



Cost-Effective Onsite LLM Training and Better Inferencing

Large Model Training and Inferencing with Your Private Data aiDAPTIV+ provides a turnkey solution for organizations to train and inference large data models on-site at a price they can afford. It enhances foundation LLMs by incorporating an organization's own data enabling better decision making and innovation.

Train and Inference Any Model Size On-Premises
aiDAPTIV+ allows businesses to scale-up or scale-out nodes to increase training data size, reduce training time and improve inferencing.



Edge/IoT/Robotics

Up to 70B Parameter
LoRA Model Training



AI Notebook PC

Up to 8B Parameter
Full Model Training



Desktop

Up to 13B Parameter
Full Model Training



Workstation PC

Up to 100B Parameter Full
Model Training



Server

Up to 405B Parameter
Full 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, 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.

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

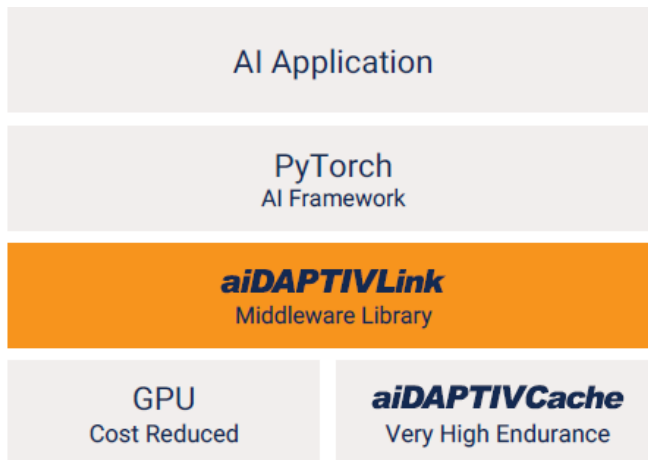


Supported Models

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

and

aiDAPTIVLink



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

aiDAPTIVCacheFamily

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.

