SAPPHIRE

AMD RADEON AI PRO R9700 | 32 GB

AMD RDNA™ 4 ARCHITECTURE



SAPPHIRE AMD Radeon AI™ Professional Graphics

For Local AI Inference, Development, and Other Memory-Intensive Workloads.

Take your AI workloads further with the AMD Radeon™ AI PRO R9700 and AMD ROCm™.
The combination of 32GB of VRAM and multi-GPU support creates a powerful foundation for large language models and other compute-heavy tasks, while delivering exceptional value in its class.

32GB VRAM for Larger AI Models Optimal VRAM Buffer for Advanced Local AI
Meet the memory demands of modern LLMs and text-to-image models on your desktop with AMD Radeon™ AI PRO R9700 Graphics featuring 32GB of VRAM.

Typical VRAM Usage by Popular Models

LLM

DeepSeek R1 Distill Qwen 32B Q6	28GB
Mistral Small 3.1 24B Instruct 2503 Q8	27GB

Text to Image

Flux.1 Schnell	24GB
SD 3.5 Medium	17GB

Massive Gains for Massive Models

Up to 5x Faster with 32GB VRAM

Phi 3.5 MoE Q4	361%
Mistral Small 3.1 24B Instruct 2503 Q8	437%
DeepSeek R1 Distill Qwen 32B Q6	454%
Qwen 3 32b Q6	447%
Qwen 3 32b Q6 Large Prompt (3000+ tokens)	496%
	100%

AMD Radeon™ AI PRO R9700 (32GB) | GeForce RTX 5080 (16GB)

Smarter Value for Your AI Projects

AMD Radeon™ AI PRO R9700 sets a new standard for cost-effective AI acceleration, delivering consistently higher value in high-VRAM workloads.

DeepSeek R1 Distill Qwen 7B	156%
DeepSeek R1 Distill Qwen 14B FP8-dynamic	148%
Qwen3 14B FP8-dynamic	139%
DeepSeek R1 Distill Llama 8B	177%
Llama-3.1 8B-Instruct	179%
DeepSeek R1 Distill Llama 8B FP8-dynamic	158%
Llama-3.2 1B Instruct	153%
Llama-3.2 3B-Instruct	170%
	100%

AMD Radeon™ AI PRO R9700 (32GB) | Nvidia RTX 4500 (Blackwell)

SAPPHIRE

AMD RADEON AI PRO

AMD RADEON PRO

MODEL NAME	 SAPPHIRE AMD RADEON™ AI PRO R9700 32GB	 SAPPHIRE AMD RADEON™ AI PRO R9700S 32GB	 SAPPHIRE AMD RADEON™ AI PRO R9600D 32GB	 SAPPHIRE AMD RADEON™ PRO W7800 48GB
GPU	AMD RADEON™ AI PRO R9700 AMD RDNA™ 4 Architecture		AMD RADEON™ AI PRO R9600D AMD RDNA™ 4 Architecture	AMD RADEON™ PRO W7800 AMD RDNA™ 3 Architecture
Interface	PCI-Express 5.0 x16			PCI-Express 4.0 x16
Compute Units	64		48	70
Ray Accelerators	64		48	70
AI Accelerator	128		96	140
Stream Processors	4096		3072	4480
Infinity Cache	64MB			96MB
Peak Half Precision (FP16 Matrix) Performance	191 TFLOPs		99 TFLOPs	90.4 TFLOPs
Peak 8-bit Precision (INT8 Matrix) Performance	383 TOPs		199 TOPs	90.4 TOPs
Transistor Count (Billion)	53.9			57.7
Texture Units	256		192	280
Memory Size/Bus	• 32GB GDDR6 • 256 bit Memory Bus • 640 GB/s Peak Memory Bandwidth • Memory ECC Supported			• 48GB GDDR6 • 384 bit Memory Bus • 864 GB/s Peak Memory Bandwidth • Memory ECC Supported
Displays	4x DisplayPort™ 2.1a		1 x DisplayPort™ 2.1a	3x DisplayPort™ 2.1a, 1x mini DisplayPort™ 2.1a
AMD Technologies	• AMD Software Adrenalin Edition • AMD FidelityFX™ Super Resolution 4 • AMD Fluid Motion Frames • AMD Radeon™ Super Resolution • AMD Smart Access Memory • AMD Smart Access Video • AMD Radeon™ Boost • AMD Radeon™ Anti-Lag • AMD Radeon™ Image Sharpening 2 • AMD Enhanced Sync Technology • AMD FreeSync™ Technology • AMD Radeon™ Chill • AMD Virtual Super Resolution			• AMD Viewport Boost • AMD Remote Workstation • AMD Radeon™ Media Engine • AMD Software PRO Edition • AMD Radeon™ VR Ready Creator • AMD Radeon™ ProRender
Cooler Type	Active blower fan, two ball bearing	Passive Cooling		Active blower fan, two ball bearing
Form Factor	2 Part Slot Occupied 266.7 (L) *111 (W) * 39 (H) mm	2 Part Slot Occupied 266.7 (L) *108.9 (W) * 39 (H) mm	1 Part Slot Occupied 241.3 (L) *111.15 (W) * 18.67 (H) mm	2 Part Slot Occupied 279.4 (L) *111.15 (W) * 39 (H) mm
External Power Connector	12V-2x6			
Power Consumption	300W		150W	260W

Scan QR code for more information:

R9700 R9700S R9600D W7800

SAPPHIRE