

# 인공 지능 (AI)

파트너 구현 패키지

인텔 기반 솔루션으로 고객의 비즈니스 과제 해결

# 프레젠테이션 노트

|           |  |
|-----------|--|
| 프리젠테이션 이름 | 인공 지능 파트너 구현 패키지   |
| 요약        | <p>이 패키지는 PSAM 및 인텔 파트너가 파트너 담당자와의 심층적인 대화에 가이드로 사용할 수 있도록 공개 및 비공개 인텔 콘텐츠로 작성되었습니다.</p> <p>SSP VX 주요 전략적 우선순위에 부합하고 파트너 판매자가 인텔의 AI 투자 및 오픈 생태계를 지지할 수 있도록 AI에 중점을 둡니다.</p> <p>이 패키지는 파트너와의 대화를 지원하도록 맞춤화되었으며 모든 대상 유형에 맞는 콘텐츠입니다.</p> <p>이 프로젝트는 현재 진행 중입니다. 새 버전이 출시될 것이며, 패키지는 정기 업데이트를 통해 최신 인텔 콘텐츠를 포함할 예정입니다. Intel Focus는 Focus는 확장 예정이며, 곧 더 많은 맞춤형 패키지가 제공될 예정입니다.</p> |
| 대상 고객     | 파트너 생태계, 최종 고객, 세일즈  |
| 예정 사항     | <ul style="list-style-type: none"> <li>• GEO 중심 패키지</li> <li>• 파트너 맞춤형 패키지</li> </ul>  |

## 콜 투 액션

1. 파트너와 대화할 때 이 지원 패키지를 활용하십시오.
2. 파트너에게는 **공개** 버전을 공유하십시오.
3. 파트너가 IPA의 회원이 아니라면 가입을 권유하십시오.
4. 지역 또는 파트너에 필요한 맞춤화 제안에 관한 피드백은 Amy Kircos(amy.kircos@intel.com)에게 전달하여 예정된 패키지에 반영하여 주십시오.

# 모든 곳에 AI 도입

모든 플랫폼에서 AI 연속체 구현하기  
클라이언트에서 에지, 데이터 센터 및 클라우드까지

# 인텔과 파트너를 맺어야 하는 이유

인텔은 지구상의 모든 사람과 모든 기업의 삶과 성과를 개선할 기회를 제공합니다.

그러나 혼자 하지 않습니다!

**인텔은 파트너와 함께 모든 곳에 AI를 도입**하고 배포 위험을 최소화하여 고객에게 실질적인 가치를 창출합니다.



인텔과 파트너 관계를 맺으면 완벽한 AI 생태계와 파트너 관계를 맺는 것입니다.  
인텔의 광범위한 AI 지원 기술 포트폴리오와 하드웨어, 소프트웨어, 시스템 통합업체와의 파트너십으로 긴밀히 협력하며 산업과 기업, 커뮤니티에 차별화된 비즈니스 성과를 제공하는 실제 솔루션을 구축합니다.  
여러분의 비즈니스 성장을 지원합니다.

**모든 곳에 AI를 도입하는 여정에 참여하십시오.**

# 인텔 AI 업계 영향력



"인텔은 수백억 달러에 이르는 일생 일대의 비즈니스 기회를 파악하고 모든 산업과 비즈니스 부문에서 퍼베이시브 AI에 필요한 인프라를 구축하느라 바쁘게 움직였습니다."



"준비하세요, 업계에서 말하는 대로라면, 2024년은 AI PC의 해이며, 모든 것이 인텔에서 시작됩니다."



"인텔은 성능이 향상된 AI PC와 새로운 데이터 센터 CPU를 동시에 출시한 이후, AI 분야에서 점차 자체 이점을 만들어가고 있다는 것을 감지하기 시작했습니다."



"무엇보다, 인텔의 새로운 칩도 예정대로 출시되었으며, 회사의 턴어라운드가 순조롭게 진행되고 있음을 입증합니다."



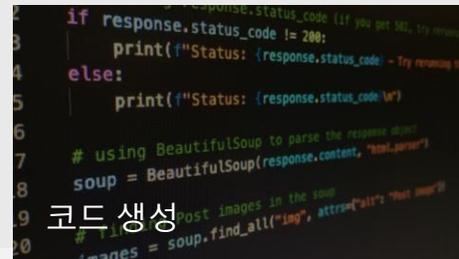
"인텔은 AI가 2024년까지 모든 분야에 적용될 것임을 알고 있으며, 인텔 프로세서가 Windows와 같은, 인터넷 및 컴퓨터 운영 체제에 넘쳐날 모든 소프트웨어 기술의 기반이 되기를 원합니다. 이를 통해 몇 번의 클릭만으로 좋아하는 노래를 재편집하거나 컴퓨터에서 쉽고 빠르게 여행 사진을 모델링할 수 있게 됩니다. 인텔® 코어™ Ultra는 모든 사람을 아티스트, 작가, 음악가로 만들어 줄 것입니다."



"인텔은 인상적인 수치와 주장을 제시할 뿐만 아니라, 새로운 실리콘이 구현할 AI 워크로드의 종류에 대한 구체적이고 현실적인 사례를 제공했습니다. 예를 들어, 레스토랑은 개인 예산과 식습관 요구 사항에 따라 저녁 메뉴 선택을 안내할 수 있고, 제조업체는 공장 현장에서 품질 및 안전 문제를 파악하는 새로운 시스템을 구축할 수 있게 될 것입니다. 인텔의 실리콘 기반 고급 AI는 의사가 놓칠 수 있는 문제를 포착할 수 있는 더 효과적인 초음파 시스템 개발로도 이어질 것입니다."

# 전 세계 비즈니스를 혁신하는 AI

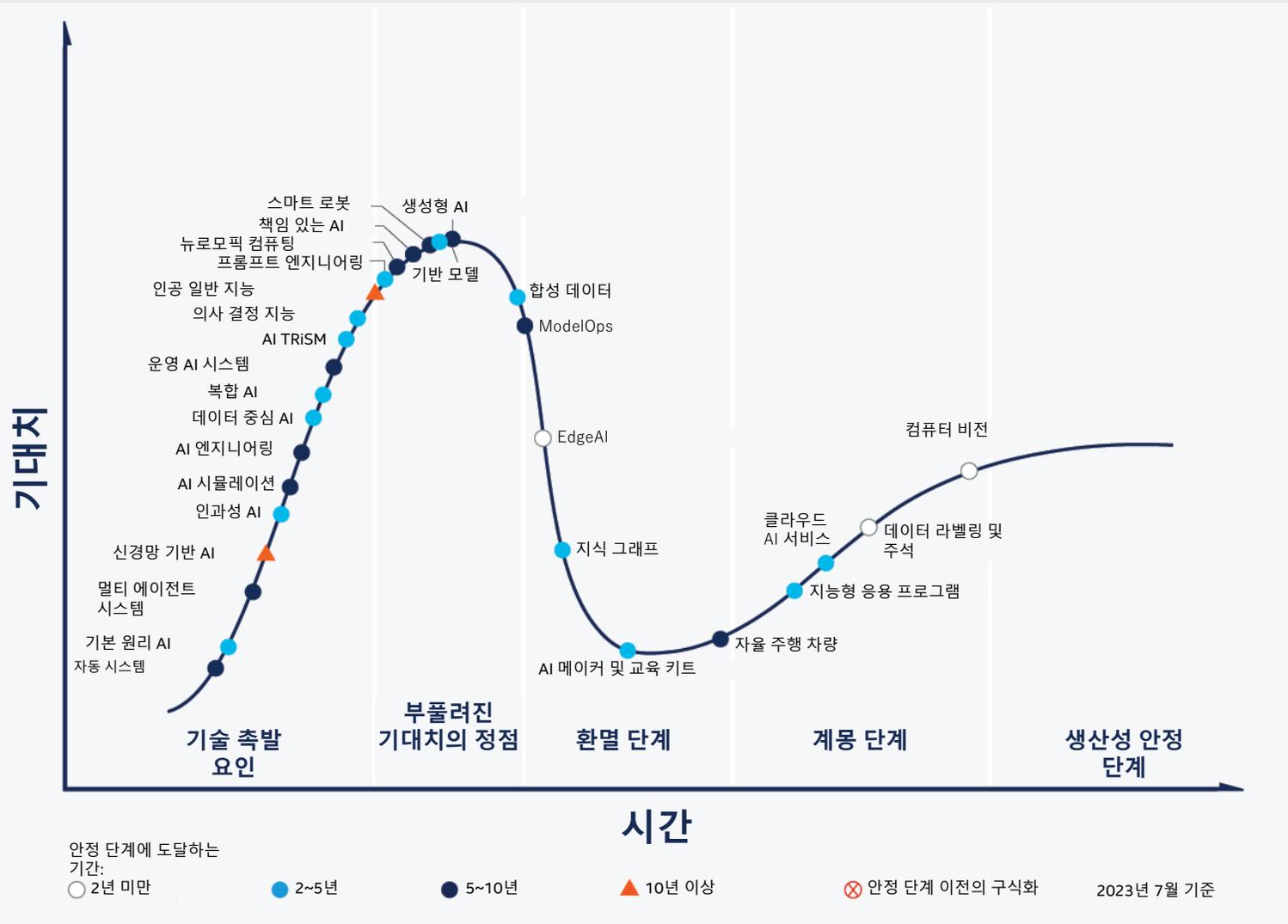
기업을 위한 이점은 무엇입니까?  
비즈니스는 AI를 활용하여 수익을 높이고  
효율성을 향상할 수 있습니다.



# Gartner AI Hype Cycle

2023년 인공지능(AI)용 Gartner Hype Cycle™은 중요하고 혁신적인 이점을 제공하는 동시에 오류가 발생할 수 있는 시스템의 한계와 위험을 해결하는 혁신과 기술을 식별합니다.

AI 전략은 어떤 것이 투자에 가장 신뢰할 수 있는 사례를 제공하는지 고려해야 합니다.



"이러한 혁신을 초기에 도입하면 상당한 경쟁 우위를 확보할 수 있으며, 비즈니스 절차 내에서 AI 모델을 활용하는 것과 관련된 문제를 완화할 수 있습니다."  
 Gartner 이사 분석가, Afraz Jaffri



# 빠르게 진화하는 AI

기본 데이터 기술:



3000억  
달러

2026년까지 전 세계 GenAI 지출 3000억 달러 초과 예상

모든 분야에서 AI 활용

2026년까지

50%

이상

데이터 센터 또는 클라우드 외부에서 생성 및 처리되는 엔터프라이즈 관리 데이터의 비율

50%

AI를 포함할 에지 배포의 비율

58%

AI에 적극적으로 투자하는 주요 대기업 CEO의 비율

인터넷만큼이나 파괴적인 AI

생성형 AI가 2040년까지 세계 경제에 최대 4.4조 달러의 가치를 창출할 것으로 예상<sup>2</sup>

컴퓨팅 비용을 증가시키는 AI 추론, 무어의 법칙 속도를 초과

대형 모델의 성장(IT+ 매개변수 모델)

더 작고 빠른 모델의 성장(~100억 개 매개변수)

<sup>1</sup> <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights>  
<sup>2</sup> Worldwide Artificial Intelligence Spending Guide (IDC)

# 비즈니스 기회 사용 사례 예시



## 교육

교사 보조

학생 학습 지원

학부모 채팅 포털



## 보건

약물 검색

의사 협력 진단

환자 가족 챗봇



## 금융

알고리즘  
트레이딩

고객 포트폴리오  
보조

위험/신용 평가



## 리테일

제품 프로모션

고객 인터페이스  
및 감성 도구

이미지 쇼핑 지원



## 정부

정부 서비스 챗봇

문서 검색 요약

라이브 언어 번역



## 에너지

소비 예측

운영 성능

에너지 트레이딩  
지원



## 자동차

자동차 개발

차량 내 다국어  
지원

공급망 최적화



## 제조

공장 자동화

예측적 유지보수

정밀 농업



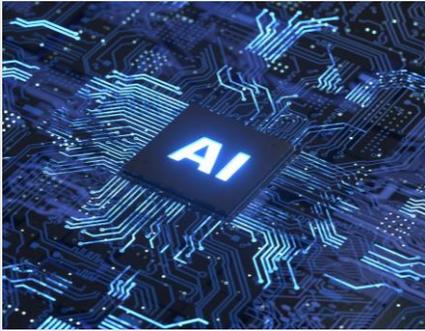
## 통신

맞춤형 고객  
서비스

네트워크 자동화

운영 성능

# 오늘날 AI가 직면한 과제는 무엇입니까? 인텔과 파트너를 맺어야 하는 이유



## GPU 가용성

글로벌 GPU 부족에 대한  
인텔 CPU 대안

Nvidia의 GPU 가격 인상과 GPU의 글로벌 부족에 대한 글로벌 정보 기술 기업들의 불만이 점점 커짐에 따라 [Naver는 AI 서버를 전환](#)하였습니다.



