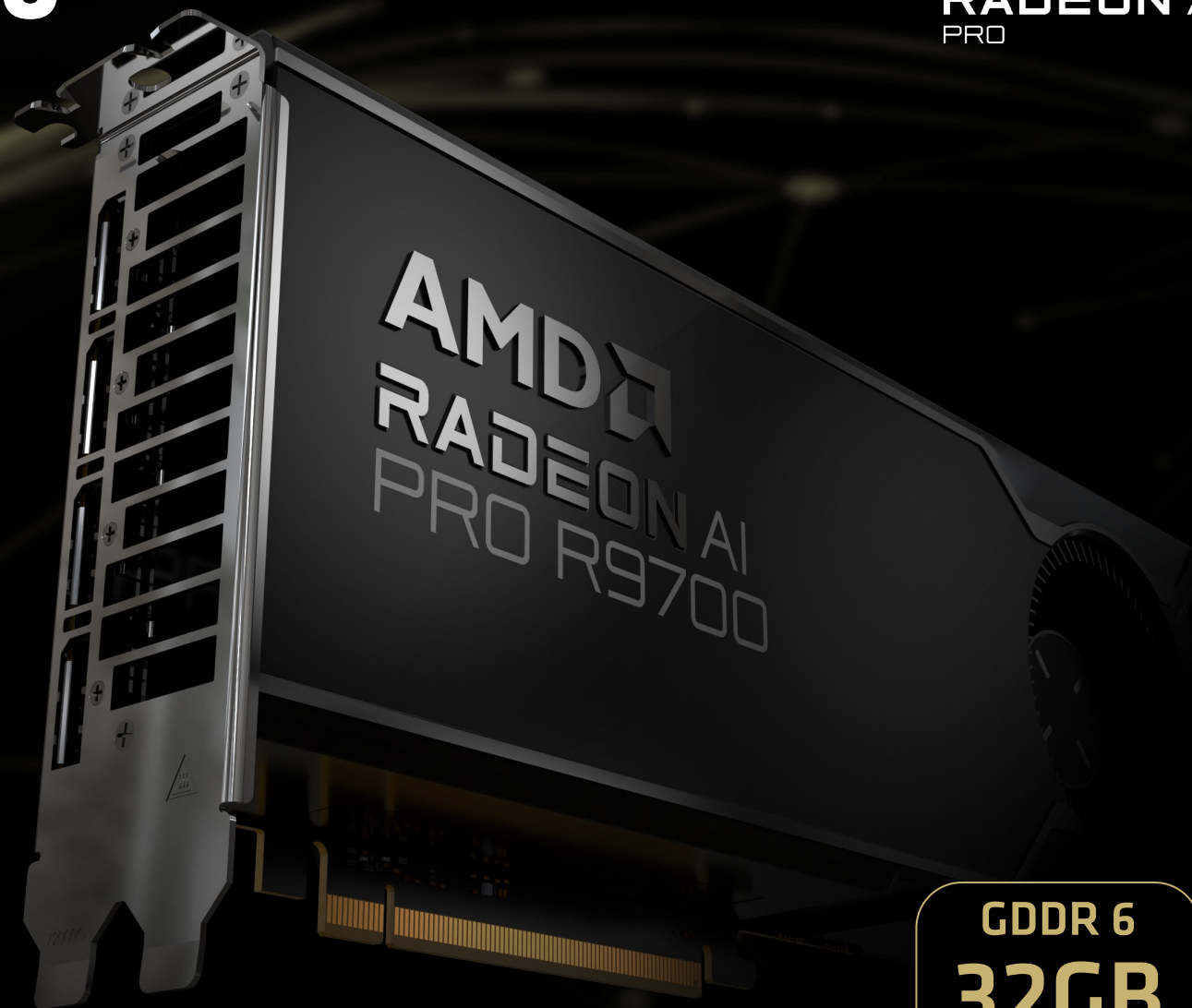


AMD RADEON™ AI PRO R9700

GRAPHICS

AMD
RADEON AI
PRO



HIGH PERFORMANCE ON THE EDGE

GDDR 6
32GB

Key Features

AMD Radeon™ AI PRO R9700 Graphics



AMD RDNA4™ Architecture

Dual Slot Design

32GB GDDR6

256-bit Memory Interface

640 GB/s Memory Bandwidth



64 Compute Units

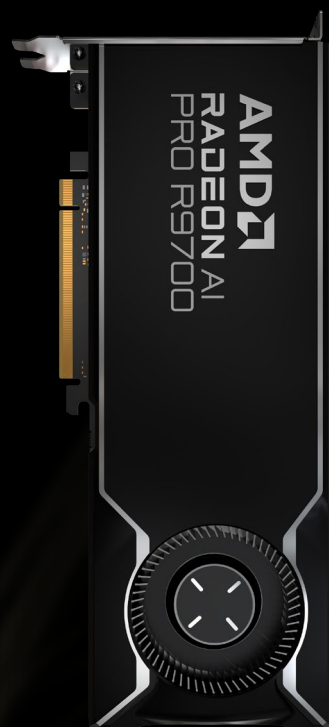
128 AI Accelerators

Up to
191 TFLOPS
FP16 Dense

Up to
1531 TOPS
INT4 Sparse

300W TDP

AMD RADEON AI PRO R9700



Technical Specifications

GPU Architecture

AMD RDNA™ 4

Hardware Raytracing

Yes

Ray Accelerators

64

ROPs

128

Stream Processors

4096

Compute Units

64

Dedicated Memory Size

32 GB

Dedicated Memory Type

GDDR6

Memory Interface

256-bit

Peak Memory Bandwidth

Up to 640GB/s

AMD Infinity Cache™

64 MB

Transistor Count

53.9 Billion

Total Board Power (TBP)

300 W

PSU Recommendation

750 W

External Power Connectors

12V-2x6

Peak Half Precision (FP16 Vector)

Performance

95.7 TFLOPS

Peak Single Precision (FP32 Vector)

Performance

47.8 TFLOPS

Peak 4-bit Precision (INT4 Matrix)

Performance

766 TOPS

Peak 8-bit Precision (INT8 Matrix)

Performance

383 TOPS

Peak Half Precision (FP16 Matrix)

Performance with Structured Sparsity

383 TFLOPS

Peak Half Precision (FP16 Matrix)

Performance

191 TFLOPS

Peak 8-bit Precision (FP8 Matrix)

Performance with Structured Sparsity

(E5M2, E4M3)

766 TFLOPS

Peak 8-bit Precision (FP8 Matrix)

Performance (E5M2, E4M3)

383 TFLOPS

Peak 4-bit Precision (INT4 Matrix)

Performance with Structured Sparsity

1531 TFLOPS

Peak 8-bit Precision (INT8 Matrix)

Performance with Structured Sparsity

766 TFLOPS

Cooling

Active

Displays Type(s)

