intel ai

Edge AI

Partner Enablement Package

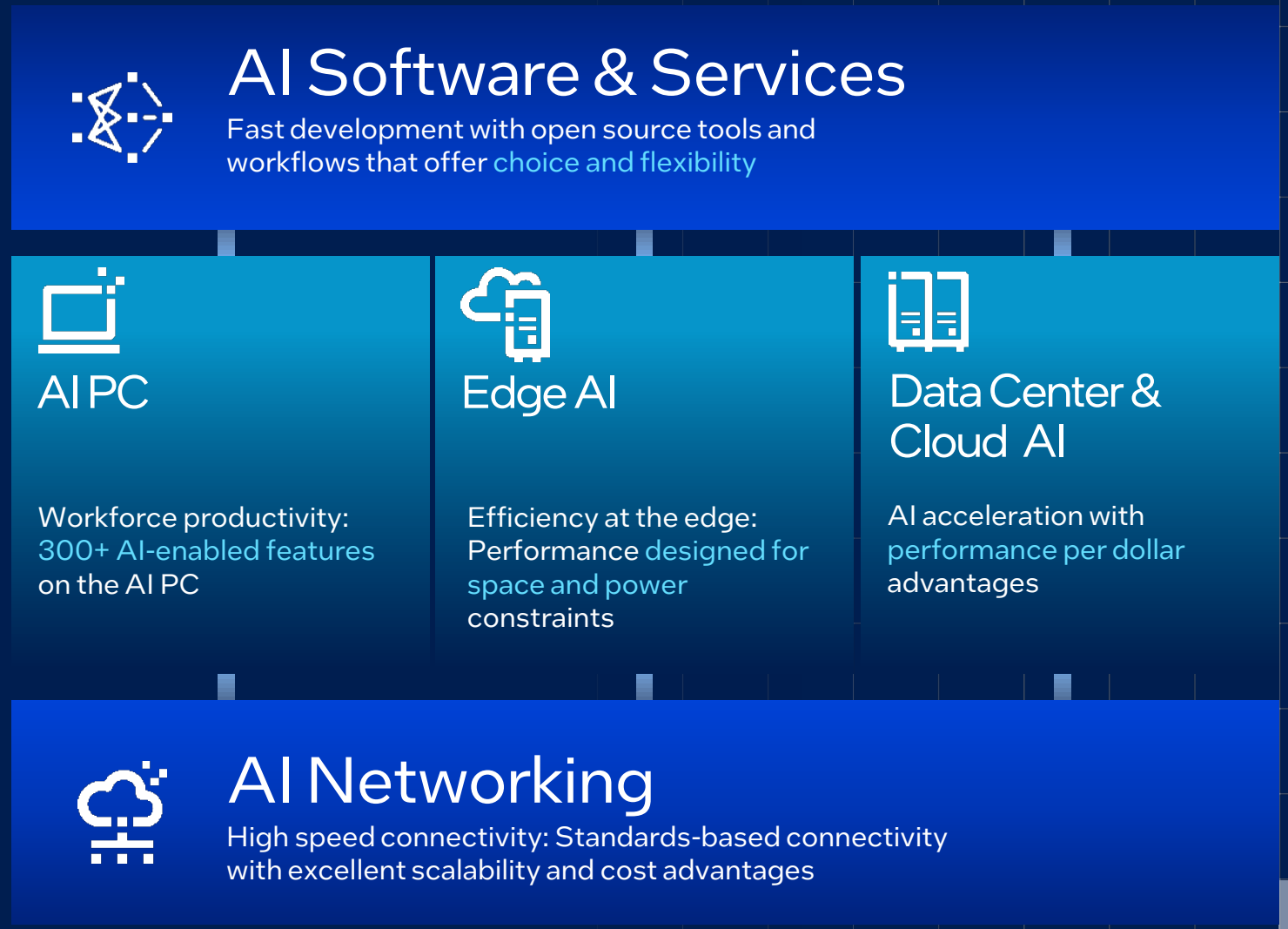
Learn how the combination of AI and edge computing delivers near-real-time value to businesses

Contents

intel

Bringing AI Everywhere

Scalable AI computing platforms



intel ai

Fast Development in an Open Ecosystem

Open source tools and workflows offer choice and flexibility

AI Software & Services

intel ai Software

Accelerate on an open ecosystem

Open frameworks and libraries to accelerate performance—includes PyTorch optimizations and GenAI models on Hugging Face

OpenVINO

Develop once, deploy everywhere

Deploy across diverse hardware with minimal code modification

intel Open Edge Platform

Bring AI to the edge

Open source, secure, and scalable, Intel-driven and partner-enabled platform for optimized edge and AI solution development, deployment and management

oneAPI

No vendor lock-in

Optimized libraries for all major AI frameworks, with tools to migrate code from CUDA

