

AI 통합 플랫폼을 통한 금융 혁신

: 개방형 아키텍처와 MLOps 구현 전략

김종규 (jonggyou@redht.com)

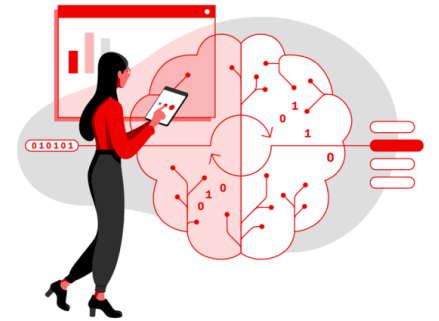
Red Hat Korea



1. AI 세대
2. 금융에 최적화된 AI를 육성하는 방법
3. 금융을 위한 AI 전략
4. Red Hat 의 AI 전략
5. 요약



AI 시대



AI 시대

최신뉴스 전체 | 정치 | 경제 | 사회 | 세계 | IT/과학 | 스포츠 | 연예 | 칼럼

은행들, '생성형 AI' 주도권

브리핑경제 원문 | 기사전송 2024-11-17 09:44

0 0

뉴스

[2024 금융AI 혁신①] '생성형AI'로 또 한

격동... 은행권, 'AI 조직' 공격적 확대

경제 > 경제

시중銀, 생성형 AI 연계 혁신금융 도전장

발행일 : 2024-10-03 15:53 | 지면 : 2024-10-04 | 2면

가 가

금융 · 증권 / 금융일반

"은행원 없이 시만"...

금융권 최초 '미래은행 점포' 구현...생성형
"무인 점포로 운영 예정"...신한은행 '들들

경제

최신뉴스 | 생활 경제 | 부동산 | 금융/증권 | 산업/기업 | 취업/창업 | 국제경제 | 경제일반

금통위 대신 챗GPT?...글로벌 중앙은행도 AI 열공 중

아주경제 원문 | 기사전송 2024-11-14 06:01

AI챗으로 요약

0 0

가속화되는 생성형 AI 금융...데이터산출물 시장변동성 리스크 대비 필수

이정민 기자 leejm0130@csnews.co.kr | 승인 2024.11.13 15:10 | 댓글 0

'AI 국가 총력전' 선포... 금융AI 판 커진다

[Future & 2024] 2027년까지 AI 3대 강국 도약... 망 분리 개선 속도낸다

강한빛 기자 | 1,079 | 2024.10.08 | 04:30:00

용분야 AI 응용, 목표 수준 95%까지 올라와"

발행일 : 2024-11-14 16:00 | 지면 : 2024-11-15 | 3면

가 가

생성형AI-RPA

증권사, 디지털 금융 전환 위해 'AI' 도입 박차..."기술 고도화 추진"

박영선 기자 | 입력 2024.11.15 07:53 | 댓글 0

연계 시스템' 구축

24-11-13 | 3면

가 가

AI 솔루션 및 서비스에 대한 수요 증가

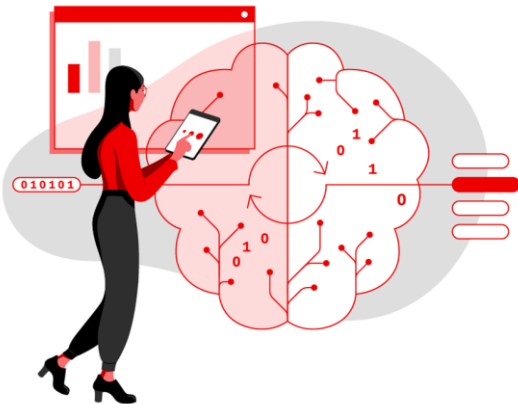
전 세계 AI 소프트웨어 시장은
2026년까지 약 7,900억 달러
규모로 성장할 것입니다¹

