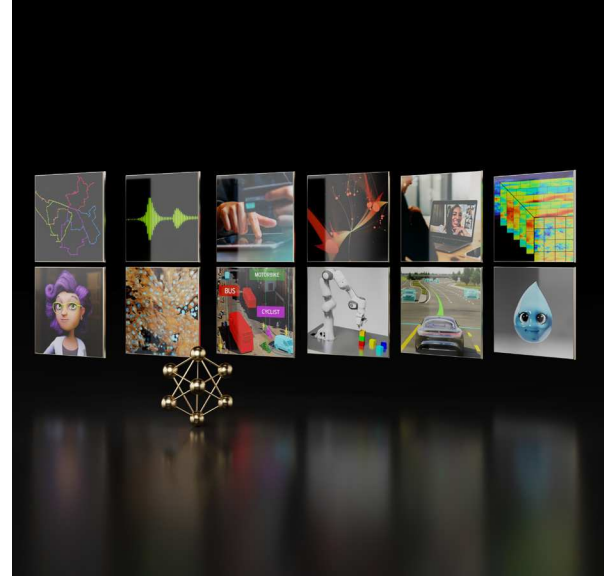




# NVIDIA AI Enterprise

엔터프라이즈급 소프트웨어 플랫폼으로  
프로덕션 AI 구축

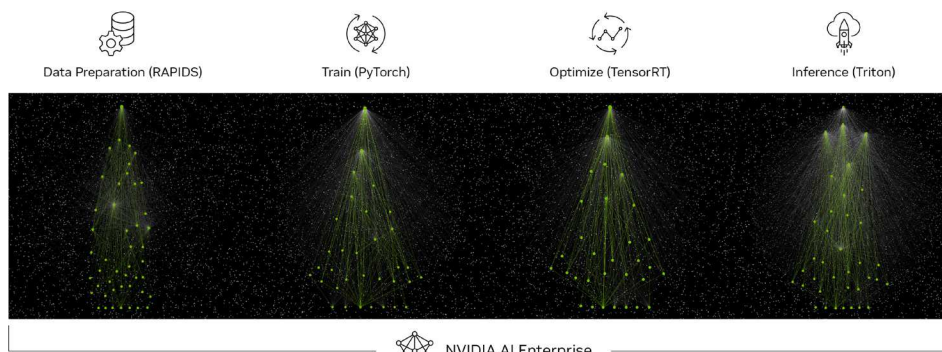


## AI 소프트웨어 플랫폼 구축 및 유지 관리의 당면 과제

AI가 비즈니스 방식을 근본적으로 변화시키고 있습니다. 기업들은 AI의 혁신적인 잠재력을 잘 알고 있지만 빠르게 진화하는 AI 기술을 구현하기란 쉽지 않습니다.

또한 AI를 구축하고 실행하기 위해 일관되고, 안전하며, 안정적인 소프트웨어 플랫폼을 유지하는 것은 대다수의 기업에게 복잡한 작업입니다. AI에 사용되는 기반 소프트웨어에 4,500개 이상의 서로 다른 소프트웨어 패키지가 포함되어 있는데, 그중 64개가 NVIDIA CUDA® 라이브러리고, 4,471개 이상이 타사 및 오픈 소스 소프트웨어(OSS) 패키지라고 추정할 수 있습니다.

이렇게 많은 소프트웨어 패키지를 사용하면 보안 취약점이 발생할 위험이 높습니다. 그뿐만 아니라 API 안정성과 서로 다른 소프트웨어 패키지 간의 9,000개의 종속성 유지 문제까지 생각하면 기업이 배포 환경에 최신 오픈 소스 버전을 안정적으로 사용하기가 사실상 불가능합니다.



NVIDIA AI Enterprise 기반 계층 이 그래프는 각각의 소프트웨어(점)와 종속성(선)을 보여줍니다.

## AI 가속화를 위한 엔터프라이즈급 소프트웨어

NVIDIA AI Enterprise는 고성능의 안전한 클라우드 기반 AI 소프트웨어 플랫폼을 자체적으로 구축하고 유지하려는 기업이 직면하는 복잡성을 해결해 줍니다.

NVIDIA AI Enterprise에는 데이터 사이언스 파이프라인을 가속하고 생성형 AI, 컴퓨터 비전, 음성 AI 등 프로덕션 AI의 개발 및 배포를 간소화하는 데 사용되는 풀스택

### 프로덕션 AI의 당면 과제

- > **복잡성:** 서로 다른 오픈 소스 소프트웨어에서 엔드 투 엔드 AI 소프트웨어 스택을 한데 모아 기존 인프라와 통합하는 것은 어려운 일입니다.
- > **위험성:** AI 소프트웨어 스택은 수천 개의 오픈 소스 패키지와 종속성으로 구성되어 있어 보안 패치를 적용하기가 쉽지 않습니다.
- > **신뢰성:** 투자 보호와 비즈니스 연속성을 위해서는 고성능의 AI 플랫폼을 유지하고 스택 전반에 걸쳐 API 안정성을 관리하는 것이 매우 중요합니다.