## 공급업체 종속

오픈 소스 표준 기반  
소프트웨어로 공급업체  
종속 방지

[인텔은 모든 업계 표준 개방형 프레임워크](#) 및 라이브러리와 협력하여 최고의 성능을 최적화하고 인텔 기술에서 즉시 사용할 수 있는 고품질 경험을 보장합니다.



## 비용

인텔은 4세대 인텔® 제온®에서 더 나은 가격과 성능을 제공합니다.

[인텔은 실제 업무 상황 적용](#)에서 더 나은 성능과 저렴한 가격, AI 추론을 위한 보다 균형 잡힌 플랫폼을 제공하여 업계의 파괴적 혁신을 주도하고 AI를 대중화하고 있습니다.

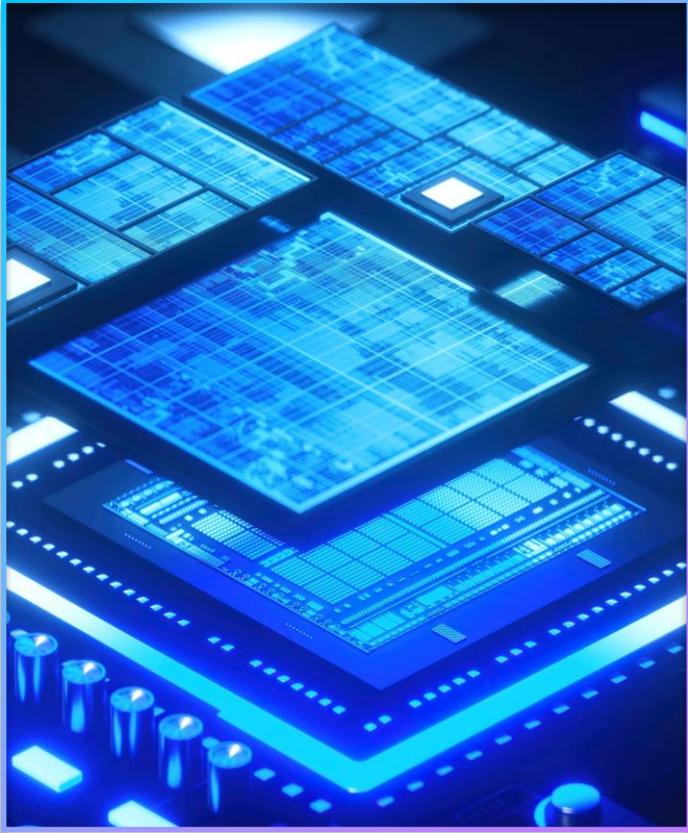


## 보안성 AI

가장 포괄적인 솔루션을  
제공하는 인텔  
보안 포트폴리오

[인텔 보안 기능](#)을 통해 워크로드에 적합한 신뢰 경계를 설정하여 고급 공격, 변조 및 도난으로부터 민감한 데이터, 콘텐츠 및 소프트웨어 IP를 보호할 수 있습니다.

# 인텔이 AI에 제공할 수 있는 것: 탁월한 엔지니어링 AI 플랫폼



- 모든 산업에서 AI를 향한 경쟁이 치열합니다. 인텔은 모든 종류의 하드웨어 및 소프트웨어 플랫폼을 갖춘 유일한 회사로 기하급수적으로 성장하여 모든 분야에 AI가 있는 시대에 파트너에게 필요한 경쟁력 있는 TCO 및 가치 창출 시간을 위한 개방형 모듈식 솔루션을 제공합니다.
- 인텔 기술에 AI를 접목하여 오늘날의 GenAI 워크로드를 지원하고, AI PC와 에지에서의 AI와 같은 새로운 용도를 촉진하며, 향후 10년 동안 AI의 미래를 발전시킬 혁신을 개척하고 있습니다.
- IP, 프로세스, 패키징, 보안, 소프트웨어, 서비스, 제조 및 파운드리 서비스 분야의 리더십을 통해 AI의 전체 잠재력을 최대한 실현하고, 산업을 혁신하며, 세계에서 가장 큰 과제를 해결할 것입니다.

# 인텔이 AI에 제공할 수 있는 것: 개방형 및 하드웨어에 구애받지 않는 소프트웨어 접근 방식



- 개방적인 접근 방식과 개발자 생태계의 본격적인 참여는 진입 장벽을 낮추고 개발자와 고객을 위한 AI 혁신을 실현하는 데 필수적입니다. 인텔은 독점적인 벽으로 둘러싸인 정원을 허무는데 필요한 개방형 AI SW 생태계를 촉진하고 있습니다.
- 인텔은 파트너와 개발자에게 인텔® Developer Cloud와 통합되고 확장 가능한 하드웨어/소프트웨어 시스템 및 솔루션을 통해 AI 솔루션을 확장할 수 있는 조기 액세스와 가장 빠른 경로를 제공합니다.

# 인텔이 AI에 제공할 수 있는 것: 신뢰할 수 있는 플랫폼 및 솔루션



- AI는 윤리적이고 책임감 있게 사용될 때만 모두가 진정으로 이용할 수 있습니다. 인텔은 업계와 협력하여 AI를 더 안전하게 만들고, 개인정보 보호가 기하급수적으로 확장됨에 따라 개인정보 문제를 해결하는 데 도움이 될 혁신적인 생태계 도구 및 솔루션을 제공하고 있습니다.
- 인텔은 파트너가 데이터 센터, 클라우드, PC, 에지 전반에서 다양한 AI 워크로드를 안전하게 보호할 수 있도록 AI와 보안을 융합하는 플랫폼과 기술을 구축하고 있습니다.

# 인텔이 모든 분야에 AI를 도입하는 방법

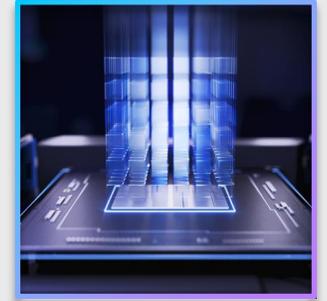
지구상에 있는 모든 인류의 삶을 개선하여 세상을 바꾸는 기술

## 인텔의 고유한 가치

- 개방형 접근 방식
- 하드웨어 및 소프트웨어 전반의 전문 지식
- 에코시스템
- 실행



인텔의 광범위한 AI 지원 기술의 포트폴리오, 미래 AI 강화 혁신에 대한 고유한 비전, 개방형 생태계에 대한 독보적인 지원은 모든 사람에게 혜택을 주도록 **모든 곳에 AI를 도입**하는 데 도움이 됩니다.



- 모든 워크로드 전반에서 AI를 확장하여 개인과 조직이 액세스할 수 있도록 지원
- 다양한 아키텍처, 개방형 표준, 고객이 데이터 센터, 클라우드, PC 및 에지에서 다양한 AI 워크로드를 안전하게 자신 있게 보호할 수 있는 솔루션

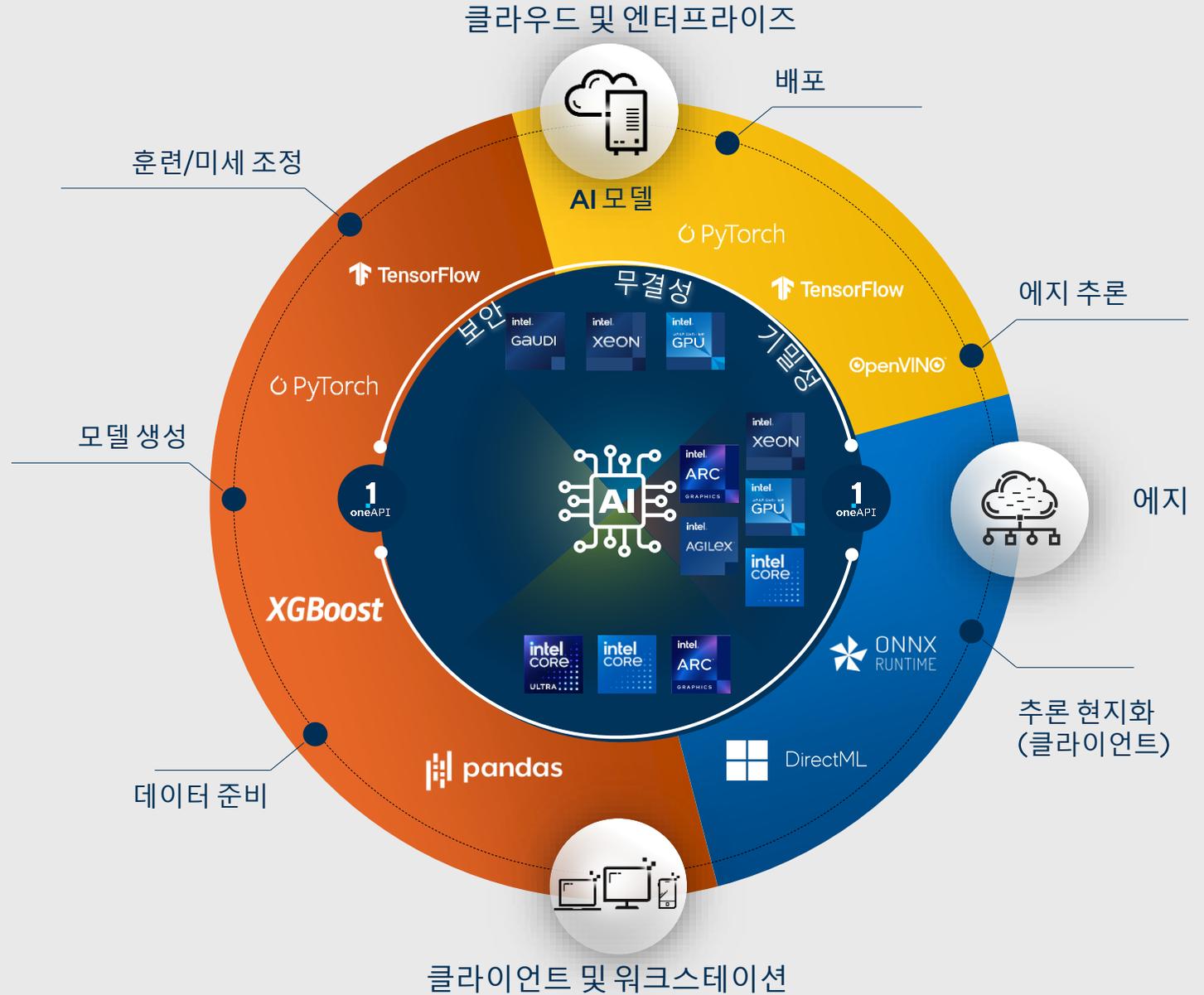


# AI 연속체

## 모든 분야에서 AI 구현

인텔은 AI 여정의 모든 단계에서 **모든 곳에 AI를 도입**하고 비즈니스를 지원하는 신뢰할 수 있는 파트너입니다.

데이터 센터, 클라우드 및 네트워크에서 클라이언트와 에지에 이르기까지



참고: 인텔® 코어™ Ultra는 Meteor Lake 이후부터 NPU 저전력 추론 엔진을 통합합니다.

# 인텔과 함께 책임지는 AI



## 접근성 향상

많은 장애인에게 독립성과 자율성은 어려운 과제일 수 있습니다. AI는 일상적인 장벽에 대한 대체 솔루션을 제공하는 제품을 만들어 이러한 변화를 돕고 있습니다.



## 환경 솔루션 만들기

연구원은 AI 기술을 사용하여 환경이 어떻게 작동하는지 더 잘 이해하고 더 나은 미래를 만들기 위한 솔루션을 개발할 수 있습니다.



