

— DISTRIBUTED BY RUTRONIK —

AI PRODUCTS








Our Portfolio to make use of AI

V1.0

Content

- Introduction 3 - 7
- AI Processors 8 - 17
- AI Platforms - Boards & Systems 18 - 39
- GPU Cards, AI Accelerators & AI Modules 40 - 47
- AI Vision 48 - 49
- Memory & Storage for AI 50 - 54
- AI Partnerships & Software 55 - 64
- Other AI Products & Applications 65 - 67

Our Product Portfolio

-  Semiconductors
-  Passive Components
-  Electromechanical Components
-  Displays & Monitors
-  Boards & Systems
-  Storage Technologies
-  Wireless Technologies

Vertical Markets

- 
- 
- 
- 
- 
- 

Committed to excellence

Consult – Know-how. Built-in.

The Technical Competence from RUTRONIK
Worldwide and individual consulting on the spot by competent sales staff, application engineers & product specialists.

Components – Variety. Built-in.

The Product Portfolio from RUTRONIK
Wide product range of semiconductors, passive and electromechanical components, displays & monitors, boards & systems, storage and wireless technologies for optimum coverage of your needs.

Logistics – Reliability. Built-in.

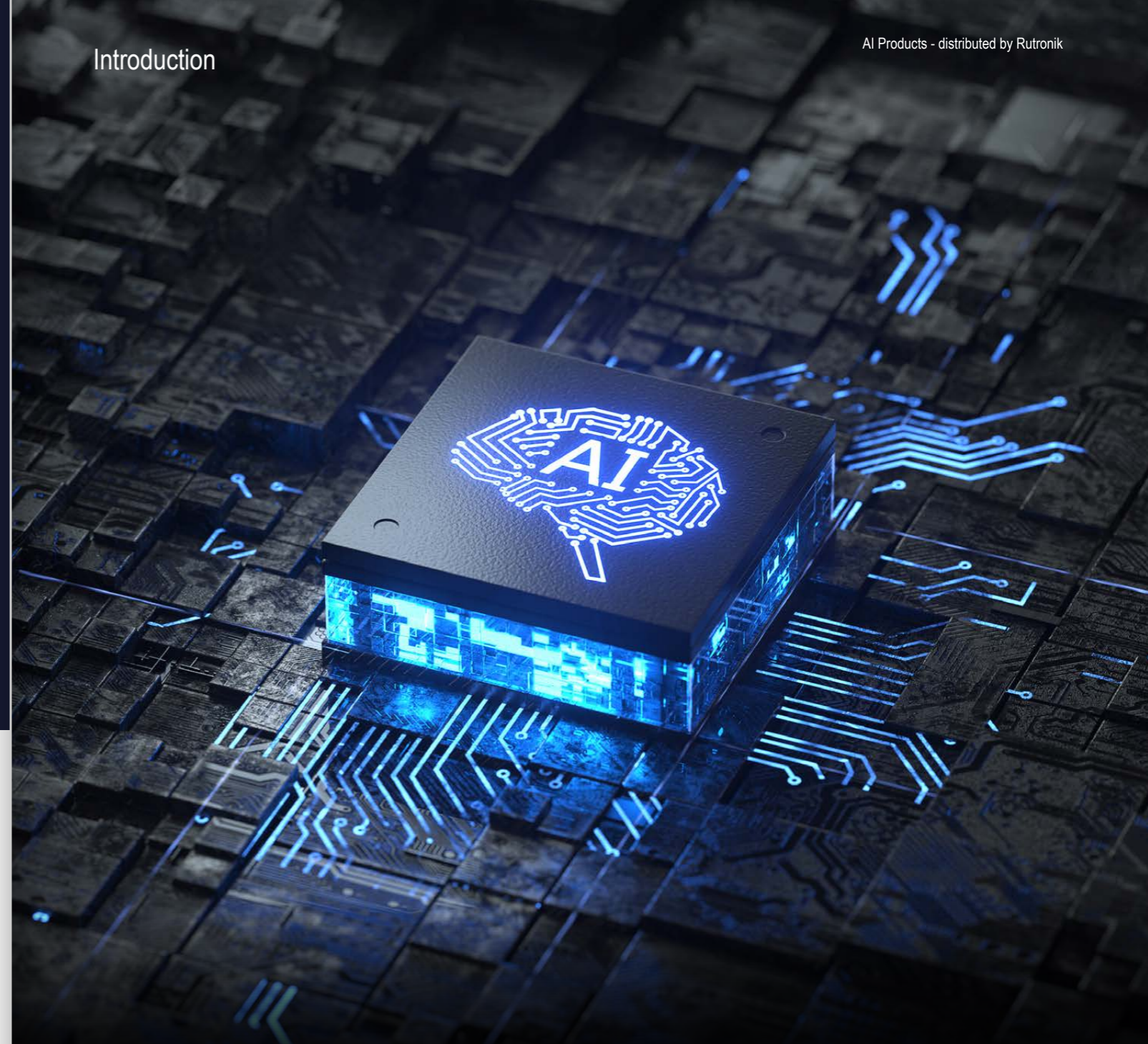
The Delivery Service from RUTRONIK
Innovative and flexible solutions: from supply chain management to individual logistics systems.

Quality – Security. Built-in.

Quality without Compromise from RUTRONIK
The integrated management system (IMS) encompasses quality control, information security, environmental protection, occupational health and safety.

Follow us

-  www.facebook.com/rutronik
-  <https://twitter.com/rutronik>
-  www.youtube.com/user/rutronik24
-  <https://rutronik-tec.com>
-  www.linkedin.com/company/rutronik



AI Hard- & Software

AI hardware refers to the specialized computational devices and components, such as **GPUs**, **TPUs**, and **NPUs**, that facilitate and accelerate the processing demands of artificial intelligence tasks, playing a pivotal role alongside algorithms and software in the AI ecosystem.

In the bustling world of artificial intelligence, it's easy to get lost in the maze of algorithms, data sets, and software libraries.

But let's not forget the unsung hero of this AI revolution: **>> the hardware.**

Think of it as the stage on which the AI performance unfolds. Without the right stage, even the best actor can't put on a show.

We've seen computers crunch numbers at astonishing speeds,

but AI, especially deep learning, demands a different kind of computational muscle. It's not just about speed; it's about parallelism, memory access, and specialized operations. The right hardware can make the difference between an AI model training for weeks or just a few hours.

As we dive into the intricate world of AI hardware, remember: it's not just about the code. The silicon, circuits, and chips play an equally starring role in this AI saga.

AI Hardware Components

AI Processors



CPU (Central Processing Unit)

- Heart and soul of computing, versatile chip for a wide range of tasks
 - Few but more powerful cores, suited for sequential tasks
- > The demands of AI have led to the rise of more specialized processors, each tailored to the unique needs of AI workloads.

GPU (Graphic Processing Unit)

- Initially designed for demanding computer graphics rendering tasks
 - Thousands of smaller cores, suited for parallel tasks (like in deep learning)
 - Architectural advantage ideal for training large neural networks
- > Excellent at performing many operations simultaneously

NPU (Neural Processing Unit)

- Designed exclusively for neural network computations
- Tailored to accelerate specific AI tasks, offering efficiency gains in areas like image recognition and natural language processing

TPU (Tensor Processing Unit)

- Google's custom-built solution to the computational demands of AI
- Designed specifically for neural network machine learning
- Optimized for tensor calculations, the foundational math behind AI operations

How to Choose between a CPU, GPU, NPU and TPU?

Choosing the best neural network architecture and framework is a critical first step. It impacts the required hardware for training models and running inference.

Most enterprises, do not need to train models. While only certain companies train models, millions of users run inference. Hardware requirements for inference are not necessarily the same as those for training. CPUs can suffice the inference requirements.

> Before deciding which hardware to choose:

- Start with the customer and figure out what they need
- Explore which AI algorithms can address the need and how to acquire them
- Assess the hardware requirements for training and inference in detail

What are TOPS in AI processor specifications?

TOPS (*Tera or Trillion Operations Per Second*) is a metric used to measure the computing capacity of an NPU or other AI-specialized processors. This measure indicates the number of operations the NPU can perform in one second, expressed in trillions.

Although TOPS is not a perfect metric and many variables influence a system's performance in executing AI tasks, producers rely on this parameter to advertise their products, primarily to simplify performance metrics and help customers understand what they are getting.



Boards & Systems

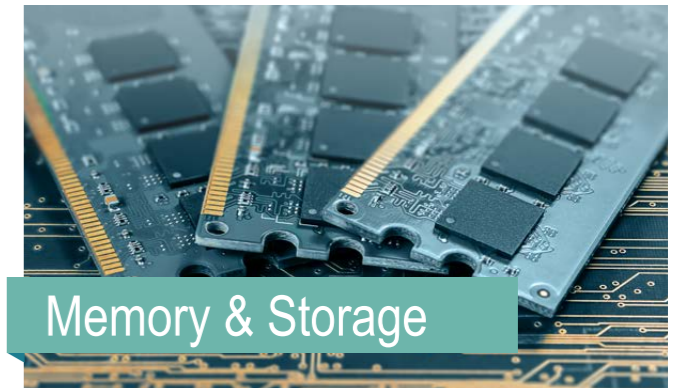
AI Platforms enabled by AI Processors

AI-enabled embedded boards and systems based on various different AI processor platforms provide the necessary infrastructure for developing, training, and deploying AI models in embedded systems. It's a key component in the AI ecosystem, enabling businesses and researchers to leverage AI capabilities in a structured and efficient manner. These AI platforms are particularly useful in real-time systems, such as autonomous vehicles or smart healthcare wearables, where every millisecond counts and can be life-changing.

Edge AI

Edge AI means that artificial intelligence (AI) algorithms are processed locally on a hardware device close to a sensor or to a signal. The algorithms are using data that are created on the device. Deep learning is a commonly used type of machine learning algorithm.

Edge AI is the opposite of cloud computing AI where you do all the machine learning processing, aka inference, in the cloud. Edge AI is a subset of the larger edge computing market. A device using Edge AI does not need to be connected in order to work properly, it can process data and make decisions independently without a connection. In order to use Edge AI, you need a device comprising a microprocessor and sensors.



Memory & Storage

AI models, especially the larger ones, require robust memory and storage solutions that can keep up with the processor's demands.

DRAM and Cache

Random-Access Memory (RAM) and cache are crucial for AI computations. They provide the processor with quick access to data, reducing wait times and ensuring smoother operations. The faster the memory, the quicker an AI model can train and infer.

LPDDR5X (Low Power Double Data Rate 5X) plays a significant role in modern AI applications. It is known for its low power consumption, making it ideal for mobile devices and edge computing where energy efficiency is critical. LPDDR5X offers data rates of up to 8,533 Mbps, which significantly improves the processing speed and efficiency of AI algorithms.

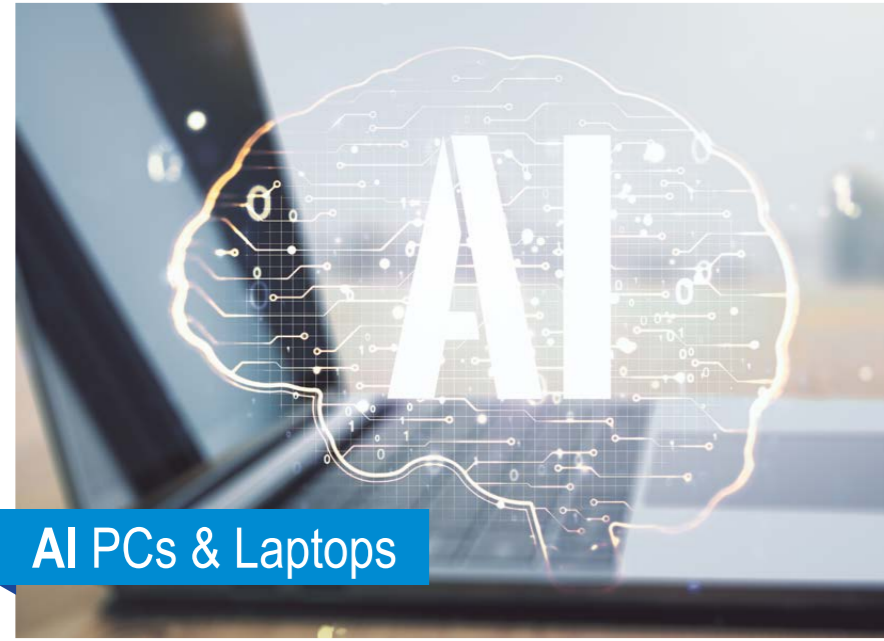
SSDs

AI applications generate and utilize vast amounts of data. Solid-State Drives (SSDs) have become the preferred storage solution in AI hardware setups. With their faster read-write speeds compared to traditional Hard Disk Drives (HDDs), SSDs ensure that data-intensive AI workloads run efficiently.



Networking Capabilities

In the world of AI hardware, it's not just about computation and storage. How components communicate is equally vital. Interconnects ensure that data flows seamlessly between processors, memory, and storage. In AI, where massive datasets and complex models are the norms, high-speed data transfer becomes critical. Efficient interconnects reduce bottlenecks, ensuring that AI systems run smoothly and without interruption.



AI PCs & Laptops

AI PCs

An AI PC, or Artificial Intelligence Personal Computer, is a type of computer specifically designed to handle artificial intelligence and machine learning tasks. Unlike regular PCs, which rely on CPUs and GPUs to handle computing tasks, AI PCs have a dedicated Neural Processing Unit (NPU) optimized for AI and machine learning tasks.

When Microsoft talks about "AI-PCs", it refers to systems with Windows 11, an NPU, and a Copilot button. The AI PC is expected to have a keyboard with a Copilot button, and logically, Windows 11 with activated Copilot should be pre-installed. The button starts the Copilot, under which Microsoft wants to bundle more and more AI functions in the future.

Intel defines all systems that have an AI unit as "AI-PC". For a system to receive the "AI PC" seal from Microsoft, it must have a CPU, a GPU, and an NPU. The type of GPU (iGPU or dGPU) doesn't matter. AI PCs are designed to run powerful AI-accelerated software with great performance and quality. Without these components, AI workloads may be very slow or may not run at all.

In the future, requirements for AI-PCs are expected to increase. For example, the second generation of AI-PCs must provide at least 40 TOPS (Trillion Operations per Second) of computing power. This means that the performance in both camps must increase significantly.

exone go Business 1490 / 1690 X14 | AI Laptops (with integrated NPU)

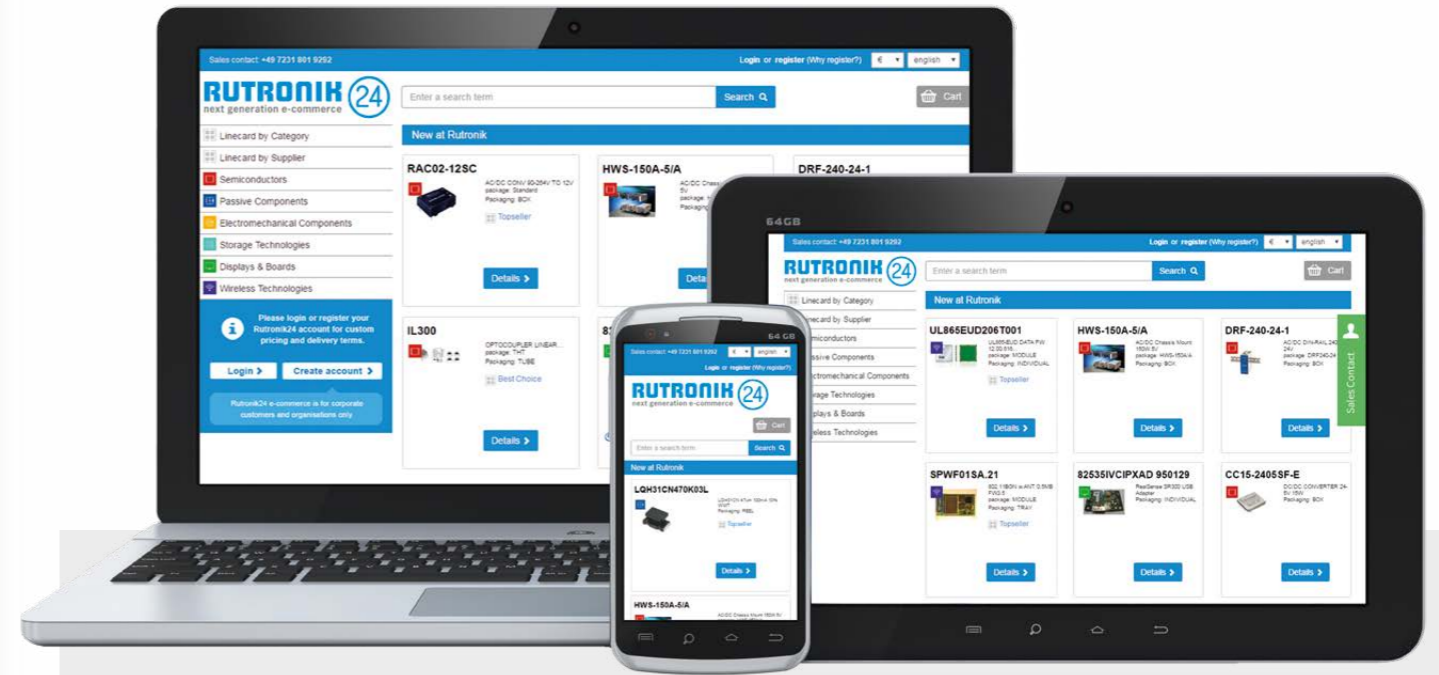


Ultra 7-155U W11Pro

Discover the exone go Business AI Laptops - a perfect partner for productive work on the go and in the office. This stylish and ultra-lightweight device combines high performance with impressive battery life, so you can stay productive all day long. The robust yet lightweight body offers a perfect balance between mobility and durability, while advanced security features keep your data safe. State-of-the-art connectivity options ensure you can stay connected anytime, anywhere.

Key Features

- Windows® 11 Pro 64-bit
- 8 GB DDR5 PC4800
- 500 GB / 1 TB SSD M.2 PCIe® NVMe®
- Intel® Core™ Ultra 7-155U
- 14,0" (35,6 cm) / 16,0" (40,64 cm)
- 16:10 format (400 cd/m²)
- 16 hours battery life
- Optional LTE upgrade possible
- Windows Hello camera
- A, C and D cover made from magnesium alloy



e-commerce made easy

FASTER. EASIER.
JUST MORE PERSONAL.

rutronik24.com



AI PROCESSORS

Intel accelerates AI Everywhere

From Data Center, Cloud and Network to the Edge and PC

- Intel® Core™ Ultra processors are ushering in the AI PC era
- Lunar Lake client processor architecture unveiled to continue to grow the AI PC category → Intel® Core™ Ultra processors (Series 2)
- Intel® Xeon® 6 processors with Efficient-cores (E-cores) → performance and power efficiency for high-density, scale-out workloads in the data center
- Intel® Gaudi® 2 and Intel® Gaudi® 3 AI accelerator kits → delivering high performance with up to one-third lower cost compared to competitive platforms
- Combine Xeon processors with Gaudi AI accelerators in a system → powerful solution for making AI faster, cheaper and more accessible
- Intel® Arc™ GPU for the Edge → Supercharge Edge AI, Graphics, and Media Processing with Intel Arc Graphics discrete GPU
- Intel® Movidius™ Vision Processing Units (VPUs)

Intel AI Processors

Get the tools and technologies you need to deliver AI innovation wherever it's required – and the freedom to choose the right technology for your workloads. Your data infrastructure is likely already running on and optimized for Intel. With our portfolio of AI processors and integrated accelerators, AI developers in all industries can enable high-performance and efficient AI solutions at the edge or in the data center.



Intel® Core™ Ultra Processors



Bring AI to the PC and unleash new opportunities, applications, and experiences for users with processors built to run the next wave of advanced AI workloads.

Intel® Xeon® Processors



From edge to cloud, boost performance for machine and deep learning training and inferring without using specialized hardware.

Intel® Gaudi® AI Accelerators



Deploy GenAI at scale via high-performance, high-efficiency deep learning processors that take the place of GPUs for large-scale AI training and inference workloads in the data center.

Intel® Max Series Product Family



Accelerate science and discovery with breakthrough CPU and GPU performance and fewer bottlenecks for memory-bound workloads.