### NVIDIA AI Enterprise의 장점

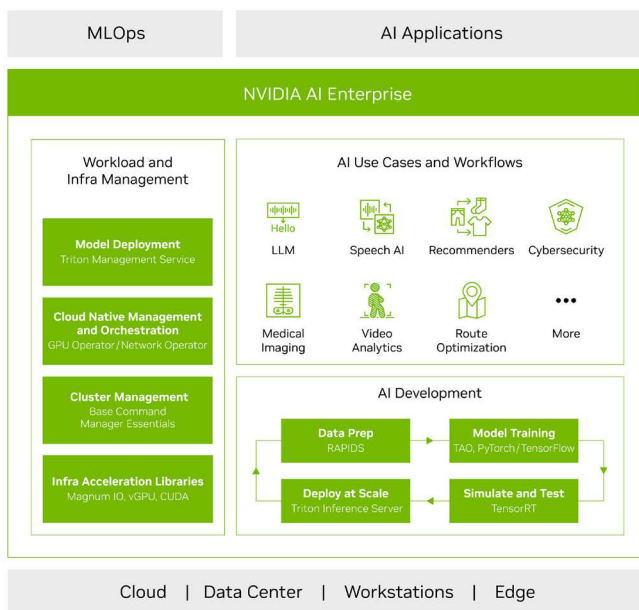
- > 가속 컴퓨팅을 통해 생산성을 높이고 비용을 절감합니다. 하루 24시간 연중무휴의 비즈니스 크리티컬 서비스
- > 엔터프라이즈급 보안, 신뢰성 및 지원을 통해 팀이 혁신적인 AI 솔루션을 자유롭게 구축할 수 있습니다. 고객 서비스 성공 관리자
- > 클라우드 기반이며, 현재 및 이전 세대 GPU 어디서나 실행되도록 인증되었습니다.
- > AI 워크플로와 사전 학습 모델을 통해 생산 시간을 단축합니다.

AI 소프트웨어가 포함되어 있습니다.

NVIDIA AI Enterprise는 다음을 포함하는 엔터프라이즈급 소프트웨어입니다.

- > 플랫폼에 포함된 엔드 투 엔드 프레임워크인 NVIDIA NeMo를 통해 기업은 NVIDIA가 사전 훈련한 기초 모델을 손쉽게 사용자 지정하여, 비즈니스 데이터를 기반으로 도메인별 사용 사례에 맞는 커뮤니티 모델을 선택할 수 있습니다.
- > 공개적으로 알려진 보안 취약점(Common Vulnerabilities and Exposures, CVE)에 대한 지속적인 모니터링 및 정기적인 보안 패치 릴리스
- > API 안정성을 보장하는 프로덕션 릴리스
- > 클라우드 및 데이터센터 환경 전반에 걸친 클러스터 관리, 자동 모델 배포, 클라우드 기반 오케스트레이션을 포함한 엔드 투 엔드 관리 소프트웨어
- > 서비스 수준 계약(Service-Level Agreement, SLA)을 통한 기업 지원 및 NVIDIA AI 전문가 서비스 제공

NVIDIA AI Enterprise는 AI의 복잡한 소프트웨어 플랫폼의 유지 관리 및 보안 부담을 덜어주어 기업이 AI를 구축하고 획기적인 인사이트를 활용하는 데 집중할 수 있게 해줍니다.



## AI 가속화를 통한 생산성 향상 및 TCO 절감

NVIDIA GPU에서 AI를 구축 및 실행하는 데 최적화된 100개 이상의 프레임워크와 사전 학습된 모델 및 개발 툴을 갖춘 NVIDIA AI Enterprise는 데이터 준비 및 모델 학습에서부터 추론 및 대규모 배포에 이르기까지 AI 여정의 모든 단계를 가속합니다.

- > Apache Spark용 NVIDIA RAPIDS™ 가속기를 사용하면 CPU 전용 플랫폼에 비해 데이터 처리 속도를 최대 5배까지 높이고, 운영 비용을 최대 5배까지 절감할 수 있습니다
- > NVIDIA TAO 툴킷으로 AI 전문 지식이나 대규모 학습 데이터셋 없이도 NVIDIA 사전 학습 모델을 조정하여 몇 달이 걸리던 맞춤형 프로덕션 레디(production-ready) AI 모델을 단 몇 시간 만에 만들 수 있습니다.

