November 12, 2026

Intel, RadixArk Collaborate to Advance Open AI Inference on Intel Platforms

SANTA CLARA, Calif.--(BUSINESS WIRE)-- Intel Corporation today announced a collaboration with RadixArk to optimize SGLang, an open-source AI inference stack, for Intel platforms that aims to help developers and enterprises deploy leading AI models and build AI applications more efficiently and cost-effectively.

“Enterprise AI is moving beyond isolated performance benchmarks to solving full-system production deployments,” said Anil Nanduri, vice president, AI Products & GTM, Intel Data Center Group. “This collaboration ensures developers have Day-0 readiness and high-performance inference software directly optimized for Intel CPUs, GPUs and accelerators with an open approach and gives them the freedom to choose the tools, frameworks and technologies that best fit their needs.”

Why It Matters

Enterprises require heterogeneous, open inference infrastructure, as AI models grow larger and more complex. To deliver AI in production it requires that every prompt is routed to the right place, matching the different workloads with the right intelligence, right compute and right economics. Open inference runtimes provide the essential connective tissue that makes full-stack deployments practical, reliable, and economically sustainable across diverse enterprise environments.

What It Delivers for Developers

This collaboration expands developer choice to deploy leading AI models on enterprise-proven Intel hardware. Model builders and software engineers can build and innovate on standard, open frameworks and interfaces while taking full advantage of Intel’s compute density and built-in hardware acceleration features, including Intel Advanced Matrix Extensions (Intel AMX). Developers gain faster time-to-production, reduced latency, and flexible deployment options across existing data center and cloud capacity.

"Our collaboration with Intel is a major milestone for open-source AI inference, enabling SGLang optimization directly to Intel’s platforms,” said Ying Sheng, CEO, RadixArk. “By upstreaming deep optimizations into the core framework, we are eliminating constraints and ensuring developers never have to compromise between open standards and high performance. Our mission to make frontier AI infrastructure open and accessible to every builder, and Intel’s enterprise-ready architecture gives development teams the flexibility and compute density they need to scale production AI workloads seamlessly."

An Open Blueprint for Production AI

Bringing hardware and software optimization together reflects Intel’s approach of

contributing improvements directly to community software, rather than treating them as separate deployment decisions. These upstream contributions to popular engines and frameworks aim to provide high-performance paths for turning breakthrough models into scalable, production-grade applications while preserving flexibility, choice and broad architectural freedom.

About Intel

Intel (Nasdaq: INTC) designs and manufactures advanced semiconductors that connect and power the modern world. Every day, our engineers create new technologies that enhance and shape the future of computing to enable new possibilities for every customer we serve. Learn more at intel.com.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20261112043035/en/

Intel Media Relations
ContactPR@Intel.com

Source: Intel Corporation