Intel® Data Center GPU Flex Series

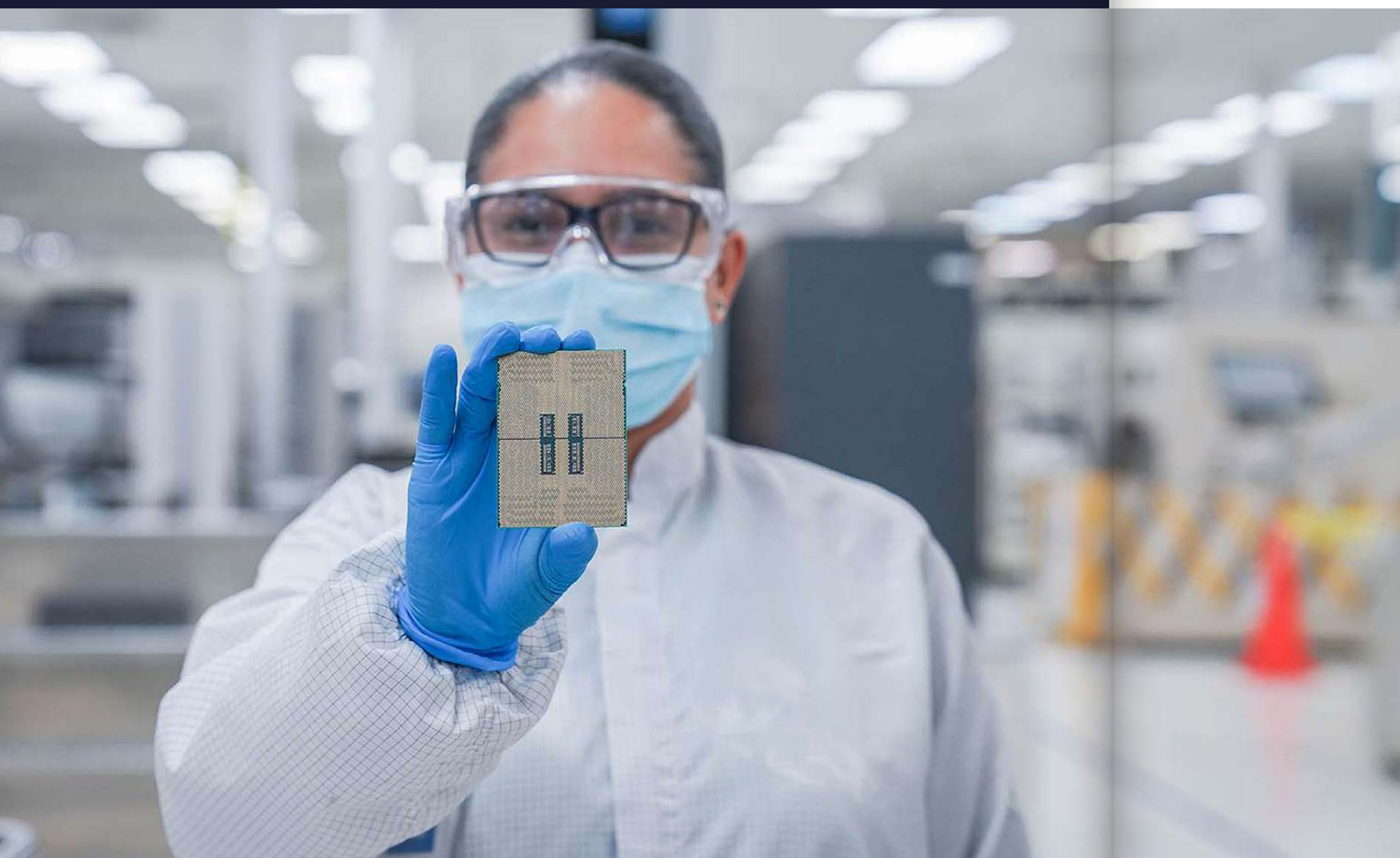


Deploy cost-effective, high-density AI visual inference at scale with the industry's most open GPU solution for the intelligent visual cloud.

Intel® Arc™ GPU



Deliver fast AI inference at the edge to create standout media experiences, apply machine vision and video analytics, and enhance customer engagement with interactive kiosks.



We add value to the future

Intel Core Ultra Processors



The industry's premier AI PC processor ([Product Brief](#))

- Up to 16 MPU Cores + Arc GPU Core + NPU Core in one housing, up to 5.1 GHz
- New AI experiences, immersive graphics, optimal battery life
- More than 500 AI Models Run Optimized on Intel Core Ultra Processors - deployed across CPU, GPU and NPU
- Available across popular industry sources, including OpenVI-NO Model Zoo, Hugging Face, ONNX Model Zoo and PyTorch
- Models draw from categories of local AI inferencing, including large language, diffusion, super resolution, object detection, image classification/segmentation, computer vision and others

Major improvements in efficiency and performance



Intel Core Ultra Processors (Series 1)

Product Name	# of Cores	Max Turbo Freq.	Cache	Processor Graphics	Vertical
Intel® Core™ Ultra 7 – 155HL	16	4.8 GHz	24 MB	Intel® Arc™ graphics	Embedded
Intel® Core™ Ultra 7 – 155UL	12	4.8 GHz	12 MB	Intel® graphics	Embedded
Intel® Core™ Ultra 7 – 165HL	16	5.0 GHz	24 MB	Intel® Arc™ graphics	Embedded
Intel® Core™ Ultra 7 – 165UL	12	4.9 GHz	12 MB	Intel® graphics	Embedded
Intel® Core™ Ultra 3 – 105UL	8	4.2 GHz	10 MB	Intel® graphics	Embedded
Intel® Core™ Ultra 5 – 125HL	14	4.5 GHz	18 MB	Intel® Arc™ graphics	Embedded
Intel® Core™ Ultra 5 – 125UL	12	4.3 GHz	12 MB	Intel® graphics	Embedded
Intel® Core™ Ultra 5 – 135HL	14	4.6 GHz	18 MB	Intel® Arc™ graphics	Embedded
Intel® Core™ Ultra 5 – 135UL	12	4.4 GHz	12 MB	Intel® graphics	Embedded
Intel® Core™ Ultra 7 – 155H	16	4.8 GHz	24 MB Intel® Smart Cache	Intel® Arc™ graphics	Mobile
Intel® Core™ Ultra 5 – 125H	14	4.5 GHz	18 MB Intel® Smart Cache	Intel® Arc™ graphics	Mobile
Intel® Core™ Ultra 9 – 185H	16	5.1 GHz	24 MB Intel® Smart Cache	Intel® Arc™ graphics	Mobile
Intel® Core™ Ultra 5 – 135H	14	4.6 GHz	18 MB Intel® Smart Cache	Intel® Arc™ graphics	Mobile
Intel® Core™ Ultra 7 – 155U	16	4.8 GHz	24 MB Intel® Smart Cache	Intel® graphics	Mobile
Intel® Core™ Ultra 5 – 135U	12	4.4 GHz	12 MB Intel® Smart Cache	Intel® graphics	Mobile
Intel® Core™ Ultra 7 – 165U	12	4.9 GHz	12 MB Intel® Smart Cache	Intel® graphics	Mobile
Intel® Core™ Ultra 5 – 125U	12	4.3 GHz	12 MB Intel® Smart Cache	Intel® graphics	Mobile
Intel® Core™ Ultra 5 – 134U	12	4.4 GHz	12 MB Intel® Smart Cache	Intel® graphics	Mobile
Intel® Core™ Ultra 5 – 164U	12	4.8 GHz	12 MB Intel® Smart Cache	Intel® graphics	Mobile



Further AI related specs

- All CPUs come with Intel® Deep Learning Boost
- All CPUs support KI software frameworks: OpenVINO™, WindowsML, ONNX RT
- All Embedded CPUs (HL/UL suffix) have AI Datatype support for: Int8, FP16, BF16, FP32

[Core Ultra](#)

[Core Ultra for edge](#)

Intel Core Ultra Processors (Series 2)



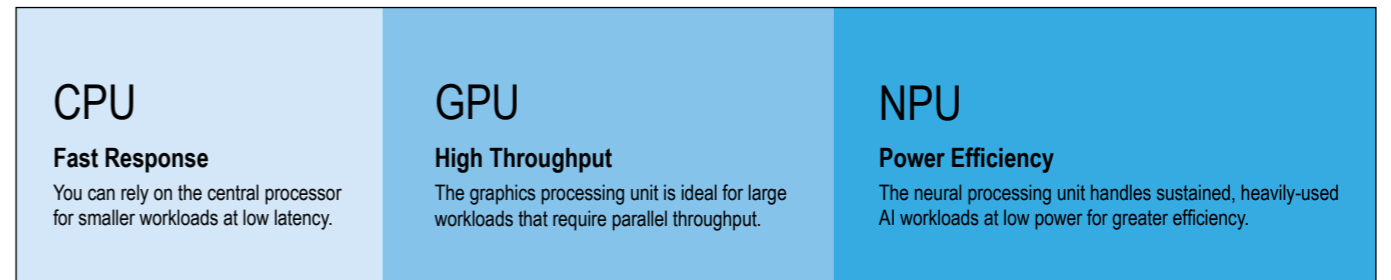
Intel's upcoming Core Ultra mobile processors (code-named Lunar Lake)

- Starting Q3 2024, Intel's upcoming client processors will power more than 80 new laptop designs across more than 20 OEMs
- Delivering AI performance at a global scale for Copilot+ PCs and Copilot+ experiences, like Recall, via an update when available.
- AI Performance > 40 NPU TOPS, > 60 GPU TOPS and 100 platform TOPS

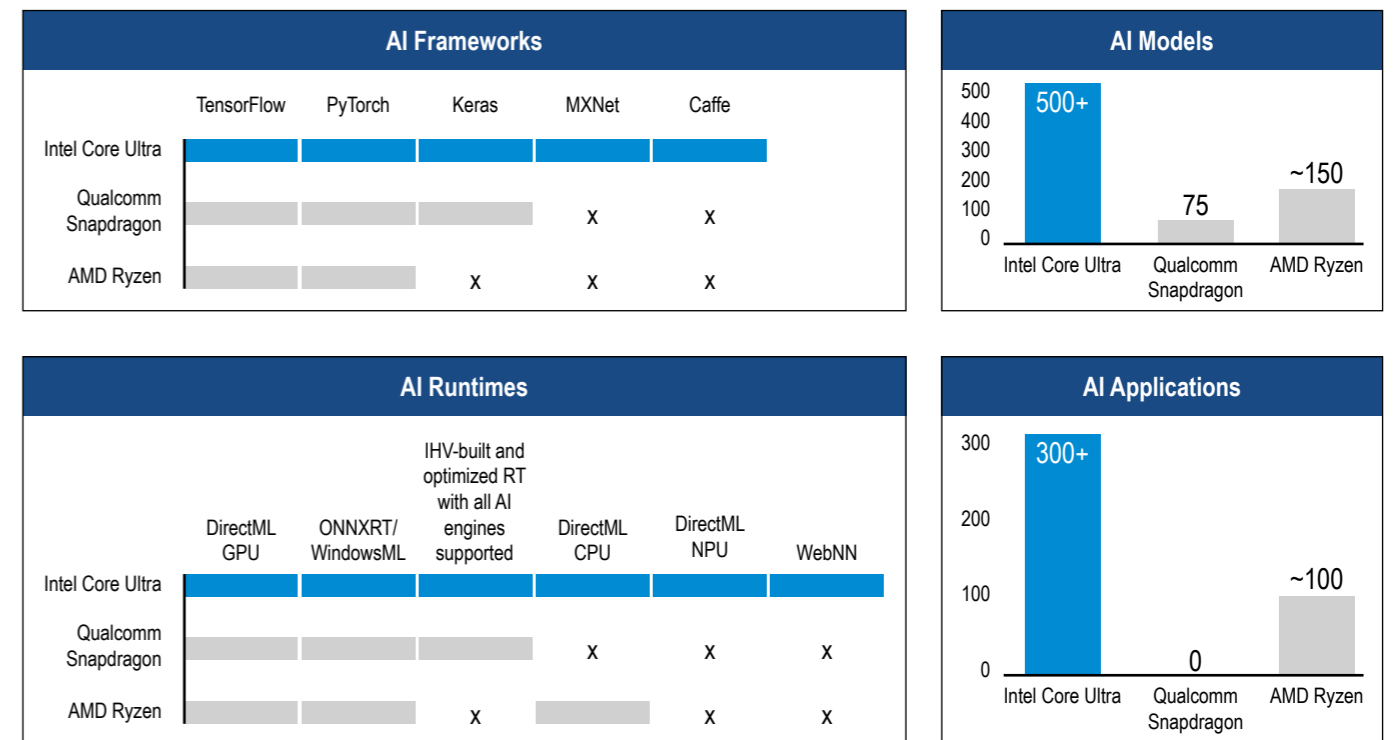


What is an AI PC?

AI PCs use artificial intelligence technologies to elevate productivity, creativity, gaming, entertainment, security, and more. They have a CPU, GPU, and NPU to handle AI tasks locally and more efficiently.



Intel Core Ultra - Full Stack Leadership in AI PC



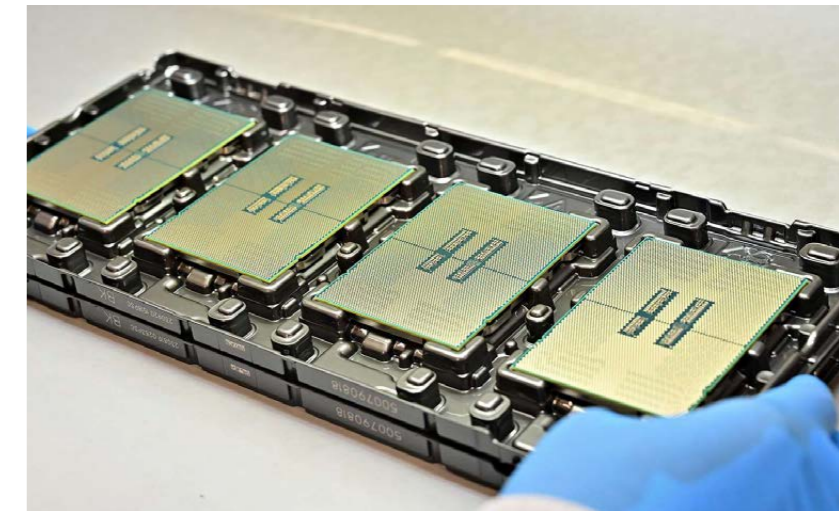


Intel Xeon 6 Processors

Meet data center performance and density goals with a broad portfolio. Intel® Xeon® 6 technology gives you a single platform with processor options focusing on high performance for AI and compute intense workloads, exceptional efficiency, or cloud scalability. A new class of Efficient-cores (E-cores) delivers high core density and exceptional performance per watt, offering distinct advantages for cloud-scale workloads that demand high throughput. Performance-cores (P-cores) excel at a wide range of workloads and give you better performance for AI and compute intense workloads than any other CPU.

Highlight Technologies

- Intel AMX provides up to 16x more multiply accumulate (MAC) operations than Intel AVX-512 for BF16 and FP16-based models to enhance AI performance (P-core-only feature).
- Intel AVX-512 encompasses unique instructions and two 512-bit fused-multiply add (FMA) units per core, boosting the speed of vector mathematics common to AI, HPC, and database workloads (P-core-only feature).
- Intel AVX2 with VNNI instructions and fast up/down convert for BF16 and FP16 enables better AI compatibility for Intel Xeon 6 processors with E-cores.



[Product Brief](#) | [Fact Sheet](#)

Intel Xeon Scalable Processors and Intel AI Engines



5th Gen Intel® Xeon® Scalable processors

- Integrated AI accelerators to run AI workloads at scale without specialized hardware
- Processor options from edge to cloud
- Enhanced performance for training and inference

Intel® AI Engines

65%
of data center AI inferencing runs on Intel® Xeon® processors

Up to **9.9x** higher real-time Natural Language Processing inference (BERT-large) performance and 7.7x higher performance/watt on 5th Gen Intel® Xeon® processors with AMX BF16 vs. 3rd Gen Intel® Xeon® processors

Up to **14x** higher real-time object detection inference performance (SSD-ResNet34) on 5th Gen Intel® Xeon® processors with AMX BF16 vs. 3rd Gen Intel® Xeon® processors

Up to **8.7x** higher batch Recommendation System inference performance (DLRM) and 6.2x higher performance/watt on 5th Gen Intel® Xeon® vs. 3rd Gen Intel® Xeon® processors

[Product Brief](#)

Deep Learning Acceleration
Intel® Advanced Matrix Extensions (Intel® AMX)

Maximize performance for deep learning training and inference workloads that rely on matrix math, including natural language processing, recommendation systems, and image recognition.

[Intel AMX Technology Brief](#)

Classical Machine Learning Acceleration
Intel® Adv. Vector Extensions 512 (Intel® AVX-512)

Accelerate AI, analytics, scientific simulations, financial simulations, and other compute-intensive tasks that involve vector-based computations.

Intel® Xeon® 6 processors

Intel Xeon 6900-series processors

Maximum performance

Introducing a new class of Intel server platform design, ideal for cloud computing, AI, HPC, software-as-a-service (SaaS), and infrastructure-as-a-service (IaaS) workloads.

Intel Xeon 6500-series/6700-series processors

Enhanced performance

A significant upgrade to established Intel server platform designs. Mainstream servers from edge to cloud for enterprise IT, digital service providers, and telco. Ideal for AI, HPC, networking and media, data services, infrastructure and storage, web, applications, and microservices.

Intel® Xeon® 6 CPUs with P-cores

- Up to 128 cores (256 threads) per CPU
- Up to 500 W per CPU
- One- or two-socket servers
- 12 channel memory
- Up to 6,400 MT/s DDR5
- 8,800 MT/s MCR DIMMs
- Up to 96 PCIe 5.0 lanes
- Six Intel UPI 2.0 links

Coming soon

- Up to 86 cores (172 threads) per CPU
- Up to 350 W per CPU
- One-, two-, four-, or eight-socket servers
- Eight channel memory
- 6,400 MT/s DDR5
- 8,000 MT/s MCR DIMMs
- Up to 88 PCIe 5.0 lanes with up to 136 lanes for 1S designs
- Four Intel UPI 2.0 links

Coming soon

Intel® Xeon® 6 CPUs with E-cores

- Up to 288 cores (288 threads) per CPU
- Up to 500 W per CPU
- One- or two-socket servers
- 12 channel memory
- Up to 6,400 MT/s DDR5
- Up to 96 PCIe 5.0 lanes
- Six Intel UPI 2.0 links

Coming soon

- Up to 144 cores (144 threads) per CPU
- Up to 330 W per CPU
- One- or two-socket server
- Eight channel memory
- 6,400 MT/s DDR5
- Up to 88 PCIe 5.0 lanes
- Four Intel UPI 2.0 links

Now available



Intel Max Series Product Family

Intel® Xeon® CPU Max Series

Maximize bandwidth with the Intel® Xeon® CPU Max Series, the first and only x86-based processor with high-bandwidth memory (HBM). Architected to supercharge the Intel® Xeon® architecture with HBM, Max Series CPUs deliver up to 4.8X better performance compared to competition on real-world HPC and AI workloads¹ and feature:

- Up to 56 performance cores constructed of four tiles and connected using Intel's embedded multi-die interconnect bridge (EMIB) technology.
- 64GB of ultra-high-bandwidth in-package memory plus support for PCI Express 5.0 and CXL1.1 I/O.
- Up to 1.14GB of HBM capacity per core.



[Product Brief](#)

Intel® Data Center GPU Max Series

Maximize impact with the Intel® Data Center GPU Max Series, Intel's highest performing, highest density, general-purpose discrete GPU, which packs over 100 billion transistors into a package and contains up to 128 Xe Cores – Intel's foundational GPU compute building block. Intel Max Series GPUs feature:

- 408 MB of L2 cache based on discrete SRAM technology and 64 MB of L1 cache.
- Up to 128 Ray Tracing units on each Intel Max Series GPU for accelerating scientific visualization and animation.
- AI-boosting Intel® Xe Matrix Extensions (Intel® XMV) with deep systolic arrays enabling vector and matrix capabilities in a single device.



[Product Brief](#)

Intel Data Center GPU

Intel® Data Center GPU Max Series

Maximize impact with the Intel® Data Center GPU Max Series, Intel's highest performing, highest density, general-purpose discrete GPU, which packs over 100 billion transistors into one package and contains up to 128 Xe Cores – Intel's foundational GPU compute building block.

- Up to 408 MB of L2 cache based on discrete SRAM technology, 64 MB of L1 cache and up to 128 GB of high-bandwidth memory.
- Up to 128 ray tracing units built into each Max Series GPU for accelerating scientific visualization and animation.
- AI-boosting Intel® Xe Matrix Extensions (XMV) with deep systolic arrays enabling vector and matrix capabilities in a single device.
- oneAPI standards-based, multiarchitecture programming and tools, which boost performance and productivity and overcome proprietary programming model lock-in.

[Product Brief](#)

Intel® Data Center GPU Flex Series

Unmatched flexibility meets powerful media processing and virtualization capabilities with the Intel® Data Center GPU Flex Series – the industry's most open GPU solution. Intel® Data Center GPU Flex Series is delivering outstanding compute density and energy efficiency for visual cloud workloads. With acceleration for visual processing and AI built into the silicon, the GPU is based on Intel® Xe HPG (high-performance graphics) microarchitecture.