## 인증된 플랫폼

- > **멀티클라우드** : AWS, Google Cloud, Microsoft Azure를 포함한 주요 클라우드 마켓플레이스
- > **하이브리드 클라우드** : VMware Cloud Foundation, Red Hat Enterprise Linux, HPE GreenLake, Ubuntu KVM, Nutanix AHV.
- > **MLOps 통합** : Azure Machine Learning에서 기술 미리 보기 및 ClearML, Domino Data Lab, Run:ai, Weights & Biases를 통한 인증
- > **컨테이너 오케스트레이션** : VMware Tanzu, Red Hat OpenShift, HPE Ezmeral, Google Kubernetes Engine(GKE), Amazon Elastic Kubernetes Service(EKS), 업스트림 Kubernetes
- > **NVIDIA DGX 플랫폼** : 대규모 언어 모델에 대한 AI 학습을 강화하기 위해 NVIDIA DGX 플랫폼의 일부인 DGX 소프트웨어에 NVIDIA AI Enterprise 라이선스가 포함되어 있습니다.
- > **NVIDIA-Certified Systems™** : 다양한 제조업체에서 제공하는 400개 이상의 NVIDIA 인증 서버 및 워크스테이션을 사용할 수 있습니다. NVIDIA AI Enterprise의 소프트웨어 라이선스는 NVIDIA H100 PCIe 또는 NVL GPU에 각각 포함되어 있습니다.

## 라이선스 옵션 살펴보기

- > 소프트웨어 라이선스의 경우 클라우드 마켓플레이스를 통해 영구 라이선스를 취득하거나, 구독하거나, 구매할 수 있습니다.
- 제품 가격 및 라이선스**에 대해 자세히 알아보세요.

- > CPU 전용 플랫폼보다 최대 40배 빠른 성능으로 인프라 및 에너지 비용을 절감하는 NVIDIA® TensorRT™ 기반 애플리케이션을 통한 추론 최적화 GPU의 신경망 및 트리 기반 모델(Tree-based model) 모듈에 대해 프로덕션 환경에서 대규모의 AI 모델 배포를 간소화하고 최적화하는 NVIDIA Triton™ Inference Server를 통한 대규모 배포
- > NVIDIA AI Enterprise만의 독보적인 NVIDIA Triton 관리 서비스를 통해 추론 배포 및 모델 오케스트레이션을 대규모로 가속화.

## NVIDIA의 파트너 에코시스템

- > AI 및 머신 러닝 파이프라인 최적화를 위한 MLOps 솔루션 제공 파트너
- > 맞춤형 최첨단 AI 솔루션을 위한 글로벌 솔루션 통합 파트너
- > 응용 AI 서비스 제공 파트너

## NVIDIA AI 워크플로우를 통한 개발 시간 단축

AI 워크플로우는 기업이 AI 솔루션 구축을 빠르게 시작할 수 있도록 사전 패키징 된 클라우드 기반의 참조 예시로, 기업의 지식 기반에서 실시간으로 정보를 검색하여 정확한 응답을 생성하는 생성형 AI 챗봇, 지능형 가상 비서, 내부자 위협을 탐지하는 사이버 보안 솔루션, 스피어 피싱 이메일 탐지 성능을 개선하는 생성형 AI 사용 등이 있습니다.

이러한 솔루션은 마이크로서비스로서 실행할 수 있으며 Kubernetes 단독으로, 또는 다른 마이크로서비스와 함께 배포하여 프로덕션 환경에서 바로 사용할 수 있는 애플리케이션을 제작할 수 있습니다. NVIDIA AI 워크플로우는 AI의 결과 도출 경로를 가속하고, 배포 시간을 단축하며, 비용을 절감하고, 정확도와 성능을 개선할 수 있습니다.

## 어디에서나 실행할 수 있는 인증을 획득한 클라우드 기반 플랫폼

NVIDIA AI Enterprise는 퍼블릭 클라우드나 가상 데이터센터, NVIDIA DGX™ 플랫폼 등 어디에 배포하든 안정적인 성능을 제공하도록 최적화되고 인증되었습니다.

이는 개발한 애플리케이션을 어디에나 배포할 수 있는 유연성을 제공하여 파일럿에서 프로덕션으로 전환할 때 서로 다른 환경 간의 인프라 및 아키텍처 차이로 인해 발생할 수 있는 위험을 줄여줍니다.

## 시작할 준비가 되셨나요?

NVIDIA AI Enterprise에 대한 자세한 정보는 [nvidia.com/ai-enterprise-suite](https://nvidia.com/ai-enterprise-suite)에서 확인할 수 있습니다.

무료 90일 평가판 라이선스를 신청하려면 [nvidia.com/ai-enterprise-eval](https://nvidia.com/ai-enterprise-eval)를 방문하세요

NVIDIA AI Enterprise를 직접 경험해보고 싶다면 [nvidia.com/try-ai](https://nvidia.com/try-ai)에서 NVIDIA LaunchPad를 통해 무료 랩을 신청하세요.

영업팀 담당자 문의: [nvidia.com/ai-enterprise-sales](https://nvidia.com/ai-enterprise-sales)