82%

조직 중 82%가
향후 2년 내 AI 워크로드
수요 증가를 예상합니다.²

65%

조직 중 65%가
현재 **생성형AI**에
투자하고 있습니다.³



45%

조직의 45%가
두 개 이상의 기능에서
생성형AI를 사용하고 있습니다.⁴

AI의 진화

비즈니스 인텔리전스에서 생성형 AI까지

- ✓ 오늘날의 실용적인 AI는 데이터 기반
- ✓ 오늘날 비즈니스를 운영하는 예측 AI
- ✓ AI의 가치를 실현하는 지름길을 제공하는 기초 모델

비즈니스 분석 및 인텔리전스

- 데이터 수집
- 데이터 저장 및 이동
- 구조화된 데이터
- 데이터 변환

데이터 웨어하우스

고급 분석 및 Predictive AI

- 데이터 과학 기술
- 비정형 데이터
- 예측 분석
- 실시간 의사 결정

빅 데이터

파운데이션 모델 및 AI 지원 앱

- 딥러닝 기술
- 모델 실험
- 모델 튜닝

AI 시대

Predictive(예측형) AI와 Generative(생성형) AI 비교

비즈니스 애플리케이션을 위한 가장 일반적인 AI 유형

Predictive AI

- 과거 데이터를 기반으로 결과를 예측
- 신용 위험 또는 사기 예측
- 수요 예측
- 제품 추천 제공
- 법적 결과 예측
- 고위험 질병 예측
- **처음부터** 모델 학습
- **적당한** 양의 데이터 필요

Generative AI

- 학습된 데이터를 기반으로 새로운 독창적인 콘텐츠 생성
- 대화형 Q&A(ChatGPT)
- 시각적 검색 및 추천
- 가상 착용
- 이미지 만들기(영감을 받은 아트)
- 음악 만들기(기존 스타일 복사)
- **파운데이션 모델** 이용
- **방대한** 양의 데이터 필요

