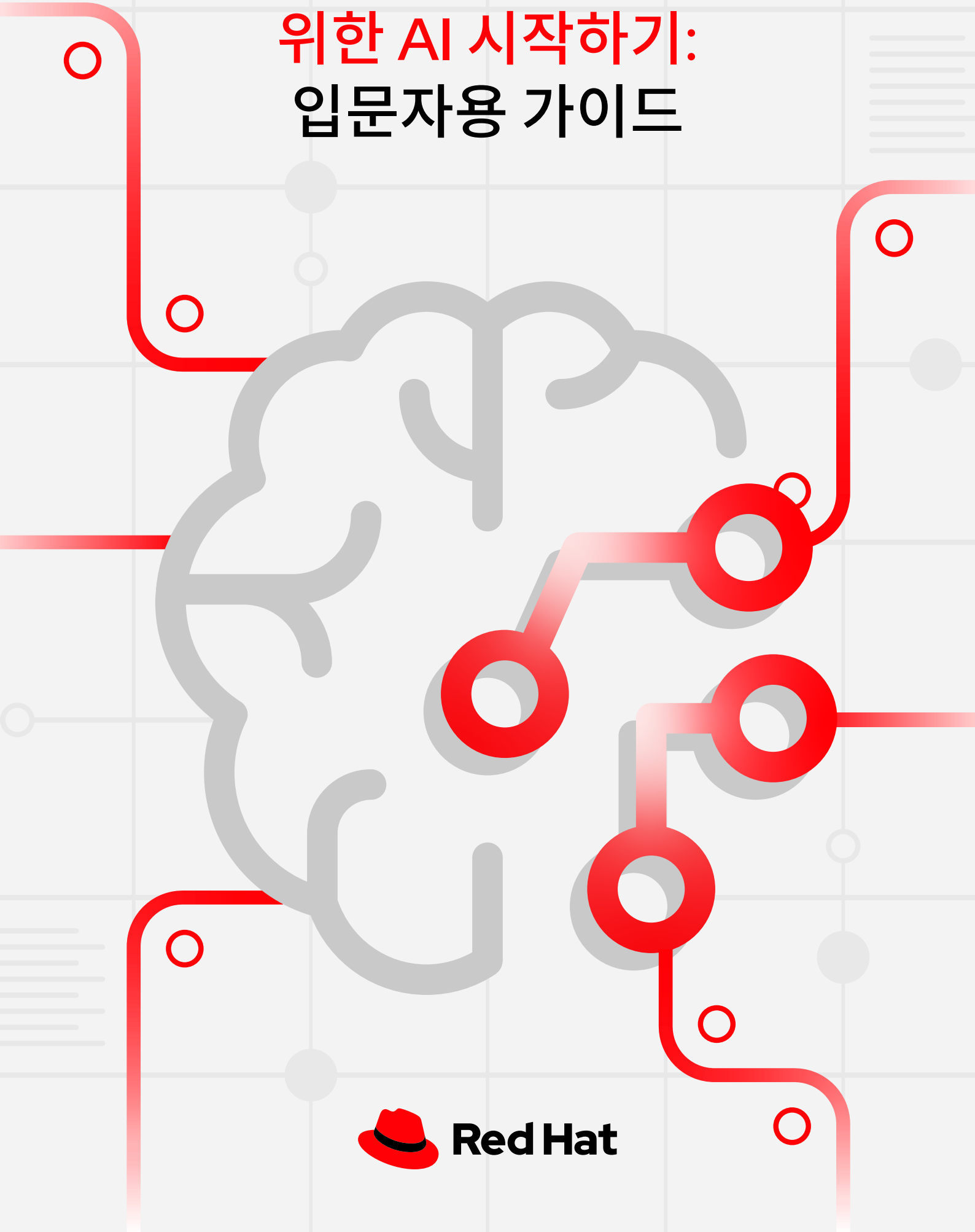
엔터프라이즈를 위한 AI 시작하기: 입문자용 가이드



Red Hat

목차

개요

3페이지

제1장

AI의 등장

6페이지

제2장

적합한 AI 모델 선택

9페이지

제3장

시작을 위한 필요조건

12페이지

제4장

Red Hat 도입과 이를 기반으로 한 확장

16페이지

자세히 알아보기

AI 도입 여정의 다음 단계로 나아갈 준비가 되셨나요?

20페이지

개요

인공지능(Artificial Intelligence, AI)이 비즈니스의 모든 측면에 제공하는 기회에 대한 기업의 인식이 날로 높아지고 있습니다.

고객 인게이지먼트, 지원, 영업부터 IT 인프라, 프로세스, 코드 개발, 솔루션 제공에 이르기까지 AI의 활용 사례가 확대되고 있으며 모든 산업에서 AI 활용이 폭발적으로 증가하고 있습니다.

IDC에 따르면 많은 기업이 운영 효율성, 고객 경험, 생산성 개선에 초점을 두고 AI 이니셔티브를 추진하면서 AI 시장이 5년 동안 매년 평균 26.9%씩 성장하여 2027년에는 4,230억 달러로 커질 것으로 예상됩니다.[1]

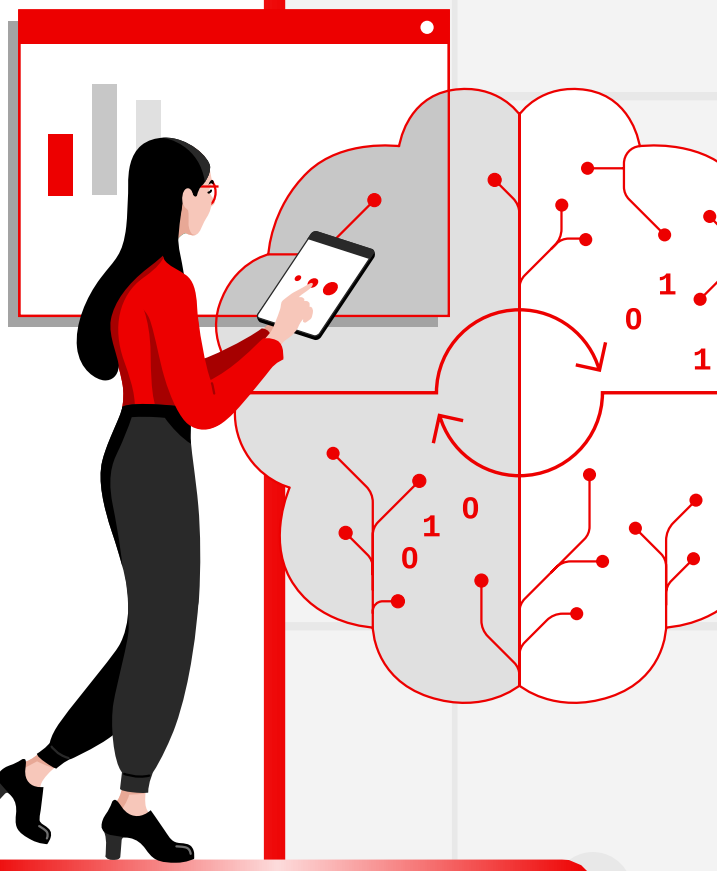
이 급변하는 상황 속에서 리더는 조직의 경쟁 우위를 높일 수 있는 AI 솔루션을 찾고, 선택하고, 빌드하고, 제공해야 한다는 과제를 안고 있습니다. 그러나 AI 혁신 속도와 대부분의 조직이 AI 성숙도를 높이는 역량의 속도는 다릅니다. 이로 인해 AI의 잠재력을 온전히 활용하지 못하고 있으며, 대부분 해답보다는 더 많은 질문을 갖게 됩니다.

AI 여정을 지금 막 시작하는 단계이든, AI가 비즈니스에 미칠 영향에 대해 알아보려고 하든, 기존에 구현한 AI를 확장하는 방법을 모색하든 이 e-book을 통해 현재 AI에 대해 갖고 있는 수많은 질문에 대한 답을 찾을 수 있을 것입니다.

