# aiDAPTIV+

AI on IoT Devices: Boost Edge Computing and Robotics Capabilities

Affordably Add LLM Intelligence to IoT Devices
Enhance edge computing and robotics processing with better inferencing and the ability train LLMs locally.

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

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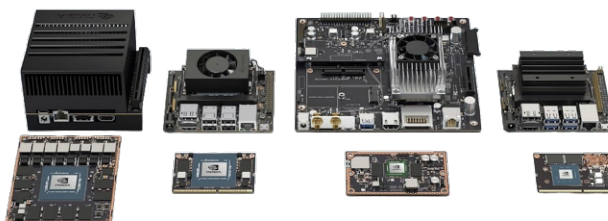
Benefits

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 date is collected or created, thus saving data transmission costs to and from the public cloud.

Simple to Use and Deploy
Offers all-in-one AI toolset enabling 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.

PHISON

Phison aiDAPTIV+ AI Processing Integrated Solution

Use a Command Line or leverage the intuitive All-in-One aiDAPTIVPro Suite to perform LLM Training and Inferencing

aiDAPTIVPro Suite

Data Ingest → RAG → Fine Tune → Monitor → Validate → Inference →
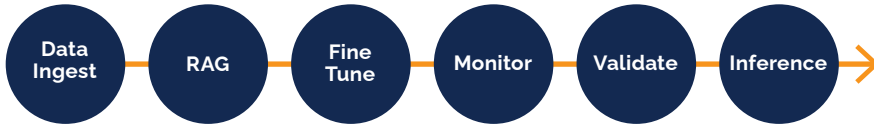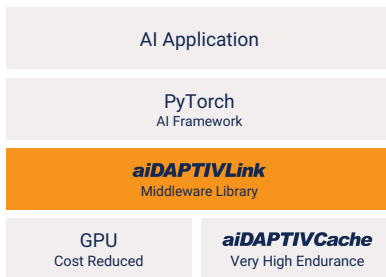

Supported Models
- Llama, Llama-2, Llama-3, CodeLlama
- Vicuna, Falcon, Whisper, Clip Large
- Metaformer, Resnet, Deit base, Mistral, TAIDE
- And many more being continually added

and/or

aiDAPTIVLink

AI Application
PyTorch AI Framework
aiDAPTIVLink Middleware Library
GPU Cost Reduced / *aiDAPTIVCache* Very High Endurance

Built-in Memory Management Solution
Experience seamless PyTorch compliance that eliminates the need to modify your AI application. You can effortlessly add nodes as needed. System vendors have access to AI100E SSD, middleware library licenses, and full Phison support to facilitate smooth system integration.

and

aiDAPTIVCache Family

Seamless Integration with GPU Memory
The optimized middleware extends GPU memory by an additional 320GB (for PCs) up to 8TB (for workstations and servers) using aiDAPTIVCache. This added memory is used to support LLM training with low latency. Furthermore, the high endurance feature offers an industry-leading 100 DWPD, utilizing a specialized SSD design with an advanced NAND correction algorithm.

AI100E M.2 SSD

PHISON

Find more information and resources at: phison.com