Advancing AI

47億人・181ゼタバイト規模のコンピューティングとは

2025年11月6日
日本AMD株式会社

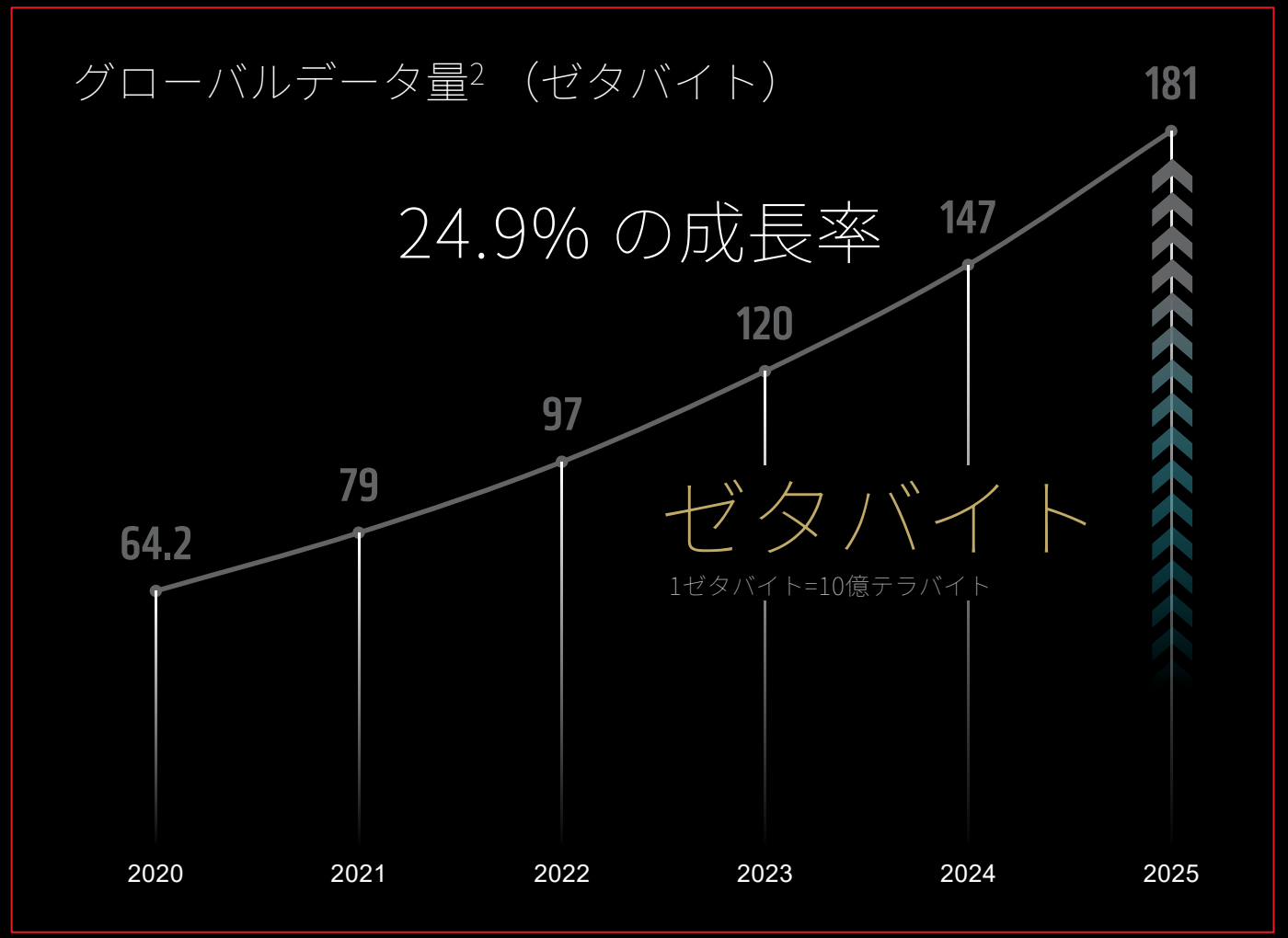
シニアソリューションスペシャリスト
大原久樹

AMD
together we advance_

AIは数十億人の日常生活を変革している

47億人

のスマートフォンユーザーが
日常的にAIの駆動ツールを
使用することによって
データの作成と消費に
貢献している[1]