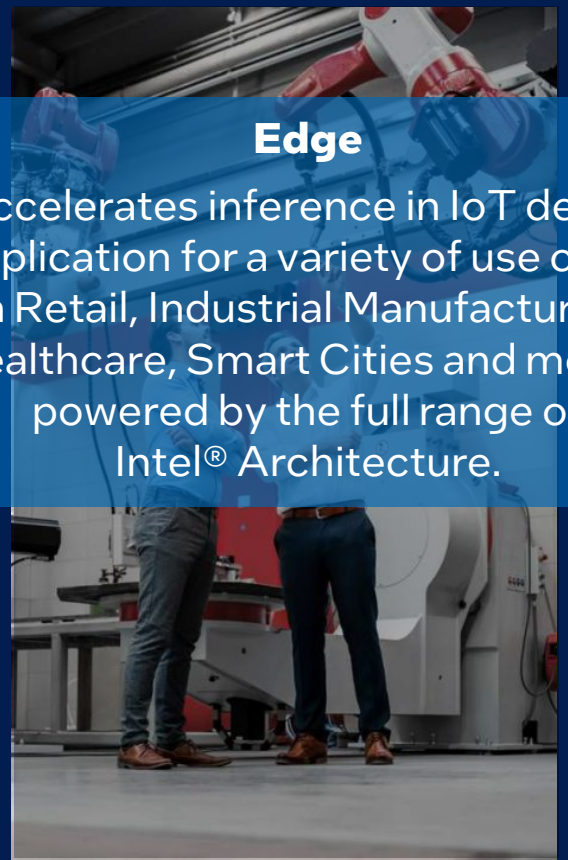
intel ai

visit openvino.ai

Deploy AI Everywhere with OpenVINO™

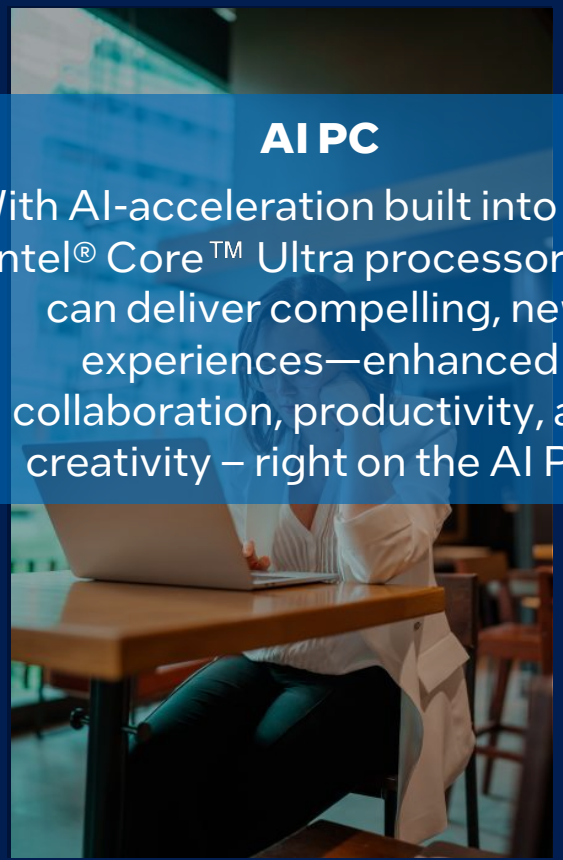
Maximize hardware with the OpenVINO™ toolkit and take advantage of integrated accelerators to run AI workloads with optimal cost efficiency

WHITE PAPER > Generative AI with OpenVINO™ toolkit



Edge

Accelerates inference in IoT device application for a variety of use cases in Retail, Industrial Manufacturing, Healthcare, Smart Cities and more—powered by the full range of Intel® Architecture.



AI PC

With AI-acceleration built into every Intel® Core™ Ultra processor, you can deliver compelling, new experiences—enhanced collaboration, productivity, and creativity – right on the AI PC.

Cloud

Support fast-growing cloud workloads powered by Intel® Xeon® Scalable processors featuring built-in accelerators for more performance-per-core and unmatched AI performance.

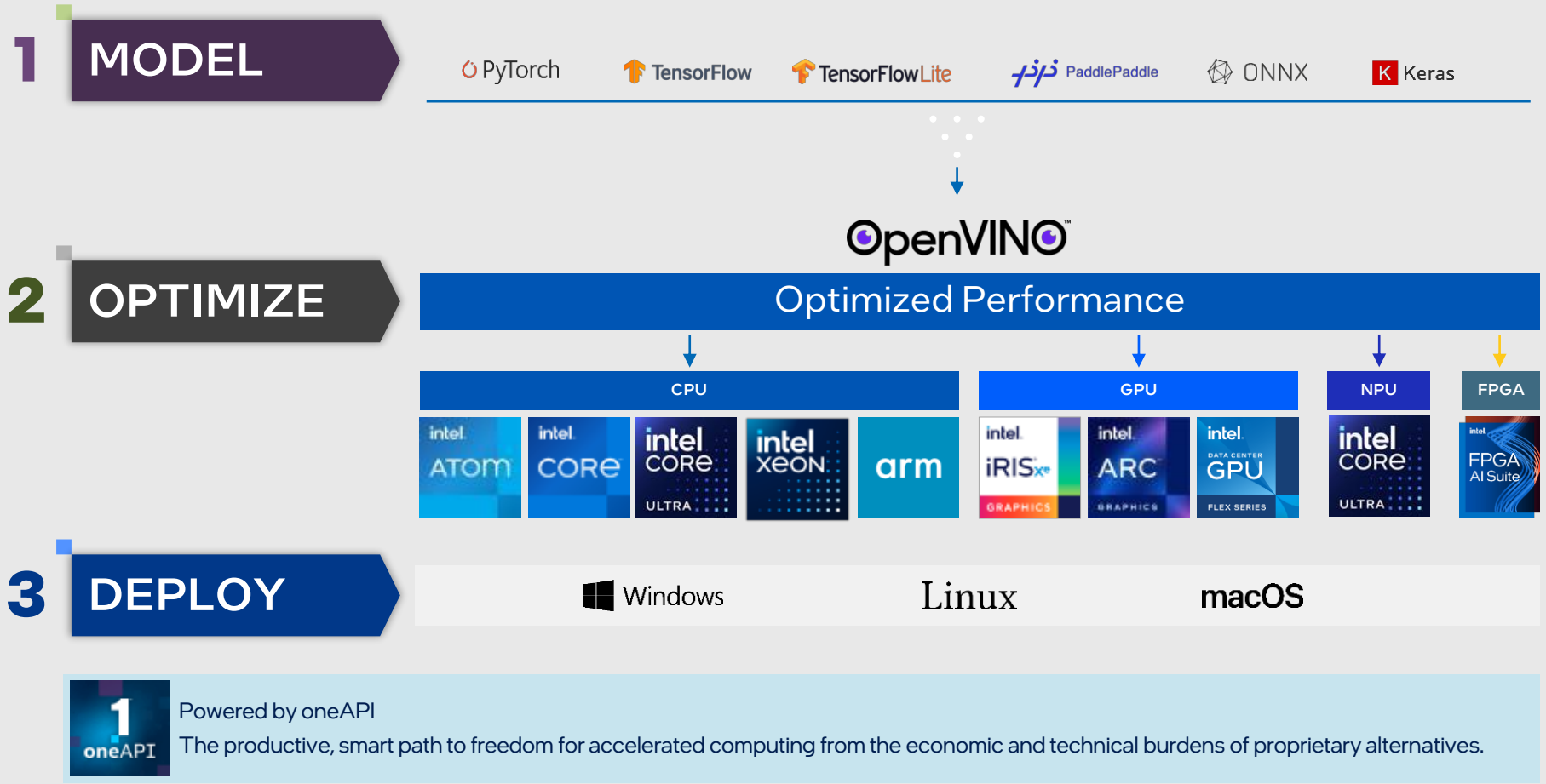
Powered by oneAPI

intel

Fast, Accurate Results with High-Performance

OpenVINO™ is a powerful toolkit that accelerates AI workloads such as computer vision, generative AI, audio, speech, language, and recommendation systems.

What sets OpenVINO™ apart is its versatility. With a "write once, deploy anywhere" approach, developers can craft an application or algorithm and execute it across various Intel architectures, including CPU, GPU, NPU, FPGA, and the ARM CPU architecture.



