IDC Innovators

IDC Innovators: Generative AI Foundation Models, 2023 — Part 2

Ritu Jyoti   Andrew Gens

IDC INNOVATORS VENDOR LIST

FIGURE 1

IDC Innovators in Generative AI Foundation Models, 2023

IDC ⓘ Innovator®	Generative AI Foundation Models	
COMPANY NAME	**FOUNDED**	**HEADQUARTERS**
◆ AI21 Labs	2017	Tel Aviv, Israel
◆ Aleph Alpha	2019	Heidelberg, Germany
◆ Hugging Face	2022	Palo Alto, California

Source: IDC, 2023

EXECUTIVE SUMMARY

IDC Innovators are emerging vendors with revenue <$100 million that have demonstrated either a groundbreaking business model or an innovative new technology – or both. This IDC Innovators study profiles four vendors in the generative AI (GenAI) foundation model market: AI21 Labs, Aleph Alpha, and Hugging Face (see Figures 2-4).
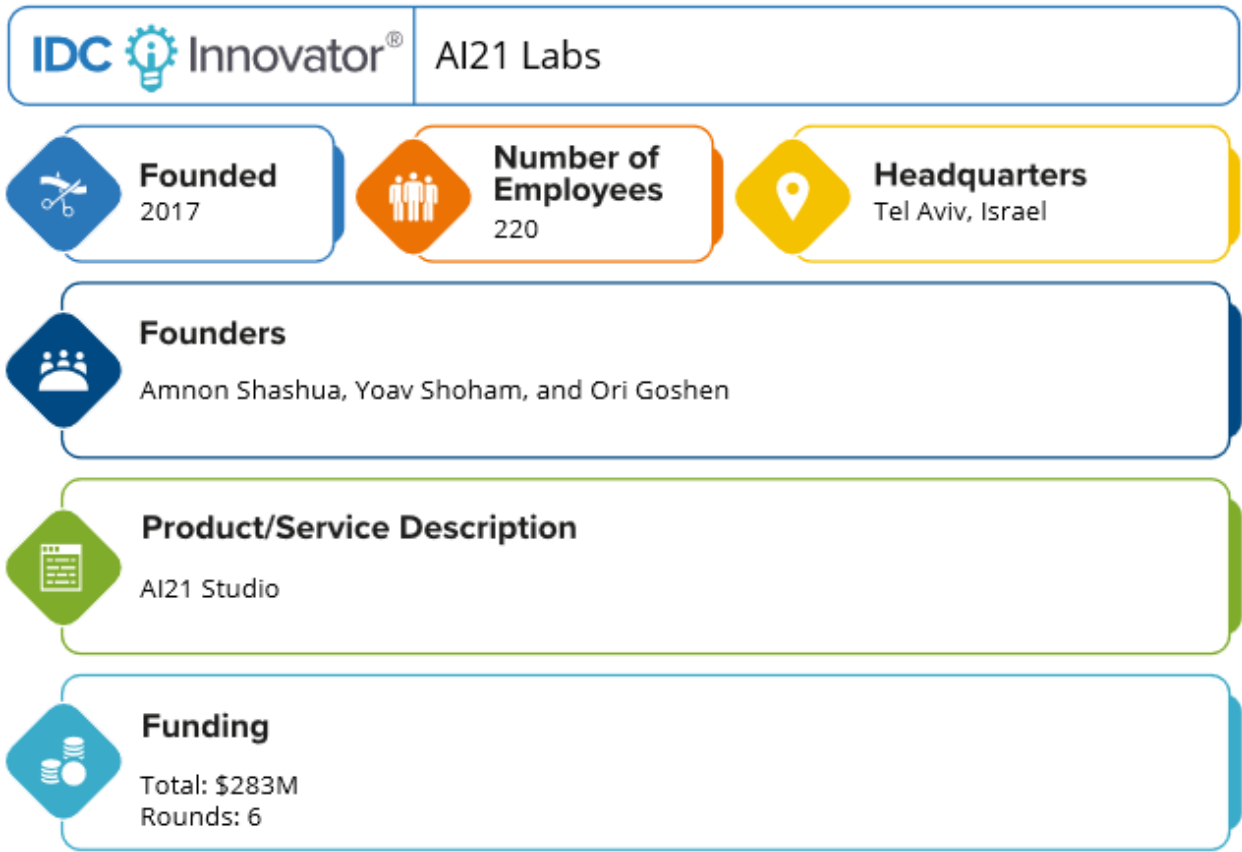
Generative AI foundation models are a rapidly emerging class of machine learning (ML) models that use large volumes of data to generate complete end products in text, image, or other modalities all via simple prompting. As demand for generative AI foundation models has grown rapidly in the past three years, the pool of vendors has expanded with many new entrants. Generative AI foundation models continue to mature quickly, and many new use cases are emerging every year prompting even more market entrants. Despite this strong overall market growth, generative AI foundation model vendors face various challenges from stiff market competition to strong pressures to differentiate products. Despite these challenges, this market is still new, and this gives vendors an opportunity to adapt their products and business models to increase their reach, relevancy, and reliability. Generative AI