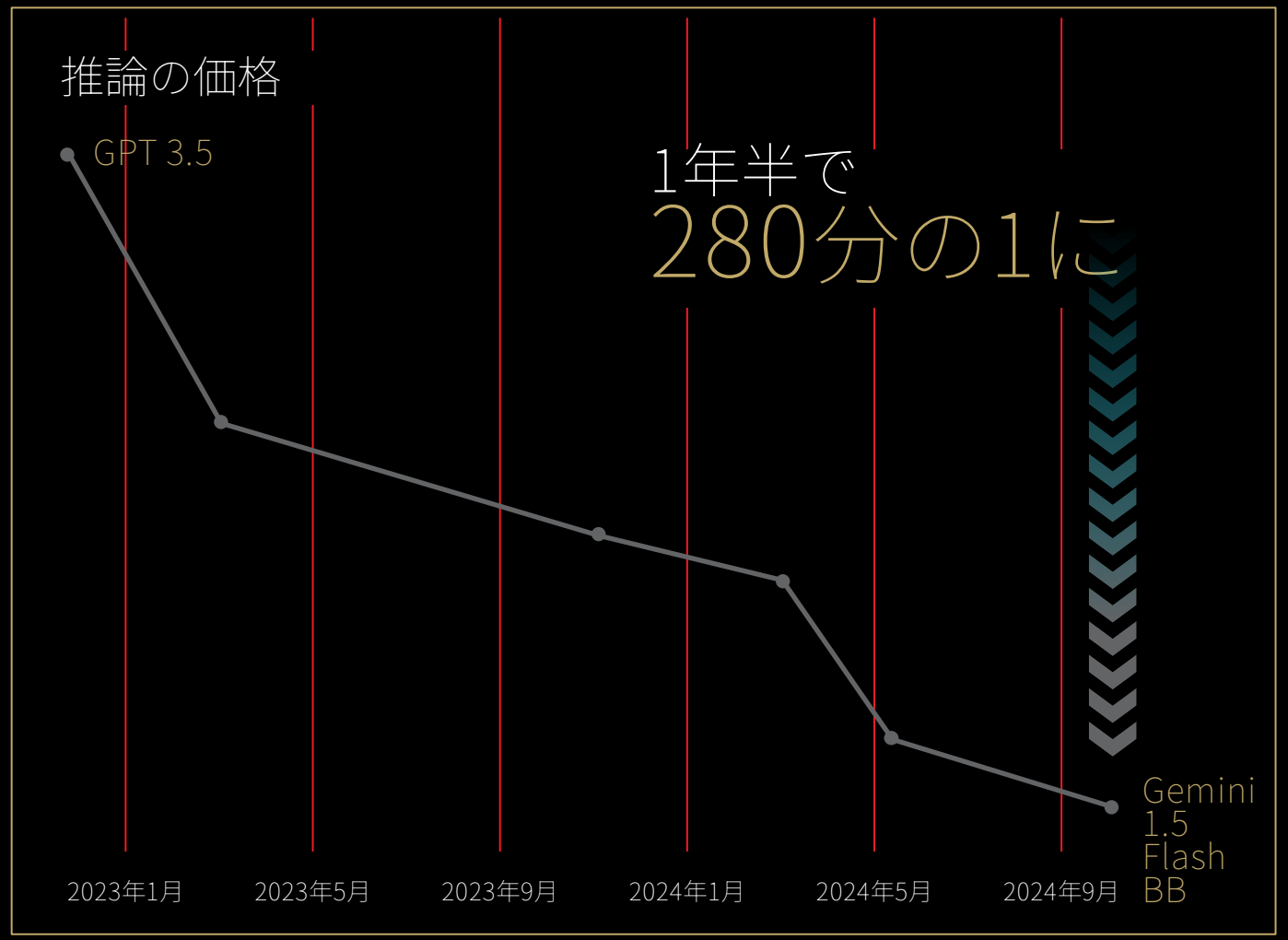


1 Statista 2025年データ；2 IDCによる調査；画像：Freepik

推論コストの急速な低価格化

$20
100万トークン
あたりのコスト
2022年

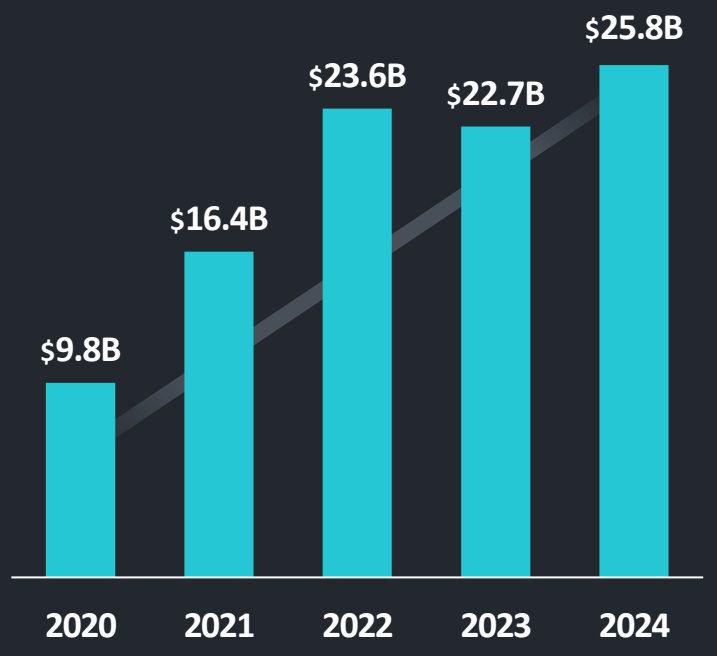
$0.07
100万トークン
あたりのコスト
2024年



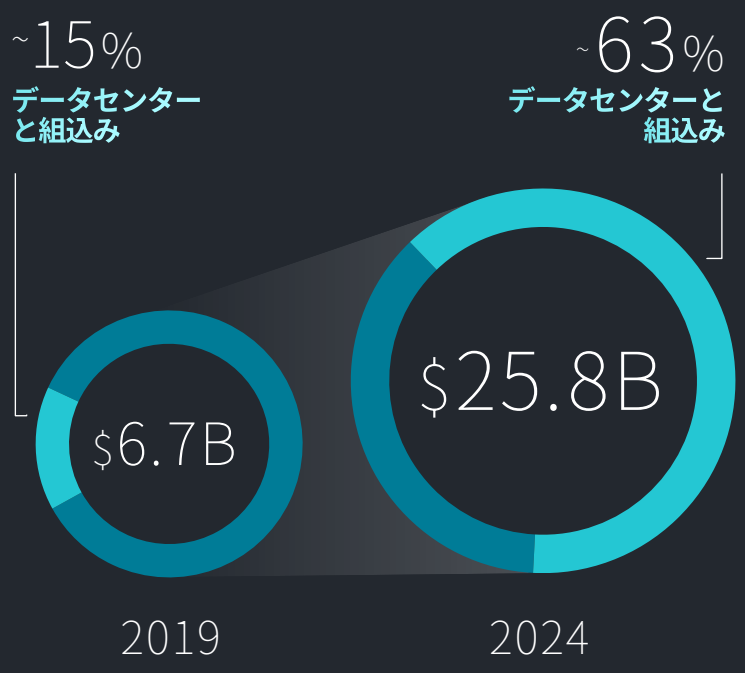
[1]出典：https://hai.stanford.edu/news/ai-index-2025-state-of-ai-in-10-charts；画像：Freepik

