NEWS RELEASE

Cloudflare Collaborates with Microsoft to Enable AI Models to Run Anywhere

9/27/2023

Building the AI railroad tracks of the future so inference tasks can move seamlessly across devices, networks, and clouds

SAN FRANCISCO--(BUSINESS WIRE)-- **Cloudflare, Inc.** (NYSE: NET), the leading connectivity cloud company, today announced that it is collaborating with Microsoft to make it easier for companies to run AI in the location most suitable for their needs. As inference tasks become more and more distributed, this collaboration will enable businesses to seamlessly deploy AI models across a computing continuum that spans device, network edge, and cloud environments – maximizing the benefits of both centralized and distributed computing models. By leveraging **ONNX Runtime** across these three tiers, Cloudflare and Microsoft can ensure that AI models can be run wherever processing makes the most sense across this three-tier architecture – from the hyperscale cloud to the hyper-distributed network edge to devices themselves – that best addresses the bandwidth, latency, connectivity, processing, battery/energy, and data sovereignty and localization demands of a given application or service.

AI model training requires significant computational and storage resources in close proximity to one another, making centralized cloud platforms the best environment for the intensive calculations needed in model training. While training will continue to be centralized, inference tasks will be increasingly performed in more distributed locations, specifically on devices themselves and on edge networks. For example, some inference tasks (e.g., an autonomous vehicle braking at the sight of a pedestrian) will run on the physical device for the lowest possible latency. However, to navigate device limitations such as compute, storage, and battery power, more and more tasks will need to run on edge networks. Edge networks – close geographical proximity to end users and devices – will provide an optimal balance of computational resources, speed, and data privacy. Some applications may require moving through three tiers of this computing continuum, with device, edge network, and cloud environments

working together to bring the best experience to the end user.

"Together, Cloudflare and Microsoft will build the railroad tracks that AI traffic and tasks will move on, to tailor AI inference to the exact needs and demands of every organization," said Matthew Prince, CEO and co-founder, Cloudflare. "Whether you're looking for speed or accuracy, dealing with energy or connectivity bandwidth challenges, or complying with regional localization requirements, Cloudflare and Microsoft can help you find the best location for your AI tasks."

"As companies explore the best way to harness the power of generative AI in unique ways to meet their needs, the ability to run AI models anywhere is paramount," said Rashmi Misra, GM of Data, AI, & Emerging Technologies at Microsoft. "With Cloudflare's global network, combined with Microsoft's experience in training and deploying the world's most advanced AI workloads through our Azure cloud, businesses will gain access to a new level of flexibility and performance for AI inference."

Together, Cloudflare and Microsoft will collaborate to make it easy for companies to run AI in the place most suitable for the workload. There are two pieces to making this happen:

1. Microsoft's ONNX Runtime creates a standardized solution that allows the same models to be deployed regardless of environment, whether on device (Windows, mobile, or in-browser), on the distributed network edge (Cloudflare), or in Azure's centralized cloud platform.
2. Cloudflare can provide the infrastructure for routing traffic across the different environments, depending on connectivity, latency, compliance, or other requirements.

Businesses want to be able to move inference tasks across this continuum of device, edge network, and cloud, depending on the performance, cost, and regulatory requirements they face. Microsoft's AI capabilities and hyperscale cloud infrastructure combined with Cloudflare's hyper-distributed edge network will empower businesses to drive innovation and efficiency across the entire AI lifecycle. As a result, businesses will be able to:

- Find the best location for AI tasks: Choose to deploy AI inference wherever processing makes the most sense to achieve the desired outcomes, maximizing the benefits of both centralized and distributed computing models. For example, a security camera system could utilize edge networks to run object detection. This overcomes the resource constraints of the device itself, without the latency of sending data to a central server for processing.
- Navigate changing needs: Run models in all three locations and adjust or fall back based on availability, use case, and latency requirements.
- Deploy on Cloudflare in a few clicks: Access easy deployable models and ML tooling capabilities on Workers AI through Microsoft Azure Machine Learning.

To learn more, please check out the resources below:

- Register to reserve your access to **Workers AI**
- Blog: **Workers AI: The inference cloud running on GPUs on Cloudflare's network**
- Blog: **The best place on Region: Earth for inference**
- Learn how to build full-stack AI applications with Cloudflare at **AI.cloudflare.com**

About Cloudflare

Cloudflare, Inc. (NYSE: NET) is the leading connectivity cloud company. It empowers organizations to make their employees, applications and networks faster and more secure everywhere, while reducing complexity and cost. Cloudflare's connectivity cloud delivers the most full-featured, unified platform of cloud-native products and developer tools, so any organization can gain the control they need to work, develop, and accelerate their business.

Powered by one of the world's largest and most interconnected networks, Cloudflare blocks billions of threats online for its customers every day. It is trusted by millions of organizations – from the largest brands to entrepreneurs and small businesses to nonprofits, humanitarian groups, and governments across the globe.

Learn more about Cloudflare's connectivity cloud at **cloudflare.com/connectivity-cloud**. Learn more about the latest Internet trends and insights at **https://radar.cloudflare.com**.

Follow us: **Blog** | **X** | **LinkedIn** | **Facebook** | **Instagram**

Forward-Looking Statements

This press release contains forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, which statements involve substantial risks and uncertainties. In some cases, you can identify forward-looking statements because they contain words such as "may," "will," "should," "expect," "explore," "plan," "anticipate," "could," "intend," "target," "project," "contemplate," "believe," "estimate," "predict," "potential," or "continue," or the negative of these words, or other similar terms or expressions that concern Cloudflare's expectations, strategy, plans, or intentions. However, not all forward-looking statements contain these identifying words. Forward-looking statements expressed or implied in this press release include, but are not limited to, statements regarding the capabilities and effectiveness of Cloudflare's products and technology, the benefits to Cloudflare's customers from using Cloudflare's products and technology, Cloudflare's expected partnership with Microsoft and the potential resulting benefits to Cloudflare customers, the potential benefits to customers from the expected ability to integrate Cloudflare and Microsoft products and technology, the potential opportunity for Cloudflare to attract additional customers and to expand

sales to existing customers through Cloudflare's expected partnership and product and technology integrations with Microsoft, Cloudflare's technological development, future operations, growth, initiatives, or strategies, and comments made by Cloudflare's CEO and others. Actual results could differ materially from those stated or implied in forward-looking statements due to a number of factors, including but not limited to, risks detailed in Cloudflare's filings with the Securities and Exchange Commission (SEC), including Cloudflare's Quarterly Report on Form 10-Q filed on August 3, 2023, as well as other filings that Cloudflare may make from time to time with the SEC.

The forward-looking statements made in this press release relate only to events as of the date on which the statements are made. Cloudflare undertakes no obligation to update any forward-looking statements made in this press release to reflect events or circumstances after the date of this press release or to reflect new information or the occurrence of unanticipated events, except as required by law. Cloudflare may not actually achieve the plans, intentions, or expectations disclosed in Cloudflare's forward-looking statements, and you should not place undue reliance on Cloudflare's forward-looking statements.

Cloudflare, Inc.
Daniella Vallurupalli
Vice President, Head of Global Communications
press@cloudflare.com

Source: Cloudflare, Inc.