MORE INFO >

OpenVINO™ Gold Deck

intel

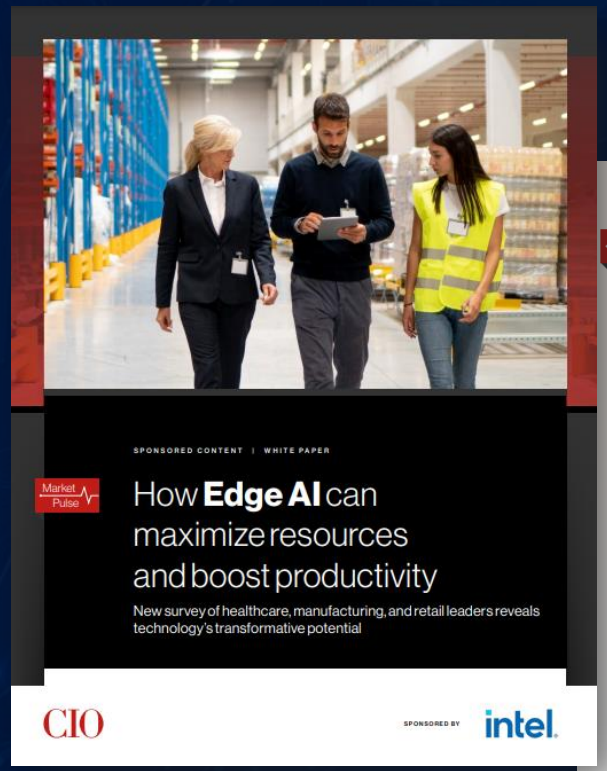

Edge AI

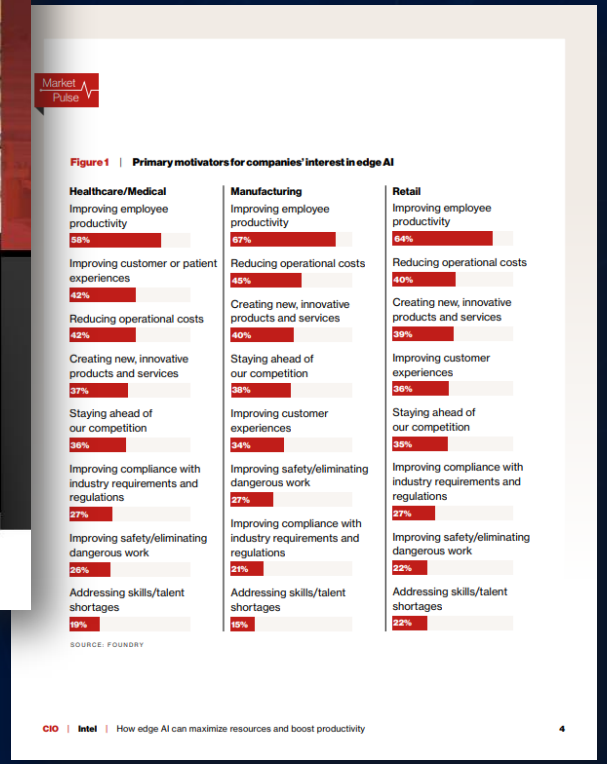
What is Edge AI?

Edge AI brings artificial intelligence (AI) to "the edge," closer to where data is generated. Unlocking new levels of business insight, efficiency, and innovation.

Innovative capabilities at the edge, facilitated by advancements in computing performance and efficiency are bringing together the physical and digital worlds.

Edge AI, which brings AI to local devices and sensors, enables rapid data analysis and action independent of the cloud or data center. This unlocks **near-real-time responsiveness and insights**, **increased efficiency**, **reduced operational costs**, and the ability to deliver **new types of customer experiences**.





LEARN MORE

- Article - What is Edge AI ?
- Report - How Edge AI can maximize resources and boost productivity

intel

Edge AI Opportunities

Processing data closer to where it is being generated enables greater processing speeds, increased volumes, and real-time, action-led results[1]

50%

of edge computing deployments will involve machine learning by 2026

Source: Gartner

"Without the enhancements that edge computing brings, enterprises will be left behind in a hypercompetitive marketplace, where speed of insight, security and expanded analytics are driving forces." – Jack Gold, Founder and Principal Analyst, J. Gold Associates, LLC

intel®

offers an edge-centric approach that has delivered over **90,000** real-world edge deployments and **200M** processors sold over the past 10 years[2]

intel

[1] Edge Computing | Accenture

[2] Internal Intel data - 2024

Empowering Industry Verticals

Accelerating digital and AI transformation

Health, Education & Consumer Industries	Health IT, medical imaging, digital signage, point-of-sale, interactive flat panel displays, multifunctional printers
Cities & Critical Infrastructure	Smart cities, safety and security, road and rail infrastructure, airports, electric vehicle charging, government edge infrastructure
Federal & Industrial	Manufacturing, robotics, commercial buildings, warehousing and logistics, utilities, integrated energy companies, military, aerospace, government

intel

Edge AI Use Cases

LEARN MORE

Edge AI Tech Talks

Edge AI unlocks use cases with applications for every industry, enabling enterprises to deploy to the far and near edges, and utilize the cloud for scalable operations

Retail

- Inventory management
- Smart retail shop
- Loss prevention
- Counterfeit detection
- Self-checkout
- Customer flow analysis
- Smart shelving

Manufacturing

- Asset tracking
- Defect detection
- Targeted performance reports
- Automated safety measures
- Predictive maintenance
- Efficiency optimization

Energy

- Equipment monitoring
- Asset tracking
- Fuel management
- Hazard detection
- Peak shaving
- Consumer & worker safety

Transportation

- Traffic intersection monitoring
- Performance monitoring
- Ticketing systems
- Passenger information
- Fleet object identification
- Theft detection

intel

MORE INFO >

Infographic - Your distinct AI edge in manufacturing: A safer workforce at a lower cost

Whitepaper - The state of edge AI in manufacturing: moving to a new paradigm

Edge AI Use Cases Resource Guide

Intel® AI Portfolio for the Edge

Intel® AI Portfolio for the Edge

Intel® Xeon® processor

Bringing AI to the edge

Incredible performance per watt for rigorous network and edge workloads, including AI, where space and power are limited

Intel® Core™ Ultra processor

Supercharge AI vision at the edge

Take on challenging AI workloads at the edge and analyze more video streams with multiple compute engines working together

Intel® Arc™ GPU

Edge GPUs deliver fast AI inference

Outstanding compute density and energy efficiency for AI inference at the edge with purpose-built acceleration

intel

INTEL SOFTWARE VIDEOS >

Edge AI Reference Kits

Choose Hardware for Specific AI Workloads

Light AI

Efficiency for sub-100W designs

CPU AI, built-in GPU, built-in NPU

intel. CORE™ + intel. ATOM™ + Xe

Medium AI

Scale up perf/W for diverse system designs

Discrete GPU & built-in CPU AI

intel. XEON™ + intel. CORE™ + intel. ARC™ GRAPHICS

Heavy AI

Optimize for peak perf and density

Optimize with high-end Discrete GPU

intel. XEON™ + intel. CORE™ + intel. ARC™ GRAPHICS

Intel® Xeon® Processors for Edge Workloads

Intel® Xeon® 6 SoC

Trusted Xeon® cores in a dense, integrated System-on-a-Chip (SoC) package designed to address space and power constraints

Acceleration

Media, network, and AI accelerators

Integration

Intel® QuickAssist Technology and Intel® Ethernet in one BGA package

Long life and power optimization

IO die with Intel 4 process for the highest efficiency and density and long-life options to support edge requirements

Intel® Xeon® 6 SoC increases RAN capacity by up to

2.4X

versus previous generation processor[1]