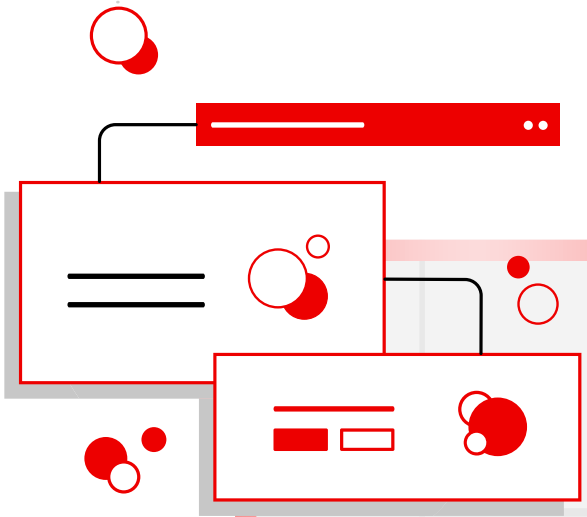
[1] IDC FutureScape 웹캐스트. 'Worldwide Artificial Intelligence and Automation 2024 Predictions(전 세계 인공지능 및 자동화 2024 예측).' 문서 번호 #US51901124, 2024년 3월.

AI에는 어떤 유형이 있나요?

AI를 최대한으로 활용하려면 AI에 대한 모든 것을 알아야 합니다. 현재 기업에서 사용하는 대표적인 AI 유형 중 2가지를 알아보겠습니다.



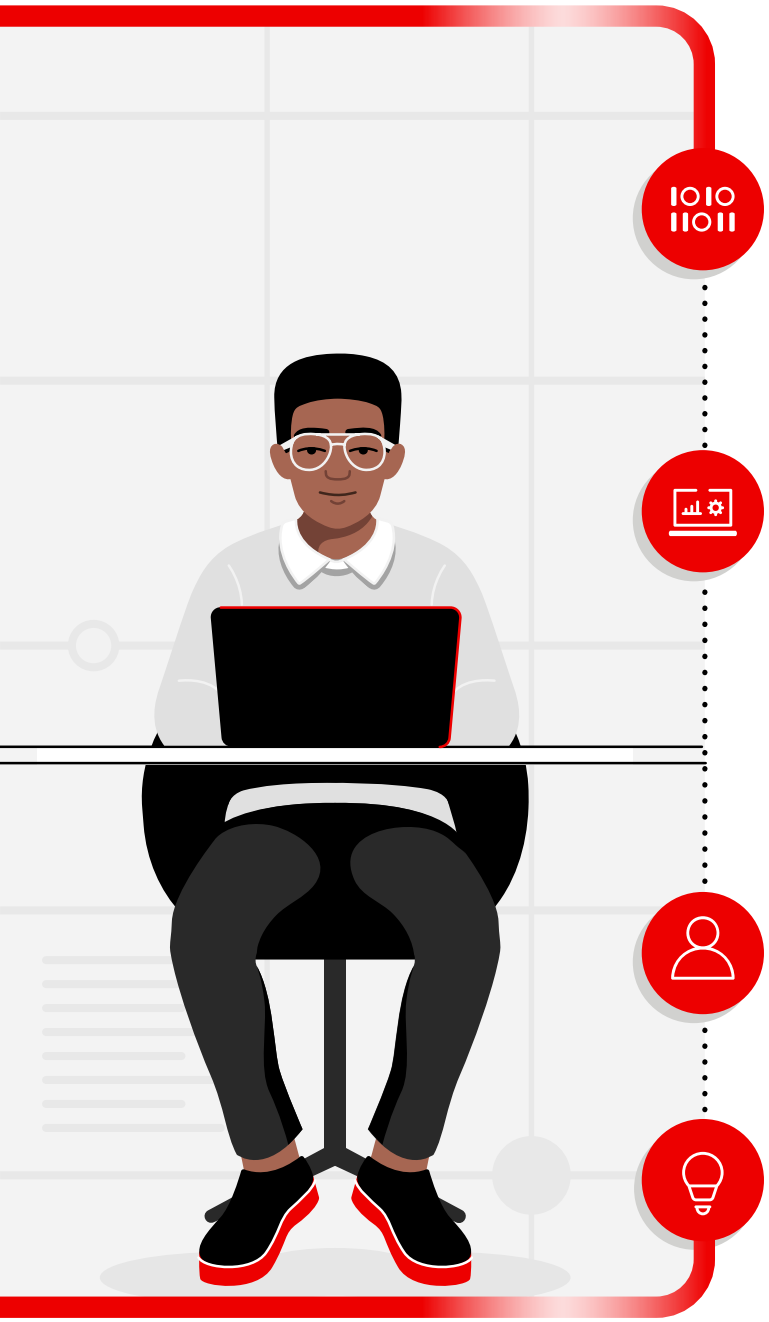
예측 AI: 예측 AI는 과거 데이터를 사용하여 조직이 패턴을 찾고 데이터를 기반으로 미래에 대한 의사결정을 내릴 수 있도록 합니다. 예측 모델은 수요 예측, 예측 유지 관리, 운영 계획과 같은 활용 사례를 지원합니다. 예측 AI는 강력한 데이터 사이언스와 머신 러닝(Machine Learning, ML) 기술을 기반으로 하여 데이터 처리량 증가에 따라 AI가 개선될 수 있도록 합니다.



생성형 AI(생성형 AI):
트랜스포머(transformer)와 같은 딥러닝 모델을 기반으로 한 생성형 AI는 텍스트, 이미지, 코드와 같은 새로운 콘텐츠를 생성할 수 있습니다. 이 유형의 AI는 챗봇, 자동화된 콘텐츠 생성, 창의적인 툴과 같은 활용 사례에 특히 유용합니다. 사전 학습한 생성형 트랜스포머(Generative Pretrained Transformer, GPT)와 같은 모델은 사람과 같은 텍스트와 이미지를 생성하여 자연어 처리와 창의적인 분야를 획기적으로 바꾸어 놓았습니다.

AI 구현의 이점은 무엇인가요?

AI가 어느 정도의 잠재력을 갖고 있는지 아직 완전히 알지는 못하지만, 이미 다양한 산업 내 모든 규모의 기업에서 급속도로 발전하는 이 기술을 얼마나 유용하게 사용하고 있는지 알아보고 조직에 AI를 통합할 영역을 결정하는 데 참고할 수 있습니다.



다음과 같은 AI의 이점을 조직에서 어떻게 활용할지 고민해 보세요.

데이터 볼륨. 데이터의 양이 기하급수적으로 증가하면서, 조직은 수집하는 대량의 정보를 관리하고 그런 정보로부터 인사이트를 얻는 것에 어려움을 겪고 있습니다. AI는 거대한 데이터세트를 신속하게 처리하고 분석하여 수작업으로는 찾기 어려운 중요한 인사이트와 추세를 발견할 수 있습니다.

운영 효율성. 많은 조직은 비효율적인 프로세스와 병목으로 인해 생산성이 저하될 수 있으며, 그 결과 이러한 문제를 해결하는 데 더 많은 시간과 노력이 든다는 것을 잘 알고 있습니다. AI 기반의 자동화를 통해 운영을 간소화하여 오류를 줄이고 프로세스 효율성을 개선할 수 있습니다. 작업 항목과 다음 단계가 적힌 회의 노트를 자동으로 생성하는 단순한 작업부터 웹사이트나 소셜 미디어에 게시할 이미지와 동영상을 더 빠르게 제작하는 사례 등 다양한 예시가 있습니다.

고객의 기대치. 고객은 원활한 맞춤형 경험을 기대합니다. AI는 고객 데이터를 분석하고 개인화된 추천과 맞춤형 상호작용을 제공하여 고객 서비스와 개인화의 정도를 향상할 수 있습니다.

