

# 개방형 생태계로 강력한 AI 통찰력 확보

AMD Instinct™ MI300X 가속기 및 ROCm™ 6 개방형  
소프트웨어 플랫폼이 탑재된 Dell PowerEdge XE9680 서버

ChatGPT 및 기타 GenAI(Generative AI) 애플리케이션의 인기가 높아지면서 AI(Artificial Intelligence)가 빠르게 보급되고 있습니다. 그리고 더 많은 조직이 이러한 흐름에 동참하기를 원합니다. 최근 IDC 연구 보고서는 다음과 같은 결과를 보여줍니다.



## 77%의 조직

AI 활용 사례를 탐구하고 있거나 이미 AI 이니셔티브에 상당한 투자를 하는 조직의 비율<sup>1</sup>

가속된 통찰력과 기타 AI 이점을 활용하고자 하는 조직은 이미 행동을 취했어야 하며, 아직 아니라면 조속히 움직여야 합니다. 특히 AI 발전에 적극적으로 참여하는 개방형 생태계와 AI 혁신 및 고객의 비즈니스를 발전시킬 준비가 된 새로운 솔루션이 마련된 지금이 적기입니다.

<sup>1</sup> IDC, Future Enterprise Resiliency and Spending Survey, Wave 6, 2023년 7월.



## 미래를 향해 열린 길

GenAI와 기타 고급 애플리케이션이 증가함에 따라 칩 기술은 더욱 강력해지고 정교해지고 있습니다. 그러나 기초 단계에서 독점 시스템은 일선 개발자들에게 장애물이 되었습니다.

이러한 상황은 곧 바뀔 것입니다. Dell Technologies와 AMD는 오픈 소스 기술을 기반으로 구축된 인프라스트럭처 솔루션을 통해 고객의 역량을 가속하기 위해 협력하고 있습니다. Dell PowerEdge XE9680 서버에서 AMD Instinct™ MI300X 가속기와 AMD ROCm™(Radeon Open Compute) 개방형 소프트웨어 플랫폼을 실행하면 기회를 창출하고 잠재력을 발휘할 수 있는 강력한 생태계 옵션, 즉 개방형 경로가 제공됩니다.

## Dell Technologies와 AMD를 통해 성능과 유연성 확보

Dell Technologies와 AMD는 개발자의 성공을 지원하는 시스템을 통해 전 세계 기업이 성과를 가속하도록 돕고 있습니다.



Dell PowerEdge XE9680 서버는 AI 작업을 위해 특별히 설계된 강력한 데이터 처리 장치입니다. ML(Machine Learning)/DL(Deep Learning) 교육 및 추론 워크로드, 특히 LLM(Large Language Model) 교육에 적합한 8개의 가속기를 지원합니다.

3세대 AMD CDNA™(Compute DNA) 아키텍처로 구동되는 AMD Instinct MI300X 가속기는 GPU 시장에서 가장 큰 고대역폭 메모리 용량을 제공하며, 최대 규모의 AI 모델 교육을 위해 서버당 1.5테라바이트의 응집 GPU 가속기 메모리를 제공합니다. 또한 더 적은 수의 서버로 추론 용량을 늘릴 수 있습니다.

## ROCm 6로 가속화 실현

ROCm은 AMD Instinct의 컴퓨팅 성능을 최대한 활용합니다. 가속 컴퓨팅을 위한 개방적이고 유연한 확장형 소프트웨어 플랫폼인 ROCm에는 AMD Instinct 가속기의 코드 최적화를 지원하고 기존 GPU 가속 코드의 이식 기능을 간소화하는 드라이버, API 및 기타 GPU 툴이 탑재되어 있습니다.

또한 고객의 경험을 대폭 간소화합니다. ROCm은 시스템과 해당 부품 발전에 따라 진화할 수 있는 자유 공유 코드 기반을 통해 디바이스 수준에서 최적화됩니다. 이러한 개방적인 유연성 덕분에 간편하게 이식할 수 있고 코딩이 거의 필요하지 않습니다.