強固な財務基盤から、成長を加速

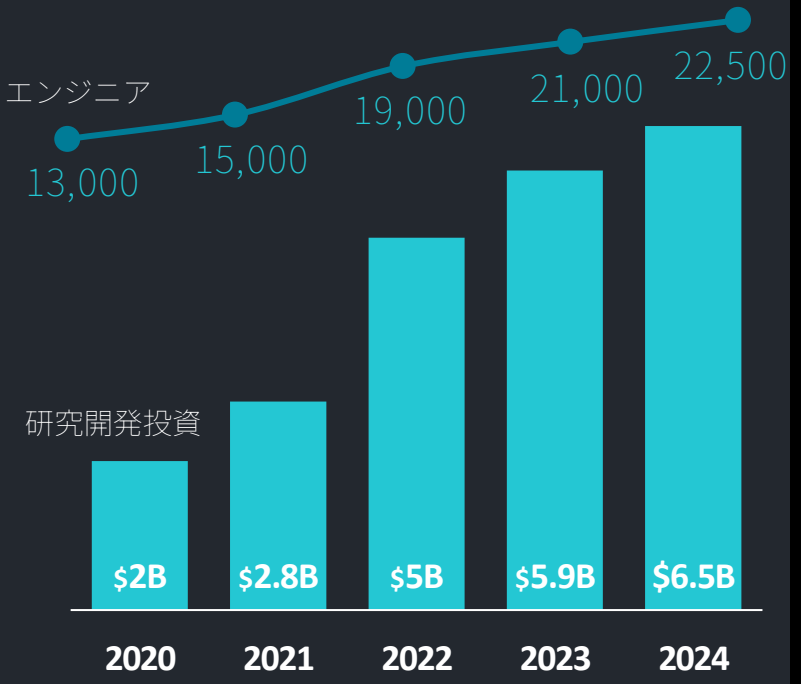
売上成長を加速



売上構成の転換



業界リーダーとしての投資



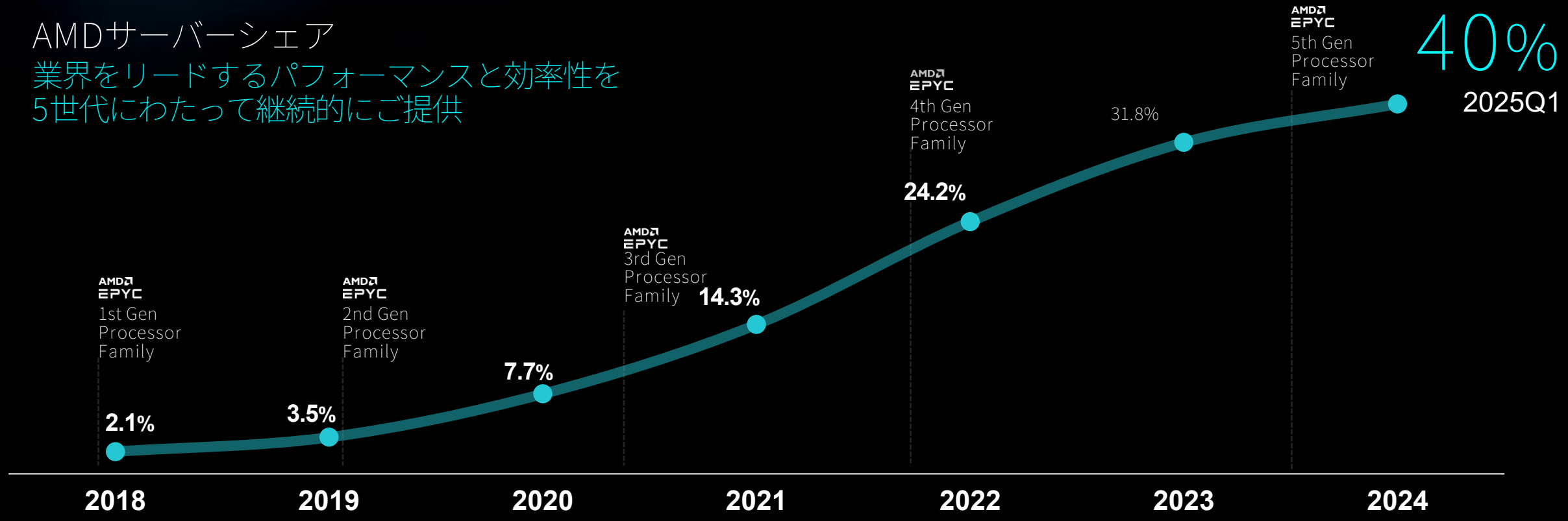
全世界サーバーの3分の1以上で使われている、信頼のAMD EPYC™

すべての主要OEM/ODMからの120を超える“Turin”プラットフォーム

主要MDCからの1000を超えるパブリッククラウドインスタンス

AMDサーバーシェア

業界をリードするパフォーマンスと効率性を
5世代にわたって継続的にご提供



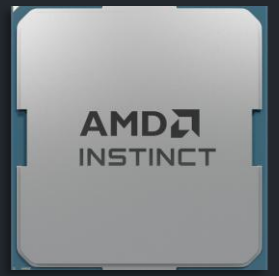
Source: Mercury Research Sell-in Revenue Shipment Estimates, 2024 Q4

