



# The Generative AI Market

1st Edition

*The Generative AI Market is a strategy report from Berg Insight analysing the latest developments and trends in the generative AI market. This strategic research report from Berg Insight provides you with 90 pages of unique business intelligence including 5-year industry forecasts and expert commentary on which to base your business decisions.*

# The Generative AI market showed triple-digit growth in 2024

Generative AI (GenAI) has popularly been compared to major technological breakthroughs such as the printing press of the 15th century, the steam engine of the late 18th, electricity in the late 19th and the emergence of the Internet in the late 20th. The GenAI hype is not without merit, since its ability to creatively generate convincingly human-like content makes it a disruptive technology with the potential to influence nearly every industry. Even though traditional AI systems have been used commercially for many years, GenAI is a more novel practice that enables computer systems to produce original content – including text, images, video, audio and software code – rather than merely analysing existing data or making predictions.

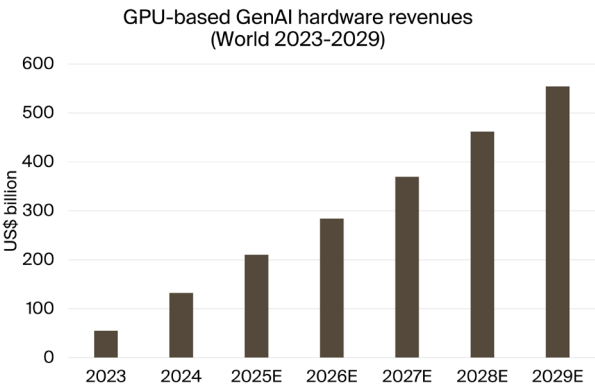
Before 2023, the use of GenAI technology was practically non-existent. The nascent market was ignited by the launch of OpenAI’s ChatGPT, which was the first widely adopted commercial product to bring GenAI to mainstream attention. Significant investments can since be observed from a diverse range of enterprises, spanning both startups and established technology giants, all trying to capitalise on the substantial market potential. However, due to the vast computational resources required to train and run AI models, the market is primarily dominated by large technology conglomerates and companies that have managed to raise significant funding.

Berg Insight has identified 31 key foundation model providers spanning LLMs, vision, audio and multimodal models. While many LLMs started as unimodal models, nearly all successful LLMs now include multimodal capabilities. Companies with multimodal LLMs or successful cross-modal offerings include US-based Anthropic, Google, Meta, OpenAI, Upstage and xAI; China-based AI.01, Alibaba, Baichuan, Baidu, ByteDance, DeepSeek, MiniMax, Moonshot AI, Stepfun, Tencent and Z.ai; France-based Mistral AI; Canada-based Cohere and Israel-based AI21 Labs. Specialised vision model developers include US-based Luma AI, Midjourney, Pika and Runway; UK-based Recraft and Stability AI; Japan-based Black Forest Labs; Canada-based Ideogram and Chinese Kuaishou. Key audio specialists include US-based Assembly AI and ElevenLabs.

The ecosystem is supported by a host of development platform providers offering streamlined environments and tools for building GenAI applications and models. In the US, these include established cloud service providers like

Microsoft, Google and AWS, as well as diversified technology companies such as IBM and Oracle. The landscape also includes hardware providers like Nvidia and SambaNova Systems, data platform specialists such as Databricks and Snowflake, model training and dataset platforms like Scale AI, the open-source model library from Hugging Face and other key players including C3.ai, Dataiku, Weights & Biases, Cloudera, Together AI, Domino and H2O.ai. Several European and Asian providers also contribute to the landscape, including Netherlands-based Nebius, Germany’s Aleph Alpha, and Chinese Alibaba, Baidu, ByteDance and Tencent.

The GenAI market grew substantially in 2024, experiencing triple-digit-growth rates in all three major segments spanning GenAI hardware, foundation models and development platforms. Hardware is currently the largest, led by Nvidia. It is driven by significant data centre investments by cloud service providers, with over US\$ 400 billion in expected AI-related spending in 2025. However, there is a significant time lag before this infrastructure spend translates into revenues from end-user AI applications. The market value for foundation models reached an estimated US\$ 4.1 billion in 2024, excluding end-user applications such as ChatGPT. The figure primarily includes income through API services or license fees as the models are used on development platforms. Meanwhile, the market value for GenAI development platforms reached an estimated US\$ 17.0 billion. Furthermore, GPU-based hardware systems used for GenAI workloads generated revenues of US\$ 132.3 billion in 2024.