## 교육 접근성 확장

인텔은 AI for Youth 및 AI for Future Workforce와 같은 프로그램을 통해 글로벌 AI 기술 격차에 대응하여 학생들이 디지털 혁명에 대비할 수 있도록 돕고 있습니다.



## 의료 서비스 발전

AI는 환자 치료 개선에서 질병 연구 개발에 이르기까지 이제 의료 서비스 및 생명 과학 분야에서 흔히 사용되고 있습니다.



## 안전성 향상

AI 기술은 자동차를 성공적으로 구동하는 것에서 아동 착취를 줄이는 것까지, 사회를 더 안전하게 만드는 데 도움이 되고 있습니다.

# 모든 단계에서 AI를 지원하는 인텔

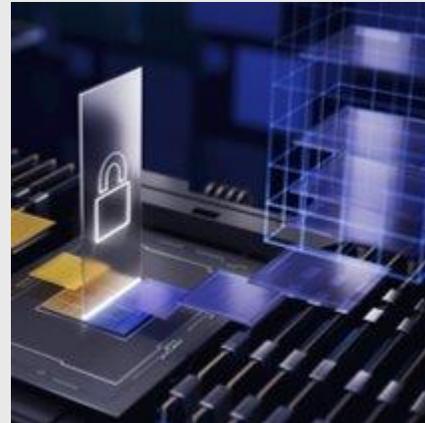
인텔은 AI 여정의 모든 단계에서 **모든 곳에 AI를 도입**하고 지원하는 신뢰할 수 있는 파트너입니다.



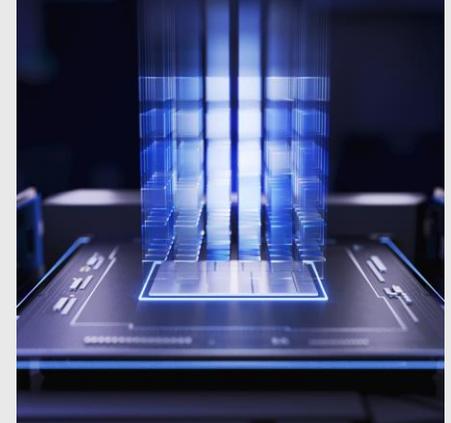
가치 극대화



어디에나 배포



보안 유지



생태계 투자

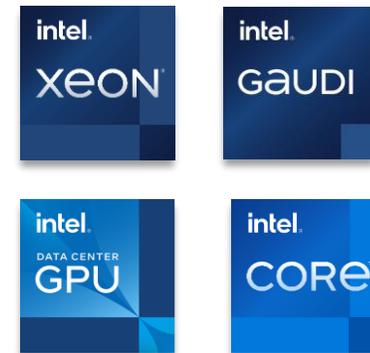


# 가치 극대화

인텔의 Open AI 접근 방식이 AI 비즈니스 요구 사항에 적합한 이유

공급업체 종속 방지  
오픈 소스 표준 기반  
소프트웨어

인텔의 하드웨어 포트폴리오 활용  
AI 사용 사례에 최적화



미래의 AI를 위해 **소프트웨어 및 개방형 표준에 의해 최적화된 하드웨어**를 통해 클라이언트와 에지에서 데이터 센터 및 클라우드까지 새로운 기회를 창출합니다.



가치  
극대화

# 인텔의 AI 전략

인텔이 AI 혁신을 가속하기 위해 제공하는 것

AI 응용 프로그램 및 소프트웨어

개방형

생산성

접근성

새로운 알고리즘

AI 데이터를 신뢰할 수 있고



인텔® Developer Cloud

하이브리드 AI

OpenVINO™



안전하게 유지하는 책임감 있는 고급

데이터 센터

스케일러블 시스템

가속기, 제온®

네트워킹

개방형 표준

네트워크 인프라

클라이언트 및 에지

AI PC

NPU, GPU, CPU

프로세스를 통해 개방형

고급 고성능 기술

Open AI 시스템 파운드리

표준 및 소프트웨어로

윤리적 리더십 기반

모든 플랫폼을 가속하여 대규모에서 성능 향상



가치  
극대화

# 인텔® AI Portfolio

모든 AI 컴퓨팅 요구 사항에 최적화된 하드웨어 및 소프트웨어 활용

개방형 소프트웨어 환경

1 oneAPI

OpenVINO™

PyTorch

XGBoost

MODIN

TensorFlow

ONNX RUNTIME

deepspeed

딥 러닝  
가속화



전용 딥 러닝 훈련 및 추론

일반 가속화



클라우드 게임, VDI, 미디어 분석,  
실시간 고밀도 비디오



병렬 컴퓨팅, HPC, HPC용 AI

일반 용도



실시간, 중간 처리량, 짧은 대기 시간,  
희소 추론



중/소규모 훈련 및 미세 조정



에지 및 네트워크 AI  
추론



클라이언트 기반 추론



# 인텔® AI 소프트웨어 포트폴리오

데이터 엔지니어링

모델 생성

최적화 및 배포



대규모 데이터 분석†



머신 및 딥 러닝 프레임워크, 최적화 및 배포 도구†



한번 작성하고 어디에나 배포

1 oneAPI

인텔® oneAPI Deep Neural Network Library

인텔® oneAPI Collective Communications Library

Intel® oneAPI MKL(Math Kernel Library)

인텔® oneAPI Data Analytics Library

CPU, GPU 및 기타 가속기를 위한 개방형 교차 아키텍처 프로그래밍 모델

클라우드 및 엔터프라이즈



클라이언트 및 워크스테이션



에지



엔드투엔드 데이터 과학 및 AI 가속



인텔® Developer Cloud 및 인텔® Developer Catalog  
최신 인텔 도구 및 하드웨어를 사용해 보고 최적화된 AI 모델에 액세스

cnvrg.io

전체 스택 ML 운영 체제

인텔® Geti

주석/훈련/최적화 플랫폼

Hugging Face

인텔 최적화 및 레시피 미세 조정, 최적화된 추론 모델 및 모델 서빙

참고: 스택의 각 계층의 구성 요소는 예상 AI 사용 모델에 따라 다른 계층의 대상 구성 요소에 최적화되었으며, 모든 구성 요소가 맨 오른쪽 열의 솔루션에서 사용되는 것은 아닙니다. † 이 목록에는 인텔 하드웨어에 최적화된 인기 있는 오픈 소스 프레임워크가 포함되어 있습니다.



# 참조 키트를 통한 AI 개발 가속화

최적화된 AI 참조 키트가 개발자 및 데이터 과학자의 혁신을 더 빠르게 지원합니다.

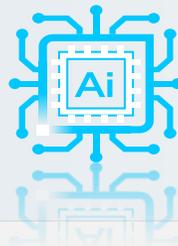
[oneAPI](#) 개방형 표준 기반 이기종 프로그래밍 모델과 [인텔® AI 분석 툴킷](#) 및 [OpenVINO™ 툴킷의 인텔® 배포](#)와 같은 인텔의 엔드투엔드 AI 소프트웨어 포트폴리오의 구성 요소를 기반으로 구축된 이들 참조 키트는 AI 개발자가 응용 프로그램에 AI를 도입하는 과정을 간소화하여 기존의 지능형 솔루션을 강화하고 배포를 가속할 수 있습니다.

그 결과 기존 모델 개발 워크플로에 비해 더 짧고 더 생산적인 워크플로를 통해 성능이 개선되는 것으로 입증되었습니다.

엔터프라이즈 대화 AI 챗봇과의 상호 작용을 설정하도록 설계된 **AI 참조 키트**를 사용하면 사용자는 [oneAPI 최적화를 통해 배치 모드에서 최대 45% 더 빠른 추론을 경험할 수 있습니다.](#)

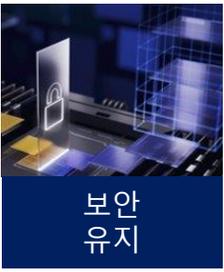


생명과학 분야의 시각적 품질 관리 검사를 자동화하도록 설계된 **AI 참조 키트**는 oneAPI 최적화를 통해 시각적 결함 감지에서 [최대 20% 더 빠른 학습과 55% 더 빠른 추론](#)을 보여주었습니다.



개발자가 유틸리티 자산의 상태를 예측하고 더 높은 서비스 안정성을 제공할 수 있도록 예측 정확도를 [최대 25%까지 높여주는 AI 참조 키트](#)가 있습니다.





# 보안 유지

내장형 보안 기능으로 AI 이니셔티브를 보호하고 규정을 준수하십시오.

보안

민감한 데이터 및  
모델 보호



규정 준수

보안 및 개인정보  
보호 규정 준수



기밀성

개인 데이터 노출 없이  
다자간 AI 활용





# 가장 포괄적인 솔루션을 제공하는 인텔 보안 포트폴리오

인텔® Software Guard  
Extension (인텔® SGX)



응용 프로그램 격리

인텔® Trust Domain  
Extension(인텔® TDX)



가상 머신 격리

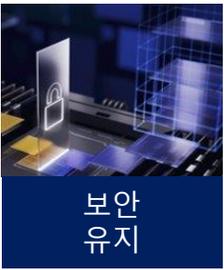
인텔® Trust  
Authority



멀티 클라우드 및 하이브리드  
클라우드에 대한 독립적인  
신뢰 검증 서비스

소프트웨어 솔루션, 클라우드, OEM 및 시스템 통합자 생태계

인텔의 보안 우선 개발 및 수명 주기 지원



보안  
유지

# 메모리의 데이터 보호

비유: 시스템 메모리

현재: 메모리를 스누핑할 수 있으면 데이터를 복호화하는 데 사용되는 개인 키를 비롯한 모든 것이 통과하는 것을 볼 수 있습니다.

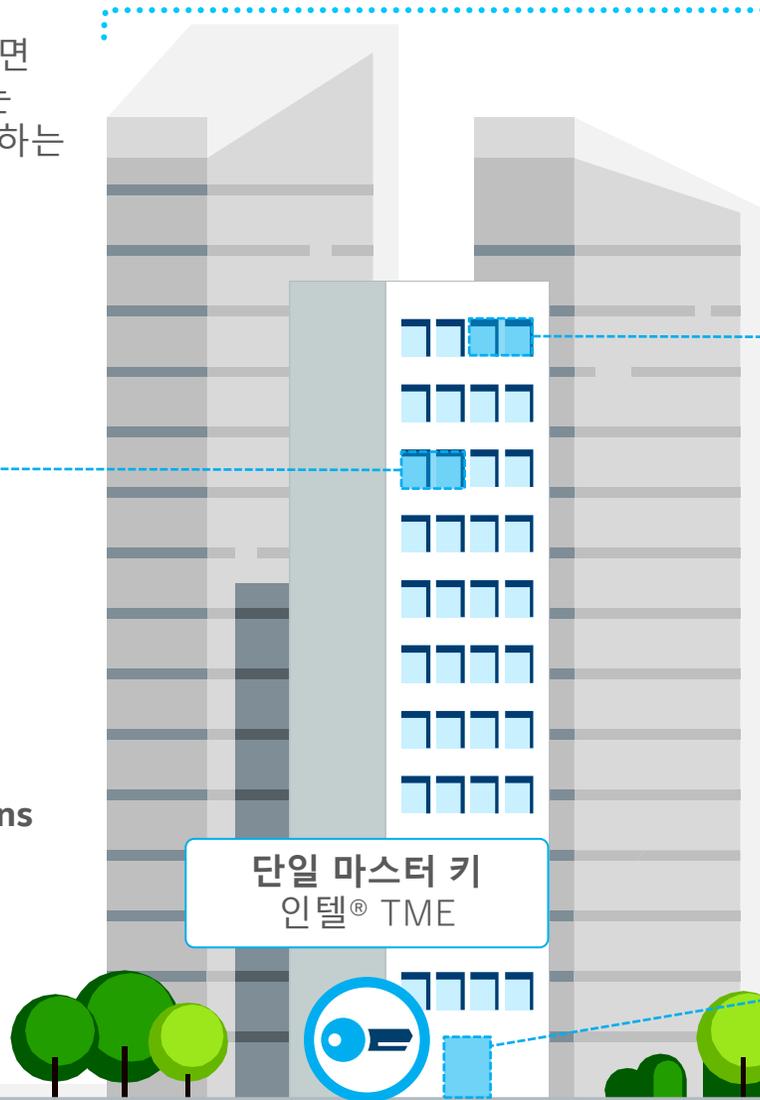
비유: VM



테넌트 키  
인텔® TDX

인텔® TDX = 인텔® Trust Domain Extensions

각 VM 공간을 별도로 암호화하기 위한 별도의 키(OS/VMM만 기능 인식 필요)



단일 마스터 키  
인텔® TME



비유:  
응용 프로그램

응용 프로그램 키  
인텔® SGX

인텔® SGX = 인텔® Software Guard Extensions

개별 응용 프로그램 데이터 공간을 위한 격리(응용 프로그램 코드 수정 또는 추상화 인터페이스 필요)

인텔® TME = 인텔® Total Memory Encryption

전체 시스템 메모리를 암호화하기 위한 단일 키(OS/앱 모드 필요 없음)



# 인텔® Security Engines로 혁신 가속화 및 데이터 보호 강화

인텔® 제온® 스케일러블 플랫폼을 통한 기밀 컴퓨팅  
비공개로 유지하면서 데이터 활용

인텔® 제온® CPU의 인텔® Security Engines로 데이터 기밀성 및 코드 무결성을 보존하면서 성능을 유지할 수 있습니다.