- Support for an open, flexible, standards-based software stack together with oneAPI unified programming.
- Industry-first hardware-based open source AV1 encoder in a GPU

[Product Brief](#)

STATS **5x** Media transcode throughput at half the power of the competition
Intel Flex Series 140 GPU compared to NVIDIA A10

2x Decode throughput at half the power of the competition
Intel Flex Series 140 GPU compared to NVIDIA A10

Up to **68** 720p30 on select game streams

Up to **46** 720p30 on select game streams



Intel Gaudi AI Accelerators

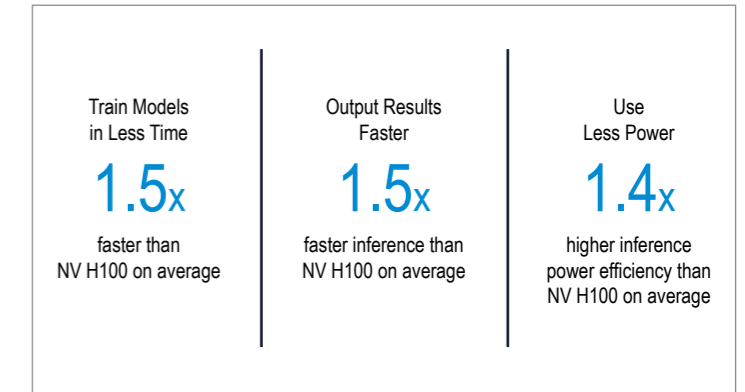


Designed to bring a new level of productivity advantages and choice to data center generative AI

- x86 infrastructure operates at scale in nearly all data center environments, serving as foundation for integrating the power of AI
- Ensuring cost-effective interoperability and the benefits of an open ecosystem of developers and customers
- Purposely designed for AI workloads, Intel Xeon processors operate in a system with Intel Gaudi AI accelerators
- Seamless integration into existing x86 infrastructure
- Only MLPerf-benchmarked alternative to NVIDIA H100 for training and inference of large language models (LLM)
- GenAI with a price-performance advantage that provides choice and fast deployment time at lower total cost of operating

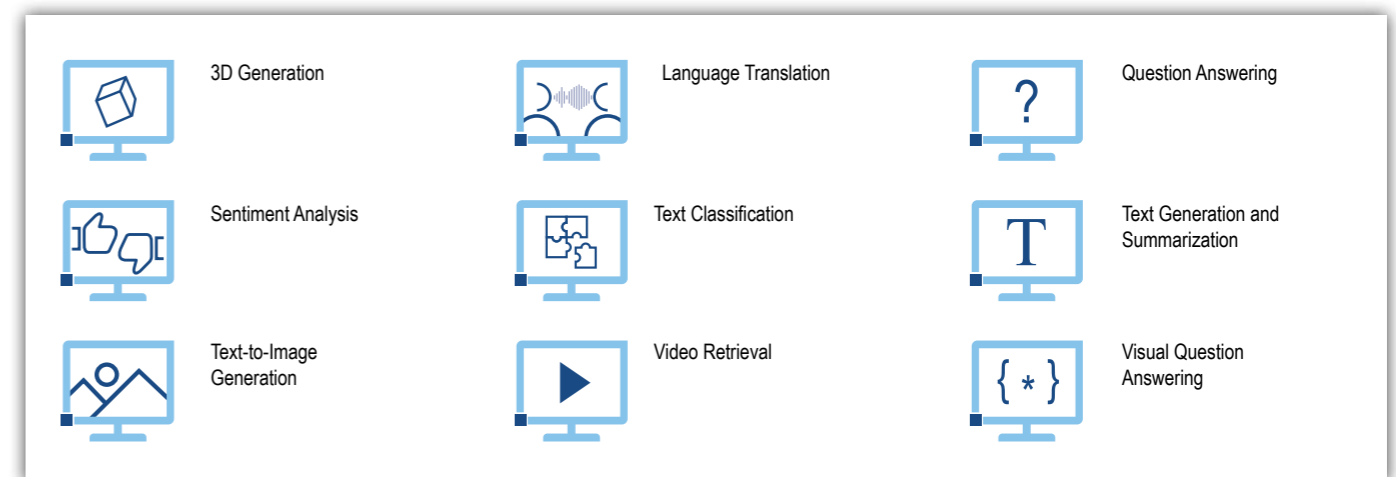
Intel Gaudi 3 AI Accelerator

- Built to handle demanding training and inference
- High-performance option for every type of enterprise AI workload
- Time and Power Saving compared to NVIDIA GPUs:



Intel® Gaudi® 3 AI accelerator	Third generation Intel Gaudi Ai accelerator brings another leap in performance and efficiency	White Paper
Intel® Gaudi® 2 AI accelerator	Second generation deep learning training and inference for leadership performance	Datasheet
Intel® Gaudi® AI accelerator	First generation deep learning training and inference for competitive price/performance	Datasheet
Intel® Gaudi® software	Software and developer tools and resources for ease of use, ease of migration, and ease of deployment	

Possible AI Use Cases



intel ARC™ 7xxE
High performance for heavy AI workloads and complex tasks

intel ARC™ 5xxE
Immersive visuals with enhanced AI inferencing capabilities

intel ARC™ 3xxE
Fits low power and small form factor requirements

intel ARC™ GRAPHICS

Intel Arc™ GPU for the Edge

Deploy discrete Intel® Arc™ GPUs at the edge for future-ready, high-efficiency AI, visual computing and media processing. GPU cards from Intel's ODM ecosystem feature diverse, edge-focused form factors with long life and optimizations for Embedded use conditions. Eliminate vendor lock-in with a more open, standards-based software stack to build high-performance AI applications and solutions.

Partner GPU Cards for the Edge (with Intel® Arc™ Graphics)



Drive AI and other GPU workloads

Purpose-built acceleration: XMX AI Engine for faster AI inference, Xe Media Engine to accelerate media-processing.

Engineered for the edge

Carve through the edge's complexities with long-life GPU card offerings from Intel's ODM ecosystem, with a range of form factors, power levels, and performance levels.

Eliminate vendor lock-in

With a more open, standards-based software stack. Code once and run across GPUs, CPUs and other hardware accelerators with the open-sourced OpenVINO™ toolkit.

[Intel® Arc™ GPU for the Edge Infographic](#)

[White Paper: Unlocking the AI Power of Intel® Arc™ GPU for the Edge: A Deep Dive into Hardware and Software Enablement](#)

Intel Movidius™ Vision Processing Units (VPUs)



Intel® Movidius™ VPU

Designed to specifically accelerate AI workloads on your PC, improving system responsiveness, efficiency, and compute performance needed for new, advanced features in collaboration and streaming.

Features and Benefits of PCs with Intel® Movidius™ VPU

Intel® Movidius™ VPUs enable demanding computer vision and AI workloads with efficiency. By coupling highly parallel programmable compute with workload-specific AI hardware acceleration in a unique architecture that minimizes data movement, Movidius VPUs achieve a balance of power efficiency, and compute performance.

Designed to accelerate Artificial Intelligence (AI) workloads on your PC, improving system responsiveness, efficiency and AI compute performance. Available on select 13th Gen Intel® Core™ Processor powered devices, the VPU will enable new Windows Studio effects that provide high quality background blur, eye contact, and automatic framing for collaboration and streaming applications. These effects can easily be applied system-wide through Windows Settings to provide consistent video effects for all apps that use the integrated camera.

Intel® Movidius™ Myriad™ X VPU

The Intel® Movidius™ Myriad™ X VPU is Intel's first VPU to feature the Neural Compute Engine – a dedicated hardware accelerator for deep neural network inference. The Neural Compute Engine in conjunction with the 16 powerful SHAVE cores and high throughput intelligent memory fabric makes Intel® Movidius™ Myriad™ X ideal for on-device deep neural networks and computer vision applications.

Key Takeaways

- Intel® Movidius™ Myriad™ X VPU delivers outstanding performance in computer vision and deep neural network inferencing applications.
- The Neural Compute Engine enables the Intel® Movidius™ Myriad™ X VPU to reach over 1 TOPS of compute performance on deep neural networks inferences.
- The Movidius family of VPUs provides a unique, flexible architecture for image processing, computer vision, and deep neural networks.
- The technology framework helps developers focus on the processing, leaving data flow optimization to the tools.

[Product Brief](#)



AI PLATFORMS (BOARDS & SYSTEMS)



Embedded Boards & Systems based on 3rd party AI Processors



Intel	Advantech	Asus	DFI	Adlink	Seco	Aaeon	iEi
Boards					X		
Systems	X	X	X	X			
GPU Cards/Edge AI Modules	X					X	X

Hailo

	Advantech	Kontron
Boards	X	
Systems	X	X
GPU Cards/Edge AI Modules	X	

Qualcomm

	Advantech	DFI
Boards	X	X
Systems	X	X
GPU Cards/Edge AI Modules	X	

Kneron

	Advantech	Aaeon
GPU Cards/Edge AI Modules	X	X
DevKits		X

Rockchip

	Advantech	Asus	Cherry
Boards		X	X
GPU Cards/Edge AI Modules	X		

Mediatek

	Via
Boards	X
Systems	X

Google Coral

	Kontron	Asus	iEi
Systems	X		
GPU Cards/Edge AI Modules		X	X

Axelera

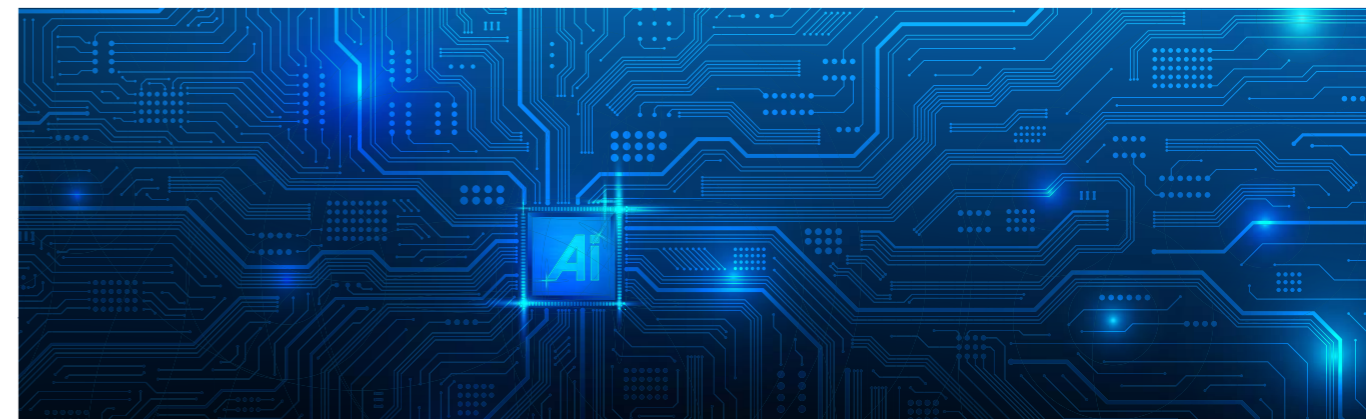
	Seco
Systems	X

NVIDIA

	Advantech	Asus	DFI	Adlink
Boards	X		X	
Systems	X	X	X	X
DevKits	X			
GPU Cards/Edge AI Modules	X			
AI Vision	X			X

NXP

	Advantech	DFI
GPU Cards/Edge AI Modules	X	X



Advantech AI Jetson Platforms

ADVANTECH

Enabling an Intelligent Planet



AI Developer Kit & Solution Kit



AI Computer Systems



AI Rugged System & AI NVR



AI IGX System



AI Camera



Compact industrial-grade solutions aimed at edge AI applications - powered by NVIDIA® Jetson™

AI is revolutionizing businesses with automation, improved data analysis, and enhanced engagement. NVIDIA Jetson™ is the leading platform for robotics and embedded edge AI applications, offering compact yet powerful computers, supported by the NVIDIA JetPack™ SDK for accelerated software development. NVIDIA Jetpack provides pre-built and cloud-native software services to fast-track development and deployment of sophisticated edge AI applications like generative AI, computer vision, and advanced robotics.

Advantech AI Developer Kit & Solution Kit



MIC-732D-AO

ADVANTECH

Enabling an Intelligent Planet

Open-frame systems designed with the test-then-design mindset, developers can use these kits to try their AI model.

Part Number	Description	Based on	Performance	Design	More information
MIC-732D-AO	AI Developer Kit	NVIDIA Jetson AGX Orin	up to 275 TOPS	Open-frame carrier design with thermal solution	Datasheet
MIC-713S-OX	AI Solution Kit	NVIDIA Jetson Orin NX	up to 100 TOPS	Open-frame solution kit design	Datasheet
MIC-713S-ON	AI Solution Kit	NVIDIA Jetson Orin Nano	up to 40 TOPS	Open-frame solution kit design	Datasheet
MIC-711D-OX	AI Developer Kit	NVIDIA Jetson Orin NX	up to 100 TOPS	Open-frame carrier design with thermal solution	Datasheet
MIC-711D-ON	AI Developer Kit	NVIDIA Jetson Orin Nano	up to 40 TOPS	Open-frame carrier design with thermal solution	Datasheet
MIC-710AILX-DVA	AI Developer Kit	NVIDIA Jetson Xavier NX	up to 21 TOPS	Open-frame carrier design with thermal solution	Datasheet
MIC-710AIL-DVA	AI Developer Kit	NVIDIA Jetson NANO	up to 472 GFLOPS	Open-frame carrier design	Datasheet



Advantech AI Computer Systems

Edge AI Inference Computers embedded with NVIDIA Jetson for video analytics in safety and security, transportation, and manufacturing applications.

Part Number	Description	Based on	Performance	More information
MIC-732-AO	AI Inference System	NVIDIA Jetson AGX Orin	up to 275 TOPS	Datasheet
MIC-733-AO	AI System	NVIDIA Jetson AGX Orin	up to 275 TOPS	Datasheet
MIC-713-OX	AI System	NVIDIA Jetson Orin NX	up to 100 TOPS	Datasheet
MIC-713-ON	AI Platform	NVIDIA Jetson Orin NANO	up to 20 TOPS	Datasheet
MIC-711-OX	AI Inference System	NVIDIA Jetson Orin NX	up to 100 TOPS	Datasheet
MIC-711-ON	AI Inference System	NVIDIA Jetson Orin Nano	up to 40 TOPS	Datasheet
MIC-710AILX	AI System (Lite)	NVIDIA Jetson Xavier NX	up to 21 TOPS	Datasheet
MIC-710AILT	AI System (Lite)	NVIDIA Jetson TX2 NX	up to 1.1.33 TFLOPs	Datasheet
MIC-710AIL	AI System (Lite)	NVIDIA Jetson Nano	up to 472 GFLOPS	Datasheet
MIC-710AIX	AI System	NVIDIA Jetson Xavier NX	up to 21 TOPS	Datasheet
MIC-710AIT	AI System	NVIDIA Jetson TX2 NX	up to 1.33 TOPS	Datasheet
MIC-710AI	AI Inference System	NVIDIA Jetson NANO	up to 472 GOPS	Datasheet
MIC-730AI	AI System	NVIDIA Jetson AGX Xavier	up to 32 TOPS	Datasheet



Advantech AI Rugged System & AI NVR

AI systems for special applications, including an AI rugged system for heavy machinery and AI network video recorder for smart city application.

Part Number	Description	Based on	Performance	Design	More info
MIC-717-OX	AI NVR Solution	NVIDIA Jetson Orin NX	up to 100 TOPS	Compact fan design	Datasheet
MIC-715-OX	Ruggedized In-Vehicle AI Fanless System	NVIDIA Jetson Orin NX	up to 100 TOPS	Ruggedized IP67 waterproof fanless design	Datasheet
MIC-730IVA	8ch AI Video System	NVIDIA Jetson AGX Xavier	up to 32 TOPS	NVR fan base design	Datasheet
MIC-710IVX	8ch AI Video System	NVIDIA Jetson Xavier NX	up to 21 TOPS	NVR fan base design	Datasheet
MIC-710IVA	8ch AI Video System	NVIDIA Jetson NANO	up to 512 GOPS	NVR fan base design	Datasheet
MIC-715-NX	Ruggedize AI Inference System Based	NVIDIA Jetson Xavier NX	up to 21 TOPS	Ruggedized IP67 waterproof fanless design	Datasheet

Advantech AI IGX System

AI systems delivers functional safety, manageability, and security without compromising any of its ultra-fast processing performance.

- Supports 1 x NVIDIA RTX A6000 for powerful visual computing
- Supports NVIDIA Metropolis SDK (MIC-735E-IO, MIC-735I-IO)
- Supports NVIDIA Clara™ Holoscan (MIC-735M-IO)



Part Number	Description	Based on	Performance	Design	More information
MIC-735E-IO	Enterprise grade server tower chassis edge AI system	NVIDIA IGX Orin	up to 275 TOPS	Open-frame carrier design with thermal solution	Datasheet
MIC-735I-IO	Industry grade short depth 4U rackmount chassis edge AI system	NVIDIA IGX Orin	up to 275 TOPS	Open-frame solution kit design	Datasheet
MIC-735M-IO	Medical grade small tower edge AI system	NVIDIA IGX Orin	up to 275 TOPS	Open-frame solution kit design	Datasheet



ASUS Edge AI Systems

The Game-Changing Platform for AI Applications

ASUS IoT edge AI systems combine GPU computing with AIoT potential, offering embedded MXM GPU modules from both NVIDIA® and Intel®, NVIDIA® Jetson-based platforms, and GPU computing platforms for diverse market needs. With unparalleled performance, they enable real-time AI inferencing at the edge, transforming industries. Designed with a rugged, fanless, anti-vibration build, wide temperature support and low power consumption, they excel in demanding edge AI applications like factory automation, machine vision, video analytics, and autonomous vehicles. ASUS IoT ensures robustness and reliability for the most challenging scenarios, driving innovation and efficiency in this new era.

Full Spectrum of Edge AI Systems and Applications

ASUS IoT offers a comprehensive portfolio of edge AI systems to address the diverse range of market applications. Powered by embedded MXM GPU modules from both NVIDIA® and Intel®, and PCI Express graphics cards based on NVIDIA® Quadro GPUs, as well as edge AI platforms utilizing NVIDIA® Jetson™ modules and other embedded form factors, these systems are designed to accelerate edge computing and AI workloads. With a focus on performance, long lifecycle, power consumption, and form factor, ASUS IoT provides a full spectrum of solutions to meet various embedded requirements. Whether it's high-performance computing or specific form factors, our edge AI systems deliver the power and flexibility needed to excel in diverse applications.

Edge AI Computer Portfolio
Real-time Decision-Making with Confidence

Models shown: PE 1000N, PE 1100N, PE 3000G, PE 400D, PE 4000G, PE 6000G, PE 8000G

Performance metrics: TOPS/TFLOPS, TOPS

Logos: NVIDIA JETSON, intel, NVIDIA GPUs Support

Attributes: Compact and Power Efficient, High Throughput



ASUS Edge AI Computer Portfolio X86 based



Flyer ASUS Edge AI Systems



		PE3000G	PE400D	PE4000G	PE6000G	PE8000G
Case	Dimension	240 x 230 x 125.7 mm	176.6 x 210 x 250 mm	225 x 190 x 350 mm	225 x 221 x 443 mm	225 x 288 x 443 mm
	Weight	8.2 kg	6.8 kg	8 kg	9.2 kg	Coming soon
	Chassis construction	Aluminium alloy with heavy duty metal	Aluminium alloy with heavy duty metal	Aluminium alloy with heavy duty metal	Aluminium alloy with heavy duty metal	Aluminium alloy with heavy duty metal
System	Processor	Intel® Core™ i7-12800HE Intel® Core™ i5-12800HE Intel® Core™ i3-12800HE	Intel® Xeon® W-1290TE Intel® Core™ i9-10900E Intel® Core™ i7-10700E Intel® Core™ i5-10500E Intel® Core™ i3-10100E	Intel® Core™ i9-12900E/ i9-12900TE Intel® Core™ i7-12700E/ i7-12700TE Intel® Core™ i5-12500E/ i5-12500TE Intel® Core™ i3-12100E/ i3-12100TE	Intel® Core™ i9-12900E/ i9-12900TE Intel® Core™ i7-12700E/ i7-12700TE Intel® Core™ i5-12500E/ i5-12500TE Intel® Core™ i3-12100E/ i3-12100TE	Intel® Core™ i9-12900E/ i9-12900TE Intel® Core™ i7-12700E/ i7-12700TE Intel® Core™ i5-12500E/ i5-12500TE Intel® Core™ i3-12100E/ i3-12100TE
	Chipset	-	W480E	R680E	R680E	R680E
	Graphics	Intel® Iris® Xe Graphics (i7/i5) Intel® UHD Graphics (i3)	Intel® UHD Graphics 630	Intel® UHD Graphics 770	Intel® UHD Graphics 770	Intel® UHD Graphics 770
	Memory	2 x SO-DIMM, up to 64 GB DDR5 SDRAM	2 x SO-DIMM, up to 64 GB ECC/ non-ECC DDR4 SDRAM (ECC only for Xeon CPU)	2 x SO-DIMM, up to 64 GB ECC/ non-ECC DDR5 SDRAM	2 x SO-DIMM, up to 64 GB ECC/ non-ECC DDR5 SDRAM	2 x SO-DIMM, up to 64 GB ECC/ non-ECC DDR5 SDRAM



ASUS Edge AI Computer Portfolio ARM based



Flyer ASUS Edge AI Systems

		PE1100N Series				PE1000N Series		
SoC/SOM		Jetson Orin Nano 4G	Jetson Orin Nano 8G	Jetson Orin NX 8G	Jetson Orin NX 16G	Jetson Nano™	Jetson TX2 NX	Jetson Xavier™ NX
System	CPU	6 x Arm® Cortex®-A78AE v8.2			8 x Arm® Cortex®-A78AE v8.2	4 x Arm® Cortex®-A57	2 x NVIDIA Denver 2 64-Bit 4 x Arm® Cortex®-A57	6 x NVIDIA Carmel Arm® v8.2 64-Bit
	GPU	512-core NVIDIA Ampere GPU with 16 Tensor Cores	1024-core NVIDIA Ampere GPU with 32 Tensor Cores			128-core NVIDIA Maxwell™	256-core NVIDIA Pascal™ GPU	384-core NVIDIA Volta™ with 48 Tensor Cores
	Memory	4 GB 64-bit LPDDR5	8 GB 128-bit LPDDR5	8 GB 128-bit LPDDR5	16 GB 128-bit LPDDR5	4 GB 64-bit LPDDR4	4 GB 128-bit LPDDR4	8 or 16 GB 128-bit LPDDR4x





DFI NVIDIA Jetson Platforms

DFI provides a wide range of tailored GPU computing solutions, such as MXM GPU modules or dedicated graphics cards, plus NVIDIA® Jetson modules for Edge AI applications, that optimize edge computing and AI tasks in response to performance requirements.

DFI AI Embedded Board Products

Deep Learning at A Small Footprint: CS181 with MXM & 5G module

Industrial mITX motherboard on the solid computing basis of 8th/9th Gen Intel® Core™ i7/i5/i3 processors tailored to infuse Deep Learning into smart edge computing.

- Accelerated graphic processing by MXM for AI vision and machine learning
- High-precision computer vision by small PC
- For demanding computer vision or AI applications, like AOI, machine vision, vision guided robotics, and AGV in factory automation



CS181-Q370

Datasheet: [CS181-H310 mITX](#) | [CS181-Q370 mITX](#)

Key features

- DDR4: 2x DDR4 SODIMM up to 64 GB
- 4K High Resolution: Supports 4K/ 2K resolution
- Multiple Displays: eDP; DP++/HDMI (auto-detection); 4x DP (MXM)
- Multiple Expansion: MXM Type-A/B/B+/C; M.2 M / B / E Key
- Rich I/O: 4x Intel GbE; 2x COM; 4x USB 3.2 Gen1; 2x USB 2.0



CS181-Q370

DFI AI Embedded System Products

AI Inference/Training systems & Medical systems.

