

Red Hat AI Enterprise를 사용하여 통합된 AI 플랫폼을 기반으로 구축하세요

AI의 급증

AI는 이미 광범위하게 채택되었으며, 다음 수치를 통해 확인할 수 있습니다.

- ▶ 55%의 조직이 생성형 AI를 사용합니다.¹
- ▶ 50% 이상의 조직이 에이전틱(Agentic) AI에 대한 개념 검증을 수행하거나 특정 활용 사례를 구현하고 있습니다.²
- ▶ 2026년 말까지 AI 기반 애플리케이션 중 30%가 에이전틱(Agentic) AI를 사용할 것으로 예상됩니다.²

AI 기회의 진화에 따른 복잡성 가중

AI는 기업에 기존의 운영 및 경쟁 방식을 새롭게 정의할 수 있는 기회를 제공합니다. AI는 오늘날 비즈니스 환경의 핵심 요소로 빠르게 자리 잡았지만, 채택이 늘어나고 있음에도 불구하고 많은 기업이 이 기술을 온전히 활용하지 못하게 하는 당면 과제를 안고 있습니다.

대표적으로 모델 비용의 증가, 복잡한 커스터마이징 요구 사항, 엄격한 배포 제약 조건, 갈수록 빨라지는 혁신 속도 따라잡기 등을 들 수 있습니다.

기업이 생성형 AI 혁신, 에이전틱(Agentic) AI 워크플로우, 소버린 AI에 더 많은 비중을 두기 시작하면서 이러한 문제를 해결하기가 갈수록 어려워지고 있습니다. 이 난제를 풀려면 추론 비용을 낮추고, 확장 및 모니터링을 간소화하고, 소버린 오픈프레미스 AI를 구축하고, 변화에 빠르게 적응하는 데 도움이 될 수 있는 AI 플랫폼으로의 전략적 전환이 필요합니다.

AI 복잡성을 해결해 주는 엔터프라이즈 플랫폼 및 툴

AI 여정을 이제 막 시작하는 기업이든, 기존의 AI 이니셔티브를 조직 전반으로 확장하는 기업이든 Red Hat® AI를 사용하여 AI 혁신을 가속화하는 동시에 AI 솔루션을 개발하고 하이브리드 및 멀티클라우드 환경 전반에 배포하는 데 드는 운영 비용을 절감할 수 있습니다.

Red Hat AI는 최적화된 모델과 효율적인 추론 능력을 갖추고, 프라이빗 데이터와의 간편한 통합을 지원하며, 확장성과 유연성을 갖춘 플랫폼에서 에이전틱(Agentic) AI를 더욱 빠르게 구현할 수 있도록 하는 비용 효율적인 솔루션을 제공합니다. 조직은 이를 통해 단일 서버 배포부터 대규모 분산 플랫폼에 이르기까지 예측 모델과 생성형 AI 모델의 라이프사이클을 관리하고 모니터링할 수 있습니다.

검증된 오픈소스 기술 기반의 Red Hat AI는 선도적인 오픈소스 AI 모델을 폭넓게 지원하여 기업이 혁신을 가속화하고 중요 툴 및 기술에 접근하는 문턱을 낮출 수 있도록 돕습니다.

AI 파트너 에코시스템은 이를 강력하게 뒷받칩니다. 파트너는 다양한 인프라 전반에서 성능, 안정성, 가속기 지원에 초점을 맞춘 검증되고 인증된 제품 및 서비스를 제공하여 기업이 AI를 통해 주요 비즈니스 및 기술 관련 과제를 해결하도록 돕습니다.

1 IDC 기술 구매자 프레젠테이션. "Worldwide Generative AI Industry Use Case Early Adoption Trends, 2025: Executive Summary"(전 세계 생성형 AI 산업 활용 사례 조기 도입 경향, 2025: 핵심 요약). 문서 번호 US53280825, 2025년 4월(유료)

2 IDC 설문조사. "Agentic AI Impact on Digital Infrastructure Strategies"(에이전틱(Agentic) AI가 디지털 인프라 전략에 미치는 영향). 문서 번호 US53418526, 2025년 10월(유료)

Red Hat AI는 다음을 제공합니다.