AMD
together we advance_

AMD | AIデータセンターを進化させる

CPUs

クラウド、エンタープライズ、AIのための世界最高クラスのCPU



GPU

健全なロードマップ,
メモリー容量・性能におけるリーダーシップ

AMD PENSANDO

Ultra Ethernet Consortium

UA Link | Ultra Accelerator Link

ネットワーク

DPUs, UALink + Ultra Ethernet

AMD ROCm

SILO AI

ソフトウェアソリューション

オープンソフトウェアスタック

zt Systems

クラスターレベルシステムデザイン（ラックスケール）

See endnotes EPYC-029C. ZT Systems announced acquisition expected to close 1H 2025

オープンな開発が、価値とイノベーションを加速

オープンな
ハードウェア

＋

オープン
ソフトウェア

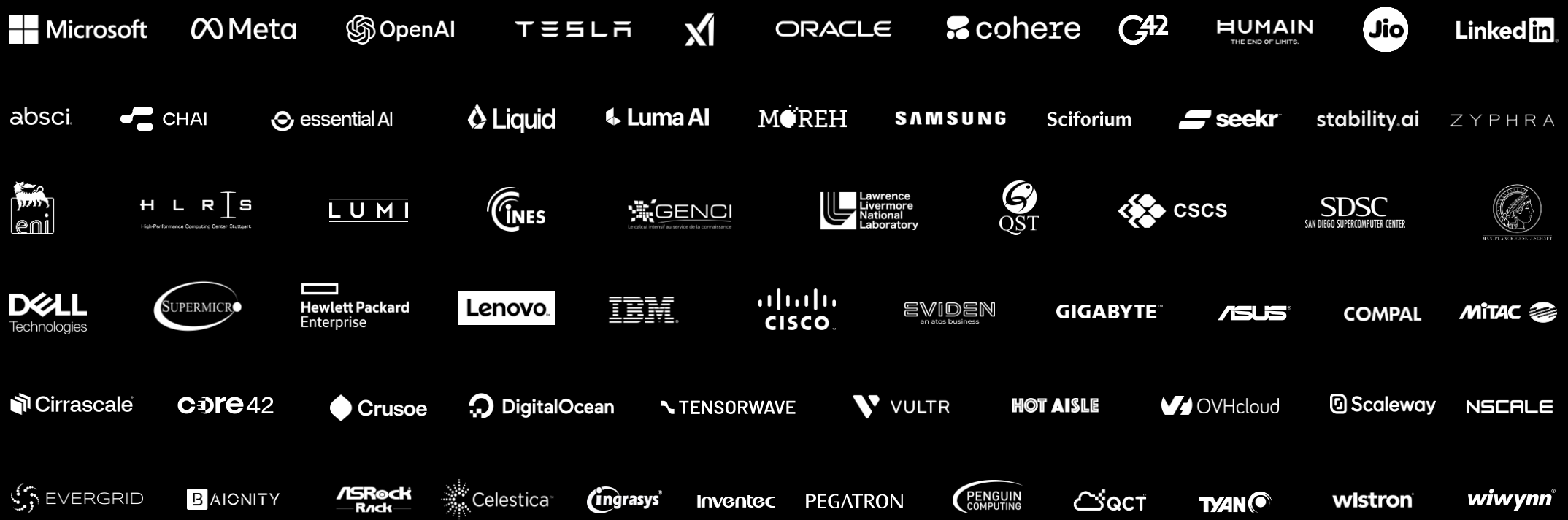
