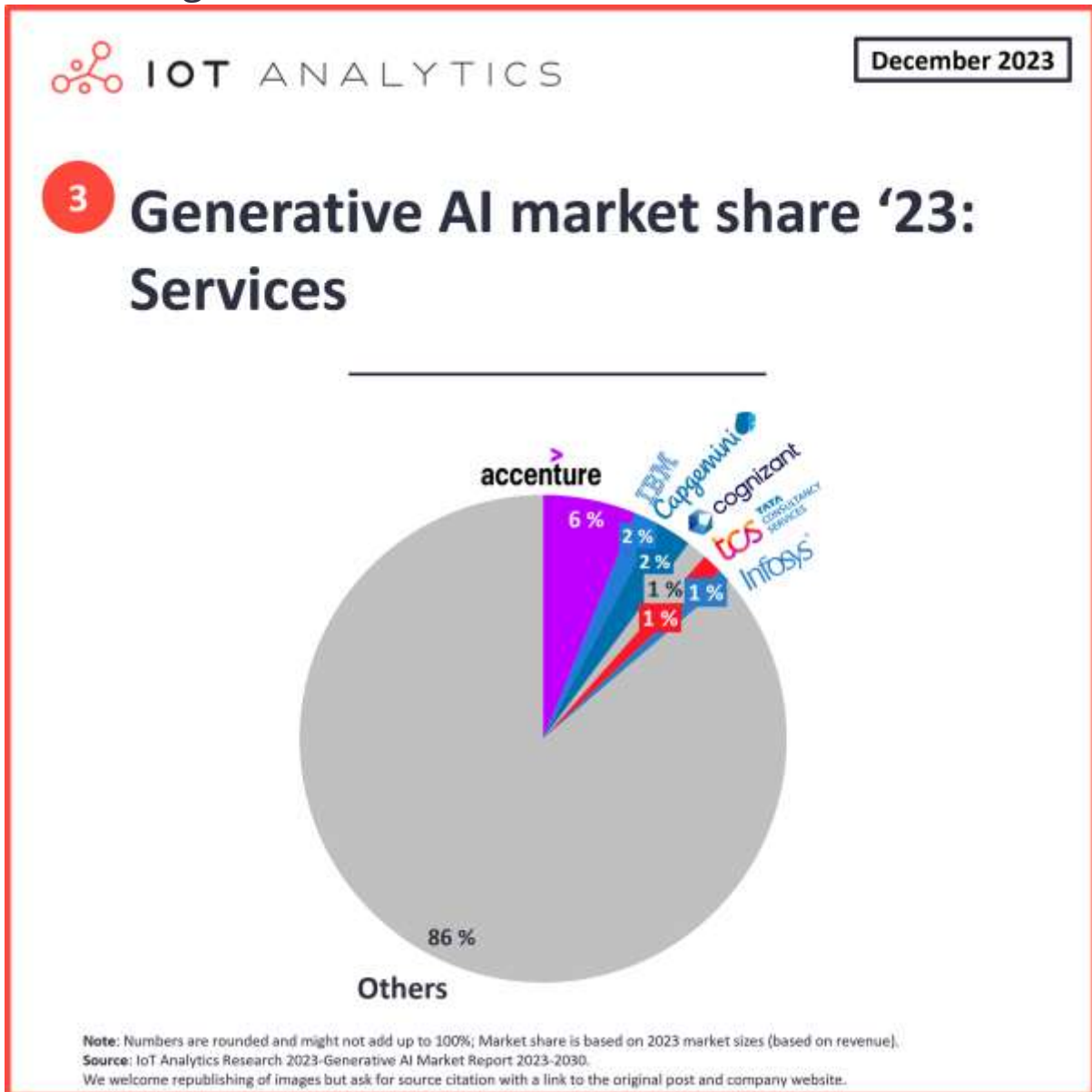
a) Market overview

The generative AI services market represents a specialized segment dedicated to consulting, integration, and implementation support for organizations aiming to integrate generative AI capabilities. With generative AI having risen as one of the <u>top discussion points in boardrooms</u>, services companies are sensing a large opportunity in helping companies formulate their generative AI strategies (e.g., what use cases to implement), advising them on technical architecture choices

IoT ANALYTICS

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

(e.g., which models to use) and helping them implement and build individual solutions.

