



AMD Radeon AI PRO R9700

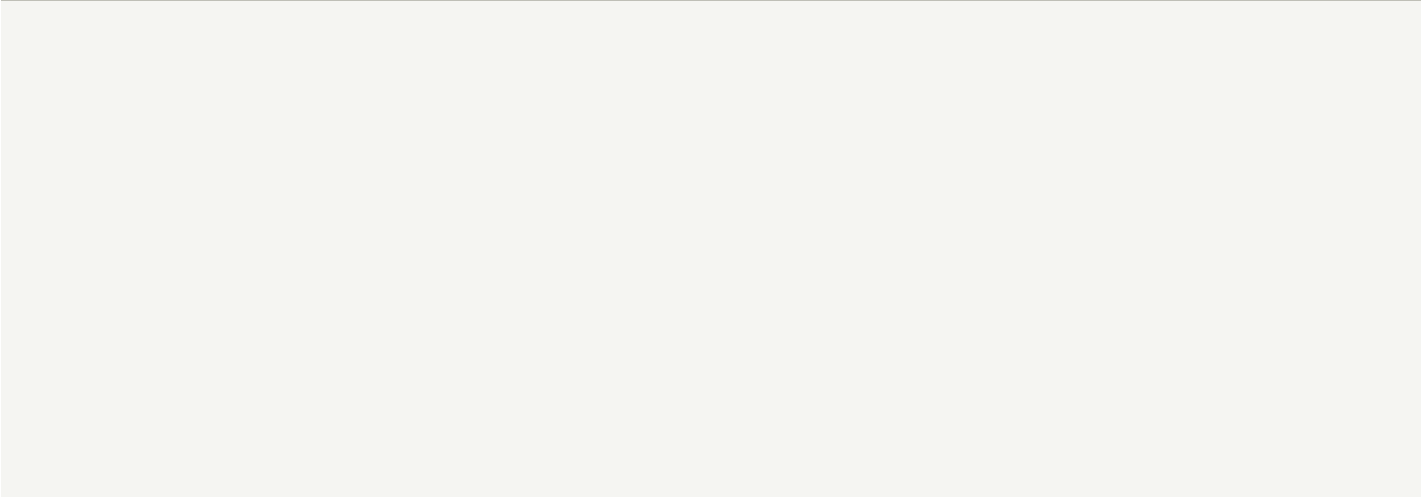
Also known as: R9700 · Radeon AI PRO R9700 32GB

The Radeon AI PRO R9700 is AMD's RDNA 4 workstation card with 32 GB of GDDR6 on a 256-bit bus, 64 compute units with dedicated AI accelerators, and a 300 W actively cooled double-slot board on PCIe 5.0 x16. It launched at \$1,299 in late 2025 and is the low-cost way to get 32 GB per card for local inference and development; two or four cards in a workstation reach 64 to 128 GB of total VRAM over PCIe. It is a desktop and workstation part, not a datacenter accelerator: no HBM, no GPU fabric, and a 12V-2x6 power connector. GPU Smith specs it for developer boxes and small private-inference nodes where price per GB of VRAM dominates.

32 GB GDDR6 Memory · 256-bit, ECC on Linux only	640 GB/s Memory bandwidth · peak	383 TFLOPS FP8 matrix · dense; 766 TFLOPS with sparsity
191 TFLOPS FP16 matrix · dense; 383 TFLOPS with sparsity	300 W Board power · 750 W PSU recommended	Double slot, active Form factor · PCIe 5.0 x16, 12V-2x6

COMPUTE		01/05
Architecture	RDNA 4	
Compute units	64	
Stream processors	4,096	
AI accelerators	128	
Ray accelerators	64	
Boost frequency	Up to 2,920 MHz · game frequency 2,350 MHz	
FP32 / FP16 vector	47.8 TFLOPS	
FP16 matrix	191 TFLOPS · dense; 383 with structured sparsity	
FP8 matrix	383 TFLOPS · E5M2/E4M3, dense; 766 with structured sparsity	
INT8 matrix	383 TOPS · dense; 766 with structured sparsity	
INT4 matrix	766 TOPS · dense; 1,531 with structured sparsity	
Transistors	53.9 billion	

