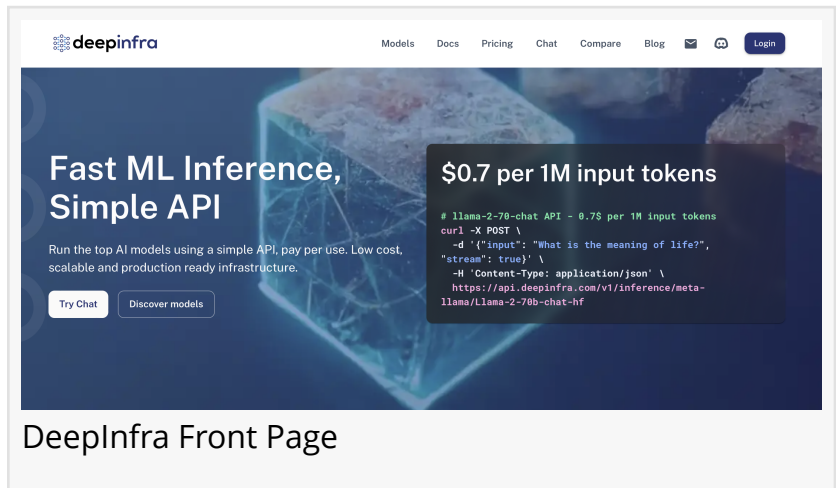


DeepInfra raises \$8M seed round to build and scale the fastest inference cloud for Large Language Models

DeepInfra allows users to harness the power of cutting-edge machine learning models with minimal effort, reducing complexity and saving both time and resources

PALO ALTO, CA, USA, December 11, 2023 /EINPresswire.com/ -- [DeepInfra](https://deepinfra.com/), a leading provider of scalable and cost-effective infrastructure for running AI models through a user-friendly API, have announced a \$8 million seed funding round from A Capital and Felicis.



This investment will propel DeepInfra's mission to revolutionize the AI landscape, providing seamless access to top AI models for businesses across the globe. DeepInfra offers an unparalleled platform that simplifies AI model integration and execution, catering to a wide range of applications and use cases.

With this round of funding, DeepInfra is poised to further enhance its technology, expand its model offerings, and optimize its infrastructure to provide users an even more efficient and cost-effective inference solutions and enhanced model customization options, including fine-tuning and other cutting-edge techniques. The company's dedication to delivering a superior experience for developers and businesses remains steadfast, and this funding will accelerate the realization of that vision.

DeepInfra's innovative approach allows users to harness the power of cutting-edge machine learning models with minimal effort, reducing complexity and saving both time and resources. The "pay per use" [pricing](#) model ensures cost-efficiency, enabling users to only pay for the actual usage of the service. DeepInfra provides access to exceptional open-source models like [Llama 2](#) and CodeLlama, offering a remarkable value proposition with significantly lower pricing of \$1 per 1M tokens when compared to alternatives like OpenAI's turbo-3.5.

"We're extremely grateful for the support and confidence shown by A Capital and Felicis," said Nikola Borisov, Founder and CEO of DeepInfra. "This funding will significantly accelerate our efforts to democratize access to AI models, empowering businesses to leverage advanced AI capabilities without the burden of managing complex infrastructure at low cost with privacy in mind."

"AI, fueled by large language models, is eating software. Soon, every app will be infused with AI, making the demand for powerful yet affordable inference services nearly infinite. The rise of outstanding open-source models only adds fuel to the fire. DeepInfra's team, having already mastered the art of scaling web services for millions globally, is perfectly positioned for this wave. Just as the web depended on CDNs and load balancers, AI is setting the stage for a huge leap in inference services. Deep Infra is right at the heart of this shift," said Ramu Arunachalam, General Partner at A Capital.

About DeepInfra:

DeepInfra is a Palo Alto-based company that offers a simple and cost-effective API to run the top AI models, providing scalable and production-ready infrastructure. With a pay-per-use model and a focus on ease of use, DeepInfra allows developers to effortlessly integrate and run machine learning models, saving both time and money.

About A Capital:

A.Capital provides resources and counsel to the creators, inventors and entrepreneurs who are building a new future. Founded by Ronny Conway in 2014, A.Capital invests in early-stage consumer, enterprise and crypto startups and offers world-class advice, connections, and team-building expertise to help those companies scale. A.Capital's portfolio includes companies like Notion, Hugging Face, Airbnb, Pinterest, Databricks, and Replit. Learn more at www.acapital.com.

Ivelina Abadjieva

DeepInfra

ivelina@deepinfra.com

Visit us on social media:

[Twitter](#)

[LinkedIn](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/673903358>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2023 Newsmatics Inc. All Right Reserved.