# Table of contents

**Executive Summary**

**1 Introduction**

- 1.1 The AI taxonomy**
  - 1.1.1 Artificial intelligence
  - 1.1.2 Machine learning
  - 1.1.3 Deep learning
  - 1.1.4 Generative AI
- 1.2 Generative AI architectures**
  - 1.2.1 Transformer-based language models
  - 1.2.2 Diffusion models, VAEs and GANs
- 1.3 The generative AI technology stack**
  - 1.3.1 Foundation models
  - 1.3.2 Databases
  - 1.3.3 Hardware infrastructure
  - 1.3.4 Development platforms

**2 Market Analysis**

- 2.1 The generative AI industry landscape**
  - 2.1.1 Foundation model providers
  - 2.1.2 Development platform providers
  - 2.1.3 GPU-based hardware providers
- 2.2 Market sizing and forecast**
  - 2.2.1 Market value for GenAI models and platforms
  - 2.2.2 Market value for GenAI hardware
- 2.3 Solution provider market shares**
  - 2.3.1 The foundation model market
  - 2.3.2 The development platform market
  - 2.3.3 The GenAI hardware market
- 2.4 Foundation model benchmarks**
- 2.5 GenAI in IoT**
  - 2.5.1 Generative AIoT use cases
  - 2.5.2 Edge vs cloud deployments
  - 2.5.3 AIoT solution providers
- 2.6 GenAI in telecom**
  - 2.6.1 AI-on-RAN
  - 2.6.2 AI-for-RAN
  - 2.6.3 AI-and-RAN
- 2.7 Market trends**
  - 2.7.1 The emergence of low-cost models and platforms from China
  - 2.7.2 LLM providers suffer profitability issues
  - 2.7.3 Large regional differences in GenAI developments
  - 2.7.4 Telecoms providers invest in sovereign AI solutions
  - 2.7.5 Moving away from tokenisation
  - 2.7.6 Agentic AI gains traction
  - 2.7.7 Physical AI nears breakthrough with GenAI
  - 2.7.8 AI regulations affecting the GenAI market

**3 Company Profiles and Strategies**

- 3.1 01.AI**
- 3.2 AI21 Labs**
- 3.3 Aleph Alpha**
- 3.4 Alibaba**
- 3.5 Anthropic**
- 3.6 Assembly AI**
- 3.7 AWS**
- 3.8 Baichuan**
- 3.9 Baidu**
- 3.10 ByteDance**
- 3.11 C3 AI**
- 3.12 Cohere**
- 3.13 Databricks**
- 3.14 Dataiku**

- 3.15 DeepSeek**
- 3.16 Domino**
- 3.17 Elevenlabs**
- 3.18 Google**
- 3.19 H2O AI**
- 3.20 Hugging Face**
- 3.21 IBM**
- 3.22 Luma AI**
- 3.23 Mistral AI**
- 3.24 Meta**
- 3.25 Microsoft**
- 3.26 MiniMax**
- 3.27 Moonshot AI**
- 3.28 Nebius**
- 3.29 Nvidia**
- 3.30 OpenAI**
- 3.31 Oracle**
- 3.32 Runway**
- 3.33 SambaNova Systems**
- 3.34 Scale AI**
- 3.35 Stability AI**
- 3.36 Snowflake**
- 3.37 StepFun**
- 3.38 Tencent**
- 3.39 Together AI**
- 3.40 Weights & Biases**
- 3.41 xAI**
- 3.42 Z.ai**

**Glossary**

## Highlights from the report

Insights from executive interviews with market leading companies.

360-degree overview of the GenAI ecosystem.

Market value forecast on GenAI models, platforms and hardware until 2029.

Market shares for 55 key GenAI providers across models, platforms and hardware.

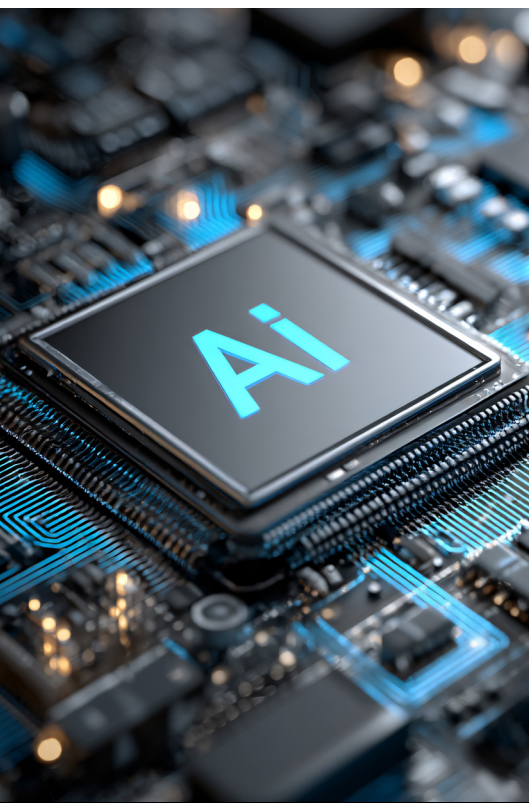
Detailed profiles of 42 key GenAI model and platform providers.