## 국경이 없는 생태계

개방형 표준과 상호 운용성을 기반으로 구축된 AMD Instinct MI300X 가속기와 ROCm은 다양한 AI 프레임워크 및 라이브러리에서 기본적으로 지원됩니다. 따라서 AMD의 성능을 활용하면서도 기존 소프트웨어 시스템과의 호환성을 유지하려는 경우 더 많은 선택권과 유연성이 제공됩니다.

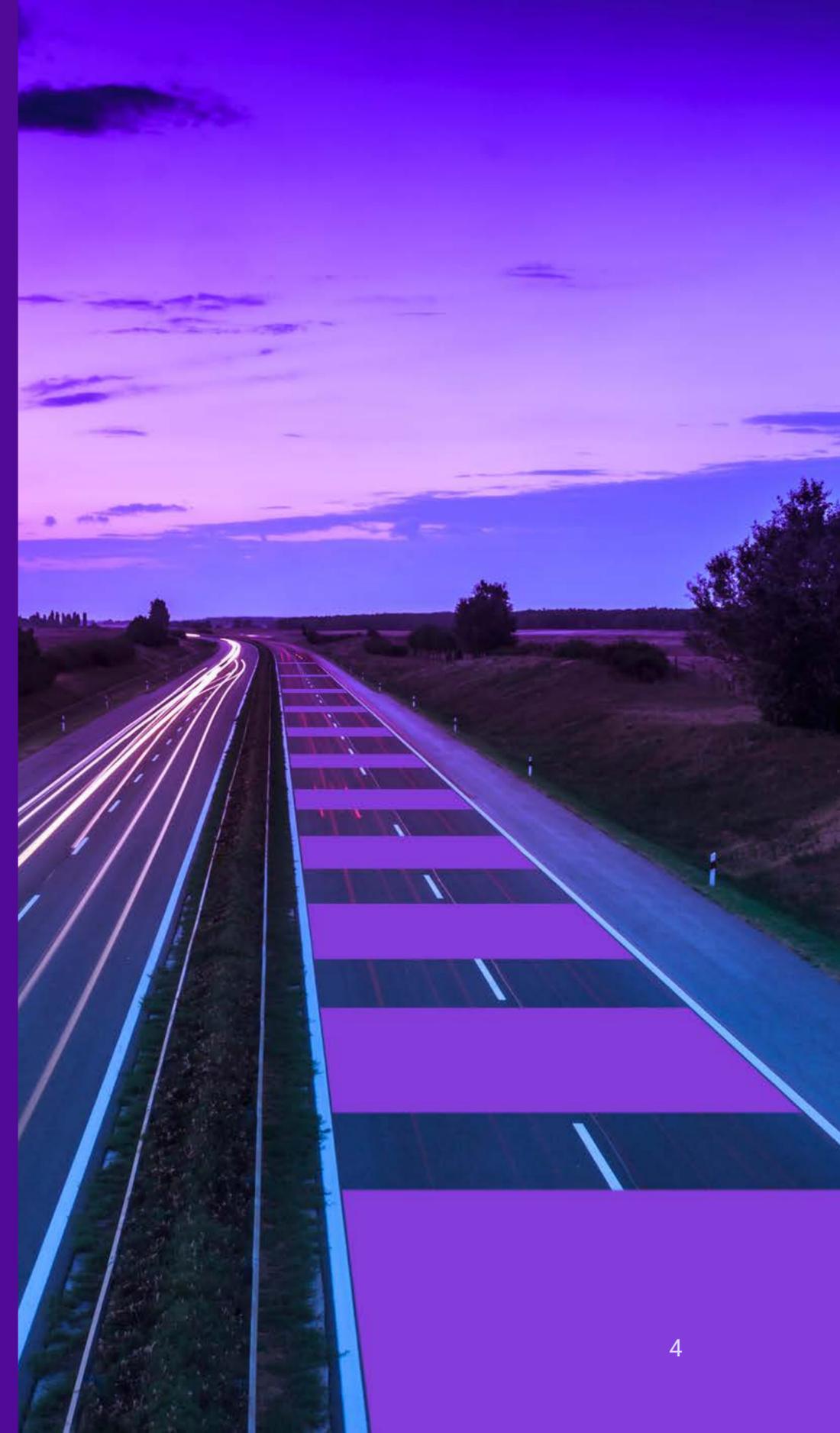


AMD는 장애물을 없애서 프로젝트를 빠르게 진행할 수 있게 해줍니다.