시장 경쟁력. 급변하는 시장에서 경쟁력을 유지하려면 지속적인 혁신이 요구됩니다. AI를 활용하면 조직이 시장의 변화에 더 빠르게 대응하고 경쟁 우위를 유지할 수 있습니다. 또한 생성형 AI는 경영진의 사고를 돕는 파트너로 활용하거나 중요한 회의를 준비할 때 접근 방식을 개선하는 데 사용할 수도 있습니다.

제1장

AI의 등장

AI는 수십 년 동안 진화를 거듭하면서 헬스케어, 금융, 제조와 같은 산업의 발전을 앞당겼습니다.

그러나 최근 등장한 생성형 AI는 사랑과 같은 텍스트, 현실적인 이미지, 심지어 소프트웨어 코드를 생성할 수 있는 능력으로 큰 주목을 받았습니다. 작업을 자동화하거나 데이터를 분석하는 기존 AI와는 달리, 생성형 AI는 창의적인 문제 해결과 고품질 콘텐츠 제작이라는 새로운 영역의 문을 열었습니다.

혁신을 가속화하는 AI 모델의 유형

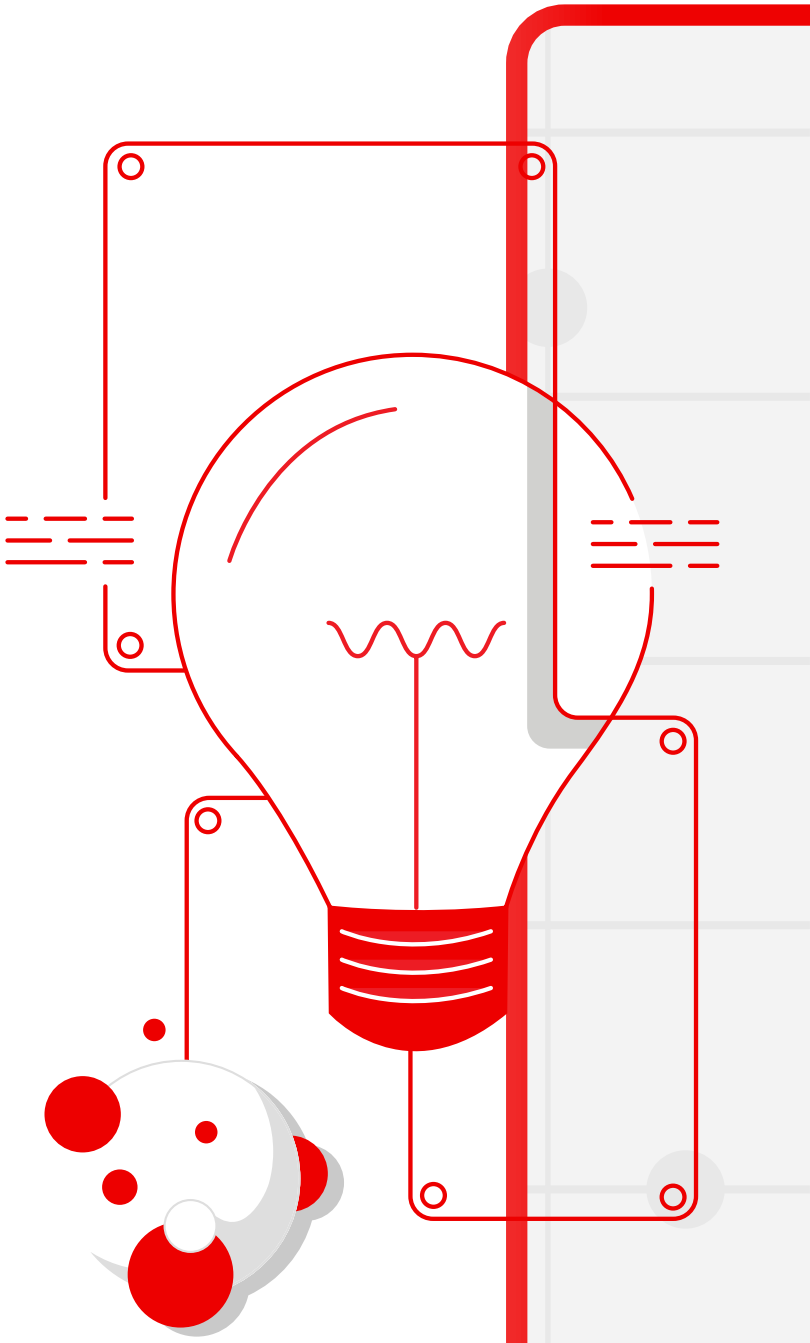
대규모 언어 모델(Large language models, LLM)과 스테이블 디퓨전(stable diffusion) 모델은 생성형 AI의 폭발적 증가에 기여하는 대표적인 AI 모델입니다. LLM은 GPT와 마찬가지로 대량의 데이터세트를 기반으로 사전 학습하여 자연어를 이해하고 생성할 수 있어 고객 지원 자동화, 마케팅 카피 생성 등에 매우 유용하게 활용됩니다. 한편 스테이블 디퓨전 모델은 초현실적인 이미지를 생성할 수 있어 엔터테인먼트, 마케팅 등의 영역에서 혁신을 가속화합니다.

고려해야 할 새로운 동향

텍스트, 이미지, 데이터 처리 능력을 하나의 모델로 통합하여 더욱 다양하게 활용 가능한 솔루션을 제공하는 멀티모달 AI를 사용하려는 기업이 늘고 있습니다. 기업 환경에서 AI의 잠재력을 최대한으로 활용하려면 이러한 동향에 한발 앞서 대응해야 합니다.

오픈소스: AI 혁신의 토대

Red Hat의 AI 전략은 오픈소스에 뿌리를 두고 투명성과 신뢰성, 저렴한 비용을 바탕으로 기업이 생성형 AI를 발전시킬 수 있도록 돕습니다. 기업은 Red Hat의 오픈 하이브리드 클라우드 플랫폼을 사용하여 AI 솔루션을 제어하면서 자유롭게 혁신할 수 있습니다.