 [제품 요약](#)

인텔을 사용하면 중요한 비즈니스 결과에 대한 더 나은 인사이트를 얻을 수 있습니다.

 [비즈니스 요약](#)

[인텔® SGX](#) 및 [인텔® TDX](#)로  
기밀 컴퓨팅 적극 활용



# AI 생태계 구현

AI 개발자가 선호하는 최신 소프트웨어 도구를 사용하여 최적화된 성능으로 새로운 기회와 주요 비즈니스 성과를 창출하십시오.

**개방형**  
프로그래밍 가능성

**선택**  
호환성

**신뢰**  
AI 추론

**보안 AI**  
보안 워크로드  
보안 모델  
사용 중인 데이터 보안

**확장**  
개발 및 테스트

인텔® Developer Cloud

[cloud.intel.com](https://cloud.intel.com)

소형, 중형 및 대형 모델

전체 시스템, 전체 클러스터

최신 인텔 CPU, 가속기 및 소프트웨어

AI를 위한 개방형, 가속, 연결 컴퓨팅

다중 공급업체      멀티아키텍처

하드웨어/아키텍처

# 인텔® 제온®에서 AI 실행

클라우드 및  
엔터프라이즈



에지



모든 곳에 AI 도입



# 5세대 인텔® 제온®: AI를 위해 설계된 프로세서

모든 코어에서 AI 가속화를 지원하는 5세대 인텔® 제온® 프로세서는 고객이 개별 가속기를 추가해야 하기 전에 까다로운 엔드투엔드 AI 워크로드를 처리합니다.

AI 추론에서 성능 향상

최대 **42%**  
이전 세대 대비<sup>1</sup>

일반적인 컴퓨팅 성능  
향상

평균 **21%**  
이전 세대 대비<sup>1</sup>

더 빠른 자연어 처리

최대 **23%**  
이전 세대 대비<sup>1</sup>

인텔 데이터 센터 및 AI 그룹  
수석 부사장 겸 총괄 매니저,  
Sandra Rivera

"AI용으로 설계된 5세대 인텔® 제온® 프로세서는 클라우드, 네트워크 및 에지 사용 사례 전반에서 AI 기능을 배포하는 고객에게 더 높은 성능을 제공합니다. 고객, 파트너 및 개발자 생태계와 오랫동안 협력한 결과, 인텔은 낮은 TCO로 빠르게 채택하고 확장할 수 있는 검증된 기반에서 5세대 인텔® 제온®을 출시하고 있습니다."

추가 정보:

[웹 사이트](#)

[제품 요약](#)



# 인텔® 제온®: 실전 AI 적용에서의 CPU 성능 리더십

인텔은 실제 실제 업무 상황 적용에서 다음과 같이 더 나은 성능, 낮은 가격, AI 추론을 위한 더 균형 잡힌 플랫폼을 제공하여 업계를 혁신하고 AI를 대중화하고 있습니다.



데이터 지역성을 지원하는 더 큰 캐시와 더 큰 문제를 해결할 수 있는 대용량 메모리



더 높은 코어 주파수, 다중 스칼라 포트 및 단일 또는 다중 스레드이지만 스칼라인 컴퓨팅을 가속하는 비순차적 실행



비DL 벡터 컴퓨팅에 도움이 되는 인텔® Advanced Vector Extensions 512(인텔® AVX-512)



AI 가속화를 위한 내장형 하드웨어 지원 인텔® Advanced Matrix Extensions(인텔® AMX)

[기술 문서 전문](#)

[인포그래픽](#)

[GPU 신화 밝히기: 내장 가속기를 갖춘 CPU로 AI를 혁신하는 방법](#)



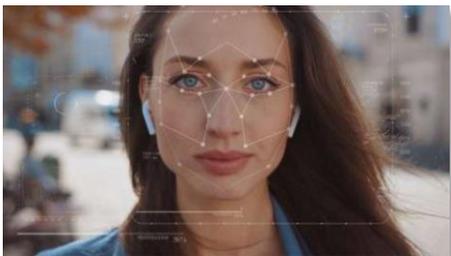


# AI 추론을 위한 가속기가 있는 4세대 인텔® 제온® 스케일러블 프로세서

인텔® AVX-512 및 인텔® AMX와 같은 가속기는 성능을 개선하고, 대기 시간을 줄이며, 메모리 대역폭을 증가시켜 까다로운 추론 AI 워크로드를 실행하는 데 적합하도록 설계되었습니다.

## 인텔® Advanced Matrix Extensions(인텔® AMX)

자연어 처리, 추천 시스템, 이미지 인식과 같은 워크로드에 이상적이며, 딥 러닝 훈련 및 추론을 크게 가속합니다.



[웹 사이트](#) | [솔루션 요약](#)  
[비디오](#) | [사용자 가이드](#)

## 인텔® Advanced Vector Extensions 512 (인텔® AVX-512)

고전적 머신 러닝 및 데이터 준비와 같은  
엔드투엔드 AI 워크플로의 기타 워크로드를  
가속할 수 있습니다.



[웹 사이트](#) | [솔루션 요약](#)  
[비디오](#) | [사용자 가이드](#) 및 [다운로드](#)



# AMD EPYC를 능가하는 인텔® AMX 탑재 4세대 인텔® 제온® 스케일러블 프로세서

더 빠른 맞춤형 AI로 매출 성장 촉진 및 고객 경험 개선



비즈니스 의사 결정에 더 나은 정보를 제공하여 매출 성장 촉진



고객 유지 및 획득 개선



참여도 증가 및 전환율 개선



비즈니스의 반복적인 작업과 비용을 줄이고 소요 시간 단축



vs.



AMD EPYC를 능가하는 인텔® Advanced Matrix Extensions(인텔® AMX)  
탑재 4세대 인텔® 제온® 스케일러블 프로세서 알아보기



# AI 워크로드: 인텔® 제온® 스케일러블 프로세서 기반 VMware

vmware®



인텔® Advanced Matrix  
Extensions(인텔® AMX)



"AI 가속화가 내장된 CPU를 사용하여 전체 엔드투엔드 AI 파이프라인(데이터 준비, 훈련, 최적화, 추론)을 실행할 수 있습니다."



"AI/ML 워크로드의 성능을 향상하기 위해 할 수 있는 한 가지는 CPU의 AMX 명령으로 AI/ML 작업 중 일부를 수행하도록 하여 비용이 많이 들고 구하기 어려운 GPU의 필요성을 줄이는 것입니다."



VMware의 직원 엔지니어 Earl Ruby의 [글 전문](#)

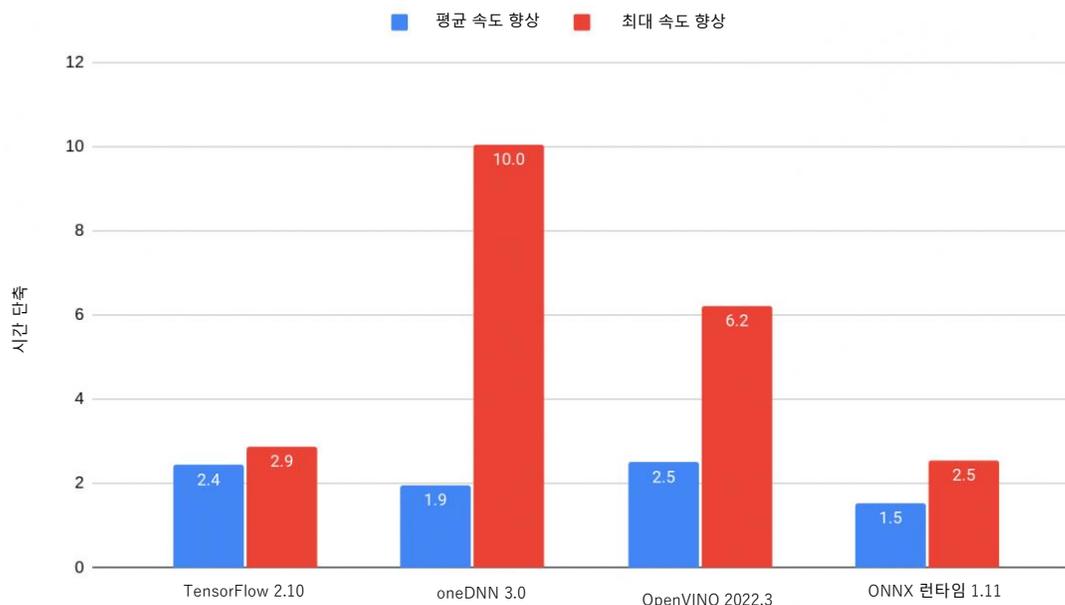


# AI 워크로드: 인텔® 제온® 스케일러블 프로세서 기반 Red Hat

Red Hat Enterprise Linux는 4세대 인텔® 제온® 스케일러블 프로세서를  
통해 상당한 성능 향상을 달성합니다.

## AMX

2P Sapphire Rapids Phoronix-Test-Suite 속도 향상 요소 vs. 4P Cooper Lake



인텔의 결과는 평균  
1.5배에서 최대 10배  
빨라지는<sup>1</sup> 4세대 속도  
향상 요소를



<sup>1</sup><https://www.redhat.com/en/blog/red-hat-enterprise-linux-achieves-significant-performance-gains-intels-4th-generation-xeon-scalable-processors>



# 사례 연구

과제

솔루션/결과

인텔 제품

추가 정보

 **Tencent Cloud**  
클라우드 컴퓨팅 서비스용 검색 엔진

대규모 쿼리를 처리하고  
검색 결과에 신속하게  
대응하는 방법

Tencent는 최적화된 BERT 모델을  
사용하여 **더 나은 서비스 경험을  
제공하고 TCO를 절감할 수  
있습니다.**

4세대 인텔® 제온® +  
인텔® AMX

[사례 연구](#)

 **Meituan**  
앞서가는 리테일 기술 회사

비용 효율적인 비전 AI  
서비스

Meituan은 온라인 리소스의  
전반적인 효율성을 **3배 이상  
개선하고 서비스 비용을 70%  
절감했습니다.**

4세대 인텔® 제온® +  
인텔® AMX + 인텔®  
IPP + PyTorch용 인텔®  
Extension (인텔® IPEX)

[사례 연구](#)

**SIEMENS**  
의료 이미지 처리

방사선 치료 전문가의  
효율성 향상

AI 기반 자동 윤곽 기술로 방사선  
치료 전문가를 지원하면 **작업  
효율성을 높이고 일관성을 개선하며  
직원의 일을 덜어주어 부가가치  
작업에 집중할 수 있습니다.**

4세대 인텔® 제온® +  
인텔® AMX +  
OpenVINO™

[사례 연구  
비디오](#)

 **Alibaba Cloud**  
앞서가는 클라우드 컴퓨팅 공급업체

주소 정제 서비스의 성능  
향상

빨라진 엔드투엔드 성능은  
**Alibaba의 물류, 전자상거래,  
에너지, 소매업 및 금융 고객의  
비즈니스 실적 개선으로 이어집니다.**  
내장형 가속기를 사용하여  
**Alibaba가 TCO를 제어합니다.**

4세대 인텔® 제온® +  
인텔® AMX + 인텔®  
oneDNN

[사례 연구](#)



# 인텔의 AI 기술에 대한 추천사



"우리는 설정 시간을 몇  
주나 단축했습니다."