Intel® Xeon® 6 SoC 36C delivers up to

14.25X

Performance/Watt gain versus Intel® Xeon® Gold 6538N 32C on Media Transcode[2]

LEARN MORE >

- 30-3-30
- Product Brief
- Website

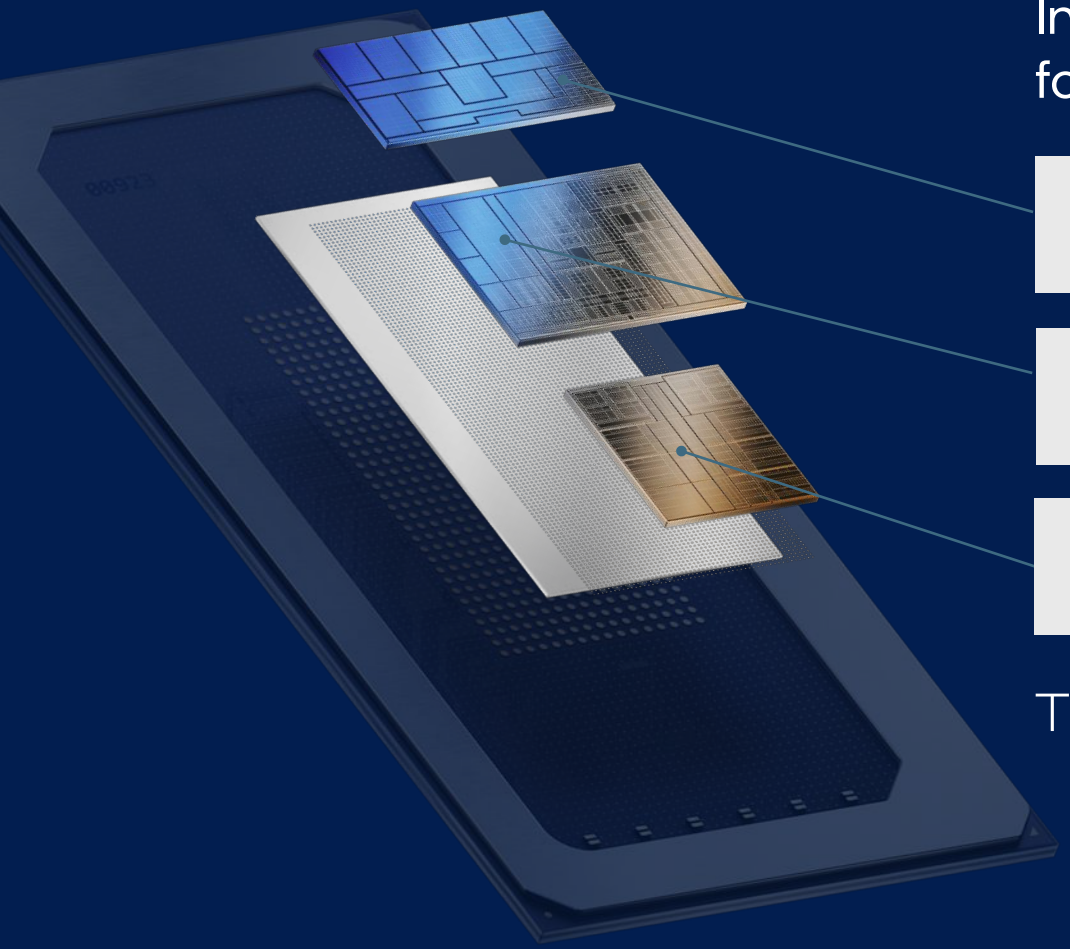
[1] See [7ND21] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

[2] See [7ND32] at intel.com/processorclaims: Intel® Xeon® 6. Results may vary.

intel ai

Intel® Core™ Ultra for the Edge

Intel® Core™ Ultra Processors



Intel® Core™ Ultra Processors provide the right engines for all AI workloads

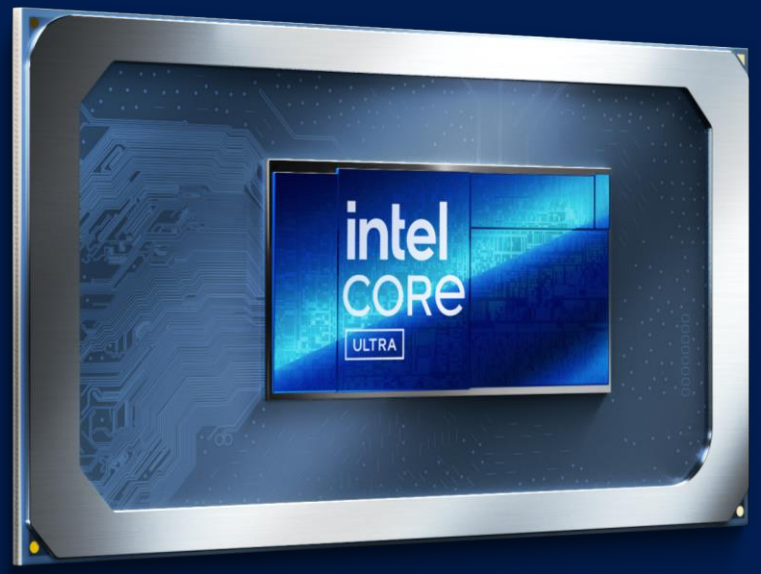
GPU	**High Throughput**	Ideal for AI-accelerated high complexity workloads
NPU	**Dedicated Low Power AI Engine**	Ideal for sustained AI and AI offload for battery life
CPU	**Fast Response**	Ideal for lightweight, single-inference, low-latency AI tasks

The right balance of power and performance for AI

PRODUCT BRIEF

Intel® Core™ Ultra Processors

Intel® Core™ Ultra Processors (Series 2)



DESKTOP >

Intel® Core™ Ultra desktop processors (series 2) are the ultimate desktop and entry workstation platform, engineered to unlock new levels of intelligent performance for the most demanding daily tasks.

LEARN MORE >

- Product Brief
- Quick Reference Guide
- How to Sell Guide

MOBILE >

Intel® Core™ Ultra mobile processors (series 2) are high-efficiency processors built to deliver next-gen AI experiences in sleek and slim mobile form factors. They feature the latest generation of P-core, E-core, and low-power E-core architectures, advanced NPU AI Engines, and available with built-in Intel® Graphics or Intel® Arc™ GPUs.[1]

LEARN MORE >

- Product Brief
- Quick Reference Guide

LAUNCH ANNOUNCEMENT CES 2025 > Intel to Power Large PC Refresh with New Silicon-Based Security

[1]**Intel® Arc™ Graphics:** Intel® Arc™ graphics only available on select Intel® Core™ Ultra processor (Series 2) powered systems; minimum processor power required. OEM enablement required. Check with OEM or retailer for system configuration.