LLM과 그 작동 원리에 대해 자세히 알아보기

오픈소스로 LLM 제어하기

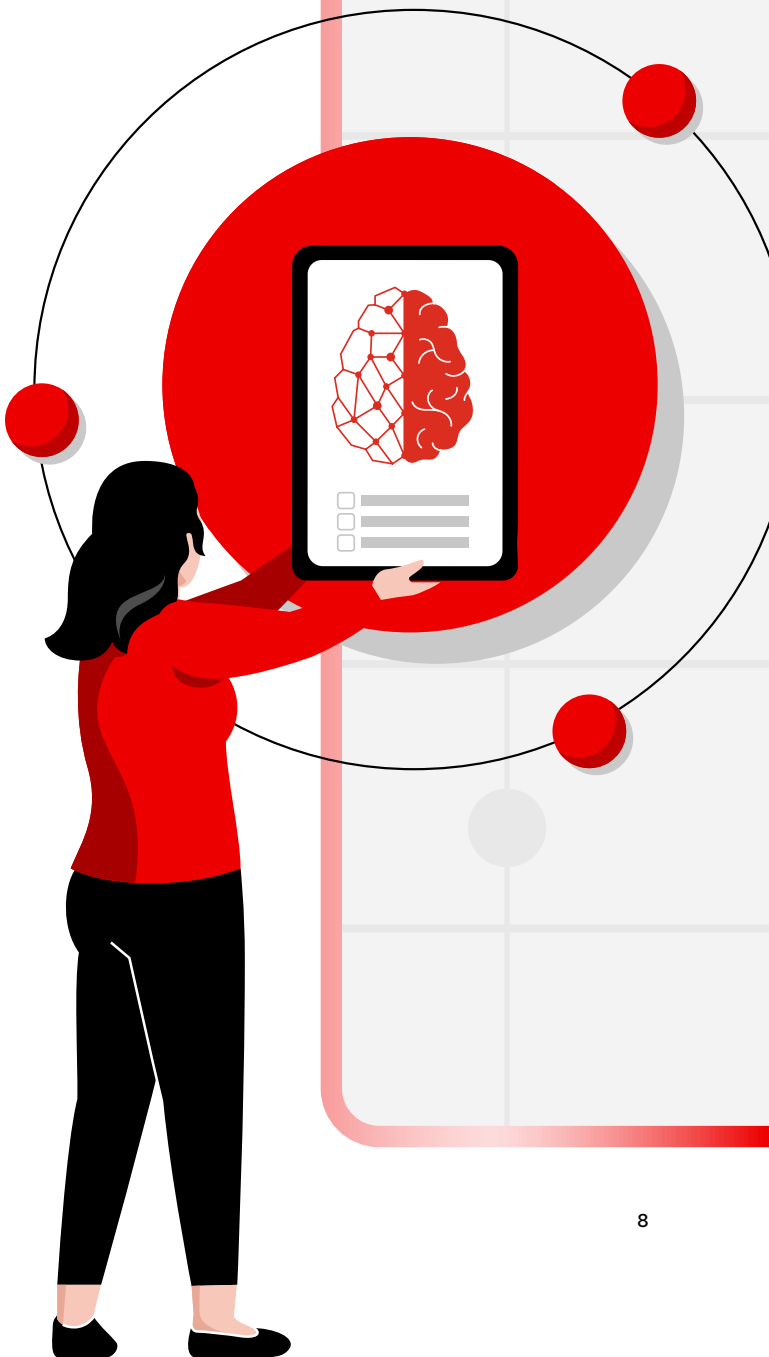
생성형 AI가 소프트웨어 개발 방식부터 사람들의 상호작용 방식에 이르기까지 비즈니스의 거의 모든 부분을 변화시키고 있지만, 서비스 공급업체가 생성형 AI 기능의 일부로 사용되는 모델(LLM 및 기타 모델)을 엄격하게 통제하는 경우가 종종 있습니다. 즉, 기업이 전문적인 스킬을 보유하고 있지 않은 경우 생성형 AI 서비스의 기능을 평가하기가 쉽지 않으며, 스킬을 보유하고 있어도 금전 및 시간 면에서 많은 비용이 소요될 수 있습니다.

모델을 만든 데이터 세트 또는 모델이 데이터를 사용하는 방식에 관한 상세 정보를 확인할 수 없어 기업은 AI 생성 콘텐츠 측면에서 잠재적 리스크에 노출될 수 있습니다. 코드 생성 모델이 저작권으로 보호되는 소스 코드를 기반으로 학습했다면 어떨까요? 해당 모델이 생성한 코드가 저작권으로 보호되는 코드의 일부라면 어떨까요? 이와 비슷한 많은 질문에 아직 완전한 답을 얻지는 못했지만 그 영향을 이해하는 것이 중요하다는 것을 깨달은 기업들은 오픈소스 AI로 눈을 돌리고 있습니다.

Red Hat의 AI에 대한 접근 방식은 오픈소스에 뿌리를 두고 있으며, 파운데이션 모델의 IBM Granite 제품군과 같은 오픈소스 모델에 대한 지원도 마찬가지입니다.

Red Hat의 AI 솔루션은 LLM 기능 향상을 위한 커뮤니티 주도형 솔루션인 InstructLab을 통해 AI 모델 개발에 직접 기여하기도 합니다.

GitHub의 InstructLab 살펴보기



제2장

적합한 AI 모델 선택

다양한 활용 사례에 다양한 AI 모델을 사용하고 적용할 수 있습니다.

하나의 애플리케이션 또는 서비스에 예측 AI와 생성형 AI를 모두 사용할 수 있습니다. 각 모델은 비용이 각기 다르고 고유한 장점을 제공하지만, 조직의 초기 기술 검증(POC) 시 시장 출시 시간을 단축하는 데 도움을 줄 수 있습니다. 이미지 세분화, 음성-텍스트 변환, 이미지 인식 모델은 일반적인 동시에 매우 효율적인 예시이지만, 중요한 부분은 비즈니스의 활용 사례에 가장 적합한 모델이 무엇인지 평가하는 것입니다.

대량의 데이터를 기반으로 학습한 파운데이션 모델은 기능 사용이 매우 유연하지만 큰 크기로 인해 비용이 늘어나고 관리 요구 사항이 증가하고 복잡성이 가중되어 모든 활용 사례에는 적합하지 않을 수 있습니다.

크기가 작고 미세 조정된 모델은 생성형 AI 제품군에 속하지만 기존 모델을 요구 사항에 맞게 미세 조정하려는 경우에 더 적합한 솔루션일 수 있습니다. 모델을 결정할 때 즉시 사용 가능하고 시스템에 손쉽게 통합되는 사전 구축된 옵션을 선택할 수 있습니다. 그러한 모델의 일반적인 예는 LLM입니다. 이는 대량의 데이터를 기반으로 사전 학습한 강력한 툴입니다.

그러나 구체적인 비즈니스 요구 사항이나 데이터 프라이버시 관련 우려 사항이 있거나 모델의 동작을 더 적극적으로 제어하기 원하는 경우 사용자 정의 모델을 구축하고 자체 호스팅해야 할 수 있습니다.

모델 구축과 모델 조정 간 비교

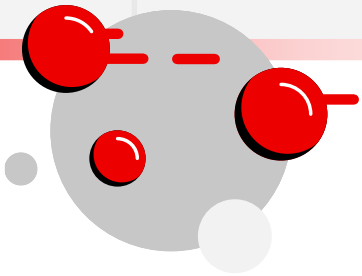
AI 모델을 완전히 처음부터 구축하려면 대단히 많은 작업을 해야 할 수 있습니다. 이 경우 조직의 비즈니스 활용 사례와 관련된 대량의 데이터세트를 수집하고 준비해야 합니다. 그런 다음 적절한 알고리즘을 선택하고 데이터를 기반으로 학습시켜야 합니다. 이 프로세스는 상당히 높은 수준의 컴퓨터 능력과 전문 지식을 요구하기 때문에 많은 시간과 리소스를 필요로 합니다. 전통적인 모델 또는 파운데이션 모델을 구축하는 것은 사용자 정의 솔루션을 제공할 수는 있지만, 모든 경우에 가장 효율적인 방법은 아닙니다.

한편 파운데이션 모델을 조정하는 것에는 사전 학습한 모델을 특정 요구 사항에 맞게 조정하는 작업이 포함됩니다. 일반적인 방식은 전이 학습(transfer learning)입니다. 여기에는 거대한 데이터 세트를 기반으로 학습한 모델을 더 작은 도메인별 데이터 세트를 기반으로 재학습시키는 것이 포함됩니다. 이 방법을 사용하면 모델이 초기 학습 중에 배운 일반적인 지식을 그대로 보유하면서 특정 데이터의 미묘한 차이에 맞게 조정할 수 있습니다.

모델 미세 조정

또 다른 방법은 미세 조정입니다. 이는 특정 태스크의 성과를 개선하기 위해 모델의 매개변수를 조정하는 것입니다. 모델의 매개변수는 선택한 모델의 변수를 의미하며, 제공된 데이터를 모델에 적합하게 맞추어 추정할 수 있습니다. 미세 조정에는 학습 속도를 변경하는 것, 모델의 아키텍처를 수정하는 것 또는 모델의 특정 계층을 다른 계층보다 더 강도 높게 학습시키는 것 등이 포함될 수 있습니다. 이러한 기술을 통해 모델의 지식을 향상하여 특정 활용 사례에 더 효율적인 모델로 개선할 수 있습니다.