foundation models will only continue to increase in relevancy in the years ahead. As new innovators emerge in this market, potential buyers and industry observers alike should remain vigilant to identify emerging trends and methods.

VENDOR PROFILE: AI21 LABS

FIGURE 2

AI21 Labs

IDC Innovator | AI21 Labs

Founded
2017

Number of Employees
220

Headquarters
Tel Aviv, Israel

Founders
Amnon Shashua, Yoav Shoham, and Ori Goshen

Product/Service Description
AI21 Studio

Funding
Total: $283M
Rounds: 6

Source: IDC, 2023

Why AI21 Labs Was Chosen as an IDC Innovator

AI21 Labs was chosen as an IDC Innovator due to its specialty in LLMs for natural language processing (NLP). Its platform offers a suite of tools and guides that empower developers and businesses to harness cutting-edge language models without requiring NLP expertise. With a focus on transforming reading and writing experiences, AI21 Labs' products, like AI21 Studio, provide exclusively text LLM solutions. Through the company's innovative approach to artificial intelligence (AI) and natural language processing, AI21 Labs is at the forefront of advancing the possibilities of language-based AI applications. Its other offering includes Wordtune.

IDC Innovator Assessment

- AI21 Labs offers integrations advertised with a simplicity of three lines of code or three clicks to connect for any user's favored stacks.

- Zero-shot instruction following is a new feature added with Jurassic-2. Zero-shot prompting is an increasingly in-demand capability that allows, in this case, Jurassic-2 to provide accurate answers to prompts without requiring examples.

- Jurassic-2 does not currently have a released training parameter count; however, Jurassic-1 was trained with 178 billion parameters, so it can be presumed that Jurassic-2 has expanded from this count.

Key Differentiator

Zero-shot prompting is a major differentiator for AI21 Labs, which was recently added with the release of Jurassic-2 in March 2023. While AI21 Labs is not the only LLM capable of zero-shot prompting, it is still far from being an industry-standard capability, and AI21 Labs is one of the smallest LLM companies, by valuation, that has been able to implement it. Beyond the added user capabilities granted by zero-shot prompting for expanding search outside of explicitly trained prompts, this also indicates that AI21 Labs, overall, is targeting some of the most advanced features of LLMs for inclusion in Jurassic-2 and future models.
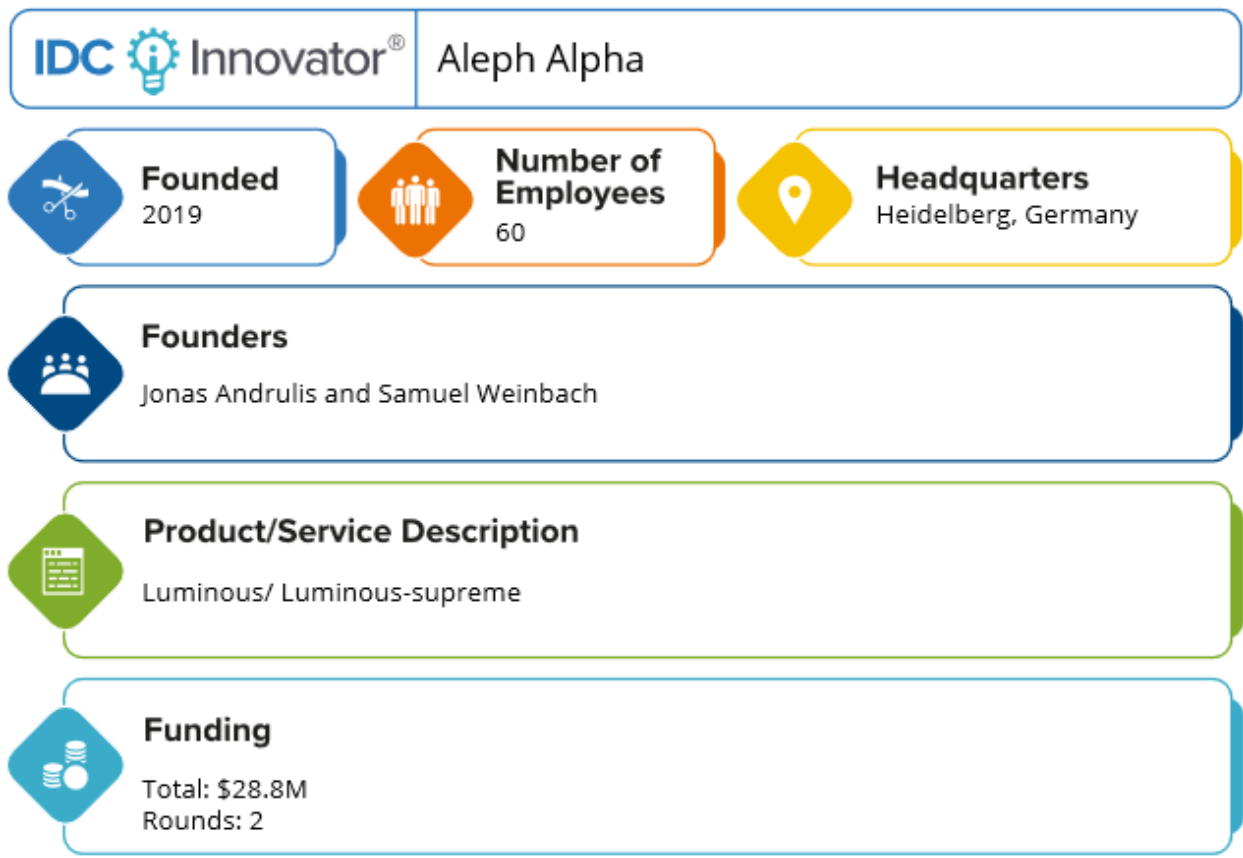
Challenges