"우리에게 인텔® 제온®  
프로세서는 기술 배포  
방법의 초석입니다. 우리는  
인텔® 제온® CPU에서만  
실행하며 이를 통해 VM,  
전용 온프레미스 베어 메탈,  
클라우드 등 어디에서나  
실행할 수 있습니다."



사례 연구

# SIEMENS

**35배** 자동차 운곽  
알고리즘의  
AI 추론 속도 향상  
(이전 세대 대비)<sup>1</sup>

**20%** 에너지  
소비 감소  
(이전 세대 대비)<sup>2</sup>



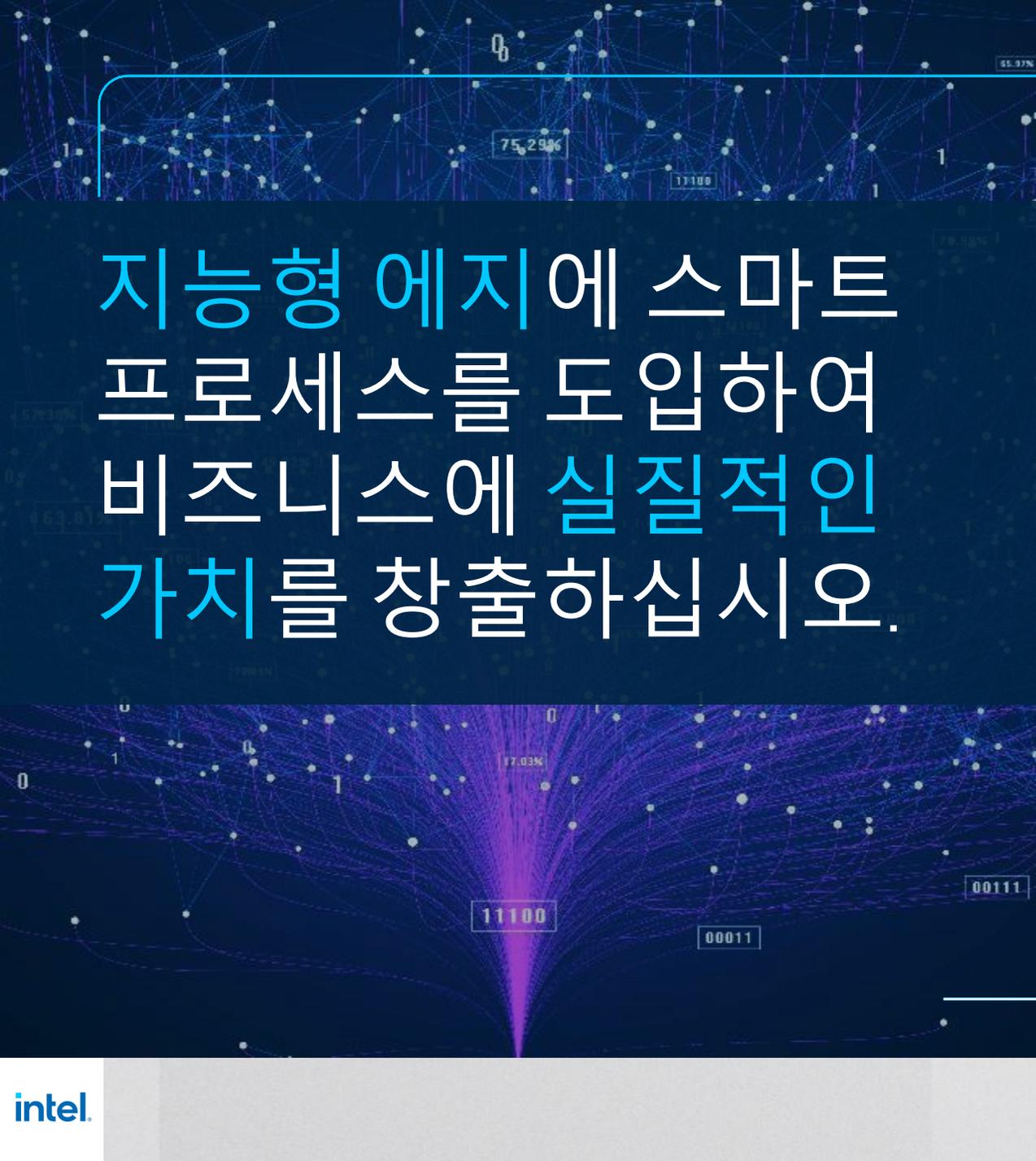
사례 연구  
비디오

# 에지 AI

에지



모든 곳에 AI 도입



지능형 에지에 스마트  
프로세스를 도입하여  
비즈니스에 실질적인  
가치를 창출하십시오.

에지 컴퓨팅과 AI의 결합된 힘을  
활용하여 기업에 더 나은 결과를,  
고객에게는 향상된 경험을  
제공합니다.



# 에지에서 데이터 처리

2025년까지 엔터프라이즈 생성 데이터의 75%가  
에지에서 생성 및 처리<sup>1</sup>

## 더 낮은 지연 시간

에지 컴퓨팅은 저장 및 처리하기 전에 클라우드로 왕복할 필요가 없으므로 인사이트를 얻는 시간이 단축되고 효율성이 향상됩니다.

## 위험 감소

에지의 데이터는 IoT 장치 자체에서 저장 및 처리되어 실시간으로 이벤트에 빠르게 대응하여 비즈니스 위험을 완화하고 보안을 강화할 수 있습니다.

## 비용 절감

에지 컴퓨팅은 더 비용 효율적인 스토리지 및 처리를 위해 데이터를 에지에 보관하면서 비즈니스 프로세스를 간소화하는 더 빠른 인사이트를 제공합니다.

에지에서의 컴퓨팅 및 처리로 인해 데이터가 생성된 곳에서 활용할 기회가 생깁니다.

<sup>1</sup>출처: <https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders>



# 업계 전반에서 혁신적인 사용 사례를 지원하는 에지 AI

## 사례 연구

- [VSBLTY](#)
- [Convergint](#)
- [Taco Bell](#)





# 내장 AI 및 보안 기능으로 핵심 에지 워크로드 가속화

4세대 인텔® 제온® 스케일러블 프로세서  
3세대 대비 성능

**1.33배**

강화된 성능

**3.01배**

더 높은 AI 추론 성능, 이미지  
분류에 인텔® AMX 사용 시

**4.25배**

더 높은 AI 추론 성능, 개체  
감지에 인텔® AMX 사용 시

IoT 에지를 위한 4세대 인텔® 제온® 스케일러블 프로세서는 놀라운 성능, 메모리, I/O, 워크로드 통합을 지원하는 리소스 관리 기능 및 에지에서의 딥 러닝 훈련 및 추론을 위한 새로운 AI 명령어를 제공합니다.

자세한 정보

생산 분야에서의 AI 성공 사례

# PC에서의 AI

클라이언트 및  
워크스테이션



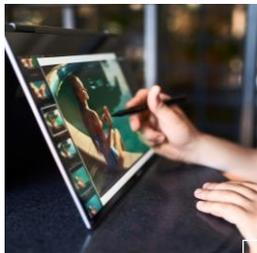
모든 곳에 AI 도입



# 사용 사례: PC에서의 AI

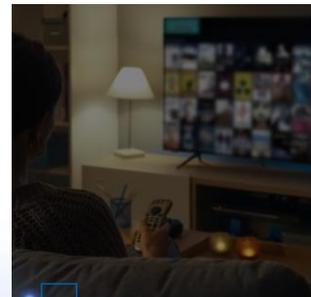
## 크리에이터: 사진 및 비디오 검색 및 편집

더 빠르고 자연스러운 필터, 더 높은  
품질의 미리보기, 더 빠른 자동  
내보내기, 더 빠른 검색



## 협업/스트리밍

차세대 비디오 회의, 스트리밍 및  
협업을 위한 새로운 AI 기능으로  
배터리 수명 보존



## 메인스트림 게임

인게임, 더한 현실감을 위한 3D  
애니메이션, 전사 및 채팅  
번역을 위한 새로운 AI 기능



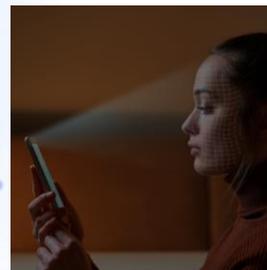
# PC에서의 AI

## 생산성

글쓰기, 제작, 코딩 및 오프라인 기능을  
위한 AI 지원, 텍스트 및 문법 예측 등

## 접근성

다양한 사용자 요구 사항을 충족하는 AI  
지원 오디오-비주얼 기능으로 PC에서 더  
쉽게 제작하고 생산성을 높일 수 있습니다.



## 크리에이터: 텍스트를 이미지로

마케팅, 광고, 디자인을 위해 몇 가지  
수식어로 이미지를 만들 수 있는  
새로운 AI 효과 및 기능

"평범함에 도전하다"



# AI PC 시대를 열어가는 인텔® 코어™ Ultra

AI PC의 핵심인 새로운 프로세서는 운영 체제 및 응용 프로그램 전반에서 AI 기능을 활용합니다.

인텔 4 공정 기술을  
기반으로 한 최초의  
프로세서  
40년 만의 가장 큰  
아키텍처 변화

최대 8개의 Xe-코어를  
탑재한 인텔® Arc™  
GPU3 내장  
이전 세대 대비 최대  
2배의 그래픽 성능<sup>1</sup>

인텔의 최신 NPU, 인텔®  
AI Boost, 저전력으로 더  
오래 실행되는 AI  
워크로드를 위해 설계  
이전 세대보다 최대 2.5배  
향상된 전력 효율성<sup>1</sup>

인텔 수석 부사장 겸  
클라이언트 컴퓨팅 그룹 총괄  
매니저, Michelle Johnston  
Holthaus

"인텔® 코어™ Ultra의 출시는 인텔이 PC에서 AI를  
지원하는 탁월한 규모와 속도를 나타냅니다. 2028년까지  
AI PC는 PC 시장의 80%를 차지할 것이며<sup>2</sup>, 인텔은  
방대한 하드웨어 및 소프트웨어 파트너 생태계와 함께  
차세대 컴퓨팅을 제공할 수 있는 최고의 위치에  
있습니다."

추가 정보:

[웹 사이트](#)

[제품 요약](#)



# 사례 연구: 인텔® 코어™ Ultra 프로세서의 AI를 통한 환자 치료 개선

CPU 기반 초음파 이미지 작업 응용 프로그램은 더 접근성이  
높고 비용 효율적인 이미지 작업 기술을 제공합니다.

## 상황

Samsung Medison은 의료 혁신 분야의 개척자입니다. 이들의 초음파 이미지  
작업 응용 프로그램은 가장 효과적인 환자 치료를 위해 AI를 사용합니다.

## 과제

이전에는 경쟁사의 별도 GPU로 가속화된 이전 세대 인텔® 코어™ 프로세서에서  
응용 프로그램을 실행했습니다.

## 솔루션

삼성은 GPU 엔진이 내장된 새로운 인텔® 코어™ Ultra 프로세서를  
테스트했습니다. 이전 세대 CPU + dGPU 콤보 대비 AI 성능이 크게  
향상되었습니다. Samsung Medison은 인텔® 코어™ Ultra를 통해 CPU만을  
기반으로 하는 차세대 초음파 장치에서 고급 AI 기능을 제공할 수 있습니다.

상세 정보:  
[자세한  
정보](#)



intel  
CORE  
ULTRA



# AI PC 가속화 프로그램

AI PC 가속화 프로그램은 독립 하드웨어 공급업체(IHV)와 독립 소프트웨어 공급업체(ISV)를 인공 지능 AI 튜체인, 교육, 공동 엔지니어링, 소프트웨어 최적화, 하드웨어, 설계 리소스, 기술 전문 지식, 공동 마케팅 및 판매 기회를 포함한 인텔 리소스와 연결하는 것을 목표로 합니다.