- ▶ 완전한 AI 모델 라이프사이클 모니터링 및 관리
- ▶ 모든 하드웨어 가속기에서 모든 AI 모델을 사용할 수 있는 유연성
- ▶ 검증되고 최적화된 일련의 AI 모델
- ▶ 하이브리드 및 멀티클라우드 전반의 유연하고 비용 효율적인 추론
- ▶ 기업이 소유한 프라이빗 데이터 자산과의 신뢰할 수 있는 통합
- ▶ 에이전틱(Agentic) AI 워크플로우 구현 간소화
- ▶ 효율적인 리소스 사용을 통해 필요에 따라 확장할 수 있는 능력

Red Hat AI Enterprise는 Red Hat AI 포트폴리오에 포함된 통합 AI 플랫폼으로, 기업이 AI 도입과 관련된 당면 과제를 해결하도록 돕습니다.

Red Hat AI Enterprise를 AI 플랫폼으로 선택해야 하는 6가지 이유

Red Hat AI Enterprise는 AI 여정의 모든 단계에서 기업을 지원하는 데 필요한 모든 톨과 기능을 제공합니다. 기업은 이를 사용하여 통합된 AI 플랫폼 기반의 하이브리드 및 멀티클라우드 환경 전반에 효율적이고 비용 효과적인 AI 모델, 에이전트, 애플리케이션을 개발하고 배포할 수 있습니다.

이 플랫폼은 기업에 바로 사용 가능한 개발 환경과 핵심 AI 기능을 제공합니다. 여기에는 모델 튜닝, 고성능 추론, 에이전틱(Agentic) AI 워크플로우 관리와 함께 모든 모델을 지원하고, 모든 하드웨어를 사용하며, 어디에나 배포할 수 있는 유연성이 포함되어 끊임없이 변화하는 데이터 위치 관련 요구 사항을 충족할 수 있습니다.

유연하고 효율적인 추론을 통해 AI 배포 간소화

AI 혁신 속도가 갈수록 빨라지고 있는 상황에서 배포의 복잡성 또한 커졌습니다. 이는 비효율적인 리소스 사용, 운영 비용 증가, 결과의 품질 저하라는 문제를 야기할 수 있습니다.

Red Hat AI Enterprise는 온프레미스, 하이브리드 및 멀티클라우드 환경, 네트워크 엣지를 포함해 다양한 환경에서 최적화된 모델 추론을 제공합니다. 기업은 이를 통해 모델 배포를 간소화하는 동시에 리소스 사용을 줄이고, 추론 비용을 절감하고, 정확성을 유지할 수 있습니다.

이는 vLLM으로 구동되는 최적화된 런타임을 통해 가능합니다. vLLM은 GPU 메모리를 더욱 효율적으로 사용해 처리량과 대기 시간을 최적화하여 생성형 AI 기반 애플리케이션의 결과 출력을 가속화하는 추론 서버입니다. 여기에는 고급 압축 기술도 포함됩니다. 이는 모델의 크기와 컴퓨팅 요구 사항을 줄여서 비용을 최적화하는 데 도움을 줍니다.

조직이 몇 개의 모델에서 수십 개의 모델로, 수십 명의 사용자에서 수백 명의 사용자로 확장하는 시점에 이를 경우 플랫폼은 llm-d를 사용할 수 있는 옵션을 제공합니다. 이 기술은 vLLM의 핵심 혁신 사항을 기반으로 구축되어 생성형 AI 모델에 비용 효율적이고 예측 가능한 성능을 대규모로 지원하는 분산 추론 프레임워크를 제공하며, 처리량이 증가함에 따라 성능이 더욱 향상됩니다.

프라이빗 엔터프라이즈 데이터로 AI 모델의 정확성 개선

공개적으로 사용 가능한 생성형 AI 모델이 다양한 기능을 제공하지만, 기업은 분야별 활용 사례를 해결하기 위해 고유한 자체 데이터를 기반으로 학습한 AI 모델을 필요로 합니다.

