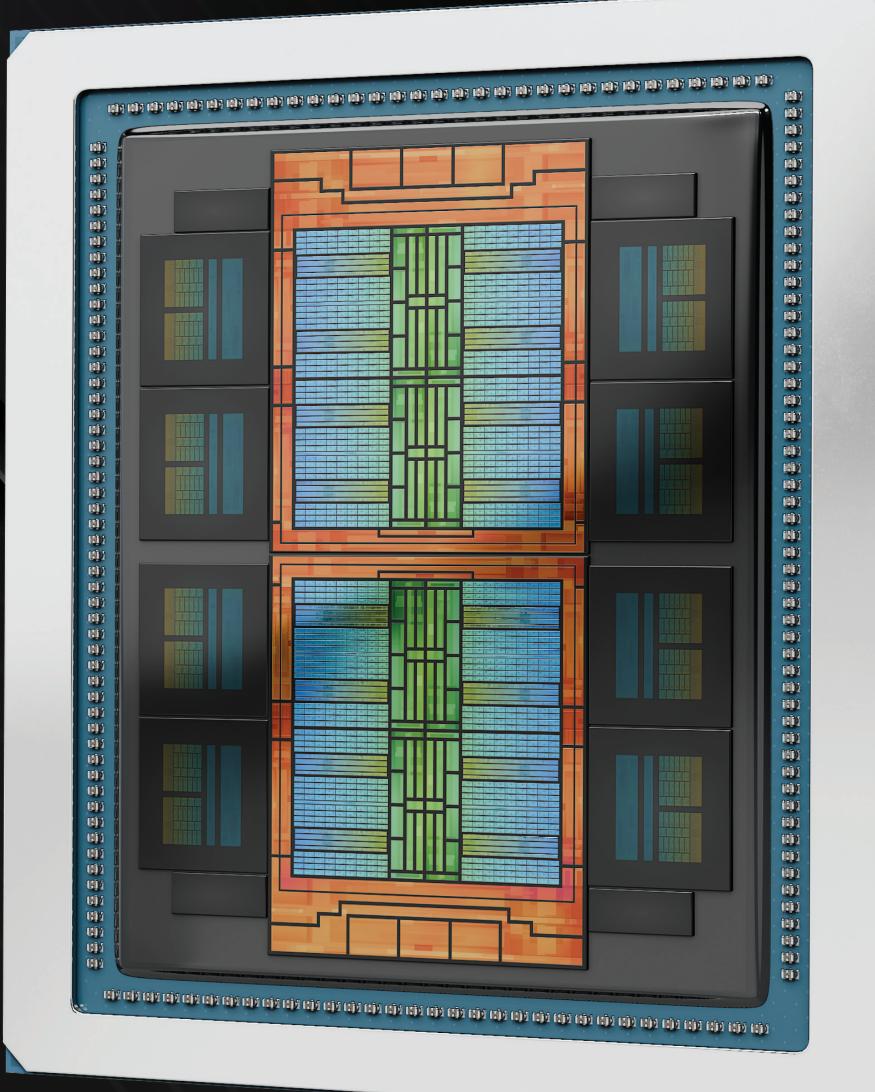
#### AMD INSTINCT

# BOOST AI AND HPC PERFORMANCE IN THE DATA CENTER WITH AMD INSTINCT™ MI350 SERIES GPUS



AMDA INSTINCT MI350 Series

As you deploy your most demanding generative AI, training and HPC applications, get the optimal performance you need. AMD Instinct<sup>™</sup> MI350 Series GPUs and the ROCm<sup>™</sup> open-source software ecosystem help take you to the next level.

# AI EVERYWHERE

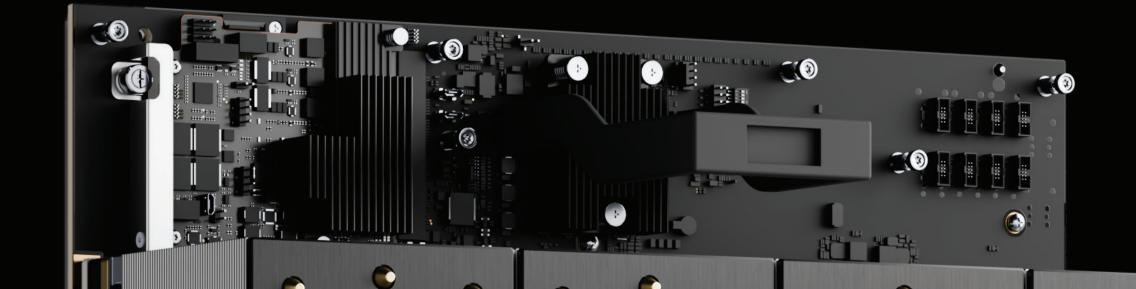
is the ANNUAL GROWTH RATE of the **global Al hardware market**—