## 인텔이 AI의 미래를 주도합니다

100+

참여

AI PC 최적화를 위한 AI ISV  
파트너

300

이상

2024년까지 AI 가속화 ISV 기능

1억

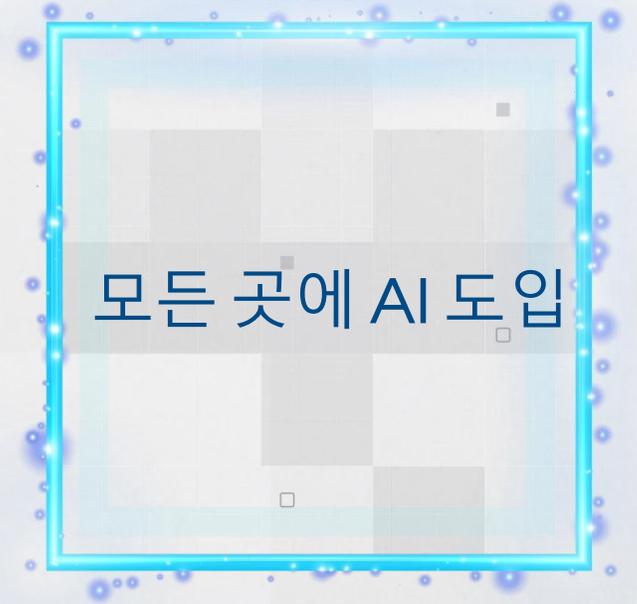
이상

2025년까지 AI 가속기가 탑재된  
프로세서

**지금 문의하여 자세한 정보를 확인하십시오!**

[ai.pc.acceleration.program@intel.com](mailto:ai.pc.acceleration.program@intel.com)

# Gaudi2에서 AI 실행





# Gaudi2: 기반 모델의 효율적인 훈련 및 추론에 적합

Gaudi2는 LLM(GPT) 및 GAI(Stable Diffusion)와 같은 대규모 기반 모델의 요구 사항을 충족하는 딥 러닝 성능, 효율성 및 확장성을 위해 설계되었습니다

| 요구사항   | Gaudi2  |
|--------|---|
| 속도     | 훈련 및 추론 모두에서 A100보다 <b>1.5~2배 빠름</b>  |
| 메모리    | 각 Gaudi2 장치는 <b>96GB 온칩 고대역폭 메모리</b> 를 갖추고 있어 대규모 기반 모델의 메모리에 쉽게 장착하고 대규모로 훈련 및 배포할 수 있습니다. |
| 확장성    | <b>24x100GbE 포트 통합형 온칩</b> 으로 효율성 확장, 서버 내 8개 카드에 모두 직접 연결, 서버 내부와 서버 간의 개방형 ROCEv2 기반 통신   |
| 사용 용이성 | SynapseAI, PyTorch 및 DeepSpeed를 통한 <b>최소한의 코드 변경</b> 으로 모델 이전 또는 구축                         |
| 전원 효율성 | <b>A100 대비 ~1.8배 더 높은 처리량/와트</b>  |
| 비용 효율성 | Amazon 클라우드에서 A100보다 <b>최대 40% 향상된 가격 성능</b> 을 제공하는 특수 목적 1세대 Gaudi 아키텍처 기반                 |



# Gaudi2: 생성형 AI 및 대규모 언어 모델 가속화

Gaudi2 딥 러닝 가속기는 Nvidia A100 대비 딥 러닝 훈련 및 추론에서 최대 2.4배 빠른 경쟁력 있는 성능을 발휘합니다.<sup>1</sup>

GAUDI<sup>2</sup>  
VS.  


[1 보도 자료](#)

Habana Gaudi2 및 4세대 인텔® 제온® 스케일러블 프로세서는 AI 훈련을 위한 최고의 성능과 최적의 비용 절감을 제공합니다.<sup>2</sup>



[2 뉴스룸](#)

[기술 문서](#)

생성형 AI 및 대규모 언어 모델(LLM)의 잠재력을 포착하는 Habana® Gaudi®2 AI 프로세서의 최첨단 기능에 대해 알아보는 인텔 웨비나 녹화본



[웨비나](#)



# 딥 러닝 혁신: 인텔, Habana Labs 및 Hugging Face

인텔의 Hugging Face와의 지속적인 협력은 최신 인텔® 제온® 스케일러블, 및 Habana Gaudi® 및 Gaudi®2 프로세서에 최적화된 훈련 및 추론 솔루션의 도입을 확대하는 데 중점을 두고 있습니다.



## Hugging Face

협력은 인텔® AI 툴킷의 가장 진보된 딥 러닝 혁신을 Hugging Face 오픈 소스 생태계에 도입하고 미래의 인텔® 아키텍처의 혁신 주도자에게 정보를 제공합니다.



[인텔, Habana Labs 및 Hugging Face가  
이론 딥 러닝 소프트웨어의 진전](#)



[Habana Gaudi에서  
트랜스포머로 시작하기](#)

더 빠른 훈련 및 추론: Habana Gaudi®-2와 Nvidia A100 80GB 비교  
[벤치마크](#)



# 대중화된 AI: 인텔, Habana Labs 및 Hugging Face



## Hugging Face

Nvidia의 A100보다 1760억 매개변수 모델에서 **20%** 더 빠른 Habana® Gaudi®2 추론 실행<sup>1</sup>

Gaudi2 서버에서 인기 있는 컴퓨터 비전 워크로드를 실행할 때 동급 A100 서버보다 와트당 처리량에서 **1.8배** 유리<sup>1</sup>

여기에서 발표를 읽고 Fireside 채팅을 시청하십시오.  
[생성형 AI의 컴퓨팅 및 지속 가능성 과제 해결하기](#)



팟캐스트

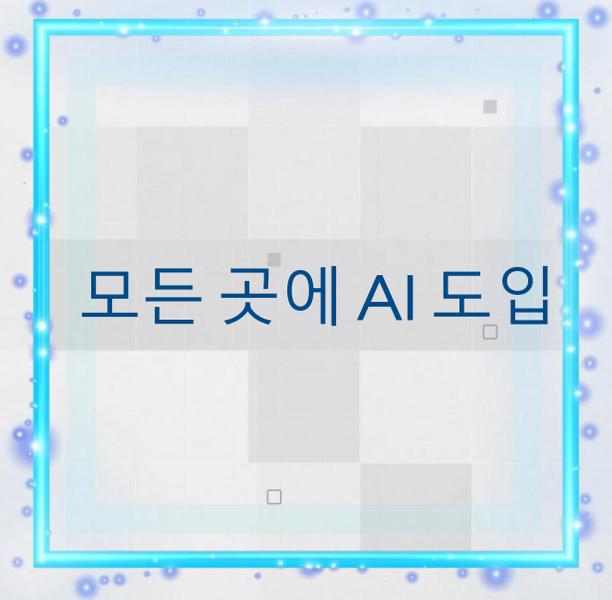
[Hugging Face 및 인텔 - 실용적이고 더 빠르며, 민주적이고 윤리적인 AI 솔루션](#)



Twitter/X  
대화

[민주화된 대규모 언어 모델이 AI 개발을 촉진하는 방법](#)

<sup>1</sup> 성능은 사용, 구성 및 기타 요인에 따라 다르며, 다음에서 워크로드 및 구성 세부 정보를 확인할 수 있습니다. Gaudi2의 Supermicro L12 검증 보고서 HL-225H SYS-820GH-THR2, 2022년 10월 20일



모든 곳에 AI 도입

인텔® Data Center GPU Max 시리즈에서 AI  
실행하기



# 인텔® Data Center GPU Max 시리즈: 혁신적인 성능

인텔의 최고 성능, 최고 밀도 별도 GPU

## 인텔의 기본 GPU 컴퓨팅 빌딩 블록은 다음을 특징으로 합니다.

- 개별 SRAM 기술 기반 최대 408MB의 L2 캐시, 64MB의 L1 캐시 및 최대 128GB의 고대역폭 메모리
- 과학적 시각화 및 애니메이션을 가속하기 위해 각 Max 시리즈 GPU에 내장된 최대 128개의 광선 추적 장치
- 단일 장치에서 벡터 및 매트릭스 기능을 지원하는 딥 시스톨릭 어레이를 갖춘 AI 강화 인텔® Xe Matrix Extensions(XMX)
- 성능과 생산성을 높이고 독점 프로그래밍 모델에 대한 종속을 방지하는 oneAPI 표준 기반 멀티 아키텍처 프로그래밍 및 도구
- 다음을 통해 강조되는 강력한 성능
  - 인텔 oneAPI 도구로 최적화되고 6개 Max 시리즈 GPU에 오프로드된 커널을 갖춘 인텔® 제온® Max CPU에서 실행된 LAMMPS(large-scale atomic/molecular massively parallel simulation) 워크로드에서 인텔® 제온® Max CPU 대비 최대

12.8배 향상된 성능<sup>1</sup>

인텔® Data Center GPU Max 시리즈는 AI 및 HPC에 사용되는 데이터 집약적인 컴퓨팅 모델에서 획기적인 성능을 제공하도록 설계되었습니다. 인텔® Max 시리즈 GPU는 SoC 구축에서 더 높은 유연성과 모듈성을 지원합니다.

[제품 요약](#)

[웹 사이트](#)

[기술 문서](#)

**1**  
oneAPI

전체 인텔® Max 시리즈 제품군은 공통적이고 개방적인 표준 기반 프로그래밍 모델을 위해 oneAPI로 통합되어 생산성과 성능을 극대화합니다.

개발자는 데이터 분석 및 머신 러닝 워크플로를 위해 oneAPI에 최적화된 딥 러닝 프레임워크 및 머신 러닝 라이브러리를 사용하여 드롭인 가속화를 실현할 수 있습니다.



# 사례 연구: 인텔® Data Center GPU Max 시리즈 기반 Aurora 슈퍼컴퓨터

세계에서 가장 어려운 문제를 더 빠르게 해결합니다.



Argonne National Laboratory(ANL)에 있는 미국 에너지부의 **Aurora** 슈퍼컴퓨터는 1엑사플롭 이상의 유지 배정밀도 성능과 2엑사플롭 이상의 최대 배정밀도 성능을 갖춘 업계 최초의 슈퍼컴퓨터 중 하나가 될 것으로 예상됩니다. 또한, 단일 시스템에 구현한 Max 시리즈 GPU와 CPU의 결합 성능을 처음으로 선보일 예정입니다. Aurora는 각각 6개 Max 시리즈 GPU와 2개 제온® Max CPU를 포함하는 10,000개 이상의 블레이드를 갖추었습니다.

[머신 러닝용 Aurora 블레이드 데모](#)

# 콜 투 액션

# 콜 투 액션

## 교육



인텔 기술이 AI 요구 사항에 어떻게 사용될 수 있는지, 인텔® 제온® 제품군이 더 많은 비즈니스를 성사하는 데 어떤 도움이 될 수 있는지 알아보십시오

자세한 정보  
[AI 교육 자선](#)

## 관심 사로잡기



기술 도메인 미팅을 시작하십시오

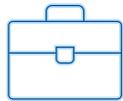
기술 공개 일정을 예약하려면 다음 주소로 이메일을 보내십시오.  
[cloud.insider.program@intel.com](mailto:cloud.insider.program@intel.com)

# 인텔® 파트너 얼라이언스의 지원

# 인텔® 파트너 얼라이언스 시작하기

인텔 파트너 얼라이언스 멤버십은 인텔의 글로벌 마켓플레이스 입점, 고급 교육 수강, 프로모션 지원을 비롯하여 독점적 비즈니스 구축 기회를 제공하며, 이 모든 것이 귀사의 필요에 맞춤화됩니다.

## 교육 및 역량



인텔® 파트너 유니버시티에 입학하면 고급 기술에 관한 전문 교육, 역량 프로그램 및 학습에 대한 보상이 제공됩니다

## 마케팅 리소스



인텔® 솔루션 마켓플레이스 및 인텔® 마케팅 스튜디오에 입장하여 귀사의 제품 및 서비스에 대한 수요를 더욱 많이 창출할 수 있습니다

## 가치 있는 보상



적격 활동으로 포인트를 적립하고, 멤버십 등급을 높이고, 비즈니스 구축에 도움이 되는 추가 리소스에 액세스해 보십시오

아직 회원이 아니라면  
[지금 가입하십시오](#)

# 멤버십 혜택

## 포인트 적립



인텔® 파트너 얼라이언스 내에서 가장 인기 있고 차별화된 혜택 중 하나는 인텔과의 비즈니스 성과, 그리고 우선순위가 높은 활동에 대한 참여를 인정하기 위해 인텔이 파트너에게 지급하는 포인트입니다. 인텔 파트너 얼라이언스 내에는 1,000가지가 넘는 포인트 적립 방법과 수백 가지의 포인트 사용 기회가 있습니다.