Intel® Core™ Ultra Benefits for the Edge

AI-ready performance

- Multiple integrated compute engines for AI — P-cores, E-cores, Intel® Arc™ GPU[1] and Intel® AI Boost, a built-in neural processing unit (NPU) for increased edge AI capabilities at low power
- Enable/accelerate AI inferencing cost-effectively without discrete accelerator

Immersive graphics and media

- Up to 8 Xe-cores (128 graphics execution units) for graphics/media-intensive workloads at the edge
- Built-in GPU reduces power consumption, lower build of material costs and enables smaller form factor design

12W to 65W TDP in a Ball-Grid Array (BGA) package

- As low as 12W in thermal design power (TDP) option for fanless design
- Full performance with 65W
- Simplify design with integrated platform controller hub (PCH)

Optimized for Edge

up to 1.5X AI performance[2]

up to 1.81X Graphics performance[2]

up to 2.56X AI performance/watt[2]

READ WHITE PAPER Unveiling the Hardware and Software Foundation of Intel® Core™ Ultra Processors for the Edge

intel

1. Intel® Arc™ GPU only available on select H-Series, Intel® Core™ Ultra processor powered systems with at least 16GB of system memory in a dual-channel configuration. OEM enablement required; check with OEM for system configuration details.
2. Performance varies by use, configuration, and other factors. Learn more at intel.com/processorclaims: Intel® Core™ Ultra processors, Edge. Results may vary.

Industry Applications

Unlock the potential of Edge AI and Computer Vision with Intel

READ BRIEF

Retail

- AI POS
- Digital Signage
- Interactive Kiosk
- Video Wall

Education

- Interactive Whiteboard
- Remote Classroom
- Video Conference

Hospitality

- Digital Signage
- Interactive Kiosk
- In-Store Analytics

Gaming

- Slot Machine
- Electronic Table Game
- Lottery Ticket Kiosks

Healthcare

- AI-augmented Imaging with Ultrasound, X-Ray and Endoscopy

Smart Cities

- License Plate Recognition
- Traffic Management
- Network Video Recorder

Industrial

- Autonomous Mobile Robot
- Vision based Defect Detection
- AI-Augmented Process Control

Key Features

Display: 4 concurrent 4K displays, Pipelock, EDID, Bezel Compensation
Media: Integrated HDMI capture
GPU virtualization with SR-IOV

AI-Capable: Intel Deep Learning Boost with Int8 support, Inferencing with CPU+iGPU+NPU, up to 50 simultaneous 1080p streams ingestion

Intel® Core™ Ultra Processors (Series 1): Up to 16 Cores / 22 Threads, 8 Xᵉ-Cores, 8 lanes PCIe 5.0, 20 lanes PCIe 4, LPDDR5-6400 memory

intel

Case Studies with Intel® Core™ Ultra

Luminar Neo Ramps Up More AI-Driven Photo Effects Faster

READ MORE

AI enhances user experiences and delivers richer photo editing capabilities, building on the new functionality of Intel® Core™ Ultra processors and the OpenVINO™ toolkit

Intel's initiatives in the AI space, both on the CPU and GPU side, as well as the company's decision to add NPUs to all its new processors, allow us to improve the user experience for our creators significantly." Dmytro Mykhalchuk, VP of Product & Engineering, Skylum

VEGAS Pro Brings Out the Best of the New Intel® Core™ Ultra Processors

WATCH VIDEO READ MORE

Powered by Intel® Core™ Ultra processors, VEGAS Pro activates hardware-accelerated, AI-assisted video editing, taps the 2X boost in graphics performance and offers integral support for HDR visuals

"Our apps heavily use the GPU for demanding media processing workloads. MAGIX appreciates the integration of an NPU in the new Intel® Core™ Ultra processor as the next step towards an even more energy-efficient heterogenous compute environment." Jägen Hirche, CTO, MAGIX

Intelligent Safety Solutions Leveraging AI at the Edge

READ MORE

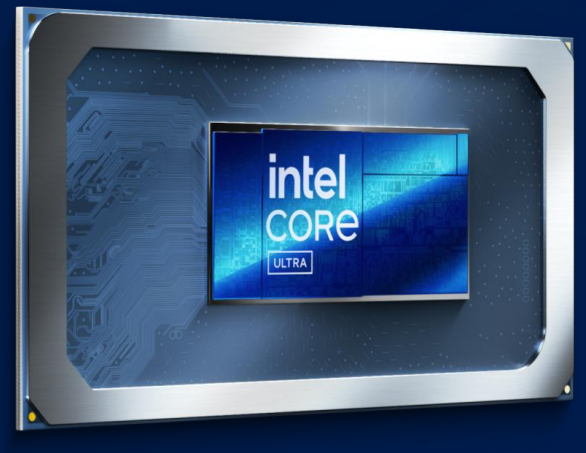
We are seeing an impressive 75% and 100% boost in video analytic workload capacity for our flagship SecurOS® Auto and SecurOS® Tracking Kit applications compared to the 11th Gen Intel Core processor, thanks in part to the significant improvements in the built-in Intel Arc GPU."

Aluisio Figueiredo, CEO Intelligent Security Systems

intel

Competitive Benchmarks

Deploy integrated AI and graphics in a BGA package



up to 3.4X
faster than NVIDIA Jetson AGX Orin in media performance[1]

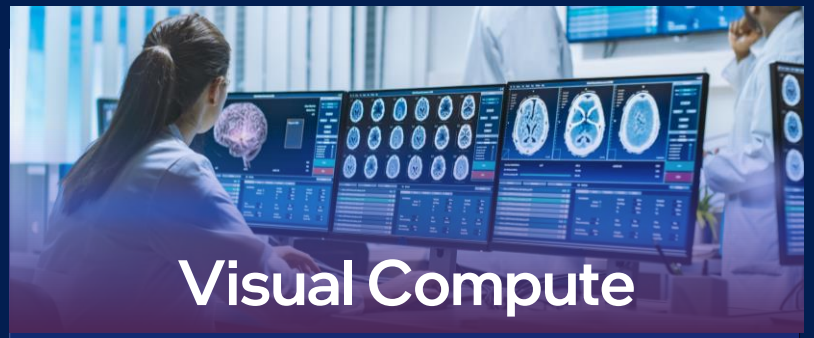
up to 2.7X
faster than NVIDIA Jetson AGX Orin in video analytics end-to-end AI pipeline performance[1]

up to 8.3X
better performance/W/$ than NVIDIA Jetson AGX Orin in video analytics end-to-end AI pipeline performance[1]

intel [1]Performance varies by use, configuration, and other factors. Learn more at intel.com/processorclaims: Intel® Core™ Ultra processors, Edge. Results may vary.