InstructLab은 필요한 AI 관련 사전 지식을 줄이고 기존 Granite 파운데이션 모델에 기업의 지식을 간편하게 추가할 수 있게 하는 것을 목표로 미세 조정 접근 방식을 사용합니다.



모델 미세 조정에 대한 대안

연구원들은 파운데이션 모델을 더욱 빠르고 효율적으로 조정하는 방법을 연구하고 있습니다. 몇 가지 공통적인 기술 중 하나는 **검색 증강 생성(Retrieval-Augmented Generation, RAG)**입니다. 이는 추가 지식(컨텍스트)이 인코딩된 외부 소스에서 팩트를 검색하는 데 사용되는 기술입니다.

RAG는 하나 이상의 외부 데이터베이스(벡터 데이터베이스)를 사용하며, 이 데이터베이스는 생성형 AI 모델이 받는 질문에 추가 컨텍스트를 제공합니다. 또 다른 새로운 방식은 에이전틱 AI(Agentic AI) 시스템입니다. 이는 여러 생성형 AI 에이전트를 하나로 결합하여 내부 데이터베이스, 회사 인트라넷 또는 인터넷과 같은 외부 시스템에 지식을 쿼리한 다음 생성형 AI 모델에 가장 정확한 최신 정보를 제공합니다.

마지막 사례는 **프롬프트 튜닝(prompt tuning)**으로, 이는 AI 모델이 수신하는 추가 단어나 AI 생성 숫자와 같은 신호 또는 프론트엔드 프롬프트를 튜닝하여 모델이 바람직한 결정을 내리도록 유도하는 것을 말합니다. RAG 쿼리의 결과는 프롬프트 튜닝 및 RAG와 함께 사용되어 프롬프트의 추가 컨텍스트를 구성합니다. 기존 LLM의 미세 조정, RAG 기술, 프롬프트 튜닝을 함께 사용하면 데이터가 제한적인 조직도 좁은 범위의 태스크에 맞게 파운데이션 모델을 조정할 수 있습니다.

AI 모델의 기반 인프라도 모델만큼이나 중요합니다. 작업마다 필요한 하드웨어가 다릅니다.

중앙 처리 장치(Central Processing Unit, CPU)

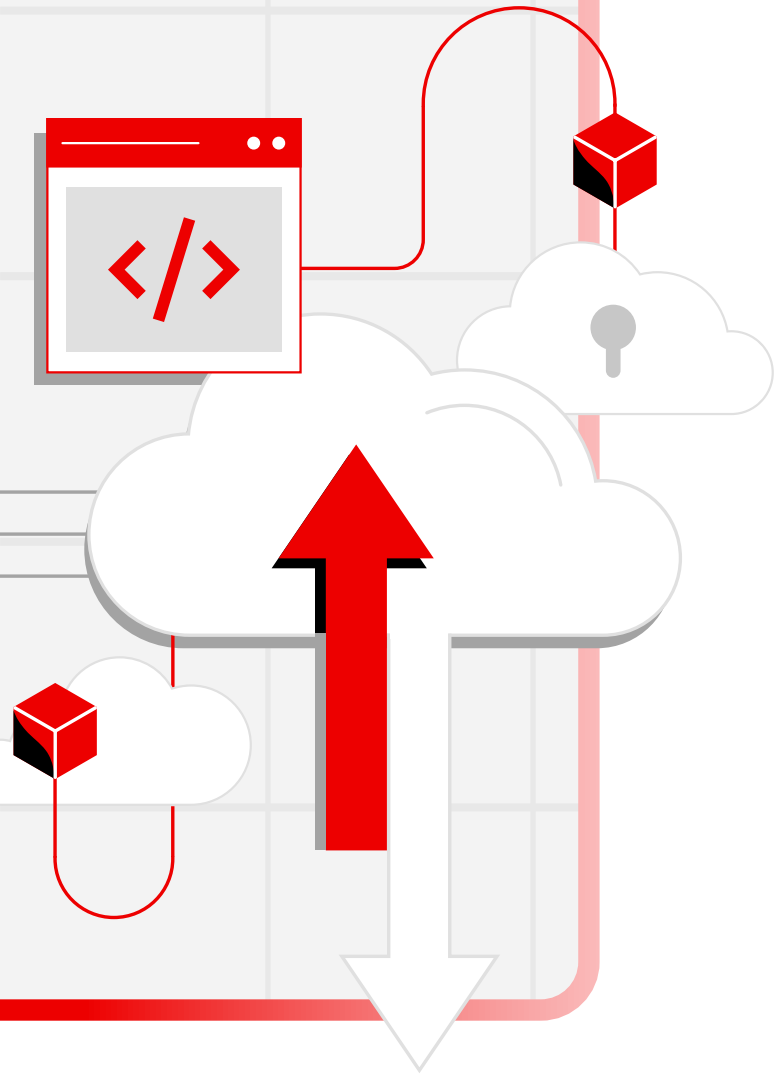
일반적인 컴퓨팅 작업을 처리하는 전통적인 프로세서입니다. 다양한 작업에 사용할 수 있지만 대규모 AI 워크로드에 사용하기에는 비효율적입니다.

그래픽 처리 장치(Graphics Processing Unit, GPU)

병렬 처리 작업을 다루도록 설계된 특수한 프로세서로, 대량의 데이터를 동시에 처리해야 하는 딥러닝 모델을 학습시키는 데 적합합니다.

신경망 처리 장치(Neural Processing Unit, NPU)

AI 작업을 위해 특별히 설계된 최신 유형의 프로세서로, 특정 유형의 모델에 더 뛰어난 효율성과 속도를 제공합니다.



기업의 AI 도입에서 하이브리드 클라우드가 하는 역할

하이브리드 클라우드 환경은 AI 도입에 중요한 역할을 합니다. 하이브리드 클라우드는 온프레미스 인프라와 퍼블릭 및 프라이빗 클라우드 리소스를 결합하여 AI 워크로드를 배포하고 관리하는 방식과 위치에 대한 유연성을 제공합니다. 예를 들어 강력한 클라우드 기반의 GPU를 사용하여 AI 모델을 학습시킨 다음 보안 또는 컴플라이언스와 관련된 이유로 이를 온프레미스 또는 프라이빗 클라우드에 배포할 수 있습니다. 따라서 하이브리드 클라우드 접근 방식을 사용하는 경우, 선택하는 툴과 플랫폼의 일관성을 중요하게 고려해야 합니다.

조직에서 Red Hat의 오픈 하이브리드 클라우드 접근 방식을 사용하면 다양한 환경에 걸쳐 AI를 통합하여 일관성과 확장성, 유연성을 개선할 수 있습니다. 이 방식으로 여러 클라우드 환경, 온프레미스 또는 네트워크 엣지에 걸쳐 AI 워크로드를 관리하고, 데이터 배치를 최적화하고, 원활한 데이터 마이그레이션을 지원하여 엔터프라이즈 규모에서의 AI 도입을 더욱 용이하게 할 수 있습니다.

AI 모델, 데이터, 인프라에 대해 이해한다면 AI 도입에 따른 복잡성을 더 효율적으로 해결하고 AI의 잠재력을 최대한으로 활용할 수 있습니다.

제3장

시작을 위한 필요조건

모든 새로운 기술을 도입할 때와 마찬가지로, AI를 성공적으로 도입하려면 조직이 극복해야 할 여러 과제가 있습니다.

다음 고려 사항을 참조하여 조직의 준비 상태를 평가하고 AI 도입 속도를 높이기 위해 중점적으로 노력해야 하는 영역을 확인하세요.

데이터 품질 및 가용성 평가. AI의 성능이 우수하려면 고품질의 관련성 있는 데이터에 액세스할 수 있어야 합니다. 데이터 품질은 AI 모델을 정확히 학습시키는 데 필수적이기 때문에, 데이터의 완전성, 정확성, 관련성을 평가하는 것이 중요합니다.