## Cloud Insider 커뮤니티



인텔® Cloud Insider 커뮤니티는 계속 새로워지는 세계 최고 수준의 클라우드 콘텐츠와 도구를 제공합니다. 회원에게는 동료 및 생태계와 연결하여 혁신적인 퍼블릭 클라우드 솔루션을 시장에 출시할 수 있는 기회가 주어집니다.

[자세한 정보](#)

## 업계 인사이트



Gold 및 Titanium 회원은 성장을 촉진하는 데 도움이 될 수 있도록 특별히 선별된 분기별 업계 인사이트에 액세스할 수 있습니다.

[자세한 정보](#)

## 재정적 인센티브



회원은 강력한 마케팅 개발 기금 및 인센티브 프로그램을 활용하여 제품 마케팅의 성공을 가속할 수 있습니다. 인텔 담당자에게 문의하여 인텔® 파트너 얼라이언스 가속 이니셔티브 등 자세한 재정적 인센티브에 대해 알아보십시오

# 리소스

# 인텔® 파트너 얼라이언스 고객 지원에 액세스하는 방법

## 인텔 Virtual Assistant

각 파트너 얼라이언스 웹 페이지에서  
오른쪽 하단에 있는 이 채팅 봇은 대부분  
질문에 셀프 도움말 또는 빠른 실시간 지원  
담당자 연결을 제공합니다.



## '도움 받기' 블레이드

[온라인 지원 요청](#)을 제출하십시오.

파트너 얼라이언스 웹 사이트 내 대부분의  
페이지 꼬릿말에 이 링크가 표시되어 있습니다.

Get Help

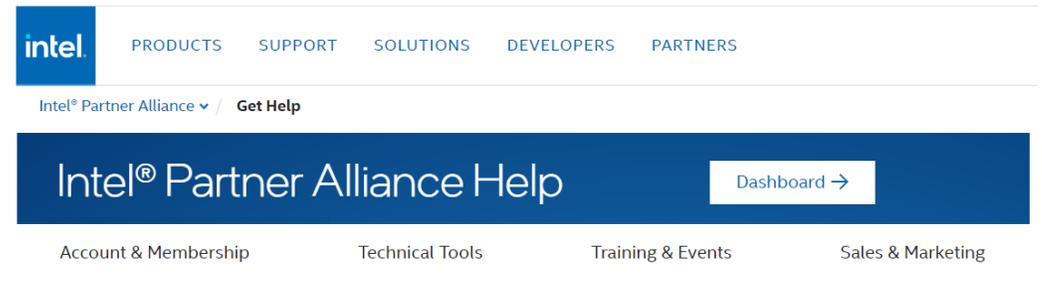
### Request Support

Contact us anytime to create a  
support request.

[Submit request >](#)

## 파트너 얼라이언스 '도움말' 페이지

[도움말](#) 페이지는 파트너 얼라이언스 회원이 사용할 수  
있는 대부분의 도구 및 혜택에 대한 자세한 셀프 도움말  
가이드를 제공합니다.



# Cloud TV

인텔® Cloud TV에서 성공을 촉진하기 위한 클라우드 컴퓨팅 뉴스, 동향 및 전략을 살펴볼 수 있습니다.



클라우드의  
Sapphire Rapids



클라우드로  
AI 강화하기



모든 곳에 AI를 확장하는  
가장 빠른 경로



클라우드 기술을 사용한  
AI 추론



클라우드에서의 AI

# 4세대 인텔® 제온® 스케일러블 프로세서를 탑재한 AI 정보 및 리소스



## 제품 개요

[4세대 인텔® 제온® 스케일러블 프로세서](#)

[인텔® 제온® CPU용 인텔® AI Engines 성능 향상  
전체 AI 파이프라인의 성능](#)

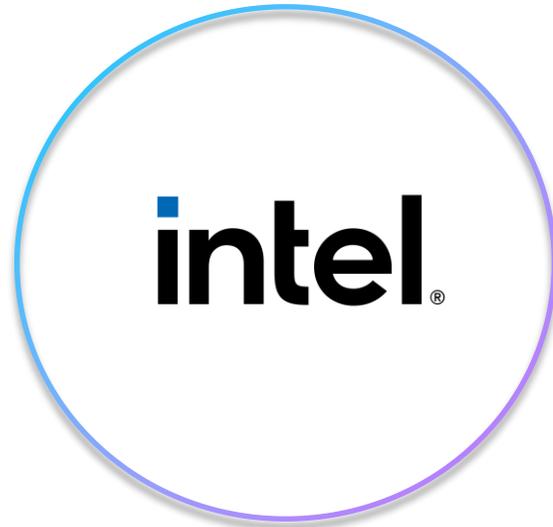


## 기술 문서

[기밀 컴퓨팅을 통한 가속화된 AI 추론](#)

[4세대 인텔® 제온® 기반 확장 가능한  
엔드투엔드 엔터프라이즈 AI](#)

[기술 혁신가와 인텔® 기술로 AI  
이니셔티브 간소화](#)



## 인포그래픽

[빠르고 비용 효율적인 고성능 AI 배포](#)

[AI로 더 빠른 투자 이익 회수](#)



## 사례 연구

[Fujitsu](#) | [Siemens](#) | [BCM](#) | [ai.io](#)



## 비디오

[인텔 AI 파이프라인 비디오](#)

[인텔® AMX: AI의 다음 큰 단계](#)

[인텔 AI 가속기 동영상](#)

[4세대 제온® 클라우드 AI 비디오](#)



## 서류가방

[기술 혁신가와 인텔® 기술로 AI 이니셔티브  
간소화](#)

# 추가 리소스



## 성능 지표

[4세대 인텔® 제온® 스케일러블 프로세서](#)



## 카탈로그

[AI 추론 소프트웨어 및 솔루션 카탈로그](#)



## 비즈니스 보고서

[Hype Cycle: 2022년 인공지능](#)

[디지털 우선 경제에서 디지털 혁신 실현: 인공지능 Disrupter 되기](#)

[데이터 센터 성능 및 기능을 가속하는 데 중점을 둔 4세대 인텔® 제온® 스케일러블 프로세서](#)



## 추가 교육

[온라인 교육에 대한 데크 내 링크](#)

# 파트너 리소스 가이드

인텔 외부에서 사용할 수 있는 리소스 페이지를 시장 출시 및 시장 출시 활동을 위한 하나의 참조 문서로 통합합니다



**목적:** 이 프레젠테이션을 IPA 파트너 및 CNDA에 속한 파트너와 함께 사용하여 전반적인 인텔 참여 유도

**대상 고객:** PSAM 및 SDM

**PRG 액세스:** [goto/resourceguide](#)(내부 액세스)

# 내부 리소스

## 비공개 콘텐츠

| 자산 유형          | 제목 및 링크   |
|----------------|---|
| 골드 데크          | <a href="#">AI Gold Deck</a>                      |
| 골드 데크          | <a href="#">SPR AI Gold Deck</a>                  |
| 판매 데크          | <a href="#">인텔 Gaudi Non-NDA 판매 데크</a>            |
| 개발자 리소스        | <a href="#">인텔 Gaudi 개발자 리소스</a>                  |
| 스냅샷            | <a href="#">Max Series GPU</a>                    |
| 골드 데크          | <a href="#">AMX 딥 다이브 데크</a>                      |
| 판매 Wiki        | <a href="#">인텔® 제온® 요청 시 판매 Wiki</a>              |
| 판매 서류 가방       | <a href="#">4세대 인텔® 제온® 우선 사용 사례와 가치 제안</a>       |
| 피치 카드          | <a href="#">인텔® AMX 피치 카드</a>                     |
| 판매 서류 가방       | <a href="#">4세대 인텔® 제온® 프로세서로 가속하여 워크로드 성능 실현</a> |
| 판매 가속화<br>대시보드 | <a href="#">SPR 판매 가속화</a>                        |

# 교육 자산

# AI 교육 자산

## 인공 지능

[인공 지능: 4세대 인텔® 제온® 프로세서를 통한 워크로드 가속화](#)  
ALL

[Fortanix 기밀 AI로 주문형 AI 워크로드를 보호하는 방법 자세히 알아보기](#)  
DevOps, 클라우드 설계자

[클라우드에 인텔 AI가 필요한 이유](#)  
DevOps, 클라우드 설계자

[AI 클라우드 배포 옵션](#)  
클라우드 설계자, 최고 경영진

[CSP AI 포트폴리오](#)  
클라우드 설계자, 최고 경영진

[데이터 센터에서 에지까지 AI 성능 달성](#)  
DevOps, 클라우드 설계자

[4세대 인텔® 제온® 플랫폼 소개](#)  
ALL

# 법적 고지 및 면책 사항

## 고지 및 면책 조항.

© Intel Corporation. 인텔, 인텔 로고 및 기타 인텔 마크는 인텔사 또는 그 자회사의 상표입니다. 기타 명칭 및 브랜드는 해당 소유업체의 자산입니다.

intel®

The Intel logo is centered on a dark blue background. It features the word "intel" in a white, lowercase, sans-serif font. A small, bright blue square is positioned above the letter "i". To the right of the word "intel" is a registered trademark symbol (®). The background is a solid dark blue with several faint, semi-transparent squares of varying shades of blue scattered across it, creating a subtle geometric pattern.

intel®

# 정책

# 하드웨어 및 소프트웨어의 복합적인 효과

## [복합적인 효과 슬라이드 보기](#)

|               |                                      |                                    |                                  |
|---------------|--------------------------------------|------------------------------------|----------------------------------|
| 시스템 보드        | 인텔® 서버 보드 S2600STB                   | M50CYP2SB1U Coyote Pass            | 인텔사/Archer City                  |
| CPU           | 인텔® 제온® Platinum 8270 CPU @ 2.7GHz   | 인텔® 제온® Platinum 8380 CPU @ 2.3GHz | 인텔® 제온® Platinum 8490H @ 1.9GHz  |
| 소켓, 물리적 코어/소켓 | 2, 26                                | 2, 40                              | 2, 60                            |
| 하이퍼스레딩/터보 설정  | 활성화/켜기                               | 활성화/켜기                             | 활성화/켜기                           |
| 메모리           | 12x16GB DDR4 2933MHz                 | 16x16GB DDR4 3200MHz               | 16x16GB DDR5 4800MHz             |
| OS            | UB-18.04 LTS                         | UB-22.04 LTS                       | UB-22.04 LTS                     |
| 커널            | 5.3.0-24-일반                          | 5.19.0-38-일반                       | 5.19.0-41-일반                     |
| 소프트웨어         | OpenVINO™ 툴킷의 인텔® 배포판 2021.4         | OpenVINO™ 툴킷의 인텔® 배포판 2022.3       | OpenVINO™ 툴킷의 인텔® 배포판 2023.0     |
| BIOS          | SE5C620.86B.02.01. 0013.121520200651 | SE5C620.86B.01.01.0006.2207150335  | EGSDREL1.SYS.9409.P31.2302280828 |
| BIOS 출시 날짜    | 2020/12/15                           | 2022/7/15                          | 2023/2/28                        |
| BIOS 설정       | 최적화된 기본 설정, 저장, 종료를 선택합니다.           | 최적화된 기본 설정, 저장, 종료를 선택합니다.         | 최적화된 기본 설정, 저장, 종료를 선택합니다.       |
| 테스트 날짜        | 2021 6/18                            | 2023/6/20                          | 2023/5/25                        |
| 정밀도 및 배치 크기   | int8/배치 1                            | int8/배치 1                          | int8/배치 1                        |
| 추론 요청 수       | 52                                   | 80                                 | 120                              |
| 실행 스트림 수      | 52                                   | 80                                 | 120                              |
| 전력(TDP)/소켓    | <a href="#">205W</a>                 | <a href="#">270W</a>               | <a href="#">350W</a>             |

워크로드(모델: HxW 입력):

Inception-v4: (299x299); Resnet-50: (224x224); Unet-camvid-onnx-0001: (368x480); Yolo-v3-tiny: (416x416)



