July 18, 2024

From Athletes to GenAI Developers: Intel Tackles Real-World Challenges with AI Systems

Intel empowers GenAI developers and athletes around the globe with open and accessible AI systems.

SANTA CLARA, Calif.--(BUSINESS WIRE)-- **What's New:** Today, Intel shares exciting details on its collaboration with the International Olympic Committee (IOC) and on an industry-driven generative AI (GenAI) retrieval-augmented generation (RAG) solution. These announcements demonstrate how open AI systems and platforms using Intel® Gaudi® AI accelerators and Intel® Xeon® processors put the power in the hands of developers and enterprises to tackle challenges created by the AI boom.

This press release features multimedia. View the full release here: https://www.businesswire.com/news/home/20240718825653/en/

An illustration shows a retrieval-augmented generation solution flow. (Credit: Intel Corporation)

> *"Through our partnership with the International Olympic Committee, we are demonstrating our dedication to making AI accessible. We're fostering an open playing field that encourages innovation and creativity and enables developers and enterprises to build tailored AI solutions that drive tangible results. By embracing an open and collaborative ecosystem, Intel is transforming ways to help our athletes and pushing the boundaries of what's possible with our customers."*
> – Justin Hotard, Intel executive vice president and general manager of the Data Center and Artificial Intelligence Group

How Athlete365 Works: Qualifying for the Olympic Games is only the beginning for athletes. To help about 11,000 athletes with varying languages and cultures navigate the venue and comply with rules and guidelines, the IOC collaborated with Intel to develop a chatbot, Athlete365. A RAG solution powered by Intel Gaudi accelerators and Xeon processors, Athlete365 is capable of handling athlete inquiries and interactions and will deliver on-demand information during athletes' stay at the Olympic Village in Paris, enabling them to focus on training and competing.

Why It Matters: Deploying GenAI solutions poses challenges like cost, scale, accuracy, development requirements, privacy and security. RAG is a crucial GenAI workload because it allows companies to leverage proprietary data securely, enhancing the timeliness and reliability of AI outputs. This improves the quality and usefulness of AI applications, which is critical in today's data-driven world.

Intel's collaborative approach utilizing AI platforms, open standards, and a robust software and systems ecosystem allows developers to build customized GenAI RAG solutions tailored to each enterprise's needs. The momentum shared today underscores Intel's commitment to providing open, robust and composable multi-provider generative AI solutions.

How the GenAI RAG Solution Architecture Works: Intel works with industry partners to create an open source, interoperable solution for easy RAG deployment. The GenAI solution is an industry-driven, out-of-the-box, production-ready RAG solution built on the Open Platform for Enterprise AI (OPEA) foundation. While the GenAI turnkey solution offers a streamlined approach to deploying RAG solutions for enterprises in their data centers, it is designed to be highly flexible and customizable, integrating components from a catalog of offerings by multiple OEM systems and industry partners.

The GenAI turnkey solution integrates OPEA-based microservice components into a scalable RAG solution designed to deploy Xeon and Gaudi AI systems. It scales seamlessly with proven orchestration frameworks like Kubernetes and Red Hat OpenShift and provides standardized APIs with security and system telemetry.

- **Breaking Down Proprietary Walls with an Open Software Stack:** Nearly all large language model (LLM) development is based on the high abstraction framework PyTorch, which is supported by Intel Gaudi and Xeon technologies, making it easy to develop on Intel AI systems or platforms. Intel has worked with OPEA to develop an open software stack for RAG and LLM deployment optimized for the GenAI turnkey solution and built with PyTorch, Hugging Face serving libraries (TGI and TEI), LangChain and Redis Vector database.
- **Meeting Developers Where They are:** OPEA offers open source, standardized, modular and heterogeneous RAG pipelines for enterprises, focusing on open model development and support for various compilers and toolchains. This foundation accelerates containerized AI integration and delivery for unique vertical use cases. OPEA unlocks new AI possibilities by creating a detailed, composable framework that stands at the forefront of technology stacks.

With GenAI turnkey and the comprehensive enterprise AI stack, Intel delivers a complete solution addressing the challenges of deploying and scaling RAG and LLM applications within enterprises and data centers. Leveraging Intel-powered AI systems or platforms and optimized software in OPEA, businesses can harness the full potential of GenAI with greater efficiency and speed.

What's Next: Increasing access to the latest AI compute technology is a challenge enterprises face in enabling critical business outcomes with GenAI. Through strategic collaborations with industry partners and customers, Intel is creating new opportunities for AI services driven by GenAI and RAG solutions.

Committed to the secure and responsible advancement of AI, Intel announced today its collaboration with Google, IBM and other industry partners in a new Coalition for Secure AI (CoSAI), created to enhance trust and security in AI development and deployment.

Intel will further demonstrate its unique approach to AI systems and continued customer and partner momentum at Intel Innovation on Sept. 24-25.

More Context: Intel Welcomes the Coalition for Secure AI | Intel Brings AI-Platform Innovation to Life at the Olympic Games

Video: Empowering Individuals with Open AI Systems – During a roundtable presentation on July 17, 2024, representatives of Intel, the International Olympic Committee, Seekr and Red Hat discussed how an open AI ecosystem puts the power in the hands of the developers and enterprises tackling challenges created by the boom in artificial intelligence. Participants in the event moderated by Justin Hotard, executive vice president and general manager of the Data Center and AI Group at Intel, were Kaveh Mehrabi, director of the Athletes Department of the International Olympic Committee; Steven Huels, vice president and general manager of Business, Product and Strategy at Red Hat; Rob Clark, president and chief technology officer of Seekr; and Bill Pearson, vice president of Software Solutions and Ecosystem of the Data Center and AI Group at Intel. (Credit: Intel Corporation)

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

© 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/20240718825653/en/

Bats Jafferji
1-603-809-5145
bats.jafferji@intel.com

Source: Intel Corporation