Part Number	Description	NVIDIA AI	Performance	More Info
X6-MTH-ORN	Fanless Edge AI Box	NVIDIA Jetson Orin NX	up to 100 TOPS	Datasheet
EC180-CS	Fanless AI-Enabled Embedded System	Up to 60W NVIDIA GPU MXM module supported		Datasheet
EC102-XNX	AI-Enable Vision Embedded System	Fully support NVIDIA® Jetson Xavier™ NX module		Datasheet
EC100-XNX	AI-Enable Vision Embedded System	Fully support NVIDIA® Jetson Xavier™ NX module		Datasheet
MPC544-RPS	Medical Edge AI inference System	Support various NVIDIA® AI GPU Card and Capture card		Datasheet



EC180-CS



EC102-XNX



EC100-XNX



MPC544-RPS



ADLINK Edge AI Platforms

The Jetson-based DLAP Edge AI Platforms offer users the flexibility to cost-effectively achieve the SWaP and AI performance. Users can choose from lines of NVIDIA Jetson SoMs, form factors, and power consumption ranges to satisfy their specific application requirements.

Product Name	Description	Based on	Performance	More Info
DLAP-211-Orin Series	Edge AI Platform	NVIDIA Jetson Orin	up to 100 TOPS	Datasheet
DLAP-211-JNX	Edge AI Inference Platform	NVIDIA Jetson Xavier NX	up to 21 TOPS	Datasheet
DLAP-211-JT2	Edge AI Inference Platform	NVIDIA Jetson TX2 NX	up to 1.33 TFLOPS	Datasheet
DLAP-211-Nano	Edge AI Inference Platform	NVIDIA Jetson Nano	up to 472 GFLOPS	Datasheet
DLAP-401-Xavier	Edge AI Inference Platform	NVIDIA Jetson AGX Xavier	up to 100 TOPS	Datasheet
DLAP-411-Orin	Edge AI Platform	NVIDIA Jetson AGX Orin	up to 40 TOPS	Datasheet
RQX-580/58G	Embedded Robotic Controller	NVIDIA Jetson AGX Xavier	512-core 1.37 GHz GPU	Datasheet
RQX59 Series	Embedded Robotic Controller	NVIDIA Jetson AGX Orin	with 1792 NVIDIA CUDA cores and 56 tensor cores	Datasheet
EOS-JNX Series	Edge AI Vision Inference System	NVIDIA Jetson Xavier NX	with 384 NVIDIA CUDA cores and 48 tensor cores	Datasheet



DLAP-211-Orin Series



DLAP-401-Xavier



RQX-580/58G



EOS-JNX Series

Powered by Intel® CPU



Product Name	Description	Based on	More Info
DLAP-3000-CF Series	Embedded System	ADLINK MXM Graphics module support (Type A/B, up to 120 W) 8 th /9 th Gen Intel Core i7/i5/i3, Celeron processor	Datasheet
DLAP-3100-CF Series	Embedded System	ADLINK MXM Graphics module support (Type A/B, up to 120 W) 8 th /9 th Gen Intel Core i7/i5/i3, Celeron processor	Datasheet
DLAP-3200-CF Series	Embedded System	ADLINK MXM Graphics module support (Type A/B, up to 120 W) 8 th /9 th Gen Intel Core i7/i5/i3, Celeron processor	Datasheet
DLAP-4000 Series	Embedded System	NVIDIA® Quadro® PEG card support 8 th /9 th Gen Intel Core i7/i5/i3, Celeron processor	Datasheet
DLAP-5200 Series	Fanless Edge AI Computer	Supports MXM 3.1 Type A/B form factor GPU module Based on 12 th /13 th Generation Intel Processors	Datasheet
DLAP-8000 Series	Industrial GPU Workstation	9 th Gen Intel® Xeon®, Core™ i7/i5/i3 LGA processors with workstation C246 chipset	Datasheet



DLAP-3000-CF Series



DLAP-3200-CF Series



DLAP-4000 Series



DLAP-8000 Series





ADLINK Edge AI Platforms

AVA-5500 Series

Rugged, Fanless AIoT Platform with NVIDIA Quadro® GPU Embedded for Real-time Video/Graphics Analytics

- 6th/7th Gen Intel® Core™ i7 Processors
- GPU MXM 3.1 Type A/B module on PCI Express x16
- 8x M12 GbE (4x PoE), 4x RS-422, 4x USB 3.0, 1x DVI-I, 4x DisplayPort with lockable connectors
- Multiple storage options: 2x 2.5" SATA 6.0 Gb/s drive bays, 1x M.2 2280 slot, 1x CFast socket



Powered by



[Datasheet](#)

Artificial Intelligence (AI) technologies are being adopted globally by all industries to drive efficiency, improve productivity and reduce costs. The rail industry is no exception. Fueled by intelligence from AI-driven systems and applications, railway operations are becoming safer, smarter and more reliable, significantly enhancing passenger travel experience and freight logistics services. For rail transportation, AI can provide improvements including fast and convenient ticket-free check-in to accurate arrival-time predictions, personalized infotainment and onboard services, real-time track health diagnostics, and rapid response in an emergency.

Powered by an Intel® Core™ i7 processor and integrated NVIDIA Quadro MXM module, ADLINK's EN50155 certified AVA-5500 AIoT Platform is not only ruggedized for both wayside and onboard deployment with its wide range DC input and isolated I/O design, but also provides an ideal edge solution for real-time video/graphics analysis applications that are vital to today's increasingly complex railroad operations.

The target applications include but are not limited to:

- Passenger Information Systems
- Railroad Intrusion Detection
- Train Station Surveillance
- Onboard Video Security
- Railroad Hazard Detection

ADLINK's AVA-5500 can also be deployed for AI-enabled surveillance and video/graphics processing applications in a variety of adjacent markets including mil/aero, public safety, energy and industrial automation.

ADLINK Edge AI Platforms

MEC-AI7400SR

4U AI Edge Server

- Single 4th Gen Intel® Xeon® Scalable Processor
- 8x DDR5 DIMMs, 4800MT/s (8 channels, 1 DPC)
- Supports up to 5x full-height PCIe slots for GPU/FPGA accelerator cards
- BMC with AST2600, IPMI v2.0/ Redfish compliant
- Short depth chassis design



Powered by



[Datasheet](#)

ADLINK's AI Edge Server MEC-AI7400 series can be equipped with various acceleration cards according to application needs. The flexible and diverse combinations can meet the various changes in smart manufacturing applications, including GPU cards, motion control cards, IO cards, and image capture cards.

The AI Edge Server MEC-AI7400 series offers several advantages for smart manufacturing applications.

- **Fast delivery:** It guarantees reliable supply and high-quality products, helping projects advance on schedule and driving business growth.
- **Diverse and flexible product portfolio:** These diversified products can meet the various needs of different customers and adapt to the ever-changing demands of smart manufacturing.
- **Complete solution:** It integrates ADINK's image acquisition cards, vision/motion cards, GPU cards, and servers into a complete system. Passenger Information Systems.


AI Edge Server Applications:

- Smart Automation: AI AOI, PCB defects AOI, DIP AOI
- Digital transformation: LLM, Digital Twin, Generative AI, ChatGPT



The Most Comprehensive AI PC Lineup


Light AI PC
Heavy AI PC



Everyday AI

10+ TOPS NPU


- NUC 14 PRO
- NUC 14 PRO+



Next Level AI

45+ TOPS NPU

- Lunar Canyon



Advanced AI

NPU + RTX 40-series GPU

- ROG NUC
- NUC 14 Performance

ASUS NUC – Light AI PCs

Configured and tuned for optimal performance, ASUS NUC 14 [Pro] comes equipped with the latest Intel® Core™ Ultra processors, Intel® Arc™ GPUs, and Intel® AI Boost, Intel's newest NPU, to meet all compute needs.

ASUS NUC 14 Pro






- **Performance:** Delivers best-in-class performance among comparable mini-PC offerings
- **Manageability:** Most comprehensive mini-PC with Intel vPro® Enterprise for exceptional security, manageability, and stability
- **AI Ready:** Harnesses the prowess of Intel® Core™ Ultra processors' three AI engines for high throughput, low power, and fast response
- **Wi-Fi Sensing:** Takes advantage of Wi-Fi sensing to deliver instant availability while reducing power consumption
- **Wireless Connectivity:** Delivers seamless connectivity for certified Bluetooth® dongle-free experiences
- **Customizable:** Designed for toolless chassis access making upgrades easier, safer, and faster

[Datasheet](#)

ASUS IN SEARCH OF INCREDIBLE

Powered by

ASUS NUC 14 Pro+



- **Performance:** Superior performance besting all other competitive mini-PCs
- **Design:** Features an eye-catching 5 x 4-inch anodized aluminum chassis
- **AI Ready:** Harnesses the prowess of the Intel® Core™ Ultra 9 processor to locally run generative AI workloads
- **Wi-Fi Sensing:** Takes advantage of Wi-Fi sensing to deliver instant availability while reducing power consumption
- **Wireless Connectivity:** Delivers seamless connectivity for certified Bluetooth®
- **Customizable:** Designed for toolless chassis access making upgrades easier, safer, and faster

[Datasheet](#)



Powered by



ASUS NUC – Advanced AI PCs

Harness the prowess of Intel® Core™ Ultra processors' three AI engines - GPU, NPU, CPU - for high throughput, low power, and fast response.

ROG NUC



- With Intel® Core™ Ultra 9/7 processor deliver superior performance.
- Ultra-small form factor delivers superior performance and seamless multitasking.
- Gamers can chat, browse, stream, edit, record and play without skipping a beat.
- NVIDIA® GeForce RTX™ 4070/4060-series graphics make immersive gaming experiences possible.
- Arrange systems vertically or horizontally (vertical stand included).
- Connect fast with 2.5 Gb Ethernet, Intel® Killer® Wi-Fi 6E, and Bluetooth® 5.3.

[Datasheet](#)

ASUS NUC 14 Performance



- **Performance:** Achieve top-tier Intel® Core™ Ultra 9 performance with our mini-PC, surpassing all expectations.
- **Speed:** NVIDIA® GeForce RTX™ 4070 and RTX™ 4060 Discrete Graphics are beyond fast for business productivity.
- **Five 4K Display:** Efficiently manage your calendar, meetings, cloud files, and emails without compromising productivity.
- **Flexible:** The stand provides better versatility in various placement options, making it easier to plug and unplug cables.
- **AI Ready:** Exploits the prowess of Intel® Core™ Ultra processors' three AI engines – GPU, NPU, and CPU – for high throughput, low power, and quick reaction.

[Datasheet](#)



Certified and AI Ready

Hardware



Cloud Service Providers



Operating Systems



AI Developer Kits







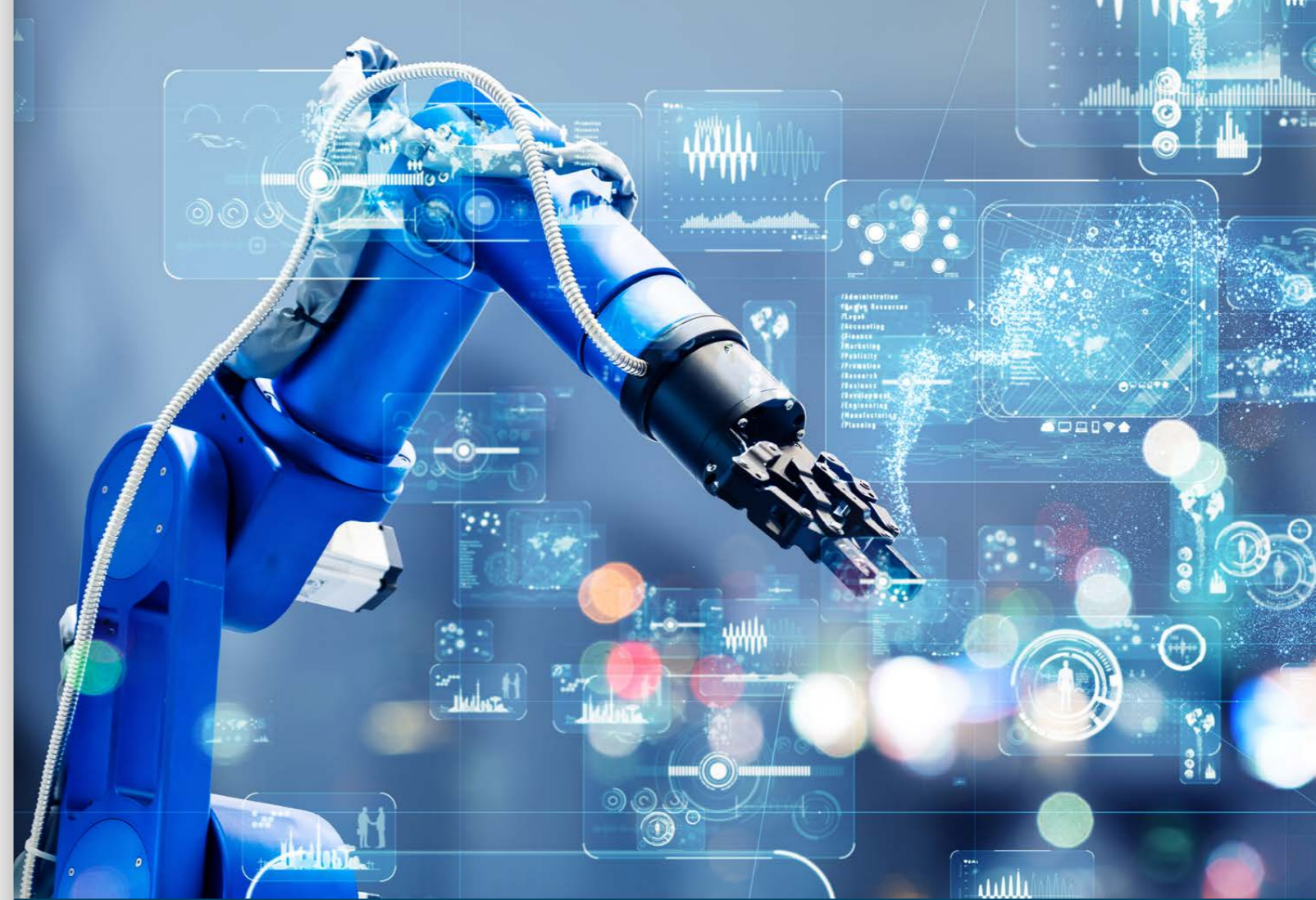
Powered by



Edge AI Inference Systems & Servers

Powered by NVIDIA Jetson family and RTX, AIR series provide scalable AI inference performance or efficient retraining on a large scale at the edge.

Part #	Description	Based on	Performance	More Info
AIR-310	MXM GPU Edge AI System	14 th Gen Intel® Core™ i3/i5/i7/i9; Support MXM 3.1 Type A GPU card		Datasheet
AIR-150	AI Inference System	13 th Gen Intel® Core™ i3/i5 with Hailo-8 M.2 AI module	Up to 26 TOPS	Datasheet
AIR-030	AI Inference System	NVIDIA® Jetson AGX Orin™		Datasheet
AIR-510	4U Edge AI Workstation	13 th Gen Intel® Core™ (RPL S) NVIDIA Certified with RTX 6000 Ada		Datasheet
AIR-520	4U Edge AI Server	AMD EPYC 7003 series processor		Datasheet
AIR-530	Edge AI workstation	NVIDIA IGX Platform; NVIDIA Qualified w/ RTX A6000 graphic card	32 TFLPOS	
AIR-020N	AI Inference System	NVIDIA Jetson Nano	512 GFLOPS	Datasheet
AIR-020T	AI Inference System	NVIDIA Jetson TX2 NX	Up to 1.33 TFLOPS	Datasheet
AIR-020X	AI Inference System	NVIDIA Jetson Xavier NX	21 TOPS	Datasheet
AIR-500D	Extreme AI Edge Server	Intel® Xeon® D-1700 Series		Datasheet
AIR-100	Edge AI Signage System	Intel® Atom™ E3950 processors Intel Movidius Myriad X MA2485 VPU x1		Datasheet
AIR-101	Edge AI Inference System	Intel® Atom™ E3950 Quad Core SoC Intel Movidius Myriad X MA2485 VPU x2		Datasheet
AIR-200	Edge AI Inference Rugged System	6 th Gen Intel® Core i5-6442EQ QC SoC Intel Movidius Myriad X VPU x2		Datasheet
AIR-300	Edge AI Inference System	Intel® Xeon®/ Core™ i3/i5/i7 LGA1151 Supports PCIe x16 high power GPU card		Datasheet
Ei-A100	4K Digital Signage Player with VPU for AI Facial Recognition	Intel® Atom™ E3950 Intel Movidius Myriad X VPU MA2485 x1		Datasheet



ADVANTECH & HAILO Partnership

ADVANTECH **NXP** GOLD PARTNER

HAILO

Advantech & Hailo Partnership
To Provide Scalable, Power-Efficient Edge AI Solutions



Hailo-8™
AI Acceleration Module



EPC-R3720



RSB-3720

Powered by NXP i.MX 8M Plus

Powered by



- Ready-to-use reference platform that enables scalable high-performance AI capabilities at the edge
- Empowering enterprises to run full-scale deep learning (DL) with 26 TOPS and best-in-class power efficiency
- Cutting-edge AI solutions to myriad intelligent market segments including robotics, medical, and transportation

RSB-3720



2.5" Pico-ITX

- Operating Temp.: 0–60 °C; -40–85 °C
- 3.5G vibration tolerance
- 7.13W (max.)
- M.2 (Key E) and mini PCIe
- 2 x MIPI-CSI interfaces

[Datasheet](#)

EPC-R3720



Edge AI System

- Operating Temp.: -40–70 °C
- 3.0G vibration compliance
- 9.2W (max.)
- Lockable DC Jack
- Plug-and-play COM/CAN port

[Datasheet](#)

EAI-1200



Hailo-8 Module is available for purchase as an add-on with Advantech's RSB-3720 and EPC-R3720.

Hailo-8™ M.2 AI Acceleration Module

- Up to 26 TOPS
- Best-in-class power efficiency at 2.5W
- PCIe Gen-3.0, 2 lanes (up to 16 Gbs)
- Frameworks: TensorFlow/Lite, ONNX, Keras, & Pytorch
- OS: Linux

[Datasheet](#)

[HAILO Product Brief](#)



Kontron Industrial AI PCs



Kontron offers a product portfolio based on standardized hardware and software platforms to enable cost-effective, industrial AI-solutions. Suitable for neural networks & Deep Learning around object recognition/classification and quality inspection of objects.

KBox A-150-WKL-AI-H8

Industrial Computer for high-performance edge AI inference solutions

Specifications

- High-performance, compact and power-efficient AI Box PC solution
- 8th Gen Intel® Core™ i or Celeron® processor
- Hailo-8™ AI accelerator
- Software compatible to common AI tools
- Fully integrated runtime system

[Datasheet](#)



Powered by
HAILO

KBox A-203-AI-GC

Industrial Computer Platform for AI and IoT Edge-Applications

Specifications

- Ideal for AI and Gateway-applications
- Based on Intel Atom® processor
- Google Coral Edge TPU
- TensorFlow Lite compatible
- Maintenance-free: no moving parts, battery-free

[Datasheet](#)



Powered by
With Coral intelligence

ASUS Single-board Computer for AI Applications



Powered by
Rockchip

Tinker Edge R

Single-board computer for AI applications, with Arm® big.LITTLE™ technology and Rockchip NPU, easy for Machine-Learning deployment.

- Arm® big.LITTLE™ A72+A53 Hexa-core SoC
- Machine-Learning capability with Rockchip NPU
- Multiple MIPI-CSI & DSI / HDMI / Type-C (DP)
- 40-pin GPIOs & mPCle for multiple expansions
- 12V~19V DC-in offers stable power delivery
- Linux & Android supported



AI accelerator (NPU)

ASUS Tinker Edge R is an Single Board Computer (SBC) specially designed for AI applications. It uses Rockchip NPU, a Machine Learning (ML) accelerator that speeds up processing efficiency, lowers power demands and makes it easier to build connected devices and intelligent applications. With this integrated Machine Learning (ML) accelerator, the Tinker Edge R is capable of performing 3 tera-operations per second (TOPS), using low power consumption. And it's optimized for Neural Network (NN) architecture, which means Tinker Edge R can support multiple Machine Learning (ML) frameworks and let lots common Machine Learning (ML) models can be easily compiled and run on the Tinker Edge R.

VIA Intelligent Edge Modules



Powered by
MEDIATEK

Bring your next-generation edge AI and embedded systems to life with VIA Intelligent Edge Solutions. Available in a wide variety of form factors, these compact and reliable platforms combine advanced compute and AI performance with low power consumption and rich I/O and connectivity features to accelerate the development of industrial, commercial, and consumer edge applications.

Product Name	Description	AI Performance	More Info
VIA SOM-7000	Powered by a MediaTek Genio 1200 Octa-Core Processor 4x Cortex-A78 cores @ 2.2GHz + 4x Cortex-A55 cores @ 2.0GHz	Integrated AI processor providing up to 4.8 TOPS	Datasheet
VIA SOM-5000	Powered by a MediaTek Genio 700 Octa-Core Processor 2x Cortex-A78 cores @ 2.2GHz + 6x Cortex-A55 cores @ 2.0GHz	Integrated AI processor providing up to 4.0 TOPS	Datasheet
VIA SOM-3000	Powered by a MediaTek Genio 350 Quad-Core Processor 4x Cortex-A78 cores @ 2.2GHz + 4x Cortex-A55 cores @ 2.0GHz	Integrated AI processor	Datasheet
VIA SOM-9X50	Powered by a MediaTek Genio 500 Octa-Core Processor 4x Cortex-A73 cores @ 2.0GHz + 4x Cortex-A53 cores @ 2.0GHz	Integrated AI processor	Datasheet
VIA SODB7	Carrier Board for VIA SOM-7000 and VIA SOM-5000 SBC form factor measuring 151mm x 134mm (5.94" x 5.28")	-	
VIA VAB-935	Carrier Board for VIA SOM-3000 3.5" SBC form factor measuring 146mm x 102mm	-	
VIA VAB-950	Carrier Board for VIA SOM-9X50 EPIC form factor measuring 140mm x 100mm	-	



SOM-7000



SOM-9X50 + VAB-950

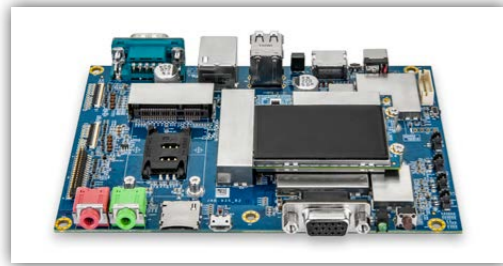




VIA Intelligent Edge Boards

Accelerate time-to-market for innovative new edge AI devices with the VIA VAB-3000 board. Featuring the low-power MediaTek Genio 350 quad-core SoC, this flexible and high-performance platform is ideal for mainstream edge AI and IoT applications such as facial recognition, object identification, and motion tracking that require vision and voice edge processing.