가치  
극대화

# 혁신 가속화

개방형 표준 도구에서 성능을 최적화하는 방법

## 더 빠르게 구축하는 최적화된 사전 훈련 모델

LLM - Bloom

LLM - Llama 2

객체 탐지

객체 인식

세분화

이미지 처리

텍스트 감지 및 처리

텍스트 음성 변환

[LLM 다운로드 >](#)

[OpenVINO™ 다운로드 >](#)

## 친숙한 모델로 최적화된 업계 표준 프레임워크

PyTorch

TensorFlow



XGBoost

MODIN

PaddlePaddle

[인텔® AI 툴킷 다운로드 >](#)

## 더 쉬워진 생태계 솔루션

snowflake

Hugging Face

HAZELCAST

ahana

ANACONDA

EROSPIKE

CLOUDERA

C3.ai

intel  
GETi

AIBLE



# 어디에나 배포

엔터프라이즈 전반에서 개념에서 운영으로 더 빠르게 진행

구축  
엔드투엔드 AI  
파이프라인  
소프트웨어

데이터 엔지니어링

모델 생성

최적화 및 배포

OpenVINO™를 사용하여  
한 번 작성하고  
어디에나 배포



최적화를 통해  
모든 아키텍처에  
배포



데이터 센터



클라우드



클라이언트



에지

# AI 워크플로 간소화



### 개방형

### 생산성

|     |                 |                    |
|-----|-----------------|--------------------|
| 솔루션 | 사전 구성된 컨테이너     | AI 도구 선택기          |
| 도구  | 확장 최적화<br>Modin | OpenVINO™<br>CNVRG |
| 참조  | AI 참조 키트        | Hugging Face 협업    |

### 접근성

에코시스템 참여

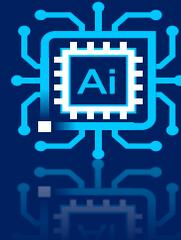
|          |            |              |
|----------|------------|--------------|
| 산업 및 학계  | 솔루션 마켓플레이스 | 하이터치 지원      |
| 개발자 교육   |            |              |
| MLOPS 교육 | 우수성 센터     | 문서 및 자습서     |
| 교육 비디오   | 서밋 및 해커톤   | Liftoff 프로그램 |

# 참조 키트를 통한 AI 개발 가속화 혜택



## 오픈 소스 머신 러닝 키트를 통한 더 빠른 혁신

오픈 소스 커뮤니티에 릴리스된 AI 모델은 사용 사례에 가장 적합한 것을 릴리스하기 위해 수천 개의 모델을 기반으로 설계, 훈련 및 테스트 과정을 거쳤습니다. 데이터 과학자는 해당 업계의 데이터를 사용하여 모델을 더욱 맞춤화하고 미세 조정할 수 있습니다.



## 머신 러닝 파이프라인을 위한 설계 및 최적화

각 참조 키트에는 엔터프라이즈의 AI 배포를 가속하기 위해 다음을 포함한 사용자 가이드가 포함되어 있습니다.

- 데이터 수집
- 데이터 전처리
- 머신 러닝 모델링
- 하이퍼파라미터 튜닝
- 모델 서빙 및 배포
- 벤치마킹



## 더 적은 컴퓨팅 리소스로 더 많은 모델 구축

모든 AI 모델은 더 적은 컴퓨팅 리소스를 사용하여 더 빠른 훈련 및 추론 성능을 달성할 수 있도록 oneAPI 기반 인텔 라이브러리, 프레임워크 및 도구를 사용하여 AI 개발 요구 사항에 맞게 최적화됩니다. AI 참조 키트는 [인텔® AI 분석 툴킷](#)과 [OpenVINO™ 툴킷의 인텔® 배포](#)를 포함한 인텔 AI 소프트웨어 포트폴리오의 구성 요소를 사용합니다.

살펴보기

[AI 참조 키트 라이브러리](#)





# AI 워크로드: Hugging Face 훈련/추론 벤치마킹



## Hugging Face

AWS에서 실행되는 4세대 인텔® 제온®  
서버 클러스터로 PyTorch 훈련 작업을  
가속하는 방법

AWS에서 세대 간 AI 추론 개선 사항  
보기



선형 확장성을 갖춘 이전 인텔® 제온® 세대 대비  
8배 속도 향상<sup>1</sup>



[4세대 인텔® 제온®]과 Hugging Face Optimum을 결합하면 코드의  
아주 작은 변경만으로 예측을 3배 가속할 수 있습니다.<sup>2</sup>



## Hugging Face

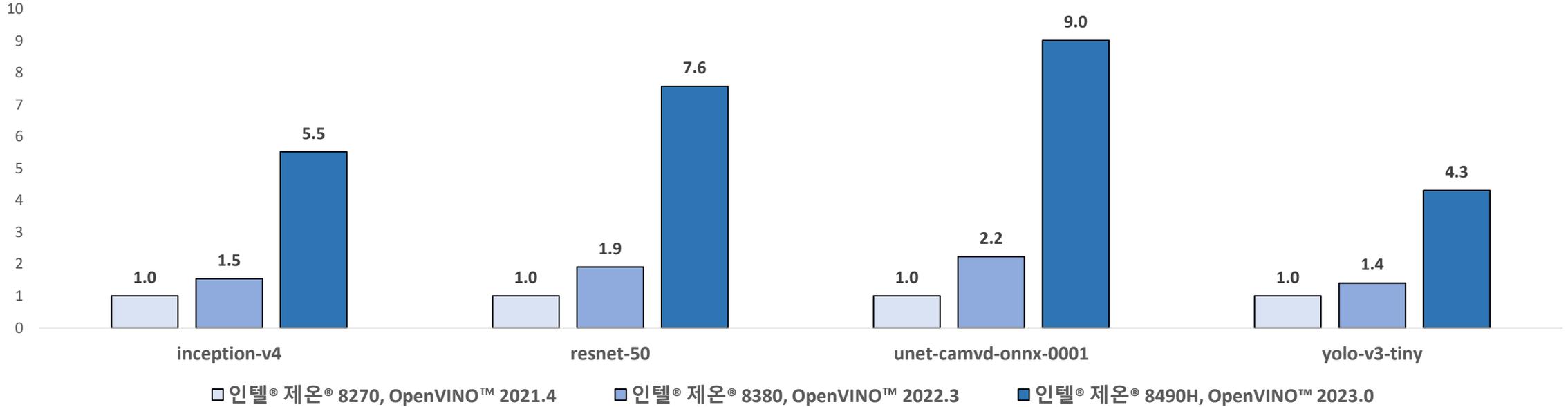
4세대 인텔® 제온® CPU는 특히 Hugging Face Optimum과 결합할 때 탁월한 추론 성능을 제공합니다. 이는  
딥 러닝을 더 접근성 있고 비용 효율적으로 만드는 또 다른 단계입니다.<sup>2</sup>

<sup>1</sup> <https://huggingface.co/blog/intel-sapphire-rapids>

<sup>2</sup> <https://huggingface.co/blog/intel-sapphire-rapids-inference>

# 계속적인 AI 추론 개선에 투자하는 인텔

인텔® AMX를 사용하는 OpenVINO™는 AI 추론 성능을 기하급수적으로 향상합니다.  
3세대 인텔® 제온® CPU의 비교(정밀도: INT8, 배치 크기: 1, 높을수록 우수)

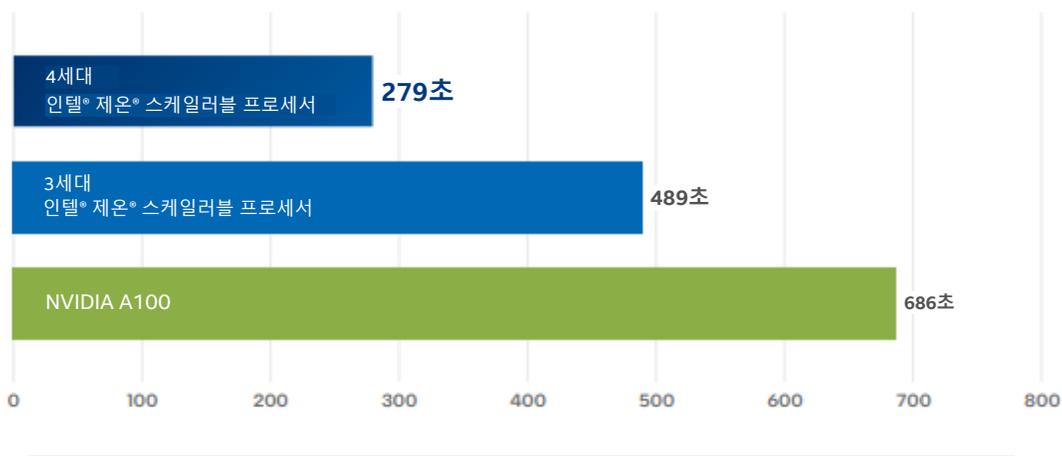




# 단일 세포 유전체학: 인텔® 제온® 스케일러블 프로세서에서 성능 향상

## 130만 개 단일 세포 유전체학 분석

초 단위로 측정되는 분석 완료 시간(낮을수록 우수)



CPU에 최적화된 소프트웨어에서 3세대 인텔® 제온® 스케일러블 프로세서는 **NVIDIA A100 GPU보다 최대 1.4배 빠른 성능**을 제공합니다.<sup>1</sup>

CPU에 최적화된 소프트웨어에서 4세대 인텔® 제온® 스케일러블 프로세서는 **NVIDIA A100 GPU보다 최대 2.5배 빠른 성능**을 제공합니다.<sup>2</sup>

## CPU 성능 이점

### 더 빠른 엔드투엔드 머신 러닝 성능

NVIDIA는 최근 130만 개 세포의 분석 결과를 발표하여 NVIDIA A100 GPU의 성능을 인텔® 제온® 프로세서와 비교했습니다. 테스트에서는 CPU에 비해 GPU가 30배의 성능 우위를 보였습니다.

인텔® [oneAPI 데이터 분석 라이브러리](#) (oneDAL) 및 [Katana Graph](#)를 통해 더 나은 병렬 알고리즘을 사용하고 성능을 기본 아키텍처에 맞게 조정하여 파이프라인을 가속한 결과, **최적화된 인텔® 제온® 스케일러블 프로세서가 1.4배~2.5배 성능 우위를 보인다는 것을 확인할 수 있었습니다.**



<sup>1,2</sup> 성능은 사용, 구성 및 기타 요인에 따라 달라집니다. 워크로드 및 구성 세부 정보는 다음 [벤치마크 요약](#)에서 확인할 수 있습니다.



# AI 추론에서 성능 향상



**2~3배** 인텔®  
AMX와 함께 Tencent  
Search 응용 프로그램에서  
사용되는 BERT 모델의 AI  
처리량 (이전 세대 대비)



Tencent는  
최적화된 BERT 모델을  
사용하여 서비스 경험을  
개선하고 TCO를 절감할 수  
있습니다.

[사례 연구](#)



**3.4배** Meituan의  
컴퓨팅 비전 플랫폼에서  
인텔® AMX 사용 Bfloat16  
최적화를 통한 AI 처리량  
(인텔® AMX  
미사용 시 대비)



Meituan은 온라인  
리소스의 전반적인  
효율성을 3배 이상  
개선하고 서비스 비용을  
70% 절감했습니다.

[사례 연구](#)



**5.7배** 인텔® AMX 탑재  
4세대 인텔® 제온®  
스케일러블 프로세서를  
사용하는  
vSphere/vSAN 8.0 기반  
자연어 처리 시간 단축



인텔은 즉시 사용할 수 있는  
최고의 성능과 엔드 투 엔드  
생산성을 제공하기 위해 다양한  
개방형 무료 도구, 최적화된  
라이브러리, 업계 프레임워크를  
제공합니다.

[사례 연구](#)

[기사 및 데모](#)



# 인텔® 제온® 스케일러블 프로세서가 가장 빠르게 성장하는 워크로드를 지원하는 5가지 방법

1 기존 하드웨어에서 더 빠르게 워크로드 실행



2 다양한 워크로드 전반에서 성능 개선



3 전력 효율 향상



4 가장 민감한 데이터 보호 강화



5 가장 데이터 집약적인 워크로드를 위한 더 많은 처리 용량 확보