### ROCm의 장점

- GPU 성능을 활용하여 컴퓨팅 작업 가속화
- 대규모 ML/DL/GenAI 워크로드에 최적화
- PyTorch®, TensorFlow™ 등의 주요 모델 및 프레임워크 즉시 지원
- JAX, ONNX(Open Neural Network Xchange), OpenAI Triton 등에서의 기본 지원을 비롯한 새로운 혁신의 신속한 도입 지원
- OSS 개발 커뮤니티와의 연계를 통해 혁신, 차별화 및 협업 촉진
- 개방형 이식성을 위한 유연한 토대 제공

[ROCm의 새로운 기능에 대한 자세한 정보.](#)



## ROCm을 사용하여 AMD Instinct용 애플리케이션을 손쉽게 이식 및 최적화

- 마찰을 최소화한 상태에서 AMD로 소프트웨어를 이식할 수 있습니다. 개발자는 HIP를 사용하여 단일 소스 코드에서 AMD Instinct MI300X용으로 이식 가능한 애플리케이션을 만들 수 있습니다.
- 일반적으로 사용되는 라이브러리 및 연산자로 모든 기능을 제공합니다.
- 대부분의 애플리케이션을 신속하게 이식 및 최적화하여 성능을 향상합니다.

## 업스트림 지원 증가

ROCm은 다음을 비롯하여 오늘날 가장 널리 사용되는 프레임워크에 대한 업스트림 지원을 제공합니다.

- PyTorch
- TensorFlow
- ONNX(Open Neural Network Xchange)
- JAX



## 개방형 이식성으로 소중한 시간 절약

ROCm을 사용하면 개발자는 새로운 생태계와 이 생태계가 기존 코드에 미칠 수 있는 영향을 걱정할 필요가 없습니다. ROCm 플랫폼의 주요 장점은 기존 코드에 대한 재작업을 최소화하면서 애플리케이션을 이식할 수 있다는 것입니다. 이식은 간단하며 시간을 엄청나게 절약할 수 있습니다.

4가지 방법으로 기존 애플리케이션을 AMD Instinct MI300X 가속기에 이식할 수 있습니다.