파운데이션 모델로 구동되는 생성형 AI 애플리케이션

파운데이션 모델을 통해 특화된 AI 지원 애플리케이션 개발 가능

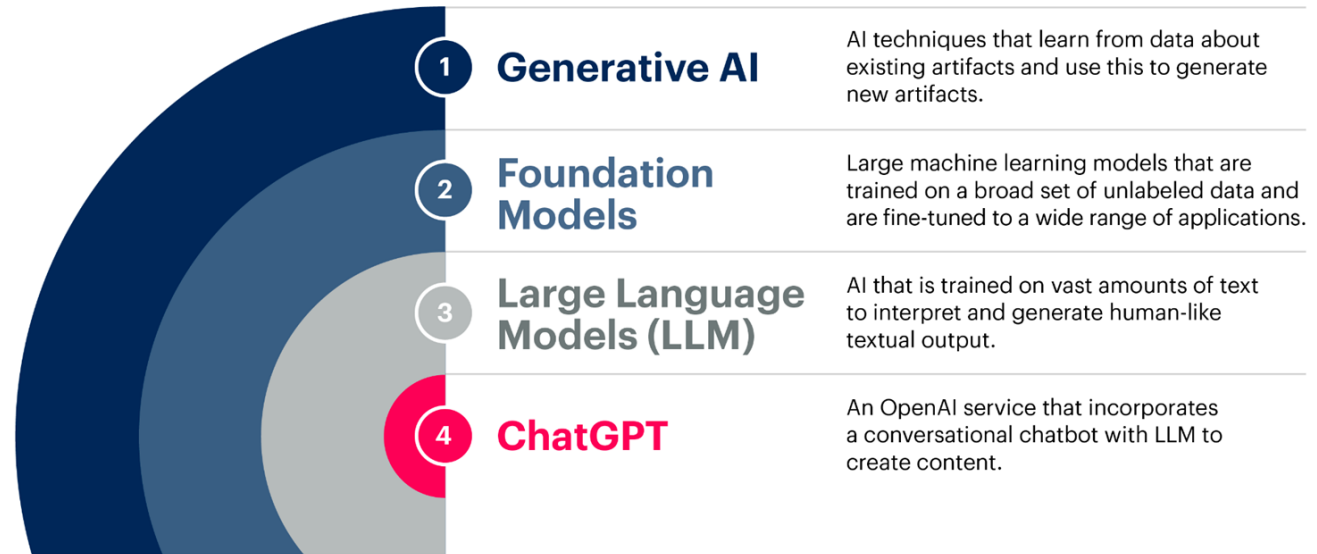
파운데이션 모델의 이점

- 가치 실현 시간
- 정확성
- 접근성
- 다용도성

가장 일반적인 생성형 AI 애플리케이션

- 텍스트 요약
- 텍스트, 코드 및 이미지 생성
- 감정 분석
- 분류
- 대화형 Q&A

What Is Generative AI?



Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. 2421958

Gartner.

생산성 향상을 위한 AI 작업



AI가 할 수 있는 작업

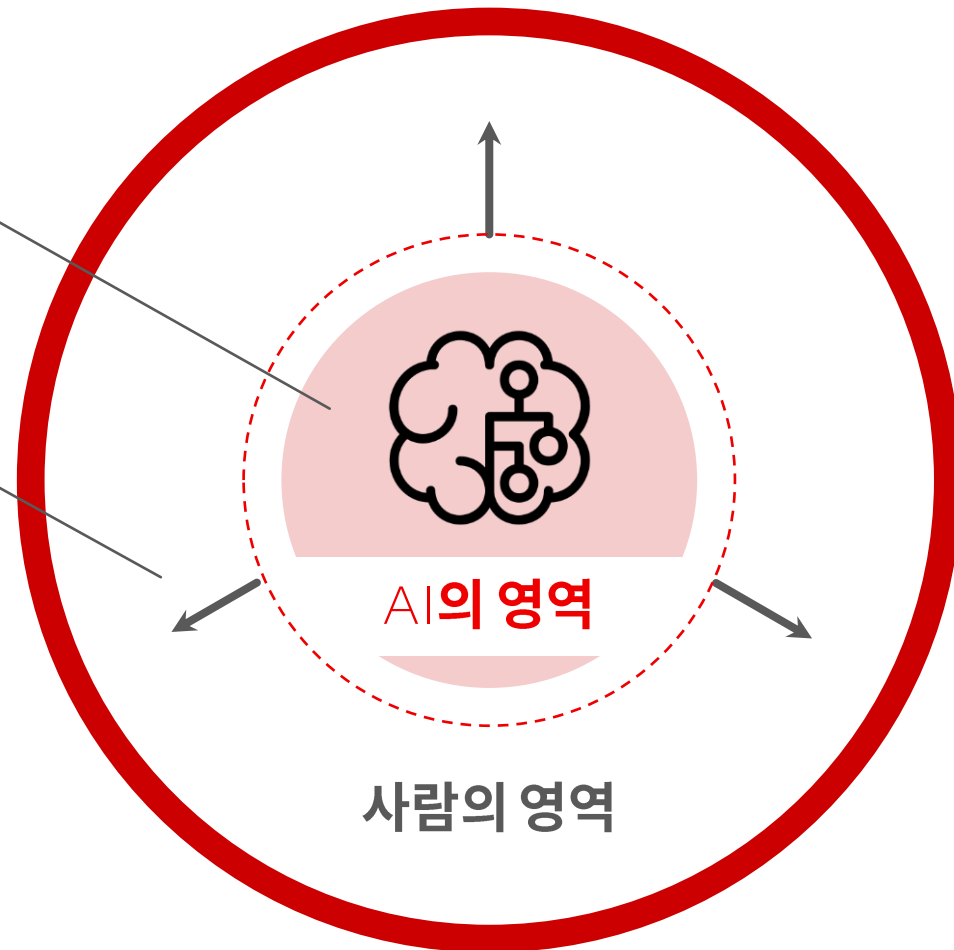


사람만이 할 수 있는 작업

AI가 할 수 있는 영역을 늘리고, 사람만이 할 수 있는 일에 집중



확장하지 못했던 **비즈니스가 확장**하는 시대로..