＋

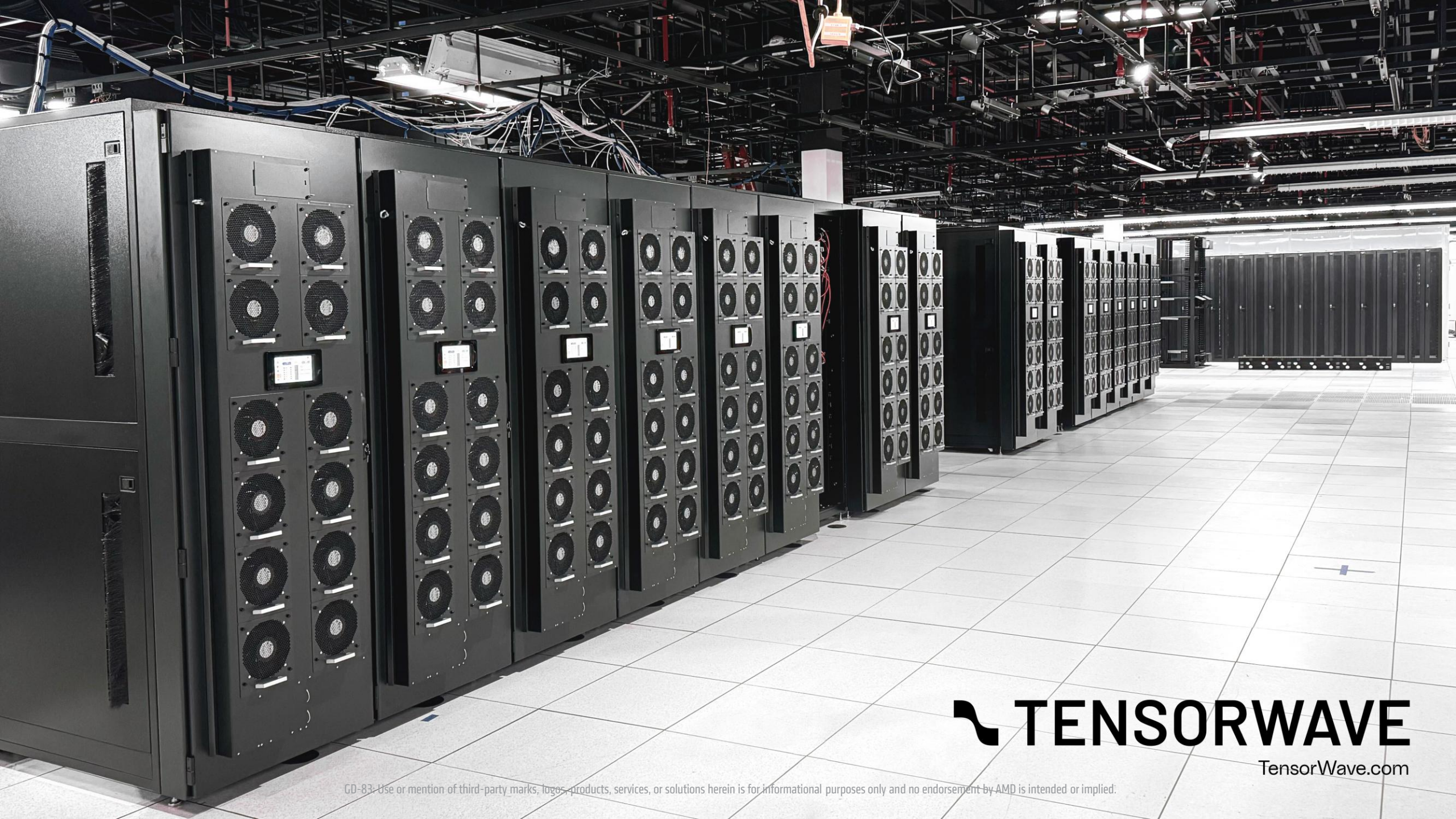
オープンな
エコシステム

選択 | 自由度 | 共創イノベーションの加速 | ポータビリティ | 実績

AMD INSTINCT | 大規模な採用が拡大中

世界のAIトップ10社中7社が AMD Instinct を採用

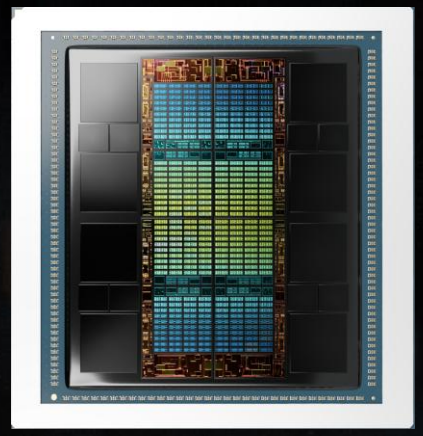




AMD INSTINCT

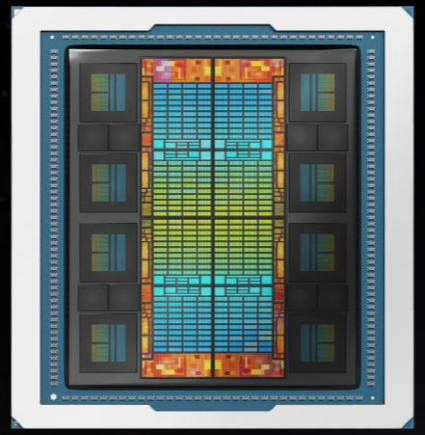
ロードマップを着実に遂行

AMD INSTINCT™
MI300A/X



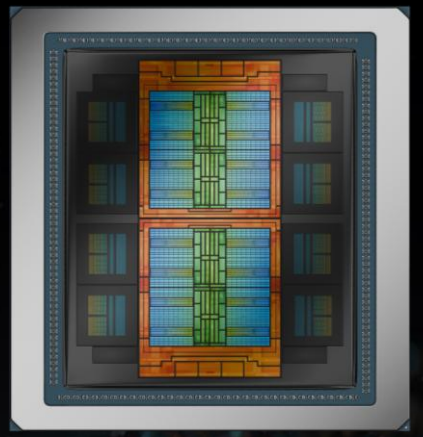
2023

AMD INSTINCT™
MI325X



2024

AMD INSTINCT™
MI350 SERIES



2025

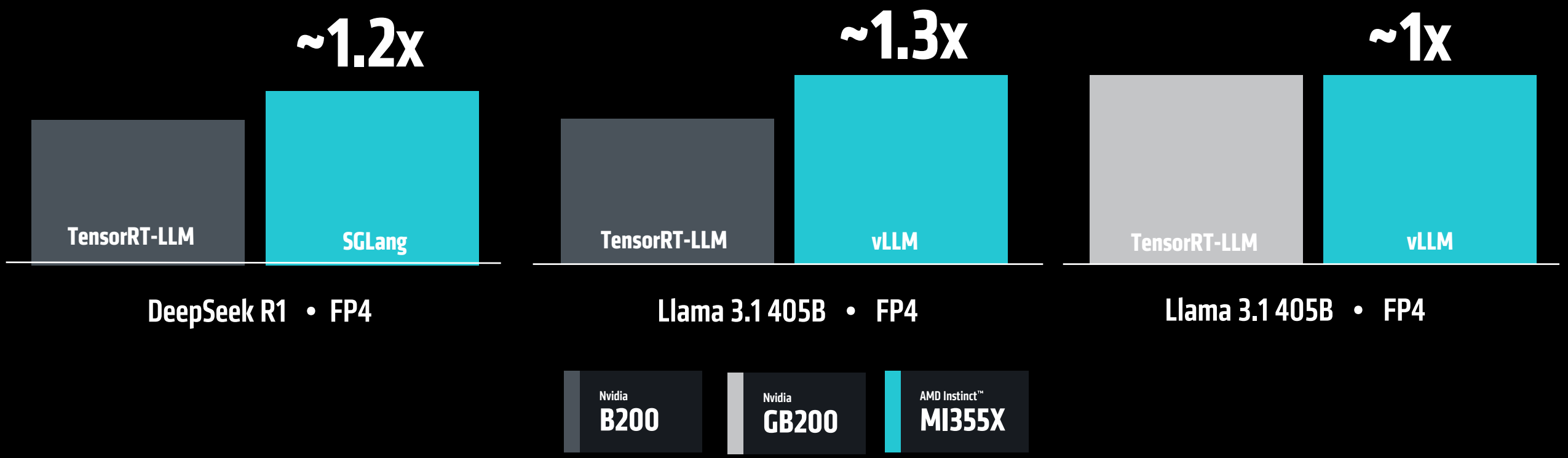
AMD INSTINCT™
MI400 SERIES

2026