**1** ROCm에 대한 업스트림을 지원하는 기존 AI 프레임워크를 활용합니다.

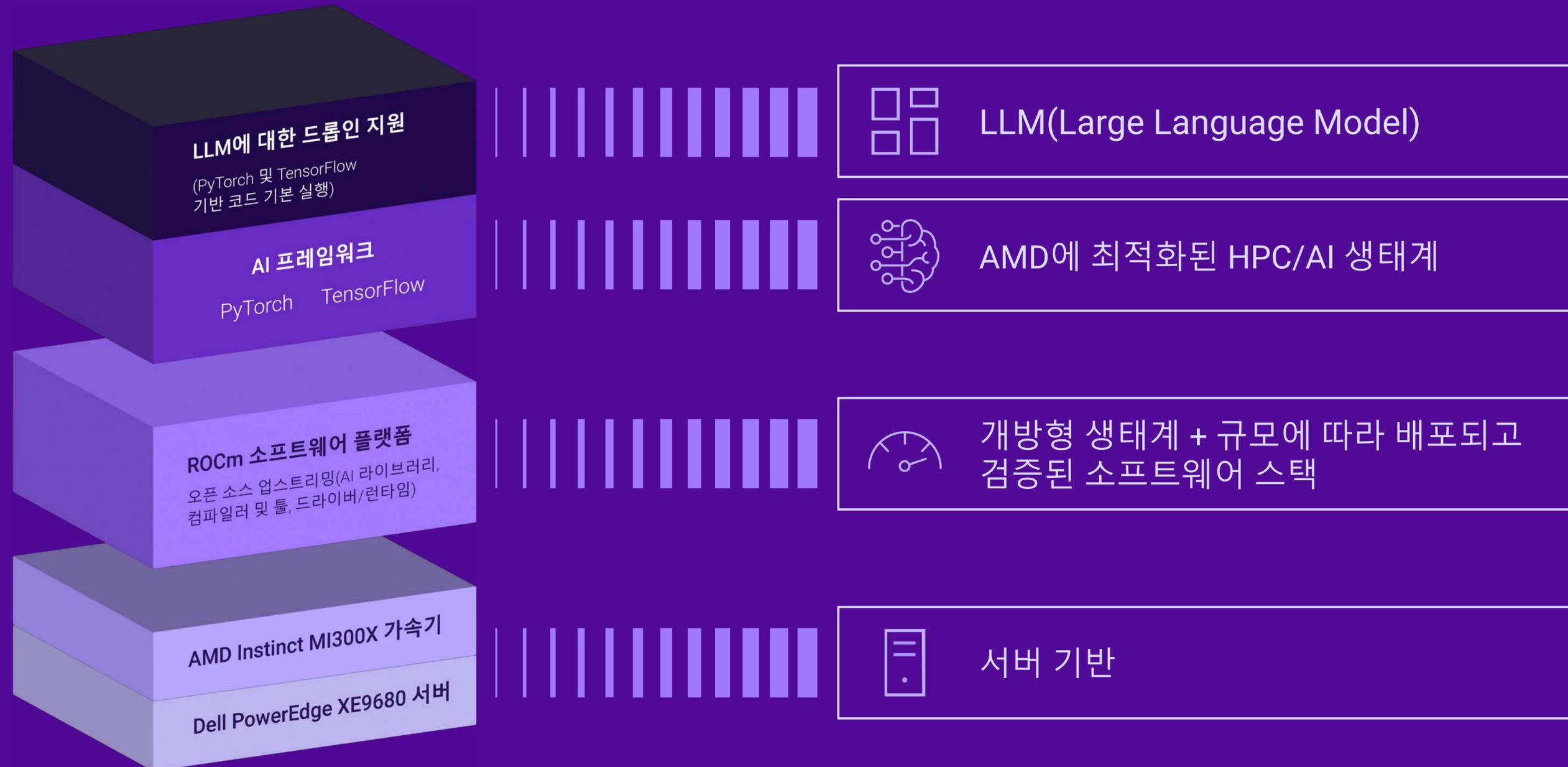
**2** AMD Instinct MI300X에 최적화된 드롭인 ML 커널을 사용합니다. 또는 ROCm 기반 GPU에서 CUDA 애플리케이션을 최적화할 수 있는 AMD HIP 변환 툴을 활용합니다.

**3** ROCm에 최적화되고 유지 관리되는 기존 CUDA 라이브러리 소프트웨어 리포지토리의 미러링 버전을 사용합니다.

**4** AMD의 사전 패키지 AI 애플리케이션 및 컨테이너에 액세스합니다.

## 지금 바로 시작

Dell Technologies 및 AMD의 개방형 프레임워크를 통해 기존 AI와 새로운 GenAI 프로젝트를 가속하십시오. AMD Instinct MI300X 가속기를 비롯한 최신 AMD 가속기를 지원하는 검증된 Dell PowerEdge 기반에서 워크로드를 강화하십시오. ROCm 개발자 허브를 방문하십시오.



## 바로 배포 가능한 AMD Instinct MI300X 가속기

획기적인 성능을 제공하는 AMD Instinct MI300X 가속기는 특히 오늘날의 기업들이 리소스, 복잡성, 레이턴시 및 아키텍처 문제를 처리하는 데 필요한 속도, 확장성 그리고 유연한 개방형 생태계를 제공합니다. AMD Instinct MI300X 가속기는 다음과 같은 측면에서 확실한 이점을 제공합니다.