Red Hat AI Enterprise는 간소화되고 일관된 경험을 제공하여 효율적인 모델 커스터마이징을 지원하고 AI 엔지니어와 데이터 과학자가 모델 응답의 정확성과 관련성을 개선하는 것을 돕습니다. 기업은 이 솔루션을 통해 검색 증강 생성(Retrieval Augmented Generation, RAG) 활용 사례를 위해 모델을 조직 데이터와 연결할 수 있습니다.

이 플랫폼은 미세 조정(fine-tuning), 지속적 학습, 강화 학습도 지원합니다. 또한 문서 처리 및 구문 분석, 합성 데이터 생성, 모델 평가 및 다양한 태스크를 위한 유연한 모듈식 접근 방식을 제공합니다.

에이전틱(Agentic) AI 워크플로우로 혁신 가속화

AI 기술은 비교적 짧은 기간에 빠른 속도로 발전했습니다. 다양한 새 활용 사례를 제공한 예측 AI를 시작으로 생성형 AI 모델이 쏟아져 나왔으며, 현재 주목을 받는 것은 에이전틱(Agentic) AI 워크플로우입니다. 이는 AI 전략을 더욱 확장하고 기업에 더 많은 비즈니스 가치를 가져다 줄 차세대 기술입니다.

Red Hat AI Enterprise는 민첩하고 안정적인 플랫폼을 제공하여 기업이 에이전틱(Agentic) AI를 통해 가치를 성공적으로 실현할 수 있도록 돕습니다. 그것을 가능하게 하는 것은 통합된 애플리케이션 프로그래밍 인터페이스(Application Programming Interface, API) 계층, 기업 활용 사례에 맞게 에이전틱(Agentic) AI 워크플로우 어셈블리를 간소화하는 즉시 사용 가능한 구성 요소, 전용 사용자 환경, 에이전틱(Agentic) AI 시스템의 배포와 관리를 지원하는 유연하고 확장 가능한 기반입니다.

이 플랫폼은 여기에 더해 모델 컨텍스트 프로토콜(Model Context Protocol, MCP)에 대한 지원도 제공합니다. 이는 에이전틱 배포를 위한 핵심 구성 요소로, 광범위한 툴 및 기능과 대규모 언어 모델(Large Language Model, LLM) 사이에서 표준화된 번역기 역할을 합니다.

유연성과 일관성을 바탕으로 하이브리드 클라우드 전반에 AI 확장

다양한 산업의 조직이 초기 테스트 활용 사례에서 AI의 가치를 입증할 수 있었습니다. 하지만 그중 대다수는 여전히 그 가치를 엔터프라이즈 전반으로 확대하는 동시에 비용 효율성을 유지하고 규제 요건, 프라이버시 및 보안 요구 사항을 충족하는 데 어려움을 겪고 있습니다. 이는 하이브리드 및 멀티클라우드 환경이나 네트워크 엣지에서의 AI 배포와 같이 다양한 IT 환경을 사용하는 경우 더욱 복잡해집니다.

Red Hat AI Enterprise는 모든 하드웨어 또는 IT 환경을 아울러 조직의 AI 워크로드 전략에 가장 적합한 위치에서 예측 AI와 생성형 AI 모델을 학습, 튜닝, 배포, 관리하고 확장하는 데 필요한 제어 기능과 확장성을 제공합니다.

이 플랫폼은 원하는 하드웨어, OEM(Original Equipment Manufacturer), 클라우드 공급업체, 데이터센터의 조합을 기반으로 원하는 AI 모델을 사용할 수 있는 유연성을 선사합니다. 이를 일련의 포괄적인 관측성 및 모니터링 기능과 함께 사용하면 성능을 최적화하고, 비용을 효율적으로 제어하고, 규모 성장에 따라 거버넌스를 유지하고 보안에 초점을 맞춰 변화하는 비즈니스 요구 사항을 충족할 수 있습니다.

온프레미스 AI 배포로 데이터 주권 유지

세계 각국의 정보가 AI 사용과 데이터 저장 및 처리 방식에 관한 규제를 잇따라 도입하고 있습니다. 이러한 현상은 많은 기업이 데이터 주권과 프라이버시를 유지하기 위해 온프레미스 AI 배포를 구축할 수 있는 방법을 모색하도록 했습니다.