Intel® Arc™ GPU for the Edge

Target Workloads



Visual Compute

- Simulation & Visualization
- Display: Video Wall, Interactive Flat Panel, Kiosk
- 3D Rendering and Visualization, e.g. medical imaging

AI Inference

- Media Analytics
- Machine Vision & Visual Inference: Objection Detection & Classification
- Natural Language Processing

Media Processing

- Media processing: Encode/Transcode & Streaming; Compression
- Content Creation
- Video Production

READ WHITE PAPER

Unlocking the AI Power of Intel® Arc™ GPU for the Edge: A Deep Dive into Hardware and Software Enablement

intel

Intel® Arc™ Pro B-Series Graphics Cards

For Design & Engineering Software

Expedite complicated design changes, speed up visualization, and ensure precise viewport linework in tools like AutoCAD, Solidworks, Maya, 3ds Max, and Inventor.

For Architecture & Construction Software

Accelerate the graphics performance of complex models to gain seamless viewport interactions and responsiveness in architectural tools like Revit, Enscape, MicroStation, and Rhino.

For Product Design Software

Create detailed 3D models with responsive lighting, textures, and shadows essential for persuasive client presentations or design exploration with tools like Catia, Fusion, or NX.

Scalable AI

Gain scalable performance with multi-GPU LLM support from the Intel Arc Pro B60 GPU, optimized to run larger, more advanced AI for higher quality and faster outputs.

Smooth Viewports

Enable smooth viewports and workspaces across commonly used workstation software with support for various industry-standard API languages and software technologies.

Energy Efficient

Satisfy today's energy-conscious demands with the compact Intel Arc Pro B50 GPU, which combines impressive performance with energy efficiency at just 70W power.

Compare Intel® Arc™ Pro B-Series GPUs

Intel Arc Pro B50 GPU

Explore a compact, dual slot traditional workstation GPU for tasks like design and engineering, offering added local AI capabilities up to 170 TOPS, and support for 4x displays all with an efficient peak power draw of up to 70W.

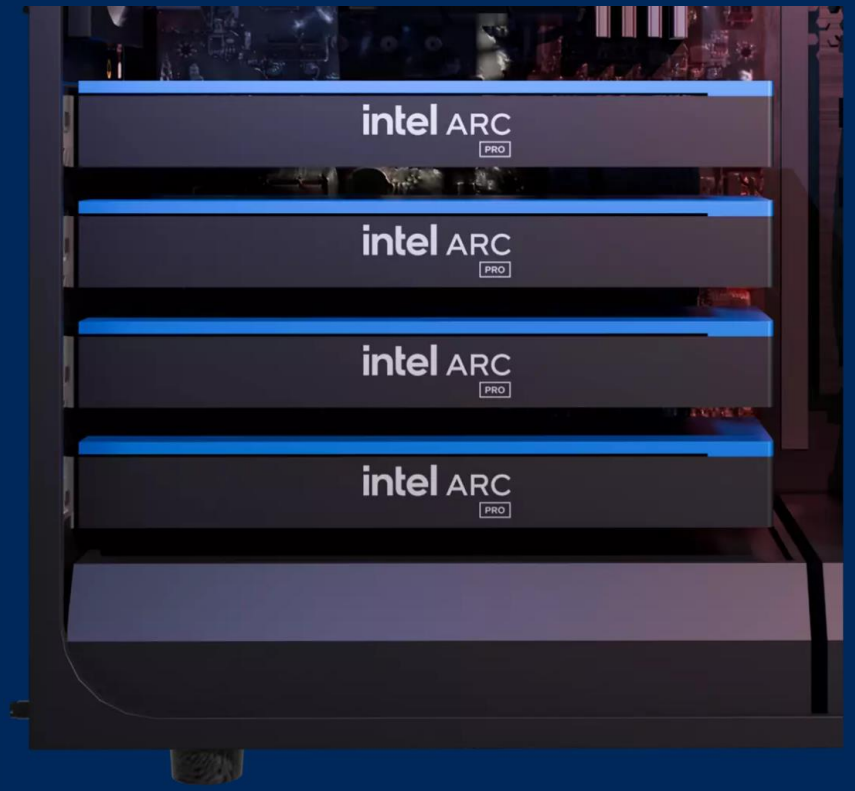
A Compact, 16GB Pro GPU

Intel Arc Pro B60 GPU

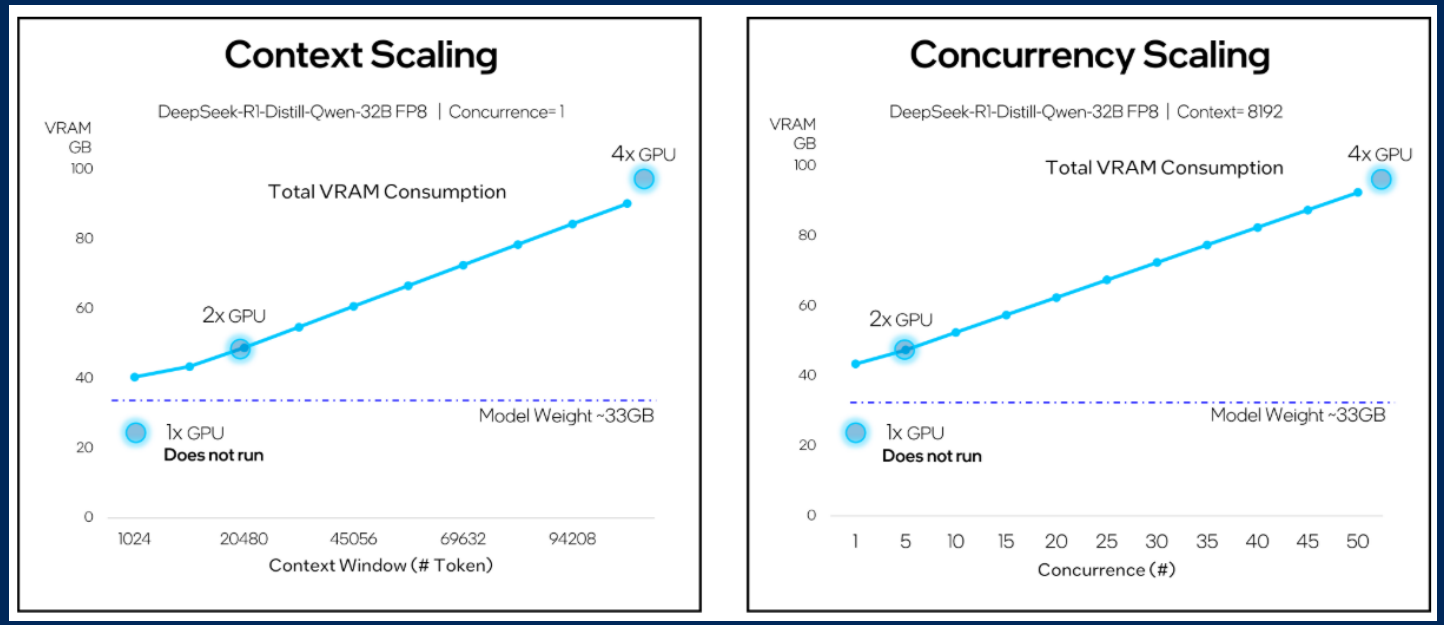
This advanced AI workstation GPU delivers up to 197 TOPS, features 24GB of memory and supports 4x displays. Depending on the board partner it is available in various form factors, with an average power draw ranging from 120W to 200W.