Based on a 3.5" SBC form factor, the VIA VAB-3000 integrates a host of advanced multimedia features and rich I/O interfaces, including support for dual MIPI-CSI cameras and HD displays, making it a compelling choice for a wide variety of consumer, commercial, medical, and educational use cases. For the full hardware specifications of the VIA VAB-3000 board, please download the datasheet to find out more.



VIA VAB-3000

- 3.5" SBC Edge AI Platform
- Compact form factor
- MediaTek Genio 350 Quad-Core SoC
- Integrated AI processor for Edge AI applications

[Datasheet](#)

VIA Intelligent Edge Systems

VIA Intelligent Edge systems combine robust compute and AI performance with low power consumption and rich I/O and connectivity features in a wide range of tough and durable form factors. With their reliable and flexible designs, the systems are ideal for industrial and commercial edge applications in the most demanding indoor and outdoor environments.

VIA ARTiGO A3000 System

Edge AI System



- Fanless ultra-compact Quad-Core ARM system for Edge AI applications
- Powered by a 2.0GHz MediaTek Genio 350 Quad-Core SoC
- Integrated AI processor for computer vision, deep learning, and neural network acceleration applications.

[Datasheet](#)

VIA ARTiGO A5000 System

Edge AI System



- Fanless ultra-compact Octa-Core ARM system for Edge AI applications
- Powered by MediaTek Genio 700 Octa-Core SoC; 2x Cortex-A78 @ 2.2GHz + 6x Cortex-A55 @ 2.0GHz
- Integrated AI processor for computer vision, deep learning, and neural network acceleration applications.

[Datasheet](#)

VIA IVT01

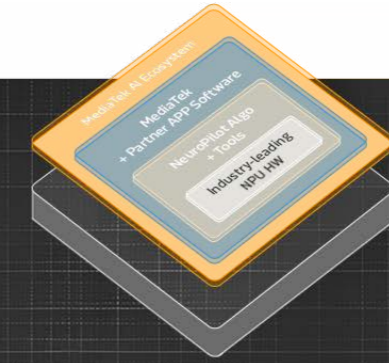
Android Tablet



- 8.0" Capacitive touch screen for in-vehicle infotainment and fleet management applications
- MediaTek Genio 500 octa-core SoC with 4x Cortex-A73 @ 2.0GHz and 4x Cortex-A53 @ 2.0GHz processors
- Integrated AI processor for computer vision, deep learning, and neural network acceleration applications.

[Datasheet](#)

The MediaTek AI ecosystem includes hardware, development tools, and a software development kit (SDK).



MEDIATEK

MediaTek AI Ecosystem

The MediaTek NPU, or processor dedicated for AI tasks, is as important as the CPU and GPU within our system-on-chips.

The MediaTek NPU provides a very efficient architecture to implement AI acceleration in edge devices, such as smartphones, tablets, AI+IoT, smart TVs, networking SoCs, and more. Using five common neural networks as a benchmark, the average power/performance curves show that MediaTek's Deep Learning Accelerator (DLA) contained with the NPU offers 27x greater power efficiency compared with a typical CPU and 15x greater power efficiency compared with a typical GPU.

MediaTek NPU

The MediaTek NPU is a highly scalable multi-core AI processor that can contain both MDLA and MVPU cores in varying quantities, depending on the application requirements. Each MediaTek NPU contains the following:

- Hardware-based multicore scheduler.
- A dedicated DMA engine that performs deep layer fusion and data compression to reduce the demand on DRAM bandwidth.
- Power intelligence and shared memory awareness.
- MediaTek designed Network on Chip (NoC) for low latency intercore communication.

MediaTek Deep Learning Accelerator (MDLA)

High performance, programmable and versatile, with an energy-efficient MAC architecture that can perform data reuse in a wide-range of NN applications. Designed to handle different types of networks, including:

- Convolutional Neural Networks
- Recurrent Neural Networks
- Long Short-Term Memory models and
- Mobile Bidirectional Encoder Representations from Transformers (BERT)

MediaTek Vision Processing Unit (MVPU)

General purpose DSP optimized for computer vision (CV) and neural network (NN) applications. Best balance of power and performance for visual processing applications such as photography, videography, and video stream playback.



Advantech RSB and Box Computer



RSB-4760
3.5" SBC form factor



- Qualcomm ARM® Cortex®-A53 APQ8016 Quad core up to 1.2 GHz
- Availability: MP
- Longevity: Q3, 2028

[Datasheet](#)

EPC-R4760
ARM based box computer

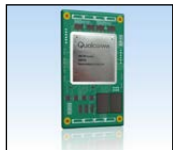


- Qualcomm ARM® Cortex®-A53 APQ8016 Quad core up to 1.2 GHz
- Availability: MP
- Longevity: Q3, 2028

[Datasheet](#)

Advantech AI Function Board

AFE-D260
Pico-ITX form factor



- Qualcomm QCS8550
- Availability: Planning
- Longevity: estimated 10 years

AIMB-293
AI-enabled Mini-ITX motherboard



- Qualcomm QCS8550 (12 Core processor)
- 45 TOPS AI acceleration
- Availability: Planning
- Longevity: estimated 10 years

Advantech AI Edge Systems

AI Edge Gateway (Product Concept)



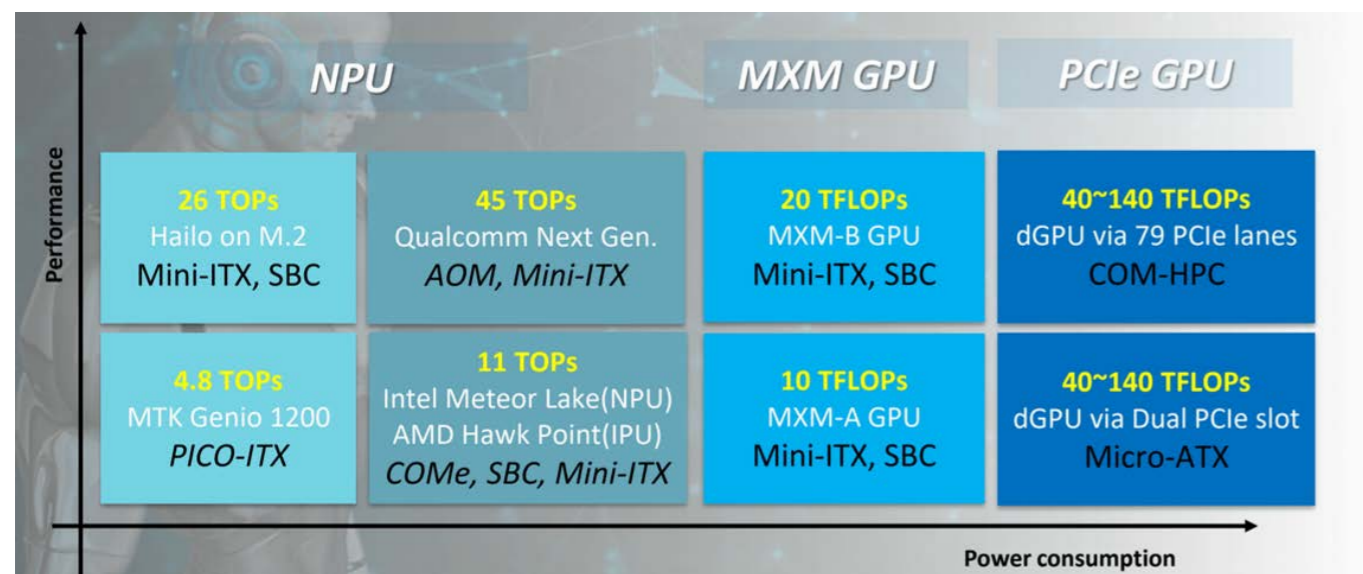
- Next Gen. SoC
- Availability: Planning
- Longevity: estimated 10 years

AI Interference PC (Product Concept)



- Next Gen. SoC
- Availability: Planning
- Longevity: estimated 10 years

Advantech Embedded AI PC Platform Portfolio



DFI Embedded AI Board Products



QCS051
2.5" Pico-ITX

- Qualcomm QCS6490 high-level platform
- Target on AMR, box PC application
- Thin client concept and ruggedize
- 2 M.2 Expansions: 1 M.2 E key 2230, 1 M.2 B key 3052
- Supports CPU Life Cycle Until 2036

AIMF51
1.8" SBC

- Qualcomm® QCS8550 high-level platform
- Target on edge AI box, AIoT application
- Thin client concept and ruggedize
- Optimized AI architecture (48 INT8, 12 FP16 TOPs)
- Supports CPU Life Cycle Until 2033



[Datasheet](#)



[Datasheet](#)

DFI Embedded AI System



VC500-CMS-MXM
In-vehicle Edge AI System based on 10th Gen Intel® Processors



Key Features

- High Performance CPU: 10th Gen Intel® Xeon/Core™/Pentium/Celeron Processors
- 802.11af PoE Ports: 4 x RJ45 or M12 X-coded type PoE ports at 15.4 W
- AI Accelerated: Support MXM GPU module
- Multiple Expansion: 1 miniPCIe and 4 M.2 slots
- Dual SIM for single modem: Two SIM cards for 5G/LTE module

[Datasheet](#)

SECO Embedded AI Systems



Titan 300 TGL-UP3 AI

Fanless embedded computer with the 11th Gen Intel® Core™ and Intel® Celeron® SoCs plus Axelera AI Chip



CPU
11th Gen Intel® Core™ processors and Intel® Celeron® SoCs

Graphics
Intel® Iris® Xe architecture with up to 96 EUs, up to 4 independent displays

Connectivity
2x 2.5 GbE, Optional M.2 WWAN and WLAN modules

Memory
2 DDR4 SO-DIMM slots supporting DDR4-3200 ECC memory with IBECC

[Datasheet](#)

- Fanless embedded computer
- 11th Gen Intel® Core™ and Intel® Celeron® SoCs (Codename: Tiger Lake UP3) plus Axelera AI Chip
- Powered by a single Metis AIPU up to 120 TOPS
- Voyager SDK for effortless deployment of AI applications

SECO Embedded AI Boards

UDOO VISION



Pico-ITX Single Board Computer with the Intel® Atom® X Series, Intel® Celeron® N Series and Intel® Pentium® N Series (code-name: Apollo Lake) Processors

[Datasheet](#)

SOM-SMARC-MX8M-Plus



SMARC® Rel. 2.1.1 Computer on Module (CoM) with NXP i.MX 8M Plus Applications Processors.

(formerly LEVY - D18)

[Datasheet](#)

SBC-pITX-EHL



Pico-ITX SBC with the Intel® Atom® X6000E Series, Intel® Pentium® and Celeron® N and J Series (codename: Elkhart Lake) SoCs.

(formerly ICARUS - D63)

[Datasheet](#)

Cherry Embedded Solutions SBC and SoM



JAGUAR SBC-RK3588-AMR

Single-Board Computer (SBC)



- Designed from scratch for mass production of cost-efficient Autonomous Mobile Robots (AMR).
- Rockchip RK3588 is a low-power, octa-core processor for Internet of Things (IoT) devices with Artificial Intelligence (AI)
- 4x ARM Cortex-A76 and 4x ARM Cortex-A55 cores
- Mali G610 GPU
- NPU with up to 6 TOPs.

[Datasheet](#)

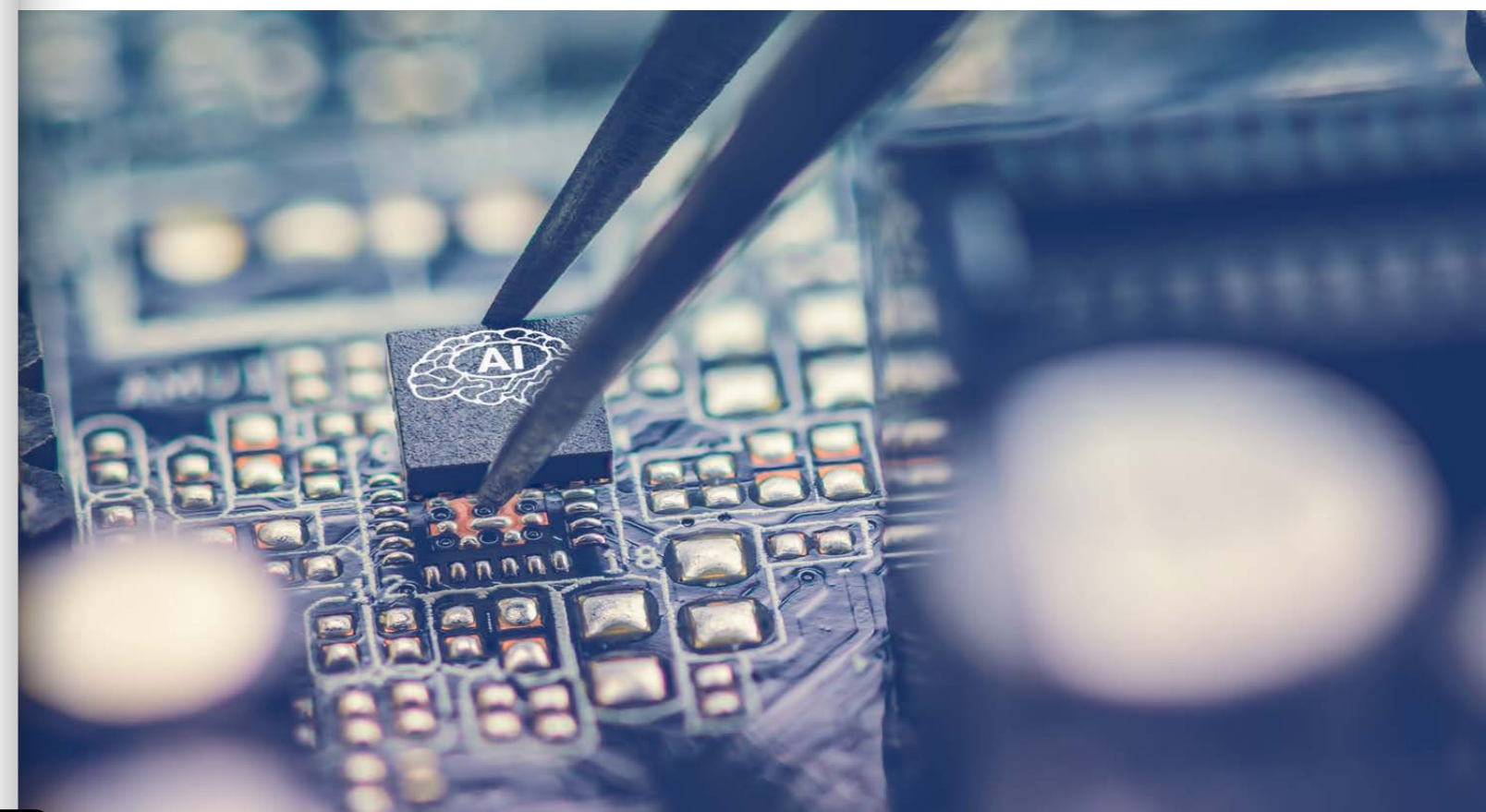
TIGER SOM-RK3588-Q7

System-on-Module (SOM)



- Cutting-edge computational performance, connects multiple high-resolution cameras, and provides deep learning capabilities to apply computer vision algorithms
- Rockchip RK3588 is a low-power, octa-core processor for Internet of Things (IoT) devices with Artificial Intelligence (AI)
- 4x ARM Cortex-A76 and 4x ARM Cortex-A55 cores
- Mali G610 GPU
- NPU with up to 6 TOPs.

[Datasheet](#)

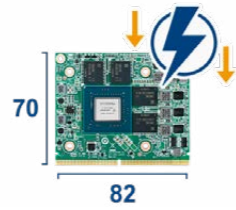


GPU CARDS, AI ACCELERATORS & AI MODULES

NVIDIA MXM GPU Cards



As an NVIDIA Elite Partner, Advantech offers MXM GPU cards powered by NVIDIA's embedded GPUs, perfect for image processing and Edge AI acceleration in manufacturing, transportation, and medical industries. Built on NVIDIA's latest architectures, Advantech MXM GPUs deliver state-of-the-art technologies giving high-performing computing and responsive capabilities for systems, as well as reducing latency, which is ideal for deploying in applications such as auto-optical inspection, driver-assistance, and surgical systems.



NVIDIA Certified Embedded MXM GPU Cards with High Performance

- Designed and manufactured by Advantech - NVIDIA's Visualization Elite partner
- MXM GPU cards are built on NVIDIA's latest Ampere and Turing architectures
- 2nd gen. RT core and 3rd gen. Tensor core accelerate ray tracing and AI inference
- Supports GDDR6 graphics memory for boosting data transfer rates + performance
- Supports 70+ professional software certified by NVIDIA

Easily integrated into diverse systems with MXM GPU cards

- High-performance, compact design, and low power consumption with Advantech strict quality control
- Fits in both fan and fanless systems with customized thermal solutions for wider operating temperatures (0°C to 55 °C) aimed at various applications + scenarios
- Ensures customization and design flexibility for realizing Edge AI inferencing with high computing performance within system constraints

Ind. Design with Longevity Supports

- MXM GPU cards feature over 5-year longevity extended product life support (better than 1-year consumer-level GPU cards)
- Advantech provides strict revision management and engineering change control to reduce re-validation and compatibility issues
- Advantech offers global and local services with 24/7 hotline AE, and global RMA systems for real-time expert advisors

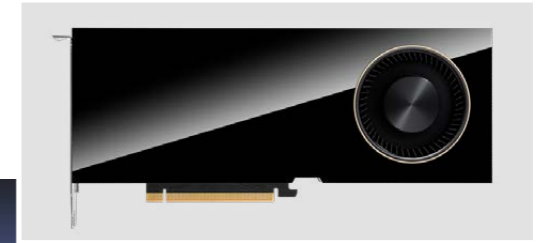
Built on NVIDIA embedded GPUs, MXM series are high performance, compact, low power GPU cards, with over 5-year longevity.

Part Number	Description	Form Factor	Performance	More Info
SKY-MXM-5000A	NVIDIA® Quadro® 5000 Ada	1) MXM 3.1 TYPE B+	Up to 9728 CUDA cores, 76 RT cores and 304 Tensor cores, 42.6 TFLOPS	Datasheet
SKY-MXM-3500A	NVIDIA® Quadro® 3500 Ada	1) MXM 3.1 TYPE B+	Up to 5120 CUDA cores, 40 RT cores and 160 Tensor cores, 23.04 TFLOPS	Datasheet
SKY-MXM-2000A	NVIDIA Ada-Lovelace architecture Quadro® 2000A	2) MXM 3.1 TYPE A	Up to 3072 CUDA cores, 24 RT cores and 96 Tensor cores, 12.99 TFLOPS	Datasheet
SKY-MXM-A4500	NVIDIA® Quadro® A4500	1) MXM 3.1 TYPE B+	Up to 5888 CUDA cores, 46 RT cores and 184 Tensor cores, 17.66 TFLOPS	Datasheet
SKY-MXM-A2000	NVIDIA® Quadro® A2000	2) MXM 3.1 TYPE A	Up to 2560 CUDA cores, 20 RT cores and 80 Tensor cores, 8.64 TFLOPS	Datasheet
SKY-MXM-A1000	NVIDIA® Quadro® A1000	2) MXM 3.1 TYPE A	Up to 2048 CUDA cores, 16 RT cores and 64 Tensor cores, 6.66 TFLOPS	Datasheet
SKY-MXM-A500	NVIDIA® Quadro® A500	2) MXM 3.1 TYPE A	Up to 2048 CUDA cores, 16 RT cores and 64 Tensor cores, 6.54 TFLOPS	Datasheet
SKY-MXM-RTX3000	NVIDIA® Quadro® RTX3000	1) MXM 3.1 TYPE B	Up to 1920 CUDA cores, 30 RT cores and 240 Tensor cores, 5.3 TFLOPS	Datasheet
SKY-MXM-T1000	NVIDIA® Quadro® T1000	2) MXM 3.1 TYPE A	Up to 896 CUDA cores, 2.6 TFLOPS	Datasheet

1) 82 x 105 mm 2) 82 x 70 mm

NVIDIA Quadro GPU Cards

Powerful visual computing GPU cards for professional and desktop workstations, are perfect for image processing and AI-accelerated computation.



As an NVIDIA Elite Partner, Advantech offers comprehensive Industrial-grade GPU Solutions, specifically focusing on NVIDIA professional graphics cards for enterprise use. As a result, Advantech is able to get early authorized access to the latest generation GPU cards and NVIDIA exclusive technical support for our customers.

NVIDIA professional RTX GPU cards are designed to accelerate professional and AI workflow, providing an unparalleled edge computing experience. These cards are highly preferred in manufacturing, transportation, and medical industries for image processing and Edge AI acceleration, making them the top choice in these sectors.

[Selection Guide](#)

NVIDIA Tesla GPU Cards

These engines of data centers can power and accelerate the highest-performing data centers for AI, data analytics, and HPC.