기업은 민감한 프라이빗 데이터를 온프레미스에서 저장하고 처리하여 끊임없이 변화하는 여러 규제를 준수할 수 있을 뿐만 아니라 공격 표면을 줄이고 보안에 더욱 집중하여 해당 데이터를 안전하게 보호할 수 있습니다.

Red Hat은 AI 보안에 대한 포괄적이고 계층화된 접근 방식을 제공하여 조직이 안심하고 AI 혁신을 수용할 수 있도록 합니다. Red Hat AI Enterprise는 신뢰할 수 있는 오픈 하이브리드 클라우드 기반 위에서 이러한 보안 접근 방식을 제공하여 AI 라이프사이클의 고유한 보안 및 안전 문제를 해결하도록 돕습니다.

이는 에어갭 환경을 포함해 모든 클라우드 및 엣지 환경에 걸쳐 모든 모델, 모든 하드웨어, 보안 중심의 GPU 공유를 지원하여 진정한 데이터 주권을 실현할 수 있도록 합니다. Red Hat은 소버린 클라우드 제공업체와 긴밀하게 협력하여 조직이 보안 중심의 프라이빗 클라우드 환경에 AI를 배포할 수 있도록 돕습니다.

또한 Red Hat AI Enterprise는 모델 설명 가능성(explainability), 공정성, 출력 정책 적용을 통해 감사 가능한 신뢰를 지원하는 톨과 함께 재현 가능한 파이프라인과 감사 준비를 통한 라이프사이클 추적성도 제공합니다.

신뢰할 수 있는 쿠버네티스 플랫폼 기반의 AI 구축

Red Hat AI Enterprise는 통합되고 즉시 사용 가능한 엔드 투 엔드 AI 스택의 일부로 이 완전한 AI 기능 세트를 제공합니다. 이는 신뢰할 수 있는 쿠버네티스 기반 위에서 구축되었습니다.

이 플랫폼은 이미 쿠버네티스와 컨테이너화를 사용하고 있는 기업에 온프레미스 데이터센터, 하이브리드 또는 멀티클라우드 환경, 엣지에서의 AI 배포에 걸쳐 친숙하고 일관된 경험을 제공합니다. 기업은 팀이 이미 알고 있는 동일한 톨, 프레임워크, 기술을 사용하여 하나의 중앙화된 플랫폼에서 현대적인 클라우드 네이티브 AI 기반 애플리케이션을 구축하고, 학습시키며, 배포하고 관리할 수 있습니다.

Red Hat AI Enterprise의 가치에 대해 자세히 알아보기

조직이 Red Hat AI Enterprise를 사용하여 AI 여정의 모든 단계를 지원하는 통합된 플랫폼으로 하이브리드 클라우드 전반에서 효율적이고 비용 효과적인 AI를 배포할 수 있는 방법에 대해 [자세히 알아보세요](#).

한국레드햇 홈페이지 <https://www.redhat.com/ko>



Red Hat 소개

Red Hat은 세계적인 엔터프라이즈 오픈소스 소프트웨어 솔루션 공급업체로서 커뮤니티 기반의 접근 방식을 통해 신뢰도 높은 고성능 Linux, 하이브리드 클라우드, 컨테이너 및 쿠버네티스 기술을 제공합니다. 또한 Red Hat은 고객이 클라우드 네이티브 애플리케이션을 개발하고, 신규 및 기존 IT 애플리케이션을 통합하고, 복잡한 환경을 자동화하고 관리할 수 있도록 지원합니다. [Fortune 선정 500대 기업에 신뢰하는 어드바이저](#)인 Red Hat은 전 세계 고객에게 [권위 있는 어워드를 수상](#)한 지원, 교육 및 컨설팅 서비스를 제공하여 모든 산업 분야에서 개방형 혁신의 이점을 실현할 수 있도록 최선을 다하고 있습니다. Red Hat은 기업, 파트너, 커뮤니티로 구성된 글로벌 네트워크의 허브 역할을 하며 고객들이 성장하고, 확장하고, 디지털 미래에 대비할 수 있도록 지원합니다.

f www.facebook.com/redhatkorea
구매문의 02-6105-4390
buy-kr@redhat.com