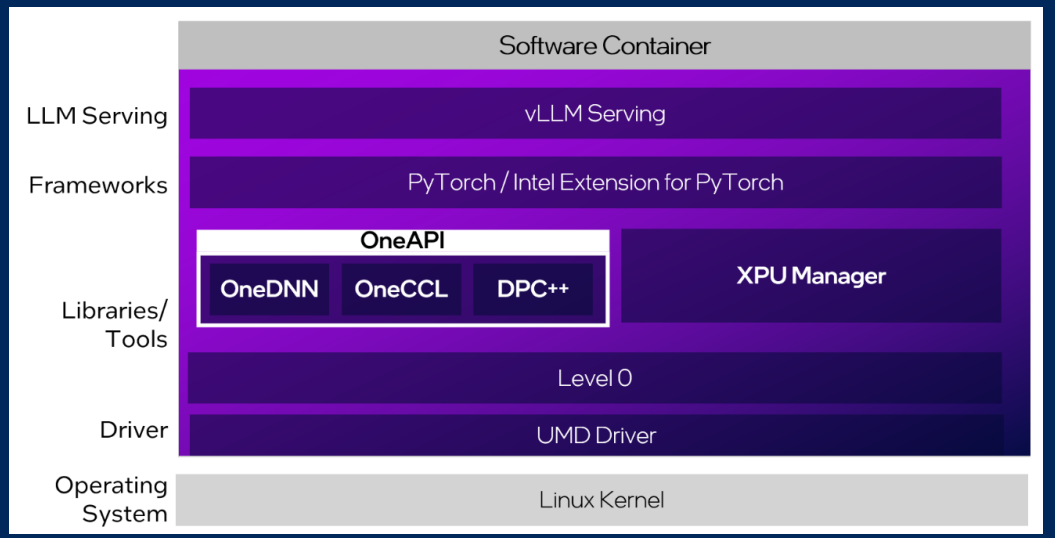
A Scalable, 24GB Pro GPU

Scalable and Accessible Inference Workstations Code-named Project Battlematrix



- Up to :
  - 8 Intel Arc Pro B60 GPUs
  - 192GB combined VRAM
  - 1576 dense INT8 TOPs





- Available Now
  - Improved parallelism
  - vLLM Serving
  - GPU telemetry
  - SRIOV
- Coming soon
  - Passthrough
  - XPU manager

Case Studies

VISTRY

Gen AI Chatbot for Restaurant Edge

“In the quick service restaurant industry, where speed and customer experience are paramount, the Intel® Core™ Ultra processor has been an exceptional asset for our edge computing needs. During our extensive testing, the Intel Core Ultra processor demonstrated exceptional performance metrics for our Employee Assist Chatbot, particularly in Time to First Token (TTFT) that **rival those of cloud-based solutions like ChatGPT 3.5,** thanks to the built-in Intel® Arc™ GPU improvement. This processor's robust capabilities ensure that our chatbot can handle peak-hour traffic with ease, maintaining swift and engaging customer interactions without compromise. Moreover, the privacy and bandwidth efficiency inherent in this edge solution are perfectly aligned with the operational needs of the restaurant industry.”

Atif Kureishy
Founder | CEO

WATCH THE DEMO
(Under Partner Spotlights)

SAMSUNG MEDISON

AI-Augmented Ultrasound Imaging

"The Intel® Core™ Ultra processor has ushered in a new era of innovation in healthcare imaging. Our tests have revealed a remarkable **22% and 25% increase in AI performance throughput** for NerveTrack and Live ViewAssist/HeartAssist real-time ultrasound imaging applications respectively, compared to previous generations Intel® Core™ processor paired with a competitive discrete GPU.

"This breakthrough, attributed in part to the built-in Intel® Arc™ GPU, allows us to offer advanced AI features in next generation mid and entry level ultrasound devices without the need for discrete GPUs, resulting in more accessible and cost-effective cutting-edge imaging technology.”

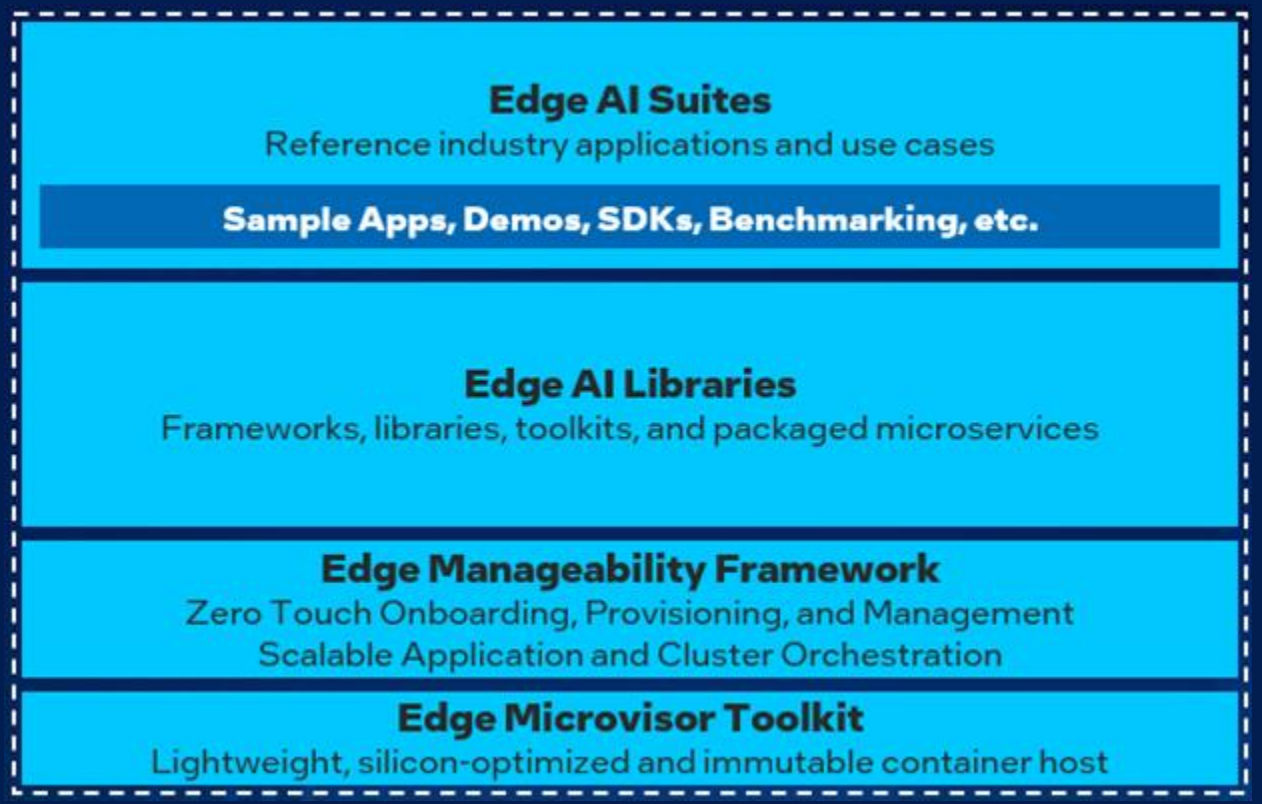
SungShik Baik,
Principal Engineer

WATCH THE DEMO
(Under Partner Spotlights)

intel

CALL TO ACTION:

Get started building, deploying and scaling AI solutions on the Edge with Intel Open Edge Platform



Intel's Open Edge Platform is a secure, optimized open platform for delivering scalable edge solutions.