Part Number	Description	Performance	More Info
SKY-TESL-L40S-48P	NVIDIA Ada-Lovelace GPU architecture; Universal Compute & Graphics GPU	18176 NVIDIA® CUDA® Cores; 568 NVIDIA® Tensor Cores; 142 NVIDIA® RT Cores	Datasheet
SKY-TESL-L40-48P	NVIDIA Ada-Lovelace GPU architecture; Universal Compute & Graphics GPU	18176 NVIDIA® CUDA® Cores; 568 NVIDIA® Tensor Cores; 142 NVIDIA® RT Cores	Datasheet
SKY-TESL-L4-24P	NVIDIA Ada-Lovelace GPU architecture; Universal Compute & Graphics GPU	7680 NVIDIA® CUDA® Cores; 240 NVIDIA® Tensor Cores; 60 NVIDIA® RT Cores	Datasheet
SKY-TESL-H100N-94P	NVIDIA Hopper GPU architecture; Compute-optimized GPU	14592 NVIDIA® CUDA® Cores; 456 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-H100-80P	NVIDIA Hopper GPU architecture; Compute-optimized GPU	14592 NVIDIA® CUDA® Cores; 456 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-H800-80P	NVIDIA Hopper GPU architecture; Compute-optimized GPU	14592 NVIDIA® CUDA® Cores; 456 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-A100-80P	NVIDIA Ampere GPU architecture; Compute-optimized GPU	6912 NVIDIA® CUDA® Cores; 432 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-A800-80P	NVIDIA Ampere GPU architecture; Compute-optimized GPU	6912 NVIDIA® CUDA® Cores; 432 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-A40-48P	NVIDIA Ampere GPU architecture; Compute + graphics GPU	10752 NVIDIA® CUDA® Cores; 336 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-A30-24P	NVIDIA Ampere GPU architecture; Compute-optimized GPU	3584 NVIDIA® CUDA® Cores; 224 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-A10-24P	NVIDIA Ampere GPU architecture; Compute + graphics GPU	9216 NVIDIA® CUDA® Cores; 288 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-A2-16P	NVIDIA Ampere GPU architecture; Compute + graphics GPU	1280 NVIDIA® CUDA® Cores; 40 NVIDIA® Tensor Cores	Datasheet
SKY-TESL-T4-16P	NVIDIA Turing GPU architecture; Universal GPU	2560 NVIDIA® CUDA® Cores; 320 NVIDIA® Tensor Cores	



SKY-TESL-L4-24P



SKY-TESL-A40-48P



SKY-TESL-A40-48P



Advantech Embedded GPU Cards

EAI-3101



PCIe x16 Intel Arc A380E Embedded GPU Card with 4x DP 1.4a

- Intel Arc A380E with 2000 MHz Base Clock, GDDR6 6GB 96 bit Memory, up to 5.018 TFLOPS
- 4x DP1.4a up to 8K UHD resolution
- 0-60°C working temperature support w/ auto smart fan
- Advantech Edge AI SDK and Intel OpenVINO support
- AV1, DirectX12, OpenGL 4.6, OpenCL 3.0, H.264/H.265 (HEVC) hardware encode/decode support

[Datasheet](#)

EAI-2100



MXM 3.1 Type A Intel Arc A370M Embedded GPU Card with DP 1.4a

- Intel Arc A370M with 1550 MHz Base Clock, GDDR6 4GB 64 bit Memory, up to 4.198 TFLOPS
- TGP 40W Low consumption with optional thermal kit, support 0-60°C working temperature
- Reserved DP 1.4a x4 displays up to 8K UHD resolution
- Advantech Edge AI SDK and Intel OpenVINO support
- DirectX12, OpenGL 4.6, OpenCL 3.0, H.264/H.265 (HEVC)/AV1 hardware encode/decode support

[Datasheet](#)

EAI-3100



PCIe x16 Intel Arc A370M Embedded GPU Card with 2x HDMI 2.0b + 2x DP 1.4a

- Intel Arc A370M with 1550 MHz Base Clock, GDDR6 4GB 64 bit Memory, up to 4.198 TFLOPS
- High Resolution 8K DP 1.4a x2 + 4K HDMI 2.0b x2, up to 4 Displays
- 0-60°C working temperature support w/ auto smart fan
- Advantech Edge AI SDK and Intel OpenVINO support
- DirectX12, OpenGL 4.6, OpenCL 3.0, AV1/H.264/H.265 (HEVC) hardware encode/decode support

[Datasheet](#)



Powered by



AAEON GPU Cards / MXM Modules

GAR-A750E

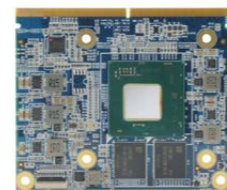


Advanced graphics card with Intel® Arc™ A750E GPU architecture

- High-Performance Graphics: 2400MHz turbo frequency, 28 Xe- Cores, and 28 Ray Tracing Units
- 16GB GDDR6 Memory: 256-bit interface, 16 Gbps memory speed.
- Broad Compatibility: PCIe 4.0 x16 interface for versatile motherboard compatibility.
- Quad Display Support: Simultaneous display outputs up to 7680 x 4320 @ 60Hz.
- Video Encoding/Decoding: H.264, H.265 (HEVC), AV1, and VP9 hardware support.
- Advanced Technologies: Ray tracing, variable rate shading (VSR), and DirectX 12 Ultimate.
- Intel Deep Link Integration: Benefit from Intel Deep Link features for enhanced AI and media processing

[Datasheet](#)

MXM-ACMA



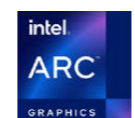
Embedded MXM 3.1 Type A Module with Intel® Arc™ A370M/A350 GPU

- Intel® Arc™ A370M/A350M Embedded GPU, up to 8 Xe-cores
- 4GB GDDR6 Memory
- PCIe 4.0 [x8] Interface
- DP++ 1.4/HDMI 2.1 x 4
- Standard MXM 3.1 Type A form factor
- OpenCL™/OpenGL®/oneAPI/DirectX12/Ray Tracing Support

[Datasheet](#)



Powered by



ASUS AI Accelerator Card

ASUS IoT AI accelerator CRL-G18U-P3DF is a full height, half length, PCIe Gen3 AI acceleration card based on Google Coral Edge TPU processor. It allows users to build intelligent devices that are efficient, private and ideal for edge AI inference.



Powered by



Powerful AI Inference Capability

- Support up to 8 x Google Edge TPU M.2 modules

Easy-to-Use Pre-trained AI Models

- Google TensorFlow Lite pre-trained ML models can be easily compiled and run on this model

Easy Installation, Common Expansion Slot

- Compatible general PCI Express Gen 3 x16 slot

Stable At High-Loading

- Optimized thermal design with high quality Copper heatsink and twin turbofans

[Datasheet](#)

Type	CRL-G18U-P3DF	CRL-G116U-P3DF
Chipset	8x Google® Coral Edge TPU Processor	16x Google® Coral Edge TPU Processor
Interface	PCIe 3.0 x16 (M.2)	PCIe 3.0 x16 (M.2)
Form Factor	Full-height, half-length, double-slot width	Full-height, half-length, double-slot width
Dimensions	42 x 126 x 186 mm	42 x 126 x 186 mm
Operating Temp.	0 to +55 °C	0 to +55 °C
Power Consumption	36 W	52 W
Power Connector	1 x 6-pin 12V External Power	1 x 6-pin 12V External Power
Technology	VLIW	VLIW
Cooling Solution	Active Fan	Active Fan
Open Vino Support	no (TensorFlow Lite)	no (TensorFlow Lite)

iEi AI Acceleration Products

TPU Accelerator Card



Powered by

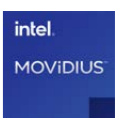


[Datasheet](#)

Type	MUSTANG-T100-T5
Chipset	5x Google® Coral Edge TPU Processor
Peak Performance	20 TOPS
Interface / Form Factor	PCIe 2.0 x4
Form Factor	HHHL, Single-slot PCIe
Operating Temperature	-20 to +55 °C
Power Consumption	~15 W
Cooling Solution	Active fan
Open Vino Sup.	no (Tensorflow Lite)



Powered by



VPU Accelerator Cards

Vision Processing Units can run AI operations faster and more efficient. With low power consumption and high performance to dedicated DNN topologies, VPUs fit perfectly for the implementation in AI edge computing devices. This is well suited for applications such as surveillance, retail, and transportation. VPU Accelerator Cards provide a flexible AI inference solution for compact size and embedded systems.



[Datasheet](#)

Type	MUSTANG-M2BM-MX2
Chipset	DL Inference Accelerator Card
Interface / Form Factor	2x Intel® Movidius™ Myriad™ X MA2485 VPU
Dimensions	M.2 2280 (B+M Key)
Operating Temperature	22 x 80 mm
Operating Temperature	-20 to +60 °C
Power Consumption	~7.5 W
Power Connector	-
Technology	VLIW
Cooling Solution	Active Heatsink
Open Vino Sup.	yes

Advantech AI Modules

ROM-2860
OSM 1.1 size L form factor



- Qualcomm QCS6490 high-level platform (1 Kryo Gold plus up to 2.7 GHz; 3 Kryo Gold at 2.4 GHz; 4 Kryo Sliver at 1.9 GHz)
- Andreno VPU 633 4K30 encode/Decode
- Andreno GPU 643, OpenGL ES3.2/OpenCL 2.0
- 12 TOPs NPU for AI applications

[Datasheet](#)



Powered by
Qualcomm

AOM-7721
AI on Module



- Snapdragon X1 Elite 12 Cores, up to 3.4GHz
- 16GB LPDDR5x, 128GB UFS and 1x 4-bit SDIO for SD card
- Andreno Snapdragon X GPU supports DX12.2, OpenGL ES 3.2 & OpenCL 3.0
- Andreno VPU 5th Gen., 4K60P Codec with H.264/HEVC/VP9

[Datasheet](#)

Advantech AI Modules



Powered by
Rockchip

AOM-3411
High-Performance, Cost Efficient Edge AI module



- Rockchip RK3576
- Cortex-A72/53 4x/4x cores
- Offload CPU usage with a dedicated AI NPU, up to 6 TOPs
- Debian, Android



Advantech AI Modules

AOM-3511
High-Performance, Safety Enabled SMARC module with AI NPU built-in



Powered by
NXP

- NXP Arm® Cortex®-A55 i.MX95 Six Cores up to 2.0 GHz
- 1 x Arm Cortex-M7 core & 1 x Cortex-M33 core
- Neural network accelerator built-in, up to 3 TOPs
- Linux, Windows

[Datasheet](#)



Powered by
NXP

ROM-2820
Powerful SoC (OSM Size L) with Edge ML



- AI Enabler
- i.MX 93 Cortex-A55, 38% better performance than Cortex-A53 in floating points computing
- Ethos-U65 microNPU design for entry-level ML applications (0.5 TOPs)
- Ethos-U65 provides ~10X better performance than Cortex-A55

[Datasheet](#)

Advantech Edge AI Acceleration Modules



Powered by
intel
MOVIDIUS

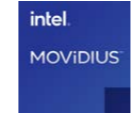
Type	VEGA-320-01A1	VEGA-330-02A1	VEGA-340-04A1	VEGA-340-08A1
Chipset	1x Intel® Movidius™ Myriad™ X; MA2485 VPU	2x Intel® Movidius™ Myriad™ X; MA2485 VPU	4x Intel® Movidius™ Myriad™ X; MA2485 VPU	8x Intel® Movidius™ Myriad™ X; MA2485 VPU
Interface / Form Factor	M.2 2230 (AE Key)	miniPCIe 2.0 x1	PCIe 2.0 x4	PCIe 2.0 x4
Dimensions	22 x 30 mm	30 x 51 mm	171 x 69 mm	171 x 69 mm
Operating Temperature	-20 to +60 °C	-20 to +55 °C	-20 to +60 °C	-20 to +60 °C
Power Consumption	3.8 W	7.6 W	6.7 W	9.6 W
Power Connector	-	-	PCIe 6pin	PCIe 6pin
Technology	VLIW	VLIW		
Cooling Solution	Passive Cooling	Passive Cooling	Active Fan	Active Fan
Open VINO Support	yes	yes	yes	yes
Datasheet	Link	Link	Link	Link







AAEON AI Edge Computing Modules



Powered by



AAEON's lineup of AI Modules brings power, scalability and flexibility to your Edge AI project. The AI Core X family, powered by Intel® Movidius™ Myriad™ X, offers a range of scalable solutions to give existing platforms an AI boost to perform at the edge. AAEON's range of AI modules is available in several form factors including mPCIe and M.2, allowing you to find AI modules that are the perfect fit.

Type	AI CORE X	AI CORE XM 2280	AI CORE XP4	AI CORE XP8
	PER-TAIC-A10-002	RS-PER-TAIX2-A10-N2280	PER-TAIX4-A10-PCIE	PER-TAIX8-A10-PCIE
				
Chipset	1x Intel® Movidius™ Myriad™ X MA2485 VPU	2x Intel® Movidius™ Myriad™ X MA2485 VPU	4x Intel® Movidius™ Myriad™ X via 2x AI Core XM 2280	8x Intel® Movidius™ Myriad™ X via 4x AI Core XM 2280
Interface / Form Factor	miniPCIe	M.2 2280 (B+M Key)	PCIe (Proprietary)	PCIe 2 x4 (HHL)
Dimensions	30 x 51 mm	22 x 80 mm		
Operating Temperature	0 to +60 °C	0 to +50 °C	0 to +50 °C	0 to +50 °C
Power Connector				PCIe 6pin
Datasheet	Link			
Technology	VLIW	VLIW	VLIW	VLIW
Cooling Solution	fanless	with fan	with fan	with fan
Open VINO Support	yes	yes	yes	yes

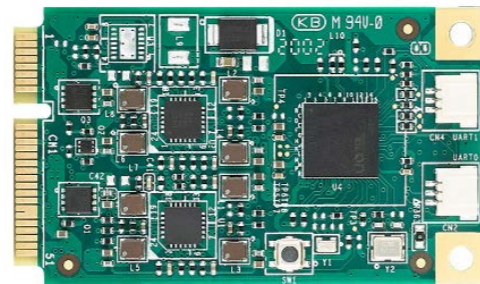
Advantech Edge AI Acceleration Modules



Powered by



Type	VEGA-230
Chipset	1x Kneron KL520 NPU
Interface / Form Factor	miniPCIe
Dimensions	30 x 51 mm
Operating Temperature	0 to 70 °C
Power Consumption	0.5 W
Power Connector	-
Cooling Solution	Passive Cooling
Open VINO Support	no (ONNX, TensorFlow, Keras, Caffe, PyTorch)
Datasheet	Link






AAEON AI Edge Computing Modules






Powered by



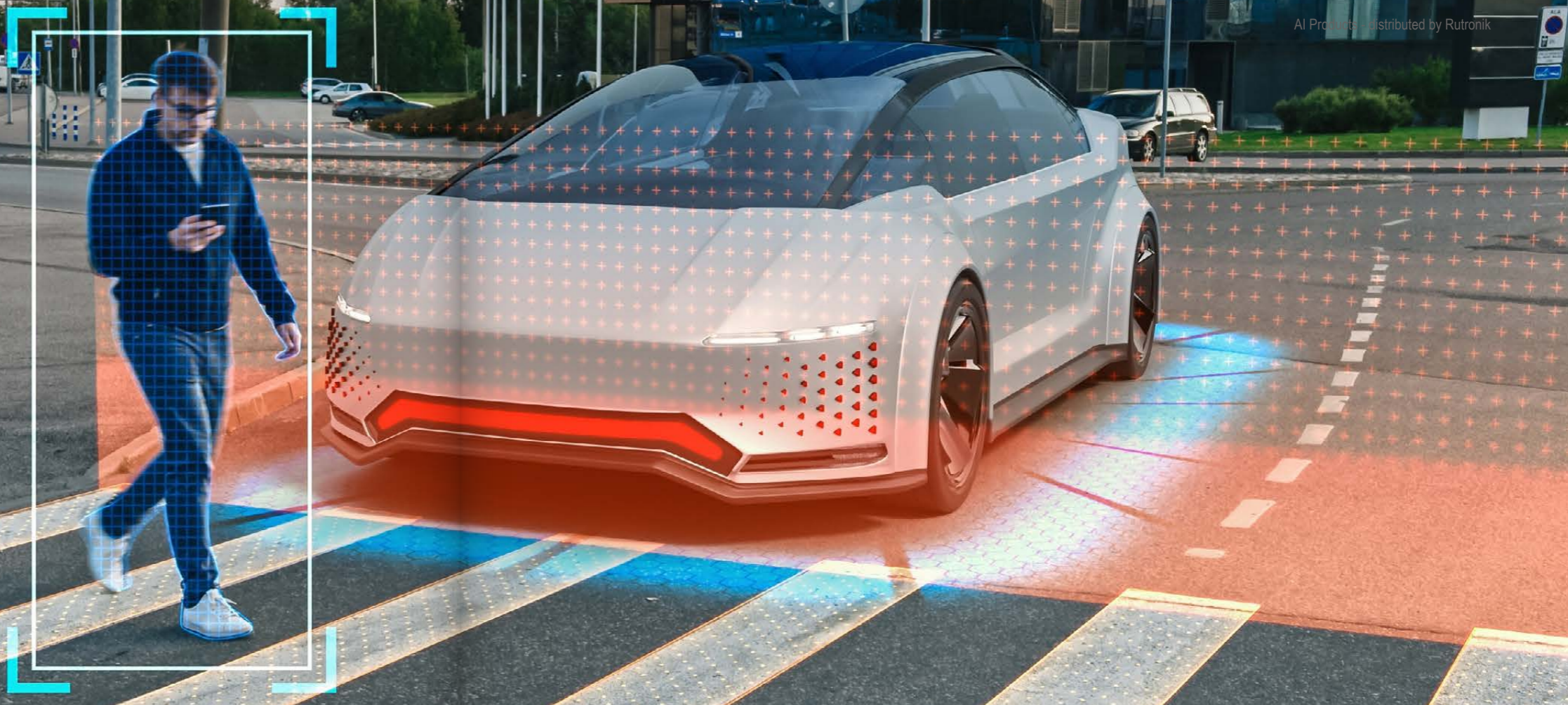
AAEON's lineup includes also Mini AI modules featuring the innovative and efficient Kneron® KL 720 and KL 520. AAEON's range of AI modules is available in several form factors including mPCIe and M.2, allowing you to find AI modules that are the perfect fit.

Type	Mini-AI-520	M2AI-2242-520	M2AI-2280-520
	PER-T520-MIAI-A11-0001	PER-T520-M2AI-A11-0421	PER-T520-M2AI-A11-0801
			
Chipset	1x Kneron KL520 NPU	1x Kneron KL520 NPU	1x Kneron KL520 NPU
Interface / Form Factor	miniPCIe	M.2 2242	M.2 2280 (B Key)
Dimensions	30 x 51 mm	22 x 42 mm	22 x 80 mm
Operating Temperature	0 to +70 °C	0 to +70 °C	0 to +70 °C
Power Consumption	0.5 W	0.5 W	0.5 W
NPU Performance	0.35 TOPS	0.35 TOPS	0.35 TOPS
More info	Datasheet	Datasheet	Datasheet
Cooling Solution	fanless	fanless	fanless
Open VINO Support	no (ONNX, TensorFlow, Keras, Caffe)	no (ONNX, TensorFlow, Keras, Caffe)	no (ONNX, TensorFlow, Keras, Caffe)

Type	Mini-AI-720	M2AI-2242-720	M2AI-2280-720
	PER-T720-MIAI-A10-0001	PER-T720-M2AI-A10-0421	PER-T520-M2AI-A11-0801
			
Chipset	1x Kneron KL720 NPU	1x Kneron KL720 NPU	2x Kneron KL720 NPU
Interface / Form Factor	miniPCIe	M.2 2242 (B-key)	M.2 2280 (B-M Key)
Dimensions	30 x 51 mm	22 x 42 mm	22 x 80 mm
Operating Temperature	0 to +70 °C	0 to +70 °C	0 to +70 °C
Power Consumption	5 W	2.3 W	5 W
NPU Performance	1.4 TOPS	1.4 TOPS	1.4 TOPS x2
More info	Datasheet	Datasheet	Datasheet
Cooling Solution	fanless	fanless	fanless
Open VINO Support	no (Pytorch, ONNX, TensorFlow 1.6, TensorFlow lite, Keras, Caffe)	no (Pytorch, ONNX, TensorFlow 1.6, TensorFlow lite, Keras, Caffe)	no (Pytorch, ONNX, TensorFlow 1.6, TensorFlow lite, Keras, Caffe)



AI VISION



Advantech AI Camera

AI camera embedded high computing power that can use AI programs to wisely deal with images and videos inside the camera.



ICAM-540



Industrial AI Camera

- 8MP 45 fps Sony ind. grade sensor
- C-mount lens compatible
- NVIDIA Jetson AI system on module
- HW ISP no GPU/CPU workload

[Datasheet](#)

ICAM-520



Industrial AI Camera

- Embedded with NVIDIA® Jetson Xavier™ NX
- 1.6MP 60 FPS, Sony ind. grade sensor
- Programmable variable focus lens
- Advanced LEDs illumination
- HW ISP no GPU/CPU workload

[Datasheet](#)

ICAM-500



Industrial AI Camera

- NVIDIA Jetson AI system on module
- 1.6MP 60 FPS, Sony ind. grade sensor
- Programmable variable focus lens
- Advanced LEDs illumination
- HW ISP no GPU/CPU workload

[Datasheet](#)

Advantech AI Machine Vision Devices

AI-enabled Machine Vision Empowers Real-time Smart Decision Making for Realization of AIoT.

Acquired production images are a common critical data type benefiting particularly from AI analytics. The capture, processing, and management process does, however, present numerous challenges, making it costly, complicated, and time consuming to implement.

A typical system comprises cameras connected to a frame grabber module in a host computer, with images acquired on the line and processed by specific vision software. Processing and management of the necessarily massive amount of captured data for machine learning requires a VPU/GPU integrated vision solution or a new heterogeneous platform.



