

November 12, 2019



Intel Speeds AI Development, Deployment and Performance with New Class of AI Hardware from Cloud to Edge

SAN FRANCISCO--(BUSINESS WIRE)-- **What's New:** Today at a gathering of industry influencers, Intel welcomed the next wave of artificial intelligence (AI) with updates on new products designed to accelerate AI system development and deployment from cloud to edge. Intel demonstrated its Intel® Nervana™ Neural Network Processors (NNP) for training (NNP-T1000) and inference (NNP-I1000) — Intel's first purpose-built ASICs for complex deep learning with incredible scale and efficiency for cloud and data center customers. Intel also revealed its next-generation Intel® Movidius™ Myriad™ Vision Processing Unit (VPU) for edge media, computer vision and inference applications.

This press release features multimedia. View the full release here:

<https://www.businesswire.com/news/home/2019112005277/en/>

A photo shows the Intel Nervana NNP-T for training packaged chip. Intel Nervana Neural Network Processors are Intel's first purpose-built ASICs for complex deep learning with scale and efficiency for cloud and data center customers. Intel demonstrated the Intel NNPs at the company's AI Summit on Nov. 12, 2019, in San Francisco. (Credit: Intel Corporation)

"With this next phase of AI, we're reaching a breaking point in terms of computational hardware and

memory. Purpose-built hardware like Intel Nervana NNPs and Movidius Myriad VPUs are necessary to continue the incredible progress in AI. Using more advanced forms of system-level AI will help us move from the conversion of data into information toward the transformation of information into knowledge."

—Naveen Rao, Intel corporate vice president and general manager of the Intel Artificial Intelligence Products Group

Why They Are Important: These products further strengthen Intel's portfolio of AI solutions, which is expected to generate more than \$3.5 billion in revenue in 2019. The broadest in breadth and depth in the industry, Intel's AI portfolio helps customers enable AI model development and deployment at any scale from massive clouds to tiny edge devices, and everything in between.

What Intel Announced: Now in production and being delivered to customers, the new Intel Nervana NNPs are part of a systems-level AI approach offering a full software stack developed with open components and deep learning framework integration for maximum use.

The Intel Nervana NNP-T strikes the right balance between computing, communication and

memory, allowing near-linear, energy-efficient scaling from small clusters up to the largest pod supercomputers. The Intel Nervana NNP-I is power- and budget-efficient and ideal for running intense, multimodal inference at real-world scale using flexible form factors. Both products were developed for the AI processing needs of leading-edge AI customers like Baidu and Facebook.

“We are excited to be working with Intel to deploy faster and more efficient inference compute with the Intel Nervana Neural Network Processor for inference and to extend support for our state-of-the-art deep learning compiler, Glow, to the NNP-I,” said Misha Smelyanskiy, director, AI System Co-Design at Facebook.

Additionally, Intel’s next-generation Intel Movidius VPU, scheduled to be available in the first half of 2020, incorporates unique, highly efficient architectural advances that are expected to deliver leading performance — more than 10 times the inference performance as the previous generation — with up to six times the power efficiency of competitor processors. Intel also announced its new Intel® DevCloud for the Edge, which along with the Intel® Distribution of OpenVINO™ toolkit, addresses a key pain point for developers — allowing them to try, prototype and test AI solutions on a broad range of Intel processors before they buy hardware.

Why It Matters: Incredibly complex data, models and techniques are required to advance deep learning reasoning and context, bringing about a need to think differently about architectures.

With most of the world running some part of its AI on Intel® Xeon®™ Scalable processors, Intel continues to improve this platform with features like Intel® Deep Learning Boost with Vector Neural Network Instruction (VNNI) that bring enhanced AI inference performance across the data center and edge deployments. While that will continue to serve as a strong AI foundation for years, the most advanced deep learning training needs for Intel customers call for performance to double every 3.5 months, and those types of breakthroughs will only happen with a portfolio of AI solutions like Intel’s. Intel is equipped to look at the full picture of computing, memory, storage, interconnect, packaging and software to maximize efficiency, programmability and ensure the critical ability to scale up distributing deep learning across thousands of nodes to, in turn, scale the knowledge revolution.

More Context: [2019 AI Summit](#) (Press Kit) | [Artificial Intelligence at Intel](#) (Press Kit) | [At Hot Chips, Intel Pushes ‘AI Everywhere’](#)

About Intel

Intel (NASDAQ: INTC), a leader in the semiconductor industry, is shaping the data-centric future with computing and communications technology that is the foundation of the world’s innovations. The company’s engineering expertise is helping address the world’s greatest challenges as well as helping secure, power and connect billions of devices and the infrastructure of the smart, connected world – from the cloud to the network to the edge and everything in between. Find more information about Intel at newsroom.intel.com and 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/2019112005277/en/>

Dan Francisco

916-377-9509

Daniel.Francisco@intel.com

Source: Intel Corporation