It offers a modular, composable software stack that brings together the open source ecosystem to help you build optimized edge and AI solutions and securely operationalize them with cloud-like capabilities.

AI Enablement Zones

intel partner alliance

Access a comprehensive resource hub designed to help grow your business and solve your customers' most pressing business challenges. Find exclusive, value-added technical and sales enablement resources to help you build and sell solutions with Intel technology.

Technical Enablement

Sales & Marketing Enablement

Technical Enablement

Sales & Marketing Enablement

Technical Enablement

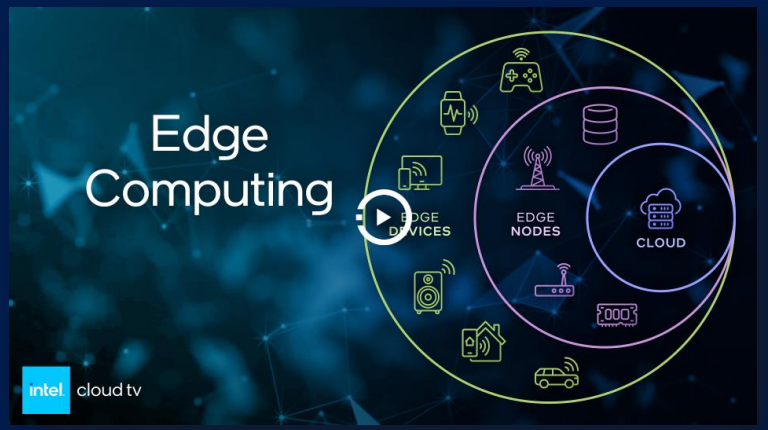
Sales & Marketing Enablement

Sign up to Intel® Partner Alliance for full access or select one of the Enablement Zones if you are already a member

intel

Training Videos

Gain Insights Using Data Inferencing at the Edge



Choosing the Right Path to Edge Computing

intel

Principles of AI Competencies

intel partner university

Principles of AI Everywhere Competency

In this curriculum, you'll delve into Deep Learning, Machine Learning, and Generative AI, and learn to navigate AI challenges using industry models tailored to data parameters. Learn how to assess customer needs effectively by applying the ADDS Methodology to offer tailored solutions from Intel's diverse portfolio, including CPU, GPUs, accelerators, technologies, software, and toolkits, for ease of AI solution deployments.

Enroll ›

Principles of AI Software & Ecosystem Competency

From this curriculum, you will learn how to expedite AI development using open standards and harness data to drive business transformation. Explore a wide range of security solutions within the broad Intel AI ecosystem to ensure data integrity and protection. Delve into the breadth of Intel's AI-based products with a deep focus on Intel's AI software stack, toolkits, and Intel Developer Cloud for ease of AI solution deployments.

Enroll ›

Edge Computing Competency

Step into the future with Intel® Edge AI Systems, where industry transformation is happening now. As AI transitions to the edge, optimizing speed, efficiency, and total cost of ownership are essential. This curriculum equips you to design, sell, and deploy solutions that meet real-world needs using Intel's optimized hardware, software, and partner-ready platforms. Gain insights into Intel's edge AI products and go-to-market strategy, explore real-world use applications and benchmarks, and access tools for faster deployment and smarter design. Get ahead in a rapidly evolving market—join us and harness the power of Edge AI with Intel.

Enroll ›

intel

Additional Training & Resources

Training Courses

- Bringing AI Everywhere… at the Edge
- Addressing Network & Edge Markets with Intel® Xeon® 6 Processors
- Top 3 Reasons to Elevate Edge AI & Graphics with Intel® Core™ Ultra Processors
- Choosing the Right Path to Edge Computing: Video
- AI on the Edge with Computer Vision
- AI from the Data Center to the Edge
- Edge AI Landscape

Additional Resources

- Dell and Intel Server Playbook for AI at the Edge
- How AI-Ready Solutions Can Help Your Business Hit the Ground Running in Today's AI Race

intel

Notices and Disclaimers

Performance varies by use, configuration and other factors. Learn more on the Performance Index site.
Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.
Your costs and results may vary.
Intel technologies may require enabled hardware, software or service activation.


intel

Configuration Details

Performance measured on Intel® Arc™ A370M GPU as proxy for Intel® Arc™ A370E GPU

System Configuration:
Intel® Arc™ A370M GPU. Advantech VEGA_P110-42A1. 8 Xe-cores, 4GB GDDR6. Intel® Core™ i7-13800HE. DDR5 5600Mhz 64GB
NVIDIA A500 GPU. ADLINK EGX-MXM-A500. 2048 CUDA Cores + 64 Tensor Cores + 16 RT Cores. 4GB GDDR6. Intel® Core™ i7-13800HE. DDR5 5600Mhz 64GB

[1] Workload & version: Unigene Superposition 2 V1.1. Windows 10. Graphics Driver: Intel 31.0.101.5186, NVIDIA 31.0.15.3799. Run Method: Warm. Iterations and result choice: 3 iterations, average.
[2] Workload & version: AI Inference Benchmark. OS: Ubunto 22.04. Kernel version 6.5.0-15-generic. GPU Driver version: Intel I915; NVDA 535.146.02. Intel Inference Framework: OpenVINO™ 2023.2.0. NVDA Inference Framework: CUDA 8.6.1.6. Intel Compute Framework: Intel OpenCL Runtime. NVDA Compute Framework: CUDA. AI Model: Resnet50, Int8, BS32. Intel AI Model Framework: TensorFlow/Caffe/Onnx. NVDA AI Model Framework: Caffe/ONNX. Run Method: Warm. Iterations and result choice: 3 iterations, average.
[3] Worklaod & version: AI Inference Benchmark. OS: Ubunto 22.04. Kernel version 6.5.0-15-generic. GPU Driver version: Intel I915; NVDA 535.146.02. Media Framework: FFMPEG 4.4.1.3. Run Method: Warm. Iterations and result choice: 3 iterations, average.
[4] Source: https://www.iotmart.com/en-en/s/product/detail/01t2y000000Gy2BAAS?language=en_US. Advertised as $345 on 5/31/24
[5] Source: https://www.mouser.com/ProductDetail/Advantech/SKY-MXM-A500-4SHA?qs=ST9lo4GX8V3tHKR4pF9fCw%3D%3D Advertised as $550 on 5/31/24