Owing to AI21 Labs' focus in the most saturated segment of LLMs — text and code — the company faces an uphill battle to achieve market awareness as well as differentiation. While AI21 Labs is an experienced foundation model vendor, it is competing with companies like OpenAI which have enjoyed explosive growth in revenue and brand recognition in the last year alone. AI21 Labs has already taken a major step to alleviate the awareness challenge this year when it entered into a partnership with Amazon to launch its Bedrock AI tool. AI21 Labs will need to continue to leverage partnerships as well as its embracing of advanced LLM features, such as zero-shot prompting to maximize its competitive momentum.

FIGURE 3

Aleph Alpha

Source: IDC, 2023

Why Aleph Alpha Was Chosen as an IDC Innovator

Aleph Alpha was chosen as an IDC Innovator due to its innovative approaches for an AI assistant for multimodal generative AI across a range of industries. Based in Europe, Aleph Alpha also has a particular focus on the European market with special considerations made for ease of use by European companies. Despite this regional focus, Aleph Alpha is a global software provider. Aleph Alpha has prioritized many high-level and low-level features in its development that aim to emphasize customer ease of use and visibility into the generative AI process. Aleph Alpha's Luminous-supreme is a multimodal foundation model, including image text to computer text, object detection, contextual image Q&A, and image captioning.

IDC Innovator Assessment

- Aleph Alpha is one of three primary recipients of SAP's strategic investment in generative AI start-ups, as well as a partner of Hewlett Packard Enterprise (HPE) to develop its GreenLake LLM cloud market.

- Luminous endpoint explainability, utilizing Aleph Alpha's AtMan method, shows users relative log probabilities of input words relating to the production of an output word. Once a "suppression" or "amplification" of a prompt word occurs, users can see how their AtMan affected the output word.

- Luminous-supreme is the company's most capable model currently utilizing approximately 70 billion parameters with a limit of 2,048 prompt tokens.