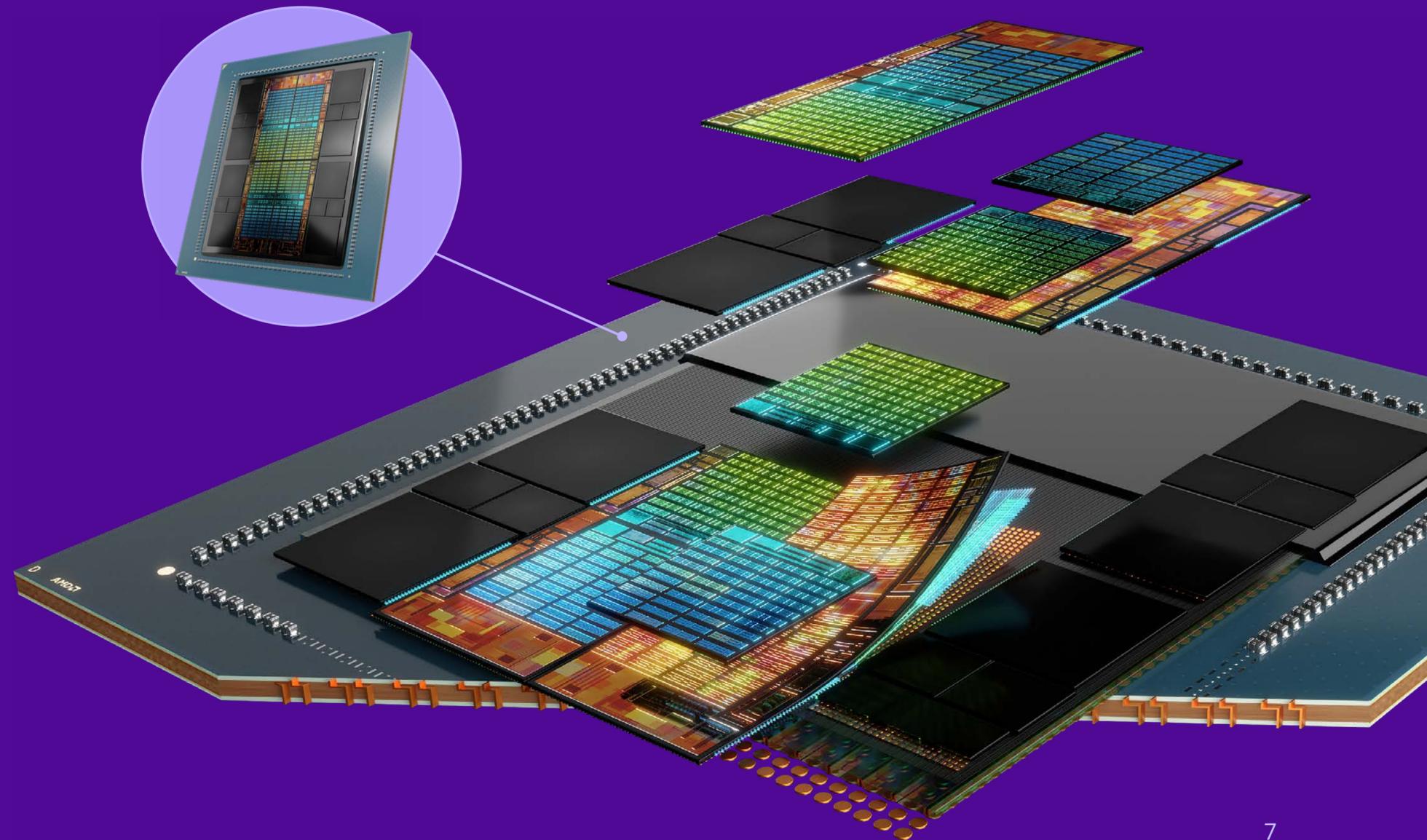
**고성능:** 초당 5.3TB의 속도로 작동하는 GPU당 192GB의 스택형 HBM3 메모리로 새로운 AI 워크플로를 지원합니다.



**집적도:** 서버당 더 많은 컴퓨팅 및 메모리 용량을 제공하는 최첨단 다이 스택킹, 칩렛 기술로 데이터 이동의 필요성을 줄입니다.



**개방성:** 접근성과 적응성을 고려하여 설계된 생태계를 통해 결과를 쉽게 얻을 수 있습니다.



## Dell PowerEdge XE9680 서버로 AI 이점 활용

AI를 위해 특별히 설계된 서버 플랫폼을 통해 더 빠르고 유용한 결과를 얻을 수 있습니다. 통찰력을 가속하고 신뢰할 수 있는 AI를 제공하며 운영을 간소화하는 솔루션으로 데이터 중심 비즈니스를 개선하고, 시각적 결과를 제공하며, 비즈니스 전체의 AI 운영을 가속하십시오.