MEMORY		02/05
Capacity	32 GB GDDR6	
Bus width	256-bit	
Bandwidth	640 GB/s · peak	
Infinity Cache	64 MB	
ECC	Linux only · per AMD	
INTERCONNECT & FORM FACTOR		03/05
Host bus	PCIe 5.0 x16	
Multi-GPU	PCIe only · no GPU-to-GPU bridge listed by AMD	
Board width	Double slot	
Display outputs	DisplayPort · up to 4; configuration varies by partner design	
POWER & THERMAL		04/05
Total board power	300 W	
Power connector	12V-2x6	
Minimum PSU	750 W · AMD recommendation	
Cooling	Active	
SOFTWARE & PLATFORM		05/05
OS support	Windows 10/11 64-bit, Linux x86-64	
Compute stack	ROCm · Linux; check release notes for supported distributions	
Launch price	\$1,299 · reported, retail from 27 October 2025	



PROCUREMENT

Typical lead time

Sold at retail and through channel partners (card partners such as XFX); available from stock at the time of verification, per reported retailer listings.

Pricing

\$1,299 at launch (reported). Street prices since have run higher, reported between about \$1,350 and \$1,900 depending on retailer and date; sources conflict, so check live listings.

- Board designs vary by partner; check slot width and display outputs for the exact SKU.
- A 12V-2x6 connector means older PSUs need an adapter; confirm the 750 W recommendation holds with the rest of the build.
- Multi-card builds need PCIe slot spacing, bifurcation and cooling planned for 300 W cards stacked side by side.

FIELD NOTES

- 32 GB per card is the reason to buy it; delivered speed is not. Phoronix measured a single card close to the RTX 6000 Ada on small FP8 models, but reviews also report slower prompt processing than comparable NVIDIA cards (reported, mixed).
- Models above 32 GB need multiple cards over PCIe, which costs bandwidth on tensor-parallel splits; test your parallelism strategy rather than assuming linear scaling.
- FP4 is not in AMD's published list for this card; the lowest listed matrix datatype is INT4. Plan quantisation around FP8, INT8 or INT4.
- ECC is Linux only, so a Windows deployment gives it up.
- Not for datacenters in the AMD sense: no passive cooling, no OEM server qualification. Use it for desk-side and workstation-class nodes.

QUESTIONS WE GET ON THIS PART

How much memory does the R9700 have?

32 GB of GDDR6 on a 256-bit bus at 640 GB/s, with 64 MB of Infinity Cache.

How much does the R9700 cost?

\$1,299 at launch (reported). Street prices of roughly \$1,350 to \$1,900 have been reported since; check live listings.

R9700 vs the Radeon PRO W7900?

The W7900 has more memory (48 GB) and bandwidth (864 GB/s) on RDNA 3; the R9700 is newer, has FP8 support and, in Phoronix testing, ran about 1.47x the W7900 performance, at 32 GB.

Can I run several R9700s together?

Yes, for example two cards for 64 GB total, over PCIe. Each is 300 W, double slot and needs 12V-2x6 power.

REFERENCES

- [1] <https://www.amd.com/en/products/graphics/workstations/radeon-ai-pro/ai-9000-series/amd-radeon-ai-pro-r9700.html>
- [2] <https://www.club386.com/amds-retail-launch-of-the-1299-radeon-ai-pro-r9700-32gb-gpu-is-imminent/>
- [3] <https://www.phoronix.com/review/amd-radeon-ai-pro-r9700>
- [4] <https://www.techradar.com/pro/return-of-the-og-amd-unveils-radeon-ai-pro-r9700-now-a-workstation-class-gpu-with-32gb-gddr6>

Latest version of this datasheet: <https://gpusmith.com/hardware/gpus/amd-radeon-ai-pro-r9700>. Specifications compiled from the references above; verify against vendor documentation before purchase.