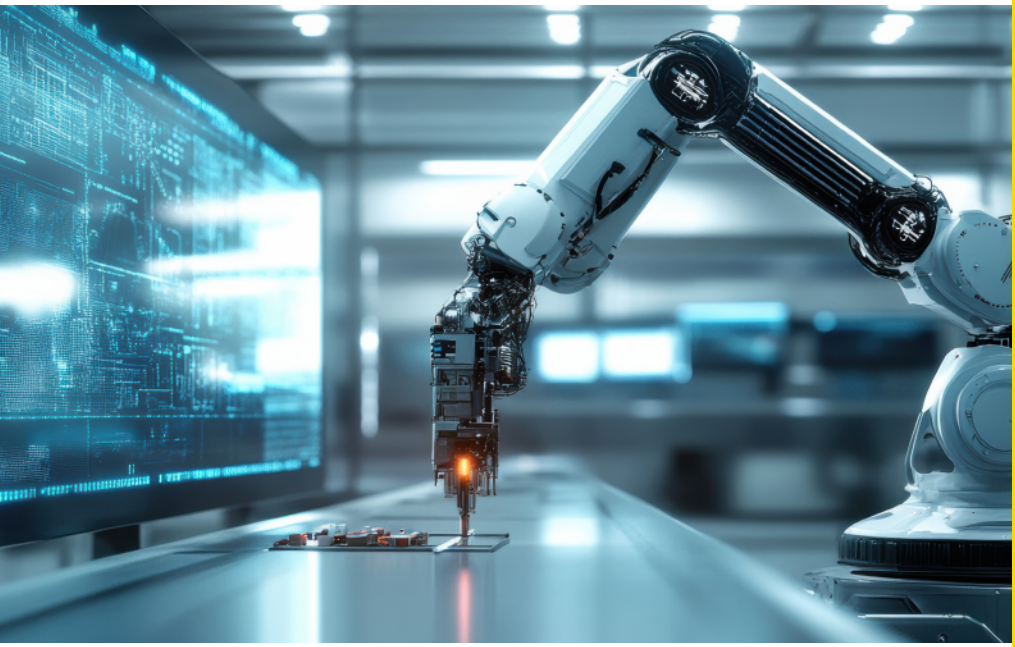
PHISON
PASCARI



AI on IoT Devices with Pascari aiDAPTIV™

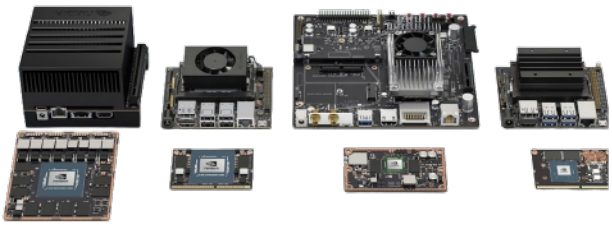
Boost Edge Computing & Robotics Capabilities

Affordably Add LLM Intelligence to IoT Devices

Enhance edge computing and robotics processing with better inferencing and the ability to train LLMs locally.

With Pascari aiDAPTIV, you can now easily add AI processing to sensor-intensive and remote computing platforms. By extending the token length, aiDAPTIV improves inference context accuracy and recall. Also, the added memory capacity enables IoT devices to train LLMs without depending on AI processing in the public cloud.

Edge/IoT/Robotics
Up to 70B Parameter
LoRA Model Training



Benefits

Fits Your Budget

Offloads expensive HBM & GDDR memory to cost-effective flash memory. Significantly reduces the need for large numbers of high-cost and power-hungry GPU cards. Keeps AI processing where the data is collected or created, saving data transmission costs to and from the public cloud.

Simple to Use and Deploy

Offers all-in-one AI toolset that enables ingest to RAG and fine-tuning to inference using an intuitive graphical user interface. Deploys in your home, office, classroom, or data center using commonplace power.

Keeps Data in Your Control

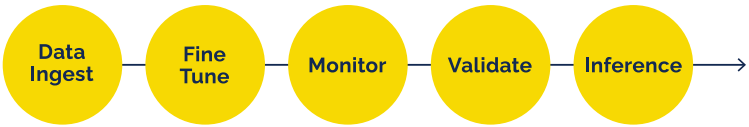
Enables LLM training behind your firewall. Gives you full control over your private data and peace of mind over data sovereignty compliance.

Find more information and resources at:
phisonenterprise.com

aiDAPTIV Technology:
LLM Training Integrated Solution

aiDAPTIV Pro Suite

aiDAPTIV Pro Suite builds on Phison's aiDAPTIV memory management middleware, while also providing higher-level tools for workflow orchestration, monitoring, and AI workload execution through both command-line and graphical interfaces.



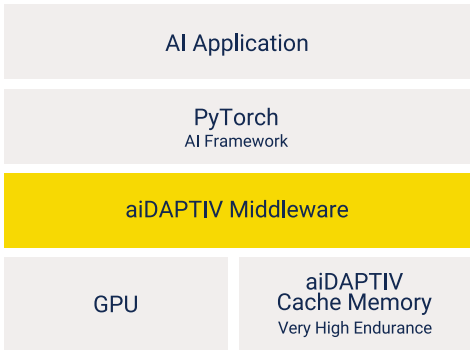
Supported Models

- LLaMA, LLaMA-2, LLaMA-3, Code LLaMA
- Mistral and other transformer-based language models
- Vision and multimodal models such as CLIP and Whisper
- Additional foundation and domain-specific models

aiDAPTIV Pro Suite

and/or

aiDAPTIV Middleware



Built-in Memory Management Solution

aiDAPTIV memory management middleware transparently manages data across GPU memory, system memory, and aiDAPTIV cache memory. It enables memory-intensive AI workloads to exceed native GPU memory limits without requiring changes to existing AI frameworks or applications.

- **AI Application**
- **PyTorch / AI Runtimes:** AI Framework
- **aiDAPTIV Middleware:** Multi-tier Memory Management
- **aiDAPTIV Cache Memory:** Extremely high-endurance SSD flash memory

and

Seamless Integration with GPU Memory

The optimized middleware extends GPU memory by an additional 2x 2TB using aiDAPTIV cache memory. This added memory is used to support LLM training with low latency. Plus, the high-endurance feature offers an industry-leading 100 DWPD, utilizing a specialized SSD design with an advanced NAND correction algorithm.

aiDAPTIV Cache Memory

M.2 AND U.2 SSDS

PHISON PASCARI

Find more information and resources at: phisonenterprise.com