기술 인프라 평가. 현재 인프라가 AI 워클로드를 지원할 수 있는지 확인하세요. 여기에는 고성능 컴퓨팅 리소스, 스토리지 솔루션, 자동화, 네트워크 기능을 사용할 수 있는지 평가하는 것이 포함됩니다.

스킬이 필요한 영역 확인. 조직 구성원이 AI 관련 전문 지식을 보유하고 있는지 평가하세요. 현재 스킬 세트를 평가하고 교육 또는 전문적인 스킬이 필요할 수 있는 영역을 확인하세요.

전략과의 방향 일치 검토. AI 이니셔티브가 비즈니스 목표 및 전략과 방향이 일치하는지 확인하세요. AI 프로젝트는 조직의 전반적인 전략적 목표를 지원하고 측정 가능한 비즈니스 가치를 제공해야 합니다.

AI를 시작하는 방법

조직 내 AI 도입의 속도와 규모는 다양한 요인에 달려 있습니다. 하지만 거의 모든 기술 현대화 프로젝트에 유효한 방식은 작게 시작하여 점진적으로 규모를 늘리는 것입니다.

다음은 조직이 AI 도입 여정을 시작하고 진행하는 데 도움이 되는 8단계입니다.

1

역량 및 목표 평가

먼저 조직의 현재 역량, 인프라, 전략적 목표를 평가하는 것으로 시작하세요. AI가 조직의 광범위한 목표와 방향이 일치하는지 확인하고 AI를 통해 가치를 높일 가능성이 있는 영역을 파악하세요. 이 초기 평가를 통해 AI 도입 여정의 방향을 명확하게 설정할 수 있습니다.

2

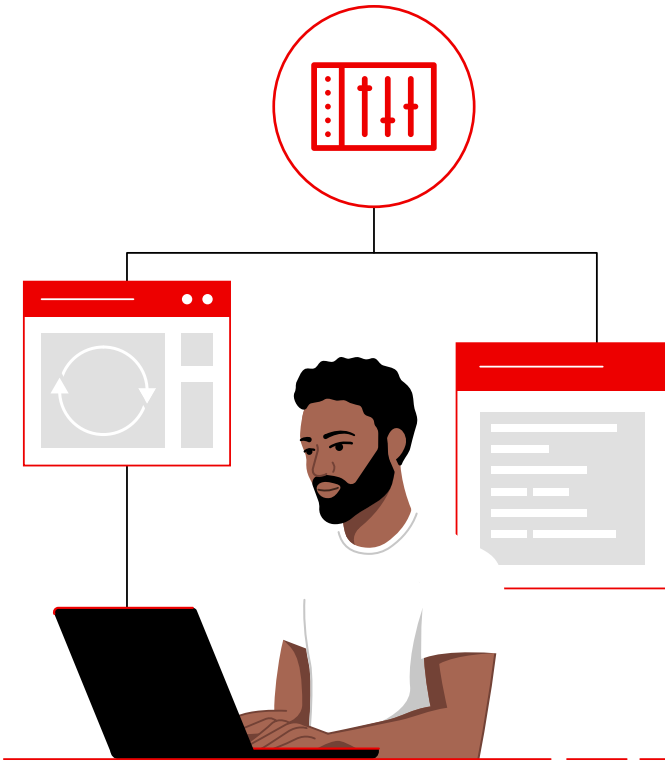
활용 사례 및 AI 팀 확인

조직 내에서 AI로 문제를 해결할 수 있는 기회를 확인하세요. 개발자, 도메인 전문가, 데이터 사이언티스트, IT 전문가 등으로 이루어진 다기능 팀을 구성하여 AI 애플리케이션을 전담하고 이니셔티브를 주도하도록 하세요. 적절하게 정의된 활용 사례에 따라 AI 도입을 추진하고 리소스를 집중적으로 할당하세요.

3

모델 선택

확인된 활용 사례에 따라 적절한 AI 모델을 선택하세요. 생성형 AI를 위한 LLM, 데이터 분석을 위한 예측 모델 등 활용 사례의 목표에 부합하는 기능을 제공하는 모델을 선택하세요. 모델의 복잡성, 확장성, 기존 시스템과의 호환성과 같은 요인을 고려하세요.



4

테스트 및 검증 루프

AI 구현의 성공 기준을 성능 메트릭, 정확도, 비즈니스 목표 등으로 명확하게 정의하세요. 테스트 및 검증 루프를 구축하여 모델의 효과를 지속적으로 평가하세요. 이러한 루프를 통해 얻은 정기적인 피드백을 바탕으로 모델을 미세 조정하고 AI 여정이 계획대로 진행되게 하세요.

5

모델 조정

조직의 데이터를 활용하여 선택한 모델을 사용자 정의하세요. 이 조정 프로세스에는 관련된 데이터를 모델에 제공하여 모델 정확성을 개선하고 비즈니스 활용 사례와의 관련성을 높이는 것이 포함됩니다. 미세 조정을 통해 조직의 고유한 컨텍스트와 요구 사항에 맞게 모델을 조정할 수 있습니다.

6

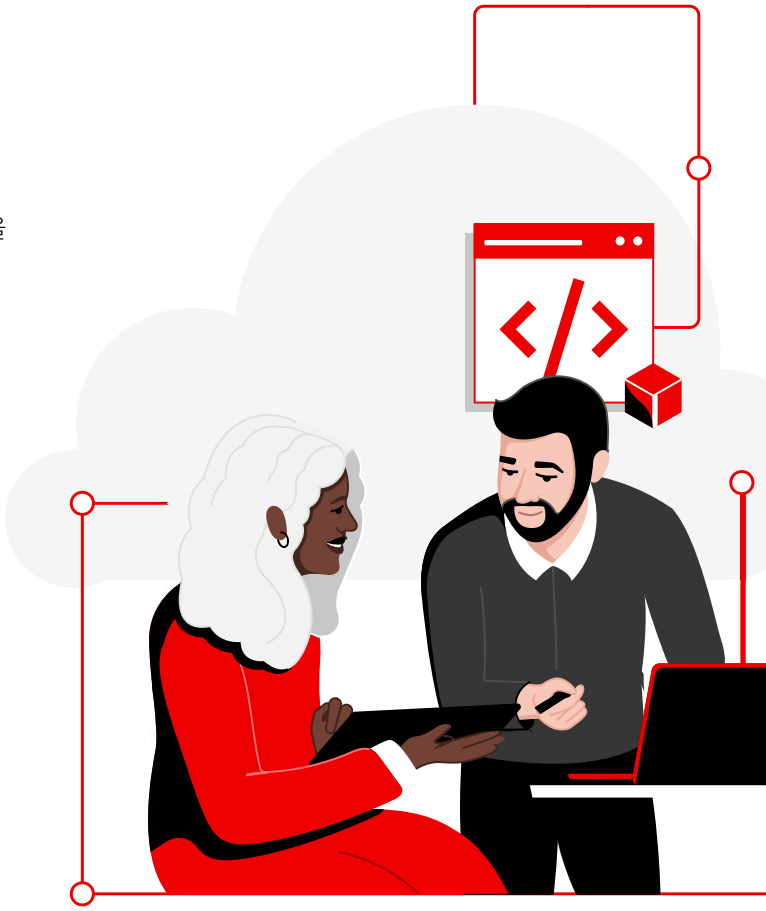
합성 데이터 학습

합성 데이터를 사용하여 AI 모델을 추가로 학습시키고 개선하세요. LLM이 마치 교사와 학생처럼 가르치고 학습하는 방법을 사용하는 이 접근 방식을 통해 실제 데이터가 부족하거나 민감한 데이터인 경우에도 고품질 학습 데이터를 생성할 수 있습니다. 합성 데이터를 사용하면 데이터를 보호하면서 모델의 견고함과 성능을 개선할 수 있습니다.