Key Differentiator

Aleph Alpha's focus on sovereign European AI is a key differentiating factor for the company's organization and technology. Aleph Alpha, as a Germany-based organization, is leading the way for a new wave of European-centric AI, targeting the specific business, cultural, and regulatory features of the broad European business ecosystem. A key example of this focus in effect is the design of Luminous to adhere to the European Union (EU) AI Act. This focus on sovereign European AI likely enables Aleph Alpha's customers to avoid a constant worry of regulatory noncompliance as Aleph Alpha already ensures this.
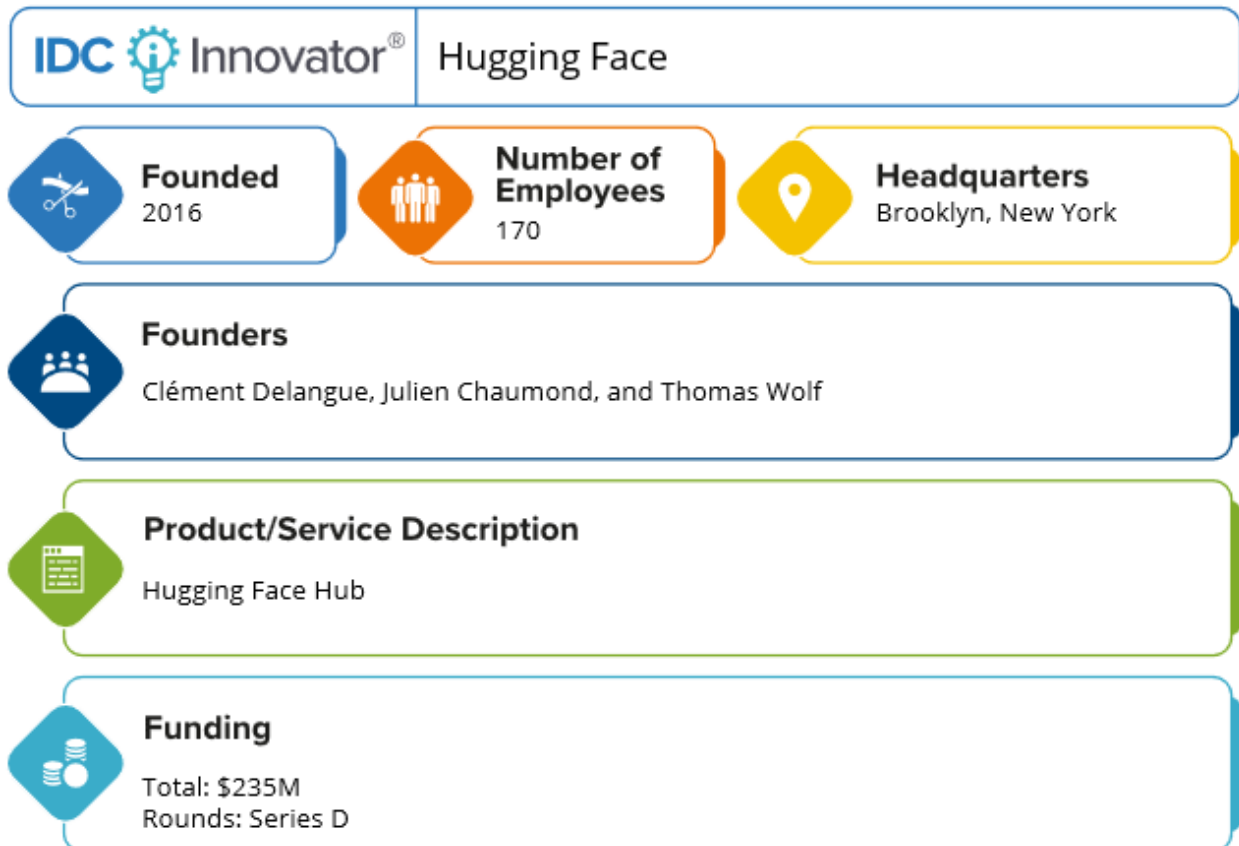
Challenges

Aleph Alpha faces the challenge of relatively late entry into, arguably, the current most cutting-edge segment of the LLM industry – multimodal models. Specifically, due to its late entry, Aleph Alpha faces competition issues such as a relative parameter deficit as well as a prompt token limit shortfall compared with other established multimodal LLMs. While there is a serious downside to pursuing the multimodal path out of the gate, Aleph Alpha will likely still benefit from its early-mover status. If Aleph Alpha can continue to improve its model accuracy across its multimodalities, it likely will reach a more stable competitive position as newer entrants emerge.

FIGURE 4

Hugging Face

IDC Innovator® | Hugging Face

Founded
2016

Number of Employees
170

Headquarters
Brooklyn, New York

Founders
Clément Delangue, Julien Chaumond, and Thomas Wolf

Product/Service Description
Hugging Face Hub

Funding
Total: $235M
Rounds: Series D

Source: IDC, 2023

Why Hugging Face Was Chosen as an IDC Innovator

Hugging Face was chosen as an IDC Innovator due to its uncommon open source hosting platform approach for NLP and ML domains as well as LLMs. Hugging Face hosts and distributes pretrained models as open source software, creating a repository for ML and a marketplace for researchers, data scientists, and engineers to share and exchange their work. In addition to this repository, Hugging Face operates proprietary AI services including a transformer library, Spaces, and tokenizers.

