



Fastly Helps Developers Build a Better Internet with New AI Accelerator

June 13, 2024

AI Accelerator underscores Fastly's commitment to developers with unique tools that produce fast, secure and compelling experiences at the edge

SAN FRANCISCO--(BUSINESS WIRE)-- [Fastly, Inc.](#) (NYSE: FSLY), a leader in global edge cloud platforms, today announced the launch of Fastly AI Accelerator, the company's first AI solution designed to create a better experience for developers by helping improve performance and reduce costs across the use of similar prompts for large language models (LLM) apps.

"AI technologies generally and large language models specifically are aggressively reshaping the technology industry, and the way millions worldwide – developers included – work every day," said Stephen O'Grady, Principal Analyst with RedMonk. "There's a lot of focus on the largest models. However, developers and enterprises alike are turning in large numbers to medium and smaller models. Whether it's to lower costs, to shorten training cycles or to run on more limited hardware profiles, they're an increasingly important option."

Fastly AI Accelerator is designed to reduce API calls and costs with intelligent, semantic caching. Built on Fastly's Edge Cloud Platform and leveraging industry leading caching technology, AI Accelerator uses a specialized API gateway to dramatically improve performance for apps using popular LLMs, beginning with ChatGPT and expanding support to include additional models.

Popular AI applications can process hundreds of thousands of API calls or questions daily. Many of the questions users ask are likely very similar and may have been asked before. Without semantic caching, each call to the LLM requires going back to the provider for the information, potentially increasing costs and latency. However, Fastly AI Accelerator's semantic caching provides a cached response for repeated queries directly from Fastly's high performance edge platform, instead of going back to the AI provider, helping to deliver a better experience by improving performance while reducing costs.

"At Fastly, we're always listening to developers, to understand both what they're excited about and what their biggest pain points are," said Anil Dash, Vice President of Developer Experience with Fastly. "Fastly AI Accelerator gives developers exactly what they want, by making the experience of their favorite LLMs a lot faster and more efficient, so they can focus on what makes their app or site unique, and what keeps their users happy."

When using Fastly AI Accelerator, developers only need to update their app to use a new API endpoint, which typically only requires changing a single line of code. Fastly AI Accelerator will then transparently implement semantic caching for OpenAI compatible APIs. This approach goes beyond traditional caching as Fastly AI Accelerator is able to understand the context of the requests and queries, and will send a similar response if two or more requests are alike.

To help developers build faster, more secure and more engaging experiences, Fastly is also now making it even easier for developers to try Fastly with an expanded free account tier that helps coders set up a new site, create a new app, or launch a new service in just a few minutes. Free tier accounts also include access to Fastly's Content Delivery Network (CDN), generous memory and storage allotments, uncapped redirects, page rules and regular expressions. Plus, the free Fastly tier includes security features such as TLS and always-on DDoS mitigation, observability tools, and much more.

To learn more about Fastly AI Accelerator and other Fastly innovations visit fastly.com/ai. Developers looking to try Fastly for free can get started [here](#).

About Fastly, Inc.

Fastly's powerful and programmable edge cloud platform helps the world's top brands deliver online experiences that are fast, safe, and engaging through edge compute, delivery, security, and observability offerings that improve site performance, enhance security, and empower innovation at global scale. Compared to other providers, Fastly's powerful, high-performance, and modern platform architecture empowers developers to deliver secure websites and apps with rapid time-to-market and demonstrated, industry-leading cost savings. Organizations around the world trust Fastly to help them upgrade the internet experience, including Reddit, Wendy's, Stripe, Neiman Marcus, Universal Music Group, SeatGeek, and Advance Publications. Learn more about Fastly at <https://www.fastly.com>, and follow us [@fastly](#).

Forward-Looking Statements

This press release contains "forward-looking" statements that are based on Fastly's beliefs and assumptions and on information currently available to Fastly on the date of this press release. Forward-looking statements may involve known and unknown risks, uncertainties, and other factors that may cause its actual results, performance, or achievements to be materially different from

those expressed or implied by the forward-looking statements. These statements include, but are not limited to, those regarding the expected benefits and functionality of Fastly AI Accelerator, including its ability to help developers accelerate performance and reduce costs or provide cached responses for repeated queries. Except as required by law, Fastly assumes no obligation to update these forward-looking statements publicly, or to update the reasons actual results could differ materially from those anticipated in the forward-looking statements, even if new information becomes available in the future. Important factors that could cause Fastly's actual results to differ materially are detailed from time to time in the reports Fastly files with the Securities and Exchange Commission (SEC), including in Fastly's Annual Report on Form 10-K for the fiscal year ended December 31, 2023, and our Quarterly Reports on Form 10-Q. Copies of reports filed with the SEC are posted on Fastly's website and are available from Fastly without charge.

Source: Fastly, Inc. 

Media Contact

Spring Harris

press@fastly.com

Investor Contact

Vernon Essi, Jr.

ir@fastly.com

Source: Fastly, Inc.