7

드리프트 모니터링

드리프트 모니터링은 일반적인 콘텐츠 기반의 모니터링을 제공합니다. 구조화된 구성 관리와 달리, 드리프트 모니터링은 로컬 파일 시스템에서 수행된 콘텐츠 변경을 추적합니다. 드리프트 모니터링을 통해 시간이 지남에 따라 모델의 정확성이 달라지거나 감소하는 것을 감지하고 해결할 수 있습니다. 지속적인 모니터링을 통해 상황이 변화하더라도 모델이 유효성과 관련성을 계속 유지하도록 할 수 있습니다.



8

전문가의 지원 활용

사내 구성원의 AI 전문 지식이 아직 충분하지 못하다면 Red Hat® Consulting과 같은 외부 전문가의 지원을 받는 것이 큰 도움이 될 수 있습니다. Red Hat 전문가는 AI 도입 중 복잡한 문제를 원활하게 해결하도록 돕고, 중요한 인사이트를 제공하고, 필요한 교육을 받도록 도울 수 있습니다. Red Hat Consulting은 귀사의 AI 여정을 가속화하고 성공 가능성을 높일 수 있습니다.

AI 도입의 성패를 결정하는 협업

성공적인 생성형 AI 프로젝트의 핵심은 다양한 역량을 갖춘 팀을 구성하는 것입니다.[2]

- **비즈니스 리더**는 솔루션을 사용하거나 솔루션의 영향을 받는 사람들을 대표합니다.
- **AI 전문가**는 생성형 AI 모델을 조정, 유지 관리, 업데이트합니다.
- **데이터 사이언티스트**는 정확하고 편향되지 않은 학습 데이터를 전처리하여 모델에 제공합니다.
- **윤리 및 컴플라이언스 담당자**는 생성형 AI 이니셔티브가 규정을 준수하는지 확인합니다.
- **IT 운영 전문가**는 솔루션을 기존 인프라에 통합하고 보안 정책을 적용합니다.
- **개발 팀과 커뮤니티**는 시작 단계에 참여하여 AI 도입을 위해 협업하고 오픈소스 툴 프레임워크 및 모범 사례를 구축, 공유, 개선해야 합니다. 이를 통해 AI 사용이 비즈니스 가치와 연결되도록 할 수 있습니다.

[2] Kearney. 'Standing up tiger teams to tackle generative AI complexity(생성형 AI의 복잡성을 해결하기 위한 강력한 팀 구축)', 2023년 11월 15일.

제4장

Red Hat 도입과 이를 기반으로 한 확장

Red Hat AI는 하이브리드 클라우드 전반에서 신뢰성과 유연한 선택, 그리고 일관성을 제공하여 기업의 AI 도입을 가속화합니다.

Red Hat AI 포트폴리오는 개별 Linux 서버 환경을 위한 Red Hat Enterprise Linux® AI와 분산된 쿠버네티스 플랫폼 환경을 위한 Red Hat OpenShift® AI를 포함하고 있으며, 통합된 머신 러닝 운영 기능을 제공합니다. 두 솔루션은 모두 오픈소스 기술과 모델을 기반으로 하여 조직이 AI 혁신을 주도하고, 더 빠르게 탐색하고, 최첨단 툴과 기술에 대한 접근 장벽을 없앨 수 있도록 돕습니다.

또한 방대한 규모의 Red Hat 파트너 에코시스템으로 AI 역량을 더욱 강화할 수 있습니다. 예를 들어 GPU 사용을 대중화한 것으로 널리 알려진 대표적인 AI 기업인 NVIDIA는 Red Hat과 파트너로서 지속적으로 협업하면서 AI 워크로드에 최적화된 엔드-투-엔드 엔터프라이즈 플랫폼을 제공하여 AI의 잠재력을 활용할 수 있도록 돕고 있습니다. NVIDIA는 기업 고객이 GPU로 가속화된 AI 컴퓨팅과 고성능 컴퓨팅 애플리케이션을 도입하는 것을 지원합니다.

> **"Red Hat과 NVIDIA는 장기간에 걸쳐 긴밀하게 협력해 왔으며, Red Hat Enterprise Linux AI는 차세대 AI 기술과 애플리케이션을 구축하는 개발자와 연구자에게 풀스택 컴퓨팅 및 소프트웨어를 제공하는 데 두 회사가 공동으로 기울이는 노력을 보여줍니다."[3]**
>
> **Justin Boitano**, 기업용 제품 담당 VP, NVIDIA

[3] Red Hat 보도 자료. "*Red Hat Enterprise Linux AI를 통해 액세스 가능한 오픈소스 생성형 AI 혁신을 제공하는 Red Hat*", 2024년 5월 7일.

Red Hat Enterprise Linux AI 자세히 알아보기

Red Hat Enterprise Linux AI는 4개의 기본 구성 요소로 이루어져 있습니다.

1

개방형 Granite 모델

Red Hat Enterprise Linux AI에는 Red Hat에서 완벽하게 지원하는 오픈소스 Granite 모델이 포함되어 있습니다. 이 유연한 모델을 통해 사용자 정의 언어 모델을 개발한 다음 공개 또는 비공개로 사용할 수 있습니다.

2

InstructLab 모델 조정

InstructLab은 Red Hat과 IBM이 주도하는 오픈소스 프로젝트입니다. 이는 특정 지식을 사용하여 AI 모델을 사용자 정의하고 학습을 위한 합성 데이터를 생성합니다. 사용자는 Git 리포지토리와 통합되는 커맨드라인 툴로 스킬을 높이고 모델을 손쉽게 학습시킬 수 있습니다.

3

최적화되고 부팅 가능한 Granite 모델 및 InstructLab용 Red Hat Enterprise Linux

Granite 모델과 InstructLab 툴은 거의 모든 하드웨어 및 클라우드 환경과 호환 가능한, AI에 최적화된 특수 Red Hat Enterprise Linux 이미지에서 실행됩니다. 이 설정은 빠른 학습 및 모델 배포에 필요한 고급 GPU를 통해 효율적인 성능을 발휘할 수 있도록 합니다.

4

엔터프라이즈 지원 및 보증

Red Hat Enterprise Linux AI 서브스크립션에는 엔터프라이즈 지원, Granite 7B 모델 및 소프트웨어로 시작하는 전체 제품 라이프사이클, Red Hat의 IP 보증이 포함됩니다.

생성형 AI 애플리케이션 개발을 돕는 Red Hat Enterprise Linux AI

생성형 AI 여정을 이제 막 시작한 조직을 위해 Red Hat Enterprise Linux AI는 단일 서버 개발 및 추론 환경에서 즉시 사용 가능한 LLM과 코드 언어 모델을 제공합니다.

이는 모델 및 툴과의 통합된 환경을 제공하여 광범위한 AI 관련 전문 지식 또는 인프라 없이도 비즈니스 데이터를 사용하여 더욱 쉽게 생성형 AI를 시작하고 모델을 사용자 정의할 수 있게 해 줍니다.

Red Hat Enterprise Linux AI는 Red Hat에서 완벽하게 지원하고 보증하여 리스크를 줄입니다. 또한 개발자와 도메인 전문가가 더욱 쉽게 접근할 수 있도록 설계된 생성형 AI에 대한 단순한 접근 방식을 통해 그들이 협업하고 더욱 빠른 시간 내에 비즈니스 성과를 달성할 수 있도록 돕습니다.

왜 Red Hat Enterprise Linux AI를 선택해야 할까요?

엔터프라이즈를 위한 LLM

오픈소스 라이센스 IBM Granite LLM은 Apache-2.0 라이센스에 포함되며, Red Hat이 전적으로 지원하고 보증합니다.

커뮤니티 협업

InstructLab을 사용하면 생성형 AI 모델 실험과 조정을 간소화할 수 있습니다.