AI Smart Camera



AI Vision System



Edge AI Vision Software



AI Camera Kits

MEMORY & STORAGE FOR AI



Swissbit – Featured Products for AI



M.2 PCIe SSD

PCIe M.2 modules for Edge AI



Product Fact Sheets

- N3000
- N3002
- N3202
- N2000
- N2600

Enterprise 2.5" U.2 PCIe SSD

PCIe U.2 SSD for Data Lakes



Product Fact Sheets

- N5200 U.2
- D2200 U.2

Enterprise E1.S PCIe SSD

PCIe E1.S SSD for Data Lakes

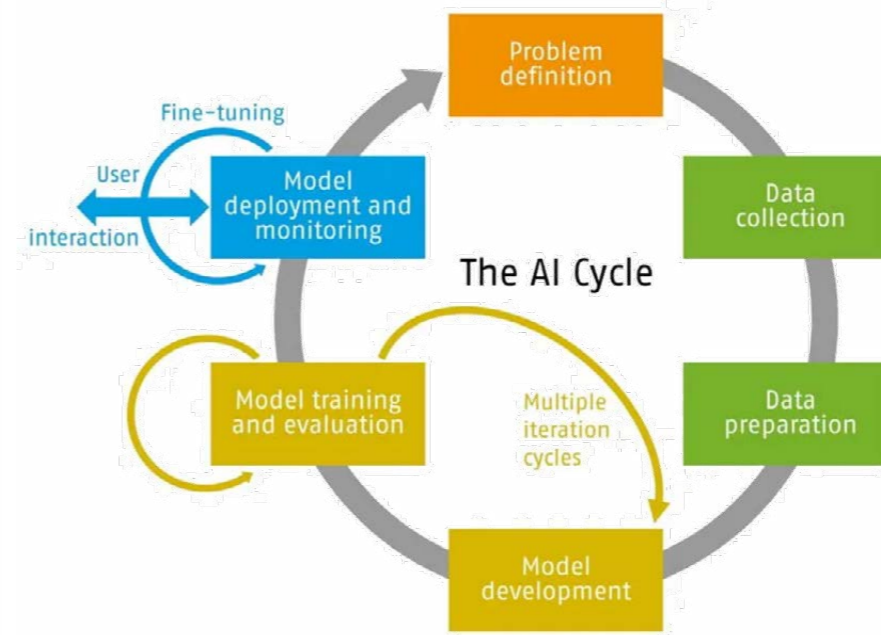


Product Fact Sheets

- N5200 E1.S
- D2200 E1.S

The Stages of the AI Cycle

Artificial intelligence is the approach to implement human like intelligence to computing systems to allow them trained decision making or generation of content. The so-called AI cycle encompasses the various stages of creating, deploying, and maintaining artificial intelligence systems. Here is a simplified representation of the different stages. This cycle is iterative, as models often need to be updated and refined based on new data, changing requirements, or evolving business needs.



Problem definition

Defining the problem and expected solution is crucial as it influences the model. Different AI types (generative, analytic, administrative) require distinct model architectures and learning processes.

Data collection

Training AI necessitates a large collection of high-quality example data, stored on large capacity SSDs like Swissbit Enterprise SSDs, in data lakes.

Data preparation

Collected data must be preprocessed to enhance or reduce it to the necessary details for model training. High-speed SSDs and high-capacity DRAM are essential for this phase.

Model development

Key aspects include coverage, response time, size, accuracy, and flexibility. Model requirements determine the platform, especially for edge implementations where capabilities must match.

Model training and evaluation

The model learns by comparing its results with correct answers, through automated or human training. High-performance AI systems with GPUs, TCUs, HMB memory, and fast SSDs are necessary. This stage may involve multiple cycles until the final model is achieved.

Model deployment and monitoring

The final model often runs on a different platform from where data collection or training occurred, usually on edge servers. User interaction can help improve the model. Swissbit offers edge server SSDs with robust environmental endurance, low power consumption, and compact form factors.



Transcend Products for Edge AI



Efficiently amplify smart deployment

Transcend's industrial SSDs deliver high-speed, reliable performance in the ever-evolving landscape of artificial intelligence. Specialized options, including Power Loss Protection (PLP), TCG Opal, and SLC Mode technologies, address common issues such as unstable power, data security, and intensive operations – making them ideal for demanding embedded computing applications.

- Robust Performance:**
 Advanced BiCS5 SSDs offer substantial capacities, enabling storage & 24/7 system operation.
- Fortified Durability:**
 Extra-thick 30µ" PCB gold fingers & anti-sulfur technology enhance endurance.
- Comprehensive Product Line:**
 Various form factors & interfaces for different systems and applications.



Transcend Products for AI Server



A Game Changer In AI Server Application

The rise of AI and machine learning is creating substantial demands for data storage. Transcend's cutting-edge storage solutions, BiCS5 SSDs and DDR5 DRAM modules, offer reduced latency, high-speed transfer rates, and ample storage capacity (up to 8TB), helping AI servers handle expansive data sets.

- Rapid Data Retrieval:**
 Fast data access improves system responsiveness, enhancing AI algorithm efficiency.
- Utmost Reliability For 24/7 Operation:**
 100% in-house tested and branded chips used for high storage resiliency.
- Valued-Added Technologies:** SLC Mode, PLP, and TCG Opal 2.0 compliance ensure SSDs' durability, security, and reliability in critical conditions.



Transcend Products for AI Vision & Recognition



Precision Storage for Cutting-Edge AI Vision

As AI technology revolutionizes image recognition, enhancing quality precision and processing speeds, Transcend's SSDs and memory cards stand out by guaranteeing high-performance speeds and ample storage for large volumes of high-resolution images. Endurance is crucial for write-intensive equipment operating non-stop 24/7. That's why Transcend's SSDs and memory cards undergo rigorous testing to ensure utmost reliability. Trust Transcend for your AI-driven image recognition needs and experience unparalleled performance and dependability.

- Direct Write:**
 Sustained write performance ensures storage integrity of high-resolution images.
- SLC Mode:**
 Extended flash lifespan and high durability for use in write-intensive equipment.
- Advanced Security:**
 TCG Opal encryption ensures critical information is well protected.



Apacer Products for AI Server



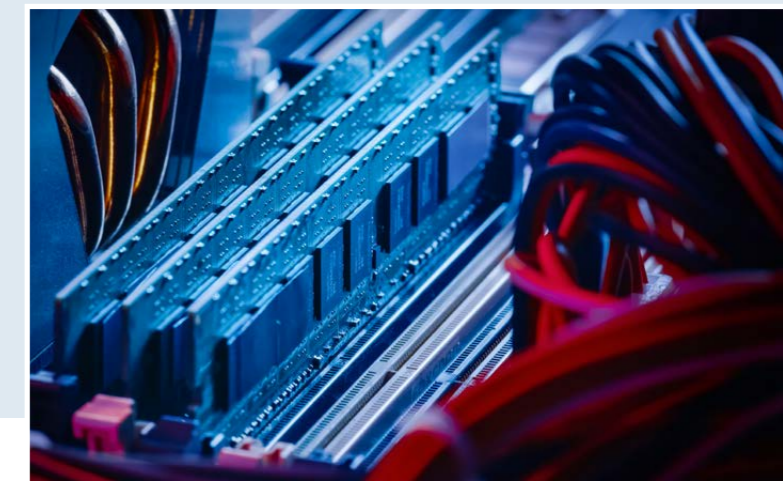
DDR5 for AI servers

In response to the recent wave of AI development, CAMM2, DDR5-6400 UDIMM and SODIMM DRAM modules specially designed for AI PCs have been launched. For the high computing requirements of AI servers, DDR5 MRDIMM and DDR5 CXL DRAM modules will be displayed. At the same time, Apacer's partners Phison (8299) and DeepMentor will present the Mentor-100 AI server equipped with Phison's patented aiDAPTIV+ solution. This collaboration will accelerate the implementation of AI applications and contribute to the growth of the worldwide product ecosystem.

In response to the data backup and rescue issues that may be required by AI applications, Apacer has developed the CoreSnapshot Series and CoreRescue technology. When a blue screen of death occurs in a networked device, the system can be quickly restarted even if maintenance personnel are not on site. As for the AI needs of edge computing, Apacer launched 64GB DDR5 UDIMM and SO-DIMM high-density DRAM modules, as well as DRAM modules that can withstand extreme temperatures and sulfur-contaminated environments. Demonstrations will also include rugged DRAM module retention straps suitable for high-vibration applications.

Storage for AI servers

AI will drive the storage demand for enterprise-class servers in the future. Apacer has four major competitive advantages: large capacities up to 30TB, high performance, low latency, and high reliability. These are coupled with the additional functions of data encryption and power outage protection. Other products showcased this year will include the world's first single-chip controller solution, the USB4 external SSD, a Gen5 x4 4TB SSD with read/write performance up to 14,000/12,000 MB/s, and the military-standard drop-certified AS723 external SSD.



LPDDR5X for modern AI applications

Low Power Double Data Rate 5X (LPDDR5X)

LPDDR5X plays a significant role in modern AI applications, especially in mobile and edge computing environments. LPDDR5X is known for its low power consumption, making it ideal for mobile devices and edge computing where energy efficiency is critical. LPDDR5X offers data rates of up to 8,533 Mbps, which significantly improves the processing speed and efficiency of AI algorithms. LPDDR5X can be supplied in packages of up to 64 GB capacity, making it suitable for applications that need to process large amounts of data. The technology offers 25% higher performance and 30% higher capacity compared to the previous generation.

Main differences between LPDDR4X and LPDDR5X

- Data rate and bandwidth:**
 - LPDDR4X offers data rates of up to 4,266 Mbps
 - LPDDR5X reaches data rates of up to 8,533 Mbps, which is a significant increase in bandwidth
- Energy efficiency:**
 - LPDDR4X operates at a voltage of 1.1 V and offers good energy efficiency
 - LPDDR5X also operates at a voltage of 1.1 V, but offers even better energy efficiency thanks to more advanced technologies such as Dynamic Voltage Scaling (DVS)
- Performance:**
 - LPDDR4X provides solid performance for mobile and edge computing applications
 - LPDDR5X offers 33% higher performance compared to LPDDR5, making it ideal for demanding AI and ML applications

Manufacturers for LPDDR5



Manufacturers for LPDDR4



ESMT / EliteMT



Nanya LPDDR5X for modern AI applications

LPDDR5/5X									
Density	Org	Voltage	Package	Speed	Availability	IT	Automotive		
							G1	G2	G3
8Gb	x16	1.05V	315-FBGA SDP	7500/8533	●	●	-	●	●
16Gb	x16	1.05V	315-FBGA SDP	7500/8533	●	●	-	●	●
	x32		315-FBGA DDP						
32Gb	x32	1.05V	315-FBGA DDP	7500/8533	●	●	-	●	●
64Gb	x32	1.05V	315-FBGA QDP	7500/8533	●	●	-	●	●

LPDDR4X									
Density	Org	Voltage	Package	Speed	Availability	IT	Automotive		
							G1	G2	G3
2Gb	x16	1.1V (VDDQ=0.6V)	200-FBGA SDP	3733/4267	Now	-	-	Now	Now
4Gb	x16	1.1V (VDDQ=0.6V)	200-FBGA SDP	3733/4267	Now	Now	-	Now	Now
	x32		200-FBGA DDP						
8Gb	x16	1.1V (VDDQ=0.6V)	200-FBGA SDP	3733/4267	Now	Now	-	Now	Now
	x32		200-FBGA DDP						
16Gb	x32	1.1V (VDDQ=0.6V)	200-FBGA DDP	3733/4267	Now	Now	Now	Now	Now
	x64		376-PoP QDP						
32Gb	x32	1.1V (VDDQ=0.6V)	200-FBGA QDP	3733/4267	Now	Now	●	Now	Now
	x64		376-PoP QDP						
64Gb	x32	1.1V (VDDQ=0.6V)	200-FBGA QDP	4267	●	●	-	●	●
	x64		376-PoP QDP						



ASUS AISDetector



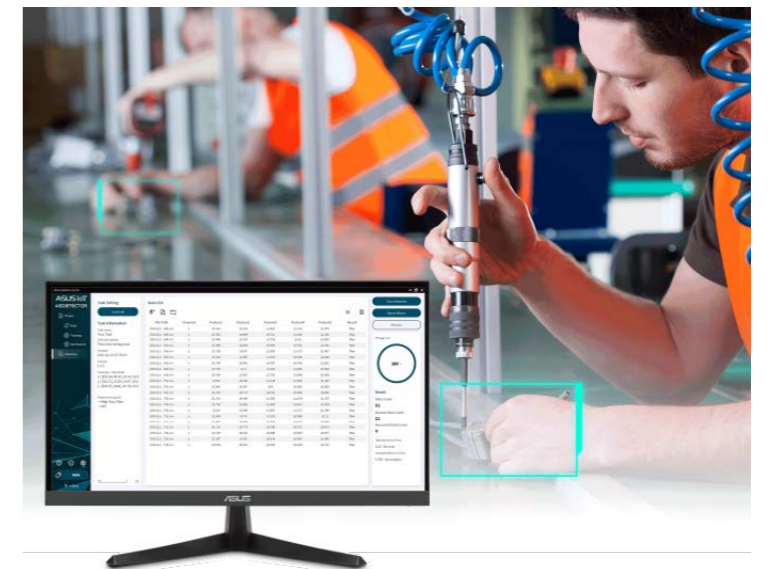
AI-based application software for product-quality inspection during production phase

ASUS IoT AISDetector is an on-premise AI signal-analysis and anomaly-detection tool empowered by AI technology developed by ASUS IoT. AISDetector facilitates the process by providing both training and inference working models, thereby simplifying AI-model development and management. Developers can utilize AISDetector for building specialized AI models, tracking training progress information, performing model training, model validation and conducting anomaly-detection testing.

AISDetector provides a user-friendly interface for model training and validation, and then leverages resulting models to test for defective products. The system includes a versatile web API to enable developers to draw signal-data inferences derived from a variety of sensors, including vibrations, sound and current – all powered by a unique and ever-evolving ASUS IoT AI algorithm.

- High-precision diverse algorithms
- Zero-code model generation
- Model validator: Adjust model sensitivity
- Model predictor: Predict with test data
- Windows 10 and Windows 11 64-bit editions supported
- English, traditional Chinese and simplified Chinese languages supported
- Notification and software-upgrade mechanism

[Datasheet](#)



ASUS AISVision



An easy-to-use AI toolkit and SDK for computer vision, suitable for model training and inference.

ASUS AISVision is an easy-to-use toolkit that uses ASUS AI techniques for the application of computer vision. AISVision offers a highly flexible training and inference architecture, with three working modes - Trainer, Scheduler and Runtime - for AI training, batch training and inference. This toolkit is designed to simplify the complexity of AI knowledge so that developers can create their own AI models in minutes, using proprietary ASUS AI techniques. Developers can also use the toolkit to retrain the model with new data input, and keep operational knowledge within their own work domain.

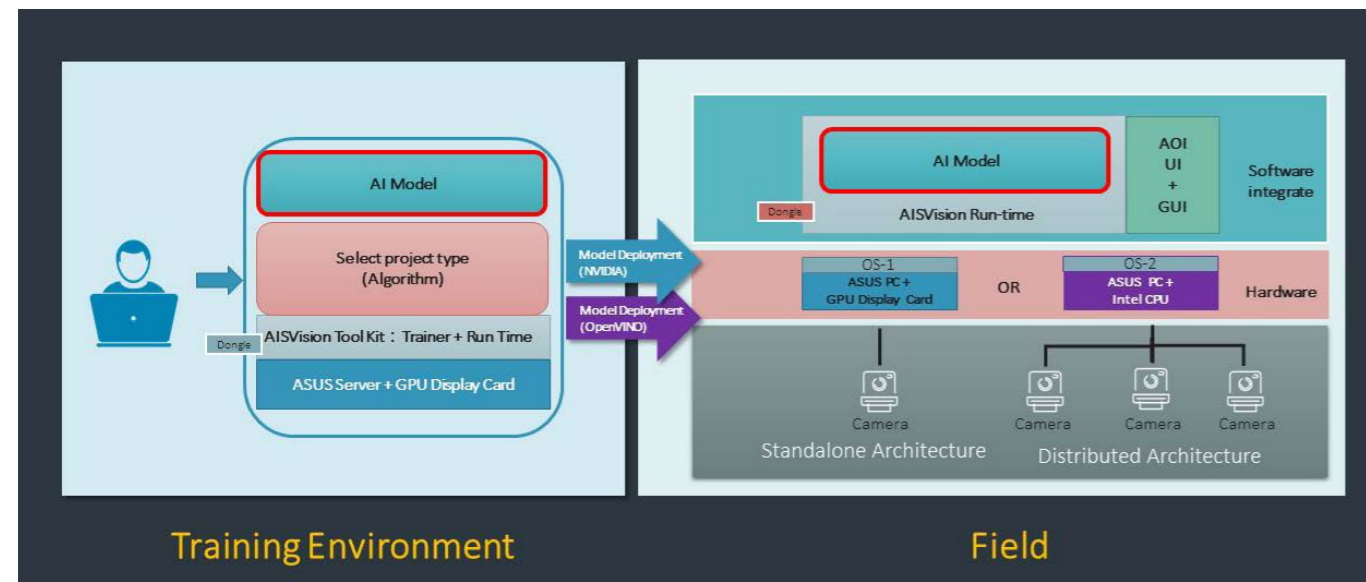
As well as the Training mode, the toolkit also includes Runtime mode as an inference engine that supports C, C# and C++ APIs to integrate existing graphical interfaces. They can also retrieve in-line data for analysis and export it for further analytics, visualization, database management, and other related tasks in the edge or cloud. As training can consume a large amount of time, the Scheduler mode can train pre-defined tasks in batches for higher efficiency.



- Zero-code machine-learning toolkit
- ASUS UX design: intuitive, simple, and quick project setup in four steps
- Diverse models (object detection, classification, segmentation, and anomaly detection) for vertical applications, and integration of various intuitive labeling tools
- Unique AI framework supporting Intel® OpenVINO architecture for inference
- A report wizard for sample filtering, with unique model training and validation
- Optimized performance for handling simultaneous training projects

- Language support for Traditional Chinese, Simple Chinese and English
- OS Support for 64-bit edition of Windows 10 and Windows 11 Professional
- API support for C, C++ and C# to ease integration into existing programs

[Datasheet](#)



ASUS AISPHEM



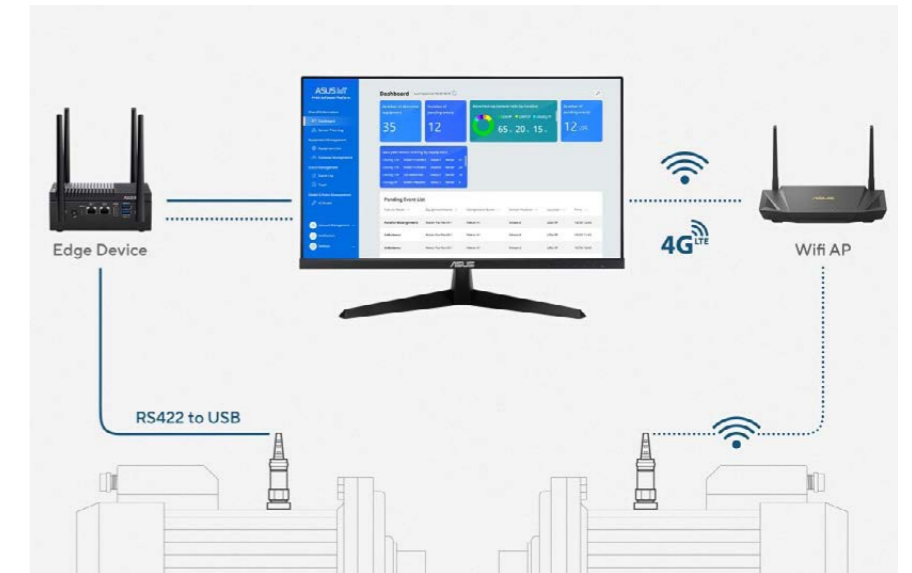
Advanced vibration-analysis solution empowers predictive maintenance and health management.

AASUS IoT AISPHEM is a revolutionary solution for predictive maintenance and health management. Leveraging real-time AI and state-of-the-art vibration analysis, AISPHEM is specially designed to detect early-stage issues in rotating equipment, ensuring unparalleled reliability and performance in the most demanding factory environments — delivering immediate, intuitive alerts and the adaptability of a flexible architecture for optimized operational efficacy, either on-premise or in the cloud.

This innovative approach empowers industries to anticipate and swiftly resolve issues, reducing downtime and maintaining uninterrupted production flow. AISPHEM stands out by preserving the health and longevity of industrial equipment, delivering continuous advancements, and providing ultimate peace of mind.

- ISO 10816-3 compliant and advanced frequency analysis
- Real-time AI analysis and on-site motor modeling
- EdgeX development framework accelerates secondary data utilization
- Intuitive notifications for motor-vibration anomalies
- Continuous software enhancements and updates
- Flexible architecture adjustment for on-premise and cloud

[Datasheet](#)





Intel Tiber Developer Cloud

Accelerate AI development using Intel®-optimized software on the latest Intel® Xeon® processors and GPU compute. With over 4,000 accelerators on the Intel Tiber Developer Cloud, customers and developers can experience first-hand how their favorite generative AI models run. What it's like to build new or migrate existing models on Intel Gaudi 2 accelerator.



Available Software

Intel® Toolkits and Libraries

- Intel® oneAPI Base Toolkit
- Intel® Rendering Toolkit
- Intel® HPC Toolkit
- AI Tools
- Intel® Quantum SDK

AI Frameworks and Tooling

- Intel® Distribution of OpenVINO™ Toolkit
- Intel® Optimization for PyTorch*
- Intel® Optimization for TensorFlow*
- SynapseAI* Software Suite
- Intel® Extension for Scikit-learn*
- Intel® Extension for DeepSpeed*
- Intel® oneAPI Collective Communications Library (oneCCL)

AI Foundation Models

- Technology Innovation Institute* Falcon LLM
- MosaicML* MPT
- HuggingFace* Bloom
- Stability.AI Stable Diffusion*
- Meta AI* Llama 2
- Databricks* Dolly

Service Tier: Standard - Free





Explore + evaluate the latest Intel AI products plus:

- Develop AI skills.
- cutting-edge learning resources.
- Get support from the Intel community.