### 신뢰할 수 있는 AI

신뢰할 수 있는 고품질 솔루션 인프라스트럭처로 AI 수명주기를 가속하여 위험을 줄이고 시간을 절약하십시오. PowerEdge XE9680은 칩 내장형 RoT(Root of Trust)부터 자산 폐기까지 모든 과정에서 내장된 보호 기능을 제공합니다.



### 운영 간소화

AI 인프라스트럭처 자동화를 통해 어디서나 AI 인프라스트럭처와 성능 워크로드를 효과적으로 제어하고 관리하십시오. Dell Technologies는 서로 연동되거나 독립적으로 작동하면서 작업을 빠르게 진행하고 사용자 환경을 단순화하는 지능형 시스템을 제공합니다.



### 통찰력 향상

비즈니스 속도에 맞춰 AI 운영을 제공하고 AI 수명주기 전반에서 향상된 성능을 이끌어내십시오. Dell Technologies는 AI를 최대한의 성능으로 활용할 수 있도록 혁신적인 컴퓨팅을 제공합니다.

# 혁신 추구

Dell Technologies와 AMD는 소중한 고객을 위한 혁신을 추진하기 위해 노력하고 있습니다. 또한 개발자에게 도움이 되는 다양한 솔루션을 통해 시장에 칩 다양성을 선보이고 있습니다. 양사의 공동 솔루션은 다음을 가능하게 만들어 줍니다.



• **시장 진출 시간 단축** - 옵션은 더 많아지고 장애물은 줄어듭니다.

• **더 빠르게 통찰력 확보** - 개방형 생태계로 방해 없이 혁신을 추진할 수 있습니다.

• **가장 적합한 옵션 선택** - Dell 공급망의 이점을 활용하여 가장 필요할 때 혁신적인 기술을 지속적으로 시장에 출시하십시오.

# Dell Technologies 및 AMD와 함께 전력으로 나아가십시오.

이미 AI가 활용되고 있지만, 그 영향력은 앞으로 더욱 확대되고 빨라질 것입니다. 이 기능을 활용하려는 기업에게는 무한한 가능성이 있습니다. 그렇기 때문에 때로는 방해가 될 수 있는 기술적 장애물로 인해 잠재력이 제한되어서는 안 됩니다. 대신 AI 워크로드 전용의 강력한 솔루션 스택으로 여러 가능성을 열어두십시오.

Dell PowerEdge XE9680 서버의 AMD Instinct MI300X 가속기 및 ROCm 소프트웨어가 그 해답입니다. 새로운 애플리케이션으로 전환하는 속도를 높이고, 시장에 더욱 빠르게 출시하며, 비즈니스와 전 세계에 중요한 영향을 미치는 성과를 창출하십시오.

[Dell.com/GPU](https://Dell.com/GPU)에서 자세히 보기

Copyright © 2023 Dell Inc. or its subsidiaries. All Rights Reserved. Dell 및 기타 상표는 Dell Inc. 또는 해당 자회사의 상표입니다. AMD, Instinct™, ROCm™ 및 그 조합은 Advanced Micro Devices, Inc.의 상표입니다. PyTorch®는 PyTorch 또는 PyTorch 사용 허가자의 상표 또는 등록 상표입니다. TensorFlow™ 및 관련 표시는 Google Inc.의 상표입니다. CUDA®는 미국 및 기타 국가에서 NVIDIA Corporation의 상표 및/또는 등록 상표입니다. 모든 상표는 해당 소유주의 자산입니다. Published in the USA 2023년 12월 eBook.

Dell Technologies는 본 문서의 정보가 해당 발행일 현재 정확한 것으로 간주합니다. 이 정보는 예고 없이 변경될 수 있습니다.