Use case examples from industries implementing GenAI.

In-depth analysis of market trends and key developments.

## Questions answered in the report

- How does GenAI technology work and what is the impact on businesses?
- How can GenAI be integrated into products and business processes?
- Which are the leading providers of GenAI models, platforms and hardware?
- What are the prices and pricing models for different GenAI solutions?
- What are the key success factors and challenges for stakeholders in the GenAI market?
- How does the GenAI landscape differ by region?
- How will the GenAI market evolve over the next five years?



## About Berg Insight's IoT market research

Our market reports offer comprehensive information and analysis on key IoT technologies and markets, addressing important concerns including total addressable market, market penetration, market shares, industry landscape, regulatory environment, market trends and forecasts. Our research portfolio today comprises more than 80 items, where each market report focuses on a specific vertical application area or cover horizontal themes. All market reports come with complementary data sets in Excel format that can be easily analysed and converted into tables and charts. We offer a range of different license options together with bundled packages and subscriptions to suit your specific needs.



HORIZONTAL THEMES

The Generative AI Market

Berg Insight estimates that the generative AI market experienced triple-digit-growth rates in all three major segments spanning GenAI hardware, foundation models and development platforms in 2024. The market is driven by significant data centre investments by cloud service providers, and over US\$ 400 billion in expected AI-related spending in 2025. The market value for foundation models reached an estimated US\$ 4.1 billion in 2024, while GenAI development platforms reached US\$ 17.0 billion. Meanwhile, GPU-based hardware systems used for GenAI workloads generated revenues of US\$ 132.3 billion in 2024.

PUBLISHED DATE	August 2025
EDITION	1st
PAGES	90
AUTHOR	Melvin Sörum

PDF & EXCEL: 1 user license	€ 1 800
PDF & EXCEL: 2-10 user license	€ 2 700
PDF & EXCEL: Enterprise license	€ 3 600

Read more and place order on [berginsight.com](https://berginsight.com)

Who should read this report?

The **Generative AI Market** is the foremost source of information about the emerging and impactful generative AI market. Whether you are a model provider, platform developer, hardware vendor, enterprise AI adopter, independent developer, investor, consultant or government agency, you will gain valuable insights from our in-depth research.

AUTHOR

Melvin Sörum



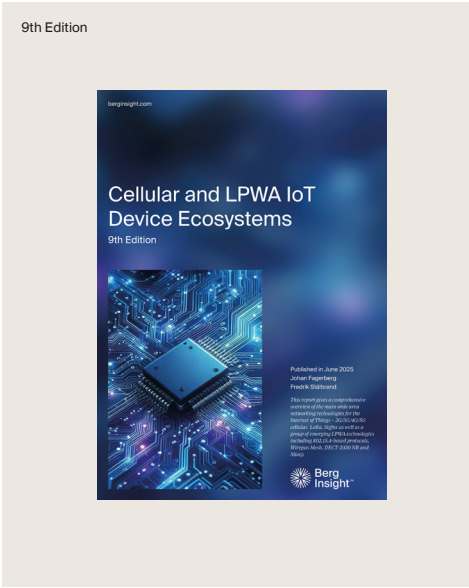
Melvin is an IoT analyst who specialises in the IoT connectivity and software markets. He contributes primarily to the horizontal research programmes, for which he produces market research about generative AI and private networks for IoT applications. Melvin holds a Master's degree in Industrial Engineering and Management from Chalmers University of Technology and joined Berg Insight in 2024.



Related products *Find them and more on [berginsight.com](https://berginsight.com)*



CATEGORY  
Horizontal Themes



CATEGORY  
Horizontal Themes



CATEGORY  
Horizontal Themes

CONTACT  
**Berg Insight AB**  
Viktoriagatan 3  
411 25 Gothenburg  
Sweden

**+46 (0)31 711 30 91**  
**[info@berginsight.com](mailto:info@berginsight.com)**  
**[www.berginsight.com](https://www.berginsight.com)**



*Berg Insight offers premier business intelligence to the telecom industry. We produce concise reports providing key facts and strategic insights about pivotal developments in our focus areas. Berg Insight also offers detailed market forecast databases and advisory services. Our vision is to be the most valuable source of intelligence for our customers.*