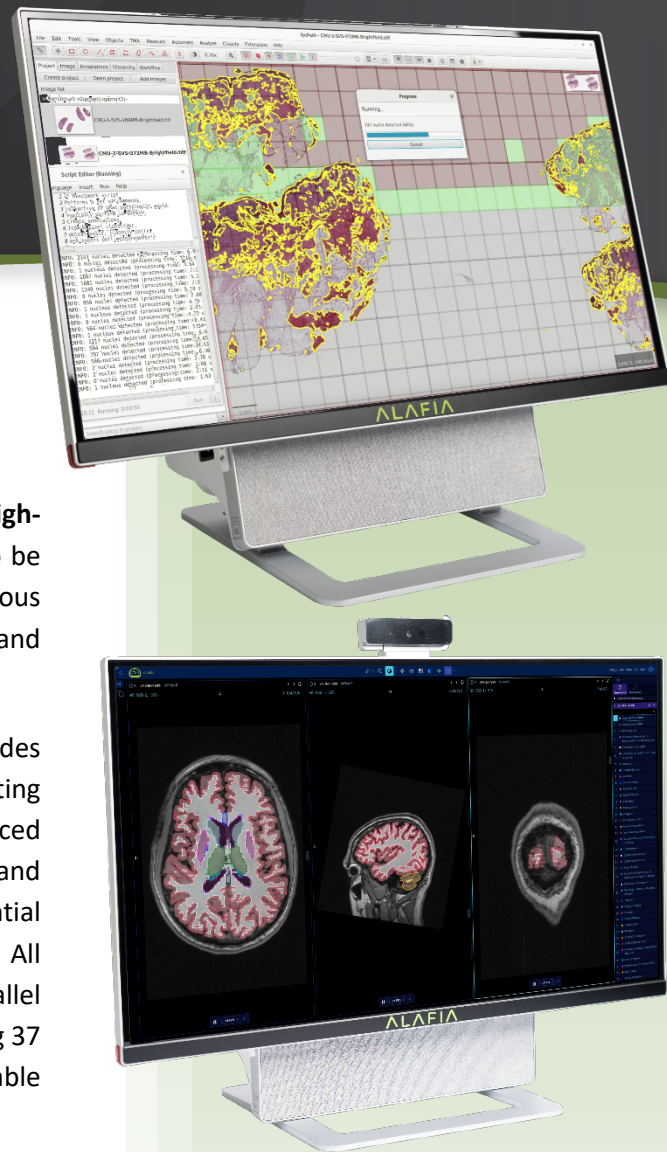




ALAFIA AIVAS Supercomputer Datasheet

Introducing **AIVAS**, the world's first **generative AI all-in-one high-performance personal computer**! Architected from the ground up to be the most powerful and easy to use massively-parallel heterogenous compute system. **AIVAS** has been designed to solve the most difficult and large-scale data intensive scientific problems.

Packing the equivalent of three full cloud native instances, **AIVAS** provides a seamless on-ramp and developer platform to any supercomputing cluster, cloud or Arm deployment environment. Leveraging 192 advanced single-instruction, multiple-data (SIMD) CPU vectorization engines and Quad-GPUs. Developers can efficiently execute both the sequential control-intensive and data-parallel phases code of any application. All while executing instructions to access memory and process data in parallel over a larger set of vectors with a unified memory architecture. Gaining 37 times more compounded performance of over state of the art comparable systems. Supporting over 13,000 AI TOPS.

Product Hardware Specifications

CPU Processor	Ampere AmpereOne 96 ARM-based v8.6+ 64-bit CPU Cores at 3.6 GHz Speed (2x128-bit SIMD units per core)
Graphics GPU	NVIDIA RTX PRO 60000 Blackwell Graphics GPU with 96GB GDDR7 error-correction code (ECC) memory
AI Accelerator GPU	NVIDIA RTX PRO 6000 Blackwell Server GPUs with 96GB GDDR7 error-correction code (ECC) memory
Memory	2048 GB (2TB) Unified Error-Correction Code (ECC) System Memory DDR5 RDIMMs at 5200 MHz
Color	Rugged Black Metal Finish Chassis and Semi-Gloss Cardinal White Cover
Display	Dell UltraSharp 32" 6K 6144 x 3456 at 60 Hz or Medical Diagnostics Screen
Side ports	(2) USB-A 3.2 Gen (3) Ethernet RJ-45 ports (2) DisplayPort
Built-in Media	8 Megapixels (3840 x 2160) Video Resolution Dual Array Microphones Stereo 14 W Speakers
Storage	8TB Gen5 PCIe 5.0 x4 NVMe Expandable up to 72TB NVMe Storage Up to 14,800 MB/s Sequential Read and 13,400 MB/s Write Speeds
Networking	(2) Broadcom BCM57416 RJ45 (10GbE)
BMC	IPMI Dedicated RJ45 Realtek RTL8211E Port
Max Power Consumption	2000 Watts
Dimensions (H x W x D)	24" x 21" x 12" Landscape Mode 26.5" x 21" x 12" Portrait Mode
Weight	33 KG 67 LBS



AIVAS AI Inference Performance

Inference Benchmarks Performance	
Model Architecture	Inferences per Second
ResNet-50 v1.5 (Image Classification)	152,582
BERT Large (Natural Language Processing)	7,399
DLRM v2 (Recommender Systems)	1,389,305

Table 1 Benchmarked with the Ampere Model Library and MLPerf™ inference

Unified System Memory	
Private L1 Cache (64KB per Core)	6,144 KB
Private L2 Cache (2MB per Core)	192 MB
Memory Bandwidth	279 GB/s

Table 3 Memory Architecture

Compute Performance	
Data Structure	Performance Throughput
Peak FP4 AI Peta FLOPS	8 PFLOPS
Single-Precision FP32	245 TFLOPs
NVIDIA Tensor Cores	1504
NVIDIA RT Cores	376
Memory Interface	512-bit
CUDA GPU Cores	48,128
GPU Memory	192 GB GDDR7 ECC
GPU Memory Bandwidth	1,597 GB/s

Table 2 Results shown with sparsity



Security & Privacy Features

Support for State-of-the-Art Encryption and Platform Security

Component Level Secure Boot & Root of Trust

- SMpro and PMpro processors
- Full disk encryption
- Secure firmware upgrade
- Firmware rollback protection

Platform Level Trusted Computing

- TPM Module support
- SHA-1 PCR Bank | SHA256 PCR Bank
- Pending Operation | Platform Hierarchy
- Storage Hierarchy | Endorsement Hierarchy

Processor Level Confidential Computing

- Memory isolation through main memory encryption
- Trusted execution environments
- Physically isolated processors

Operating System Level Privacy Enhancing Technologies and Cryptographic Support

- Differential privacy | Homomorphic encryption
- Secure multiparty computation | Zero-knowledge proofs
- Single subscription for security and compliance

More Information



Contact

For more information
please reach us at:
(341)-222-9735
contact@alafia.ai

To find out more visit:
www.alafia.ai