Roadmap subject to change

MI355X: 推論のスループットでトップクラスを実現

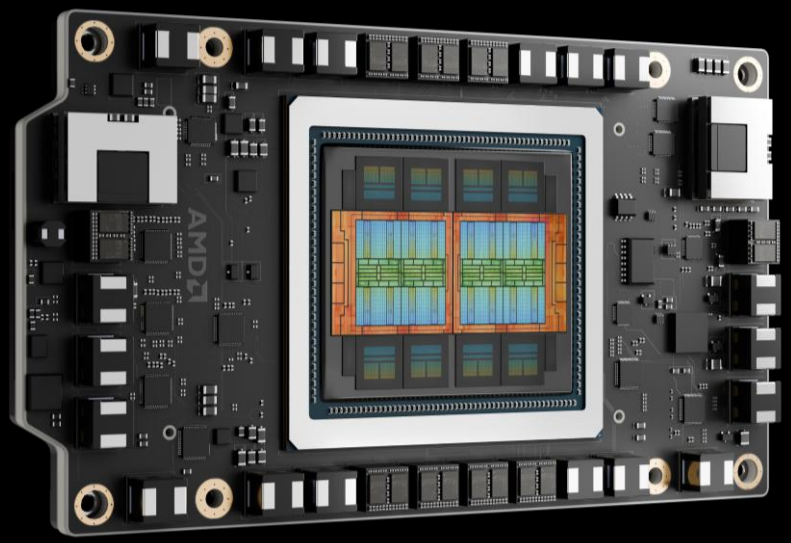
大規模モデルでの比較



*Throughput

Equal GPU Count for Performance Compare with GB200

See endnote MI350-038, MI350-039, MI350-040



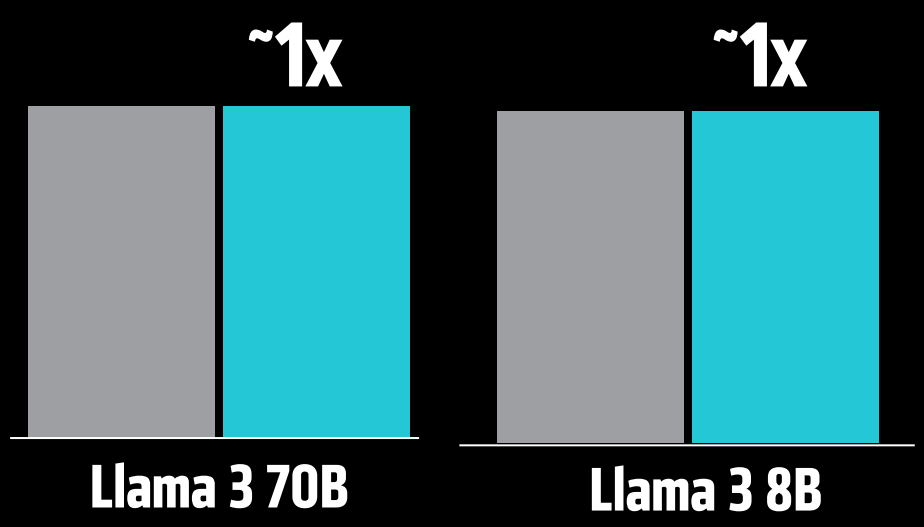
同じコストで最大40%多くのtokenを処理

Instinct MI355X の B200 に対する比較

See endnote MI350-49

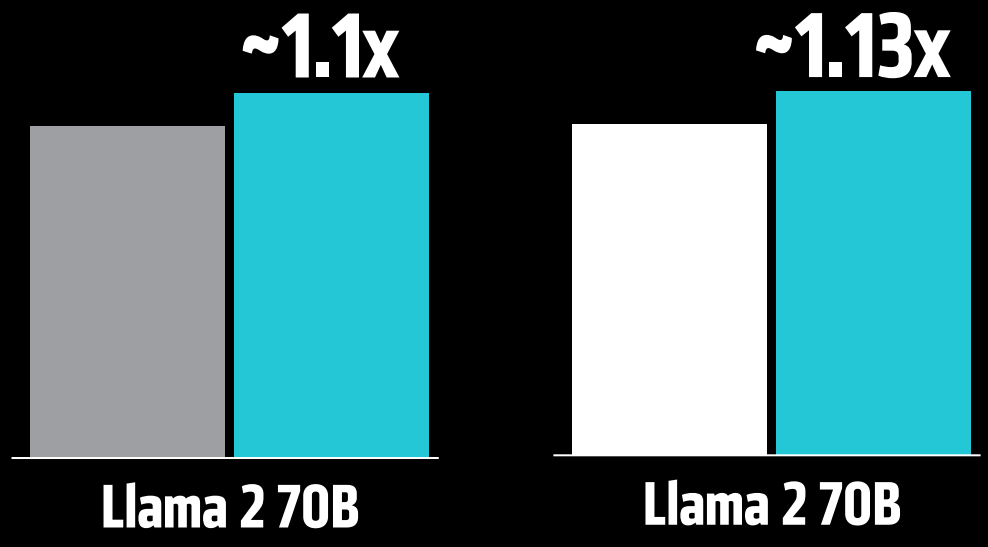
競争力あるトレーニング性能

事前学習 | FP8, BF16



ファインチューニング | FP8

MLPerf 5.0 Unofficial Results

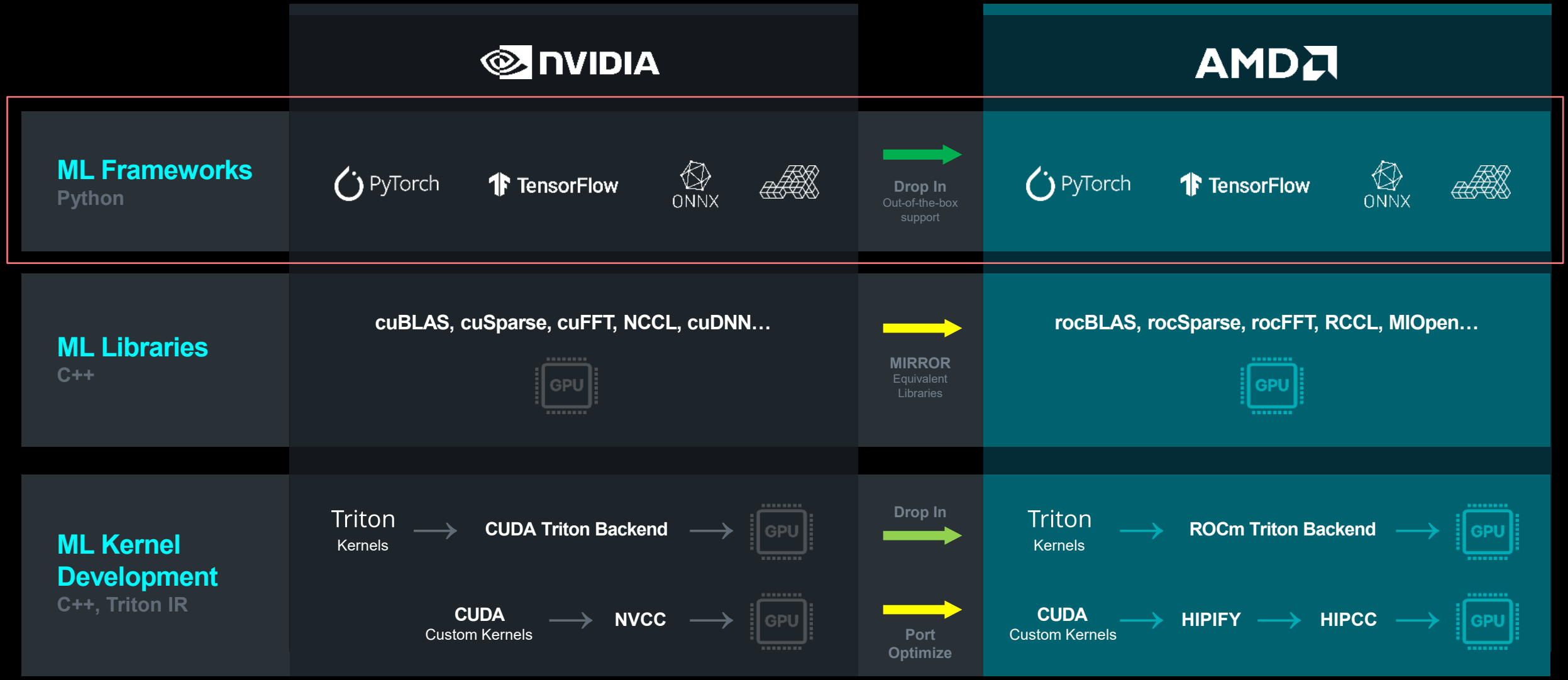