IDC Innovator Assessment

- Hugging Face Hub maintains open source repositories for LLM, NLP, and other AI models.
- Hugging Face also maintains collections of transformers, tokenizers, and data sets for users to review and select for potential configuration and use.

- Within its own services, Hugging Face utilizes several LLMs such as GPT-3 for its Space NLP application building and sharing offering; BLOOM for its text generator; StarCoder, developed in collaboration with ServiceNow for code generation; and T5 by Google AI for inferencing endpoints.

Key Differentiator

A key differentiator of Hugging Face's open source platform is the potential cost savings for customers and the potential to configure a more well-tailored solution for an organization. A growing number of organizations are gaining greater visibility into the positive impacts of LLMs for businesses but don't really know where to start. Hugging Face, although containing highly technical components in its repositories, channels its complete focus, with its proprietary offerings, toward assisting companies with getting started with low-knowledge requirement solutions, such as Hugging Face Spaces.

Challenges

As AI democratization grows and GenAI ushers in more low-code (LC)/no-code (NC) capabilities and simplified interfaces, the configuration path for acquiring a generative AI model may become less critical for companies as the savings diminish compared with SaaS offerings/managed services, which likely will be cheaper and more inclusive of various use cases as time goes on. Hugging Face's core proprietary services of onboarding AI models will need to be a continually larger driver for Hugging Face, specifically leveraging its coding assistant tool, StarCoder, and other related software to further alleviate current and future adoption-side stresses on AI software users.

TECHNOLOGY DEFINITION

Generative AI foundation models are a class of machine learning models that are trained on diverse data and can be adapted or fine-tuned for a wide range of downstream tasks. Transformers and LLMs are subsets of foundation models.

IDC INNOVATORS INCLUSION CRITERIA

An "IDC Innovators" document recognizes emerging vendors chosen by an IDC analyst because they offer an innovative new technology or a groundbreaking business model, or both, and were approved by the IDC Innovators Review Panel. It is not an exhaustive evaluation of all companies in a segment or a comparative ranking of the companies.

An IDC Innovators document highlights vendors that meet the following criteria:

- In IDC's opinion, the company exhibits innovative technology or a new business model.
- The company has annual revenue under $100 million at the time of selection.
- Customers are currently using the company's products and services (i.e., the products and services are not conceptual or are in the process of being released).
- The product, service, or business model must solve or help alleviate an IT buyer challenge.

In addition, vendors in the process of being acquired by a larger company may be included provided the acquisition is not finalized at the time of publication of the document. Vendors funded by venture capital firms may also be included even if the venture capital firm has a financial stake in the vendor's company.

Related Research

- *Worldwide Artificial Intelligence Applications Market Shares, 2022: Embedded Artificial Intelligence Is Becoming Increasingly Important, Transforming Our Daily Life and Technology Usage* (IDC #US51181823, September 2023)
- *Worldwide Artificial Intelligence Software Forecast, 2023-2027* (IDC #US50027023, September 2023)
- *Generative AI: Preparing for the Long-Term Impact - Practical Applications and Guidance for CIOs/CTOs* (IDC #US51148223, August 2023)
- *Unlocking Business Success with Generative AI* (IDC #US50789223, June 2023)
- *Generative AI Platforms and Applications: Market Trends and Forecast, 2Q23* (IDC #US50700423, May 2023)
- *Generative Artificial Intelligence: A New Chapter for Enterprise Business Applications* (IDC #US50471523, March 2023)
- *Enterprise Automation 2.0: The Connective Tissue of the Digital Business* (IDC #DR2023_GS4_RJ, March 2023)

Synopsis

IDC Innovators are emerging vendors with revenue <$100 million that have demonstrated either a groundbreaking business model or an innovative new technology — or both. This IDC Innovators document profiles software vendors providing innovative generative AI foundation models, whether text, image, or multimodal foundation models. This market is growing rapidly with many new entrants emerging and established vendors establishing new offerings in this space.

"Foundation models represent an important paradigm shift in artificial intelligence (AI) and are integral to the future of AI in the enterprise," says Ritu Jyoti, group vice president, Artificial Intelligence and Automation research with IDC's Software Market Research and Advisory Practice. "To truly harness the potential of foundation models in the enterprise, there needs to be significant investment in the entire technology stack — not just the models themselves."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology, IT benchmarking and sourcing, and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.).

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com