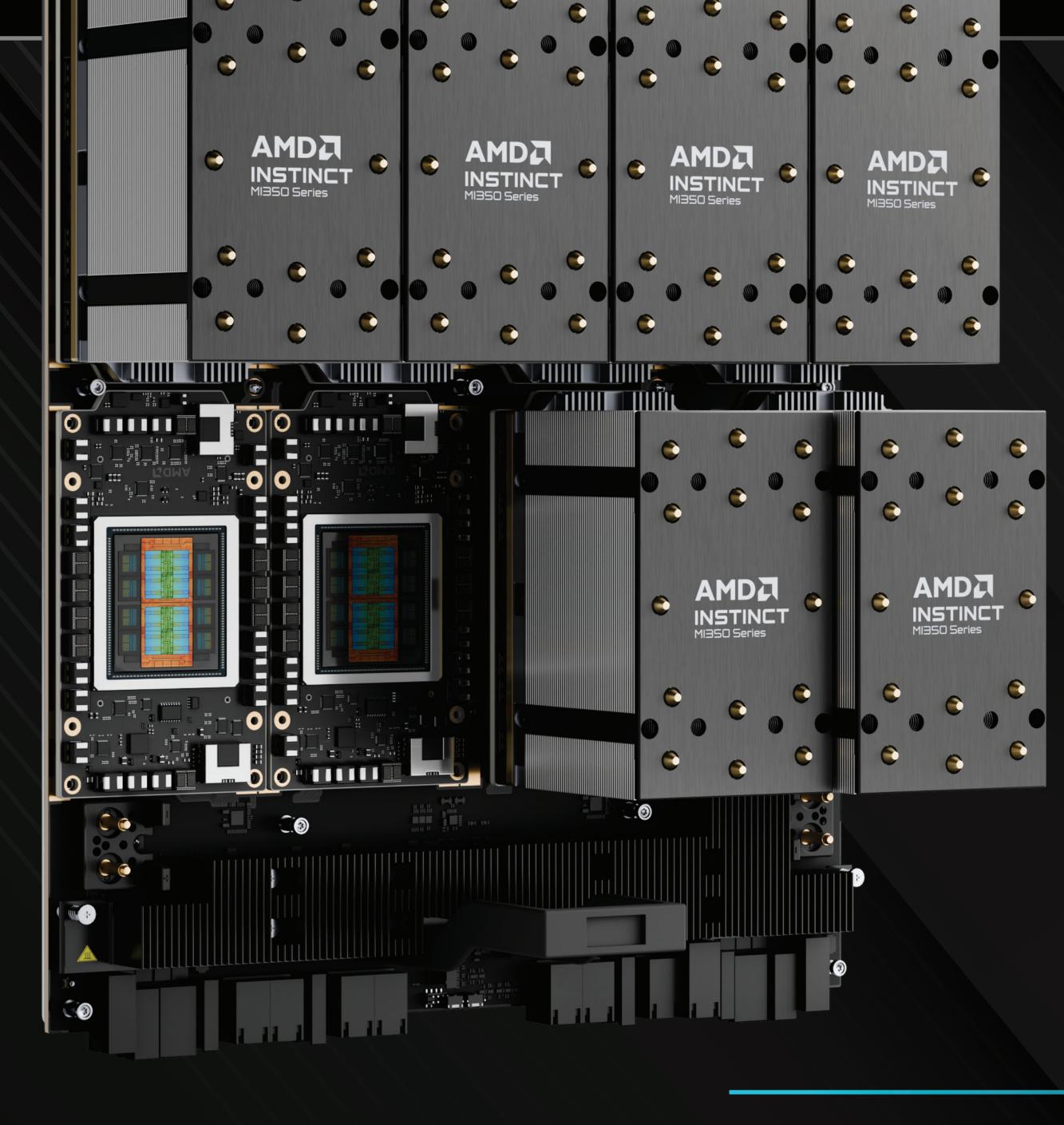
expanding from **\$53.71 billion** in 2023 to a projected **\$473.53 billion** by 2033.<sup>1</sup>

# **C-SUITE WANTS MORE AI**

of TECHNOLOGY EXECS say **investments in data and AI** are a top

organizational priority in 2025, with **57.5%** saying they are achieving *exponential* productivity gains, thanks to AI.<sup>2</sup>





AMDIA Instinct MI350 SERIES GPUS

# LEADERSHIP AI AND HPC PERFORMANCE AT SCALE

#### LEADING HPC

AMD Instinct MI350X and MI355X GPUs offer up to 2x and 2.1x higher peak theoretical FP64 matrix performance, respectively, with structured sparsity vs. the NVIDIA HGX B200 GPU.<sup>MI350-009</sup>

### TRAIN AI FASTER

AMD Instinct MI350X and MI355X GPUs offer up to 77% and 93% higher peak theoretical FP8 matrix performance, respectively, with structured sparsity than the previous-generation AMD Instinct MI300X GPU.<sup>MI350-005</sup>

#### **RESEARCH WITH EFFICIENCY**

Get up to 2.1x and 2.2x the peak theoretical FP6 matrix PFLOPS performance, respectively, with structured sparsity when comparing AMD Instinct MI350X and MI355X vs. the NVIDIA HGX B200 GPU.<sup>MI350-009</sup>