IoT Analytics assesses that the generative AI services sector's opportunity is *now*. Due to the novelty of generative AI, organizations often lack skills and experience, and the only option is to look for professional services firms that have or are in the process of building up the required expertise.

b) Leading generative AI services companies

The generative AI services market is more dispersed than the other two markets highlighted here.

1. Accenture

Accenture is estimated to have the largest generative AI services market share at 6%. The company announced in June 2023 that it is investing $3 billion in data and AI practice over three years to double its AI talent and develop new capabilities. Additionally, Accenture disclosed in its Q4 2023 earnings press release that its revenue for generative AI projects grew to $300 million for 2023.

In November 2023, Accenture announced plans to launch a network of generative AI studios in North America where companies can explore ways to integrate generative AI applications. These studies are expected to be at Accenture Innovation Hubs in Chicago, Houston, New York, San Francisco, Toronto, and Washington, DC.

2. IBM

US-based technology corporation **IBM** makes up 2% of this market. To position itself for the opportunities that generative AI brings, the company announced it had established a "Center of Excellence (CoE) for generative AI," which as of May 2023, already had over 1,000 consultants specialized in generative AI. The CoE operates alongside IBM's AI and Automation practice, which includes over 21,000 data and AI consultants.

3. Capgemini

France-based IT services company **Capgemini** also has a 2% share in this market, offering consulting services intended to help clients adopt key technologies such as the cloud and AI. In July 2023, Capgemini announced the launch of a portfolio of generative AI services, including in the following areas:

IoT Analytics GmbH   T: +49 (0) 40- 63911891
Zirkusweg 2   M: press(at)iot-analytics.com
D-20359 Hamburg   www.iot-analytics.com

- Strategy
- Customer experience
- Software engineering
- Custom solutions for enterprise

One of Capgemini's current customers is London Heathrow Airport which aims to improve traveler experiences through its "Generative AI for Customer Experience" offer. Heathrow's Director of Marketing and Digital, Pete Burns, stated that the project is intended to "assist, empower and delight passengers" with tailored customer service solutions.

4. The many others

Past this point, the remaining 86% of the market becomes a cornucopia of specialized generative AI services providers and larger general consulting and system integration companies, each taking a bite of the rapidly growing segment.

As an example, in April 2023, UK-based professional services company **PwC** announced plans to invest $1 billion over three years to not only grow its AI offerings but also transform how it works by using generative AI. Additionally, in July 2023, global consultancy firm **McKinsey & Company** announced it partnered with AI startup **Cohere** to provide customized AI solutions to its enterprise clients.

Generative AI company landscape outlook

The enterprise generative AI market is roughly a year old, and already, the generative AI companies landscape appears vast.

IoT Analytics released its first generative AI report, the Generative AI Trend Report, in March 2023. Since then, more foundational models and platforms have emerged, e.g., OpenAI's GPT4 Turbo, Google's Gemini, or Microsoft's Phi-2. At the same time, the demand for data center GPUs exploded, which is also mirrored in NVIDIA's stock performance (+231% year-to-date as of 14 December 2023). Finally, consulting giants have made investments to position themselves in the generative AI services market, such as Accenture's $3B investment in AI and its pledge to double "AI talent."

As part of this research, we talked to 30+ experts in the field and gathered information on 270+ generative AI projects and analyzed which industries and departments are fastest to adopt and which vendors are most selected today.

The coming months will reveal how many of those projects will deliver value besides just being a marketing coup or how many of those currently in the proof-of-concept stage will move forward. Most companies are only now forging their generative AI strategies and considering whether to build foundational models from scratch based on industry-specific data, use an out-of-the-box propriety model, or fine-tune open-source models. All of this comes as generative AI companies release new products at unprecedented speed.

There is still a lot of movement in the generative AI company landscape, and there will be more in the foreseeable future. IoT Analytics will stay on top of this space, with a follow-up report expected in 2024.

_____

For more information or media inquiries, please contact:

Hoang Pham Van
IoT Analytics
+49 (0) 40 6391 1891
press(at)iot-analytics.com

For further reading please visit:

www.iot-analytics.com/research-blog

About IoT Analytics

IoT Analytics, founded and operating out of Germany, is a leading global provider of market insights and strategic business intelligence for the IoT, AI, Cloud, Edge, and Industry 4.0.
Our key workstreams across the tech stack include IoT applications, IoT platforms and software, IoT connectivity and hardware, and industrial IoT.
We are trusted by 1000+ leading companies around the world for our market insights, including globally leading software, telecommunications, consulting, semiconductor, and industrial players.

###