DisplayPort™ 2.1a

Supported Video Formats

H264 Encode and Decode , H265/HEVC Encode and Decode , AV1 Encode & Decode

Supported 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

Board Height

Full Height

Board Width

Dual Slot

Product Family

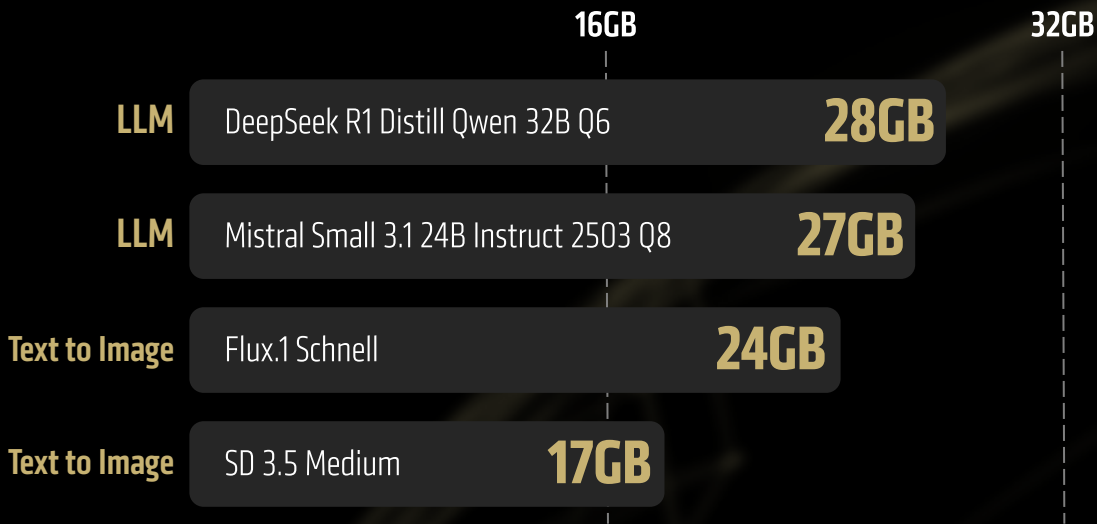
AMD Radeon™ AI PRO

Product Line

AMD Radeon™ AI PRO R9000 Series

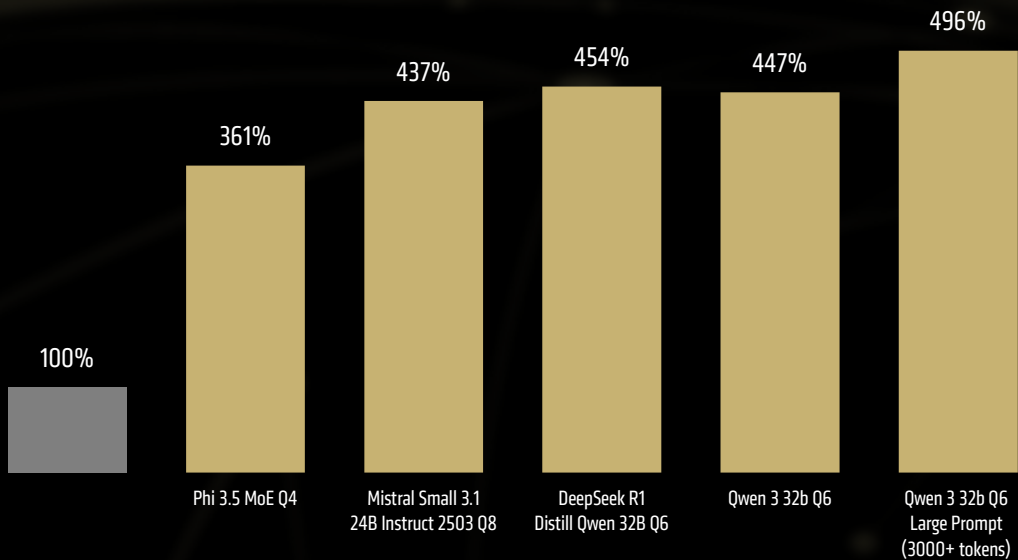
Optimal VRAM Buffer for Advanced Local AI

Typical VRAM Usage by Popular Models



32GB VRAM for Larger AI Models

Large AI Models Performance



GD-239a: Radeon™ PRO W6000 and W7000 Series and Radeon™ AI PRO R9000 Series graphics cards (and later models) are not designed nor recommended for datacenter usage. Use in a datacenter setting may adversely affect manageability, efficiency, reliability, and/or performance. GD-239a.

RPW-495: Testing as of May 2025 by AMD. Average tokens per second of three runs, dropping edge cases where the model starts spiraling (more than 2k thinking tokens) to standardize response length. No speculative decode. All tests conducted on LM Studio 0.3.15 (Build 11). Vulkan Llama.cpp 1.28 used for AMD, NVIDIA-recommended CUDA 12 llama.cpp 1.30 with Flash Attention used for NVIDIA. Short Prompt: "How long would it take for a ball dropped from 10 meter height to hit the ground?" Long Prompt: "Summarize the following in exactly five lines: [Insert Scene 1 Act 1 of Romeo and Juliet]", Models tested: Phi 3.5 MoE Q4 K M, Mistral Small 3.1 24B Instruct 2503 Q8, DeepSeek R1 Distill Qwen 32B Q6, Qwen 32b Q6 System specifications: AMD Ryzen™ 9 7900X, 32GB DDR5 RAM 6000 MT/s, Windows 11 PRO 24H2, AMD Radeon™ AI PRO R9700 32GB using Adrenalin 25.6.1 RC vs AMD Ryzen™ 9 7900X, 32GB DDR5 RAM 6000 MT/s, Windows 11 PRO 24H2 with NVIDIA GeForce RTX 5080 and GeForce 576.4 drivers. Performance may vary. RPW-495.

RPW-496: Testing as of May 2025 by AMD using DeepSeek R1 Distill Qwen 32B Q6, Mistral Small 3.1 24B Instruct 2503 Q8, Flux.1 Schnell, SD 3.5 Medium models. Tested on a System with AMD Ryzen 9 7900X CPU, Radeon AI PRO R9700 GPU, 32GB DDR5 RAM, 1TB Storage, Windows 11 PRO 24H2, Adrenalin 25.6.1 RC drivers, ComfyUI - PyTorch 2.4 on Windows. System configurations may vary yielding different results. RPW-496

DISCLAIMER: The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale. GD-18u.

© 2025 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Radeon, Radeon AI, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective owners. Certain AMD technologies may require third-party enablement or activation. Supported features may vary by operating system. Please confirm with the system manufacturer for specific features. No technology or product can be completely secure.