NVIDIA B200 | NVIDIA GB200 | AMD Instinct™ MI355X

Equal GPU Count for Performance Compare with GB200

See endnote MI350-030, MI350-031, MI350-032, MI350-033

NVIDIA から AMD へのスムーズな移行



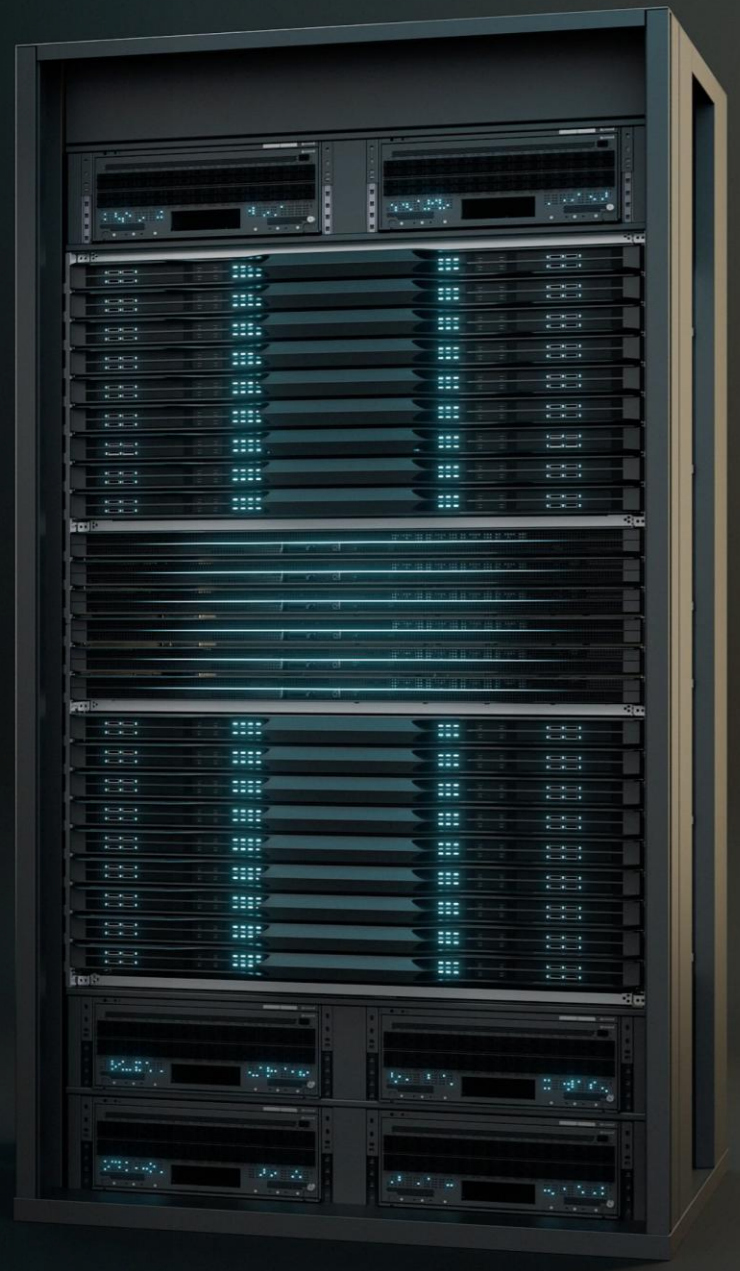
Previewing Today

AMD “Helios” AI Rack

Optimized AI Rack Solution

AMD EPYC | AMD INSTINCT | AMD PENSANDO | AMD ROCm

Available in 2026



OPEN Compute Project | Ultra Ethernet Consortium | ULTRA ACCELERATOR LINK™

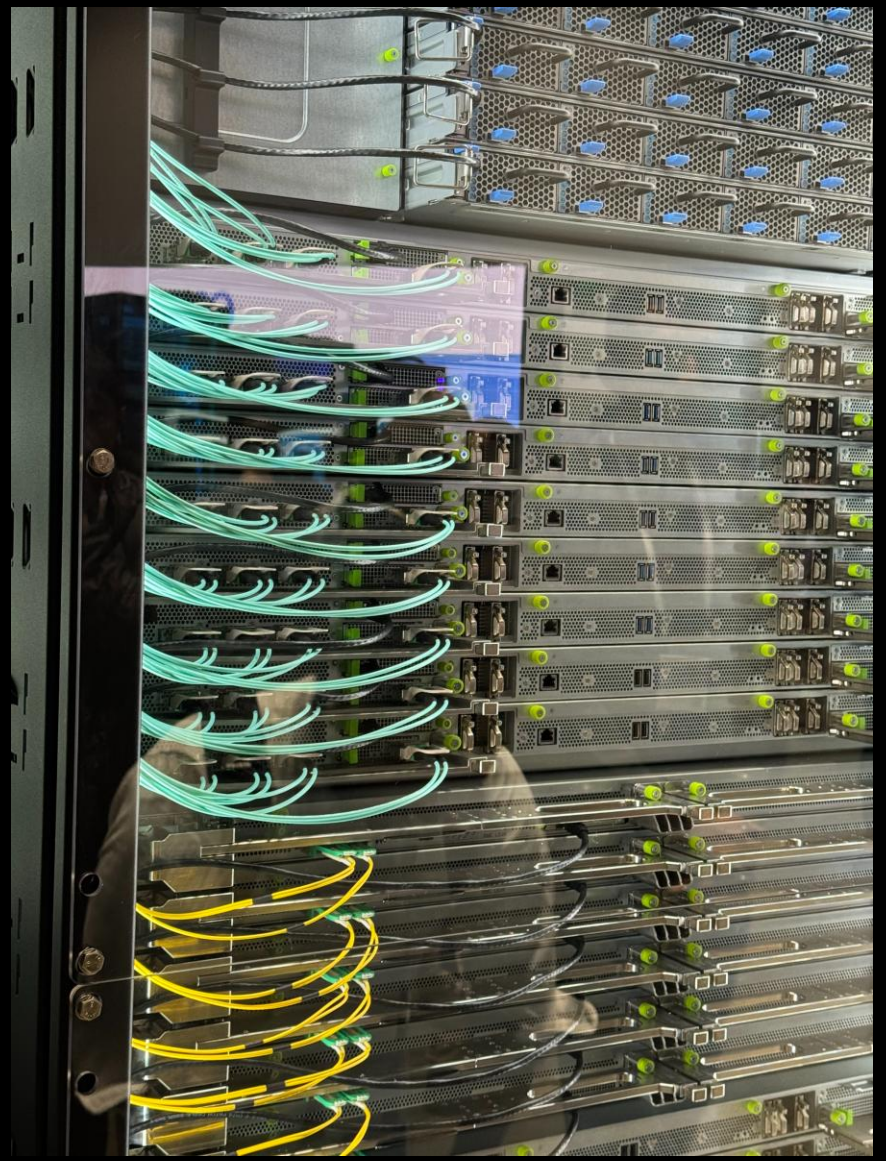
AMD “Helios” AI Rack

Rack Scale AI Performance Leadership

	Instinct™ MI400 Series AMD “Helios”	vs. Vera Rubin Oberon
GPU DOMAIN	72	1x
SCALE UP BANDWIDTH	260 TB/s	1x
FP4 • FP8 FLOPS	2.9 EF • 1.4 EF	~1x
HBM4 MEMORY CAPACITY	31 TB	1.5x
MEMORY BANDWIDTH	1.4 PB/s	1.5x
SCALE OUT BANDWIDTH	43 TB/s	1.5x

See endnote MI350-045, MI350-046A

[OCP Global Summit] Helios (AMDブース)



[OCP Global Summit] Helios (Meta社ブース)

Ultra Ethernet is the answer for scale out AI infrastructure

UEC 1.0 Specification - Q1CY25


AMD