Intel Tiber Edge Platform

Intel® Tiber™ Edge Platform streamlines the process of developing and deploying scalable edge AI solutions while giving you powerful tools to intelligently orchestrate applications and workloads. Overcome the complexity of heterogeneous edge deployments and achieve unprecedented speed and scalability with your edge applications. Intel® Tiber™ Edge Platform offers cloud-like simplicity from end to end, including deploying edge AI models and applications, provisioning resources, and configuring security.

The Tools and Resources to Scale at the Edge

 <p>Streamline End-to-End AI and Software Development</p> <p>Intel® Tiber™ Edge Platform is open and flexible, helping streamline application development. Start sooner by uploading your existing containerized apps, AI models, datasets, and solutions, or start building from scratch. You can also optimize existing apps for the edge – even if they're cloud native.</p>	 <p>Simplify Edge Application Deployment & Management</p> <p>Reduce the time required to protect and manage edge applications with streamlined, zero-trust security configuration and application lifecycle support. Deep hardware-aware telemetry enables dynamic workload deployment and orchestration.</p>	 <p>Accelerate Progress with Software Tools and Support</p> <p>Intel® Tiber™ Edge Platform offers a set of software building blocks, toolkits, pretrained AI models, and reference solutions to help deliver quick progress on your AI and application development journey.</p>	 <p>Optimize Edge AI Performance</p> <p>Run AI inference more efficiently in more places with simplified model deployment and built-in OpenVINO™ AI runtime optimizations. Intel® Tiber™ Edge Platform also supports a range of hardware architectures, model repositories, and edge AI model training platforms.</p>
---	---	---	---

Intel Geti Software

Intel's new software for building computer vision models in a fraction of the time and with less data. This software eases laborious data labeling, model training and optimization tasks across the AI model development process, empowering teams to produce custom AI models at scale.



Scaling Edge AI Solutions with Intel

Intel Geti software is part of the Intel® Tiber™ Edge Platform, a solution designed to solve edge challenges across industries. The platform enables enterprises to develop, deploy, run, manage, and scale edge applications with cloud-like simplicity, leveraging an unmatched partner ecosystem.

<p>Interactive Model Training</p> <p>Get started annotating data with as little as 20-30 images; then let active learning help teach the model as it learns.</p>	<p>Multiple Computer Vision Tasks</p> <p>Create models for AI tasks including classification, object detection, semantic segmentation or anomaly detection.</p>	<p>Task Chaining</p> <p>Train your model into a multistep, smart application by chaining two or more tasks, without the need to write additional code.</p>	<p>Smart Annotations</p> <p>Expedite data annotation and easily segment images with professional drawing features like a pencil, polygon tool and OpenCV GrabCut.</p>
<p>Production-Ready Models</p> <p>Output deep learning models in TensorFlow or PyTorch formats (where available) or as an optimized model for OpenVINO™ to run on Intel® architecture CPUs, GPUs and VPUs.</p>	<p>Hyperparameter Optimization</p> <p>Refining hyperparameters is critical to the model's learning process. With built-in optimization, Intel Geti software makes a data scientist's job easier.</p>	<p>Rotated Bounding Boxes</p> <p>Support for rotated bounding boxes extends the training simplicity and accuracy to datasets with images that are not axis-aligned..</p>	<p>Model Evaluation</p> <p>Comprehensive statistics to assess your model's performance.</p>





Protecting AI with SecEdge & Advantech



AI Model Protection

AI models represent a large investment in intellectual property that must be protected from corruption or theft. Protecting AI at the edge presents a unique set of challenges. SEC-TPM protects AI models at rest and maintains AI model integrity with secure remote management.

Device Integrity

Device integrity is ensured with trust provisioning and management of SEC-TPM, eliminating the need of a discrete TPM chip or secure manufacturing facility.

Multi-Tenant Secure Device Access

The solution enables secure connectivity across applications with independent IPsec tunnels. For example, a traffic-monitoring application having several tenants could include: a camera vendor OEM (for camera

software updates), an AI model developer (for model management), and the city (for secure access to camera feeds and model inferences).

SEC-TPM is integrated with NVIDIA JetPack 6.0 including microservices for Jetson, providing a complete set of device security functions and features, including secure boot and firmware updates, failure recovery, encryption, credentials management, and cloud integration.

"SecEdge has a unique solution that provides AI model protection from chip-to-cloud," said Magic Pao, Associate Vice President

of the Industrial Cloud & Video Group at Advantech, "Our industrial customers and partners can develop and deploy innovative edge AI systems knowing that their solution and intellectual property are safe."

"SEC-TPM is a perfect fit for Advantech, a trusted provider of innovative products and solutions," said Sami Nassar, President and Co-CEO at SecEdge. "We are excited about this first step in an exciting partnership in securing edge AI."

Related Products

Part Number	Description	Based on	Performance	More Info
MIC-710AI	AI Inference System	NVIDIA Jetson NANO	up to 472 GOPS	Datasheet
MIC-733-AO	AI System	NVIDIA Jetson AGX Orin	up to 275 TOPS	Datasheet
MIC-730IVA	8ch AI Video System	NVIDIA Jetson AGX Xavier	up to 32 TOPS	Datasheet
MIC-710AIL-DVA	AI Developer Kit	NVIDIA Jetson NANO	up to 472 GFLOPS	Datasheet
MIC-711-OX	AI Inference System	NVIDIA Jetson Orin NX	up to 100 TOPS	Datasheet



Edge AI Software



NVIDIA AI Enterprise



Best-in-class development tools, frameworks, and pretrained models for AI practitioners and reliable management and orchestration.

- Speed data processing time up to 5X while reducing operational costs by 4X with the NVIDIA RAPIDS™ Accelerator for Apache Spark.
- Create custom, accurate models in hours, instead of months, using NVIDIA TAO Toolkit, and pretrained models.
- Accelerate up to 8X LLM inference performance with TensorRT-LLM™ and up to 40X inference performance with NVIDIA® TensorRT™ over CPU-only platforms.
- Simplify and optimize the deployment of AI models at scale and in production with NVIDIA Triton™ Inference Server.

[Datasheet](#)

Intel Geti Software



Intel's new software speeds up building computer vision models, easing data labeling, training, and optimization, enabling scalable custom AI models.

- Expedite data labeling with smart annotations
- Achieve a working model with less data via active learning
- Optimize and quantize vision models automatically
- Export production-ready models for deployment
- Deploy and manage models via DeviceOn OTA and Container Management
- Quick inference benchmark and compatible runtime SDKs via Edge AI SDK

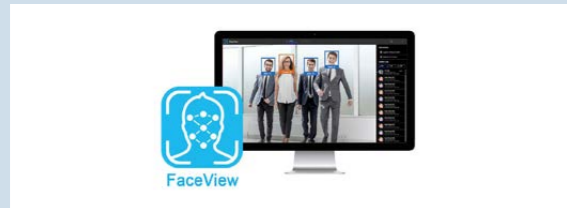
[Datasheet](#)



Edge AI Software



FaceView Facial Recognition



Fast, high-precision, scalable facial recognition in real time for AIoT applications.

- Real-time identification and notification, supports VPU/GPU acceleration
- High-precision facial recognition engine for registered ID and gender/age/ emotion with 99.7 accuracy rate
- Customer behavior statistical analysis with visualization history event dashboard for customer analytics
- Provides customer ID database which can modify VIP/Blacklist ID data for improved management
- Provides SDK for integrating FaceView functions with current solution

[Datasheet](#)

FaceView Masked Face Recognition



- Detect whether the person is wearing a mask
- Recognize masked individuals

[Datasheet](#)

FaceView Body Temperature Monitoring



This application leverages FaceMeHealth a 3-in-1 solution on facial and masked face recognition plus body temperature detection.

- Recognize masked individuals

[Datasheet](#)

nRF Edge Impulse - App



Before your AI can access LLM's and LAM's you need to collect your data out of the field and use ML already at the Edge.

Edge AI - More powerful data processing at the edge

Nordic Semiconductor use their experience in ultra-low-power to take wireless technology to the next level, allowing more powerful data processing at the edge, bringing significant benefits to your applications.

- ML Studio (by Nordic Semiconductor, powered by Edge Impulse)

Quick Start Suggestion: rutronik-tec.com/multi-protocol-iot-prototyping-with-nordic-thingy53



Integrated ML from Edge Impulse directly on the Thingy:53

nRF Edge Impulse app is developed in collaboration with Edge Impulse to train and deploy embedded machine learning models on the Nordic Thingy:53 using Edge Impulse Studio. Every Nordic Thingy:53 comes pre-installed with firmware compatible with the nRF Edge Impulse mobile app. The app allows you to upload raw sensor data via your mobile device to the cloud-based Edge Impulse Studio and deploy fully trained ML models to the Nordic Thingy:53 over Bluetooth® Low Energy (LE).

The app allows you to choose between existing projects in Edge Impulse studio data or creating a new project and can collect sensor data from the following set of sensors on the Nordic Thingy:53:

- Accelerometer
- Inertial (Accelerometer and Magnetometer)
- Microphone
- Light and Environment
- Magnetometer
- Light and Inertial
- Light
- Environment and Inertial
- Temperature, Pressure, Humidity and Air quality

The collected sensor data can be labeled as training or testing data and may be inferred directly in the app with machine learning model deployed to the Nordic Thingy:53.

A free Edge Impulse account, which can be created from the nRF Edge Impulse app or at www.edgeimpulse.com, is required to connect the app with Edge Impulse Studio.



Machine Learning for Infineon PSoC Family



World Class edge AI with Imagimob and Infineon

In May 2023, Imagimob AB became a part of Infineon Technologies AG. Since then we have been working together to deliver customers amazing edge AI solutions.

Deploy state-of-the-art edge AI models seamlessly onto Infineon hardware!

Together, Imagimob and Infineon deliver customers end-to-end edge AI solutions. We have the right solution for you, whether you want to build a model from scratch (or let us help you build one), want to customize your own model, or want to buy one ready made.

Read more about how Imagimob software is deployed seamlessly onto Infineon hardware via our integration with Infineon's ModusToolbox™ Software.

Expertise in hardware, software, and machine learning!

Imagimob's extensive Machine Learning experience perfectly compliments Infineon's knowledge and portfolio of sensors and powerful Microcontrollers.

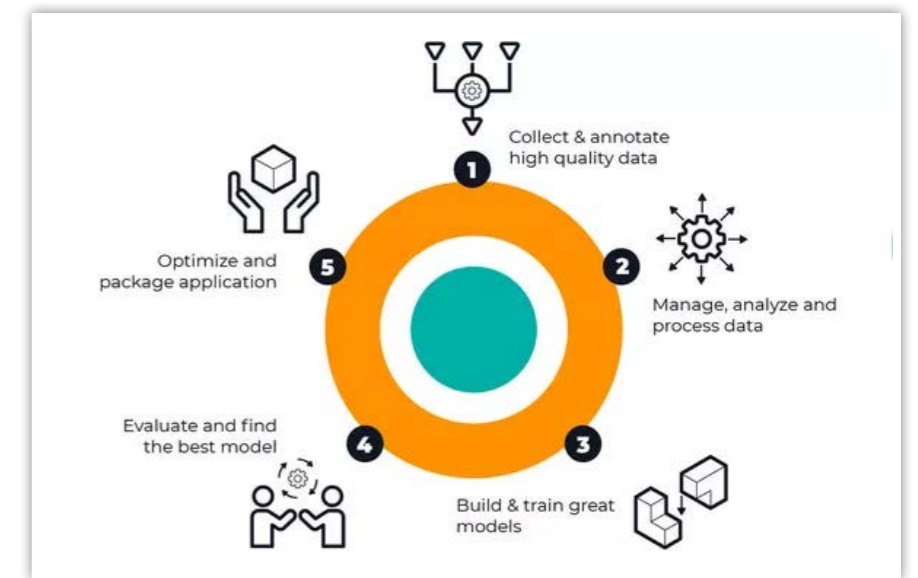
Edge AI allows you to use machine learning in your embedded devices

You know what the problem with today's machine learning (ML) is? It's reserved for the Cloud. It uses up too much memory and resources. This means it can't run on embedded hardware, even though that's where a lot of important data is generated.

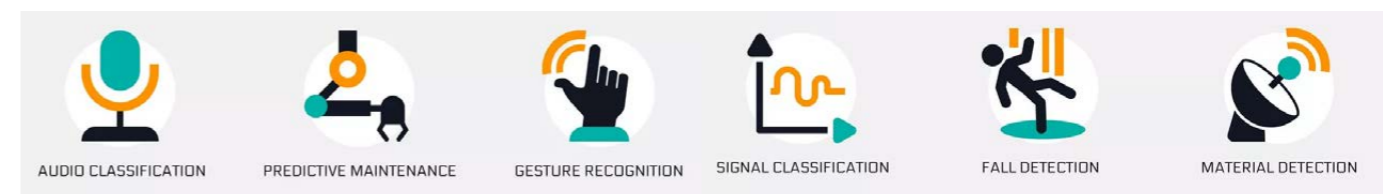
Imagine what you could do if you could bring machine learning into your embed-

ded devices. You could do audio classification without any data ever leaving your embedded device. You could sense that a machine is about to break and turn it off in a few milliseconds. Even when the connection is down.

Imagimob Studio is the Edge AI development platform that allows you to do just that. It makes it easier than ever before to develop ML models for edge devices, from data collection through deployment.

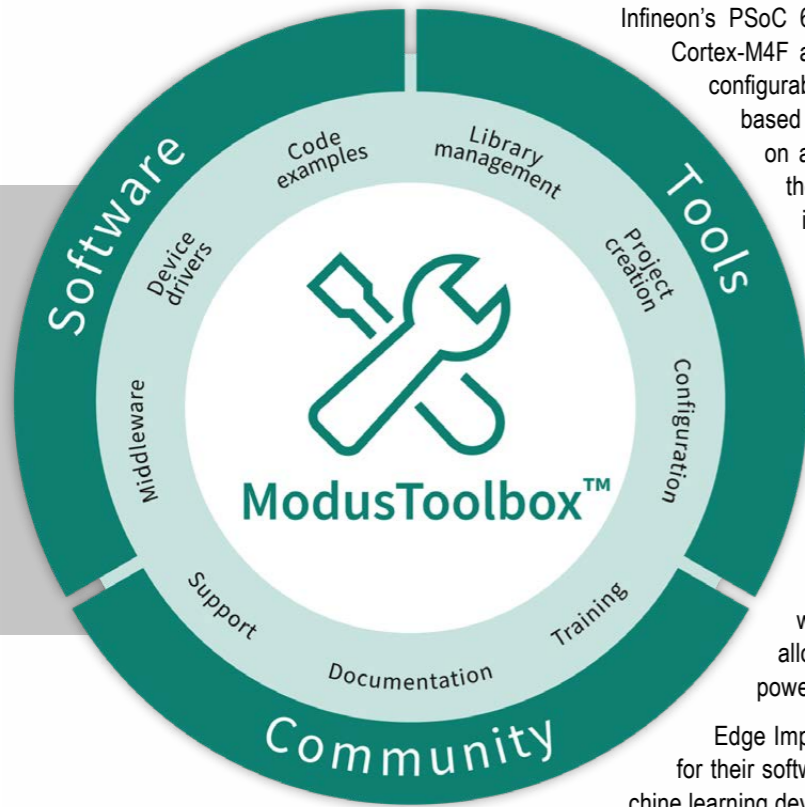


Quick Start Proposal: www.rutronik.com/rutronik-system-solutions/rutronik-development-kit-rdk3



Machine Learning for Infineon PSoC Family

Infineon further extends its edge AI capabilities and choice-of-platforms for Machine Learning-based models for Bluetooth customers by partnering with Edge Impulse,



Infineon's PSoC 63 Bluetooth LE MCU devices feature a dual-core Arm Cortex-M4F and Arm Cortex-M0+ chip architecture, Bluetooth LE 5.2, configurable voltage and frequency settings, built-in hardware-based security, state-of-the-art capacitive interfaces, and more, on a single chip. As the only 150 MHz Bluetooth LE MCU on the market, this variant of the Infineon PSoC device family is a powerful combination of power efficiency, size, and programmability making it perfectly suited for edge IoT applications that benefit from the ability to run advanced ML algorithms.

Edge Impulse's products streamline the entire process of collecting and structuring data sets, designing algorithms with pre-built building blocks, validating the models with real-time data, and deploying the fully optimized production-ready results to edge targets such as the PSoC 63 Bluetooth LE MCU.

The partnership expands Infineon's ModusToolbox™ software ecosystem to Edge Impulse's Studio environment, allowing for easy configuration of the high-performance, low-power PSoC™ 6 MCU series for edge ML applications.

Edge Impulse and Infineon have announced cross-platform support for their software environments, allowing for high-powered, flexible machine learning development on the Infineon PSoC™ 6 microcontroller series.

Quick Start Proposal: www.rutronik.com/rutronik-system-solutions/rutronik-development-kit-rdk3



ASUS Zenbo Junior II – A caring AI companion

Zenbo Junior II is our interactive robot packs various perception, cognition and loads of IoT functions. Zenbo Junior II helps you to be more efficient, more interesting and to make more of artificial intelligence in all sorts of everyday situations.

Application Scenarios

- **Education:** Use Zenbo Lab to create interactive conversations and activities to help students practice speaking + listening skills
- **Health Care:** Self-assessment in hospitals where patients can enter information to speed up assessment time
- **Much more:** Zenbo Junior II can bring lively and professional information introduction to passengers, users and customers

4 Outstanding Capabilities

- Equipped with Perception AI
- Friendly Human-Robot Interaction
- Easy-to-Use Development
- Efficient + Performance-Boosting Tools

AI Optical Sorting and Classification in Coffee Bean Processing



Application example

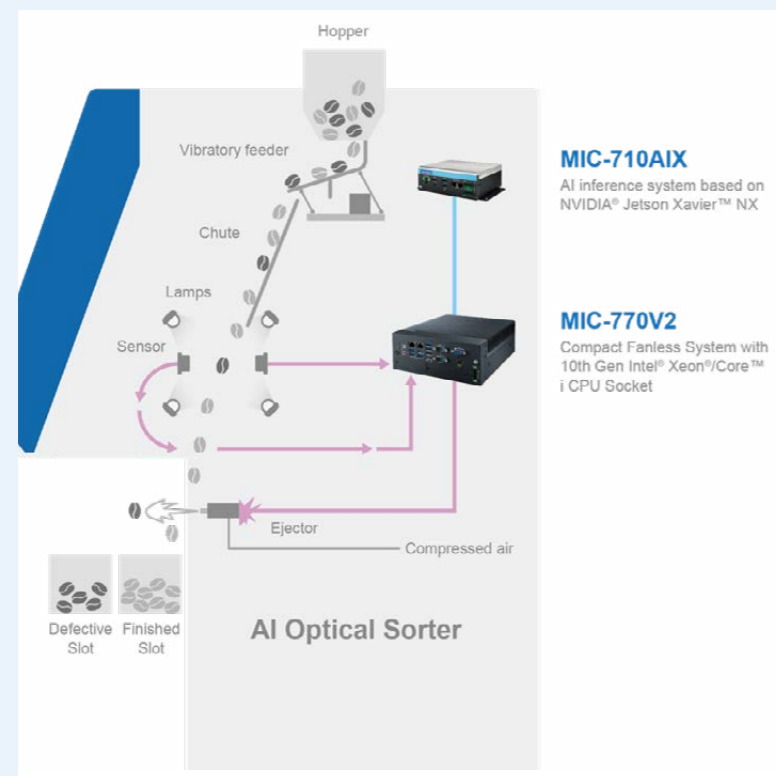
In today's market, there is a growing demand for efficient and accurate methods to select the best coffee beans for brewing premium coffee. Coffee beans of the highest quality are distinguished by color, shape, size, and morphology. This is where an automated optical inspection system using artificial intelligence is highly beneficial. By utilizing a data-driven algorithm, it becomes possible to train the system to classify objects accurately. This, in turn, enhances the quality inspection process and reduces the need for traditional inspections, resulting in significant cost and time savings.

To implement such a system, there are some requirements

- High-performance CPUs and GPUs for image acquisition and processing
- An AI-based automated optical inspection system for classification and analysis
- A tailored algorithm for determining object quality and type
- Technology to process and analyze images to meet morphology requirements



System Diagram



System Description

The AI Optical Inspection solution consists of a controller called MIC-770V2, which is fanless and x86-based. It comes with an integrated CameraLink frame grabber and camera, as well as lighting, sensor, and ejector modules. The CameraLink camera operates at high speeds, capturing images of coffee beans as they move through the feeder. This significantly speeds up the sorting process. The images are then transmitted to the MIC-710AIX which runs on the NVIDIA Jetson Xavier NX. This system segregates batches of beans, identifying specific types and superior quality using AI. Overall, the AI Optical Inspection solution is designed for high-throughput and non-destructive screening of coffee beans to select for specific quality traits.

Project Implementation

- MIC-770V2 controller with CameraLink frame grabber
- Camera, lamps, sensor, and ejector modules
- MIC-710AIX built on NVIDIA Jetson Xavier NX

Why Advantech AI Solution?

Advantech's AI-powered optical coffee bean sorter and quality classification technologies provide coffee producers with a powerful and robust set of tools for analyzing and sorting a wide variety of attributes throughout the supply and value chains. This leads to an improvement in quality and revenue while adhering to strict quality guidelines.

Sorting technologies can be customized to meet the specific sorting and processing needs of the customer, allowing them to take advantage of each machine's unique features for optimal efficiency and performance.

Fanless systems such as the MIC-770V2 and the MIC-710AIX are ideal for food industry factories. The MIC-770V2 can be integrated with add-ons and control interfaces such as COM and DIDO to control devices.

The MIC-710AIX, equipped with the powerful NVIDIA Jetson Xavier NX, provides industry leaders with faster and more efficient production methods. Systems based on these technologies can significantly increase processing capacity, particularly when using AI-based quality control methods.

Related Products

- ICAM-520 (Industrial AI Camera)
- MIC-710AI (AI Inference System)
- MIC-710AIT (AI System)
- MIC-711D-ON (Developer Kit)
- ICAM-500 (Industrial AI Camera)





www.rutronik.com



Rutronik Elektronische Bauelemente GmbH
Industriestraße 2 | 75228 Ispringen | Germany