클라우드 네이티브 확장성

Red Hat Enterprise Linux 이미지 모드를 사용하면 컨테이너 이미지로 AI 플랫폼을 관리할 수 있어 스케일링 방식을 간소화할 수 있습니다.

가속화 및 AI 툴링

오픈소스 하드웨어 가속기와 최적화된 딥러닝 기능이 더 빠른 결과를 지원합니다.

Red Hat OpenShift AI를 통한 확장

Red Hat OpenShift AI는 하이브리드 클라우드 환경 전반에서 AI 지원 애플리케이션과 예측 모델 및 파운데이션 모델을 규모에 맞게 빌드하고 학습시키고 조정하며 배포 및 모니터링하기 위한 통합 MLOps 플랫폼입니다.

Red Hat OpenShift AI는 Red Hat OpenShift 위에 구축되어 AI/ML 프로젝트의 워크로드 및 성능 요구를 다룰 때 일관적이고 능률적이며 자동화된 경험을 제공합니다. 조직은 MLOps 관행을 통해 AI 혁신에 신속하게 대응하고 AI 기반의 애플리케이션을 프로덕션에 더 빠르게 제공할 수 있습니다.

Red Hat OpenShift AI 샌드박스에서 직접 경험해 보세요

Red Hat OpenShift AI는 Open Data Hub와 기타 오픈소스 프로젝트에서 엄선된 구성 요소를 통해 데이터 사이언티스트와 개발자에게 데이터로부터 인사이트를 수집하고 AI 지원 애플리케이션을 빌드할 수 있는 강력한 오픈 하이브리드 AI/ML 플랫폼을 제공합니다.

개발자 샌드박스에서 체험하기 >

왜 Red Hat OpenShift AI를 사용해야 할까요?

모델 서빙 확장

온프레미스, 퍼블릭 클라우드 또는 엣지에서 지능형 애플리케이션으로 통합하는 데 모델 서빙을 활용할 수 있습니다. 이러한 모델은 소스 Notebook 변경 사항에 따라 재구축, 재배포, 모니터링될 수 있습니다.

AI 인프라 관리에 소요되는 시간 절감

리소스에 대한 온디맨드 액세스를 팀에 제공하여 데이터를 탐색하고 조직에 가치를 더하는 애플리케이션을 구축하는 데 집중할 수 있게 하세요.

검증되고 지원되는 AI/ML 툴링

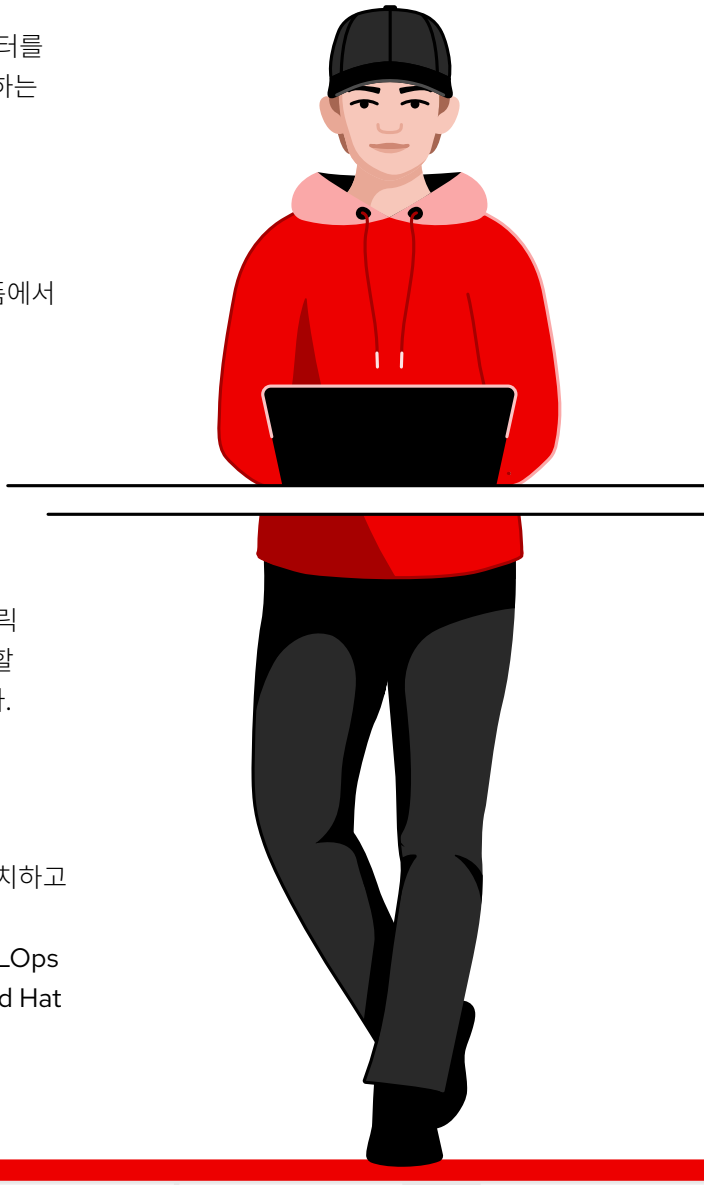
Red Hat은 Red Hat OpenShift 애플리케이션 플랫폼에서 공통 AI/ML 툴과 모델 서빙을 추적, 통합, 테스트, 지원합니다.

하이브리드 클라우드 전반의 유연성

자체 관리형 소프트웨어로 제공되거나 Red Hat OpenShift 기반의 전체 관리형 클라우드 서비스로 제공되는 Red Hat OpenShift AI는 온프레미스, 퍼블릭 클라우드 또는 엣지 등 모델 개발 및 배포 위치를 선택할 수 있는, 보안에 중점을 둔 유연한 플랫폼을 제공합니다.

모범 사례 기반의 운영

Red Hat Consulting은 Red Hat OpenShift AI를 설치하고 구성하여 최대한 활용할 수 있는 서비스를 제공합니다. Red Hat OpenShift AI 파일럿 경험이 필요하거나 MLOps Foundation을 빌드하는 데 가이드가 필요한 경우, Red Hat Consulting은 지원과 멘토링을 제공합니다.



Red Hat은 완벽한 기술 포트폴리오, 입증된 전문성, 전략적 파트너십을 제공하여 생성형 AI 목표 실현을 지원합니다. 생성형 AI 모델과 애플리케이션을 개발하고 배포하는 데 필요한 기반은 물론 신속한 도입을 위한 서비스와 교육도 제공합니다.

자세히 알아보기

AI 도입 여정의 다음 단계로 나아갈 준비가 되셨나요?

필요한 모든 위치에서 AI 애플리케이션을 실행할 수 있는 유연성을 제공하는 Red Hat의 오픈 하이브리드 클라우드 전략으로 AI 도입을 가속화하세요.

Red Hat의 전문 지식, 컨설팅 및 교육 서비스를 기반으로 AI/ML 프로젝트를 빠르게 시작하여 조직이 AI를 통해 원하는 목표를 달성하도록 도우세요.

AI/ML 서비스 자세히 알아보기:
red.ht/aiml-consulting

무료 디스커버리 세션을 예약해 보세요.
redhat.com/consulting

Red Hat AI에 대해 자세히 알아보기

Red Hat Enterprise Linux AI에 대해 자세히 알아보기

Red Hat OpenShift AI를 사용하여 확장할 수 있는 방법 알아보기

Copyright © 2025 Red Hat, Inc. Red Hat, Red Hat 로고 및 OpenShift는 미국과 그 외 국가의 Red Hat, Inc. 또는 계열사의 상표이거나 등록 상표입니다. Linux®는 미국 및 기타 국가에서 Linus Torvalds의 등록 상표입니다. 기타 모든 상표는 해당 소유자의 자산입니다.