Get up to 2x the peak theoretical FP6 matrix PFLOPS performance with structured sparsity when comparing the AMD Instinct MI355X vs. the NVIDIA GB200 GPU.<sup>MI350-008</sup>

#### **SPEED UP INFERENCING**

#### **GET SUPERPOWERS**

AMD Instinct MI350X and MI355X 8xGPU platforms offer up to 36.8 and 40.3 PFLOPs FP16 matrix performance, respectively, with structured sparsity and up to 147.6 and 161.1 PFLOPs FP6/ FP4 matrix performance, respectively, with structured sparsity.<sup>MI350-010</sup> Embrace scale and efficiency. AMD powers five of the top 10 supercomputer systems on the Top500 list<sup>3</sup> and more than half of the top 20 supercomputer systems on the Green500 list.<sup>4</sup>



#### **FULLY CONNECTED**

A massive 288GB of HBM3E memory and up to 8TB/s bandwidth deliver a huge leap in performance over previous AMD Instinct generations.

#### **OPTIMAL ARCHITECTURE**

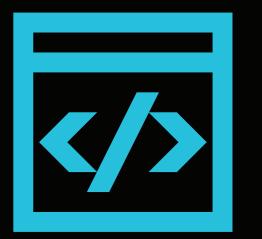
Enhanced processing combined with added next-gen FP6 and FP4 capabilities positions the AMD Instinct MI350 Series GPUs to deliver exceptional performance for generative AI models, pushing the boundaries of AI acceleration.

#### REVVED-UP PERFORMANCE PER WATT

Built on the cutting-edge 4th Gen AMD CDNA<sup>™</sup> architecture, the AMD Instinct MI350 Series GPUs feature more powerful and energy-efficient cores, enhancing performance per watt to drive the next era of AI and HPC innovation.

#### **SPARSITY SUPPORT**

Native hardware support for sparse matrices on AMD Instinct GPUs helps save power, lower compute cycles and reduce memory use during AI and ML training.



# PROVEN, OPEN, CAPABLE AI SOFTWARE

#### AMD ROCm<sup>™</sup> 7.0

The first open-source software ecosystem and programming toolset for AI and HPC/hyperscale-class GPU development and deployment.

#### **LOW-FRICTION ECOSYSTEM**

Drop-in support for major Al frameworks and HPC programming models, a key advantage for rapidly evolving software needs.

#### **POWERFUL PARTNERSHIPS**

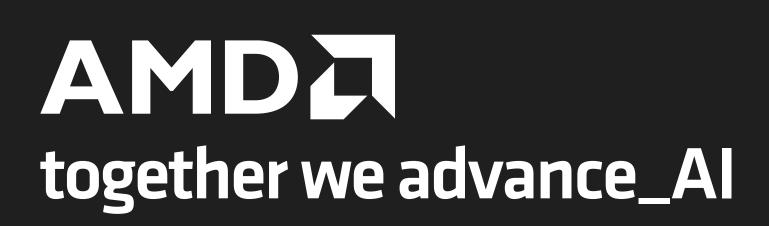
AMD works with AI industry leaders, including Hugging Face, Meta and OpenAI, and participates in open-source cross-platform initiatives such as JAX, Kokkos, ONNX, PyTorch, Raja, Runtime, TensorFlow, Triton and vLLM to enable effortless AI and HPC deployment with minimal code changes.





## POWER AI AND HPC APPLICATIONS AT SCALE WITH AMD INSTINCT<sup>™</sup> MI350 SERIES GPUs.

Learn more at <u>AMD.com/INSTINCT</u> and <u>AMD.com/ROCm</u>.



1 GlobeNewswire, "Artificial Intelligence (AI) in Hardware Market Size to Reach USD 473.53 Bn By 2033," March 6, 2024, https://www.globenewswire.com/news-release/2024/03/06/2841613/0/en/Artificial-Intelligence-AIin-Hardware-Market-Size-to-Reach-USD-473-53-Bn-By-2033.html.

- 2 Data & Al Leadership Exchange and datalQ, "2025 Al & Data Leadership Executive Benchmark Survey," 2025, <a href="https://static1.squarespace.com/static/62adf3ca029a6808a6c5be30/t/67642c0d40b42a7d7e684f49/17346181">https://static1.squarespace.com/static/62adf3ca029a6808a6c5be30/t/67642c0d40b42a7d7e684f49/17346181</a>
  25933/2025+Al+%26+Data+Leadership+Executive+Benchmark+Survey+120624.pdf.
  25933/2025+Al+%26+Data+Leadership+Executive+Benchmark+Survey+120624.pdf.
- 3 Top500 list, November 2024, <u>https://top500.org/lists/top500/2024/11/.</u>

4 Green500 list, November 2024, <u>https://top500.org/lists/green500/list/2024/11/.</u>

©2025 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD arrow, AMD Instinct, Infinity Fabric, AMD CDNA, ROCm and combinations thereof, are trademarks of Advanced Micro Devices, Inc. NVIDIA is a trademark of NVIDIA Corporation in the U.S. and/or other countries. PyTorch, the PyTorch logo and any related marks are trademarks of Facebook, Inc. TensorFlow, the TensorFlow logo and any related marks of Google Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.