데이터에 따라 달라지는 AI 모델의 정확도

선진 기업들이 제공하는 LLM은 공개된 데이터를 기반으로 만들어지고 있습니다.

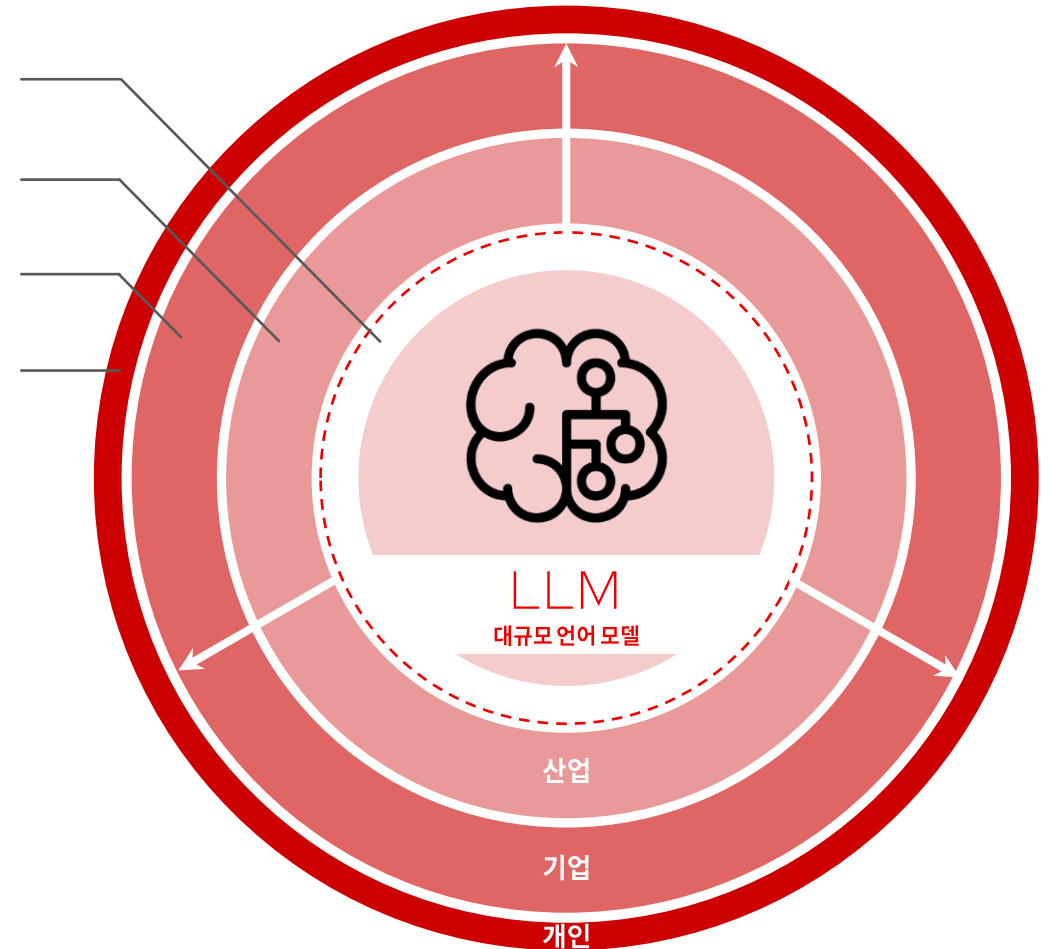
기업이 AI 활용에 필요한 많은 내부 데이터는 LLM에 포함되지 않습니다.

시장 데이터

산업 데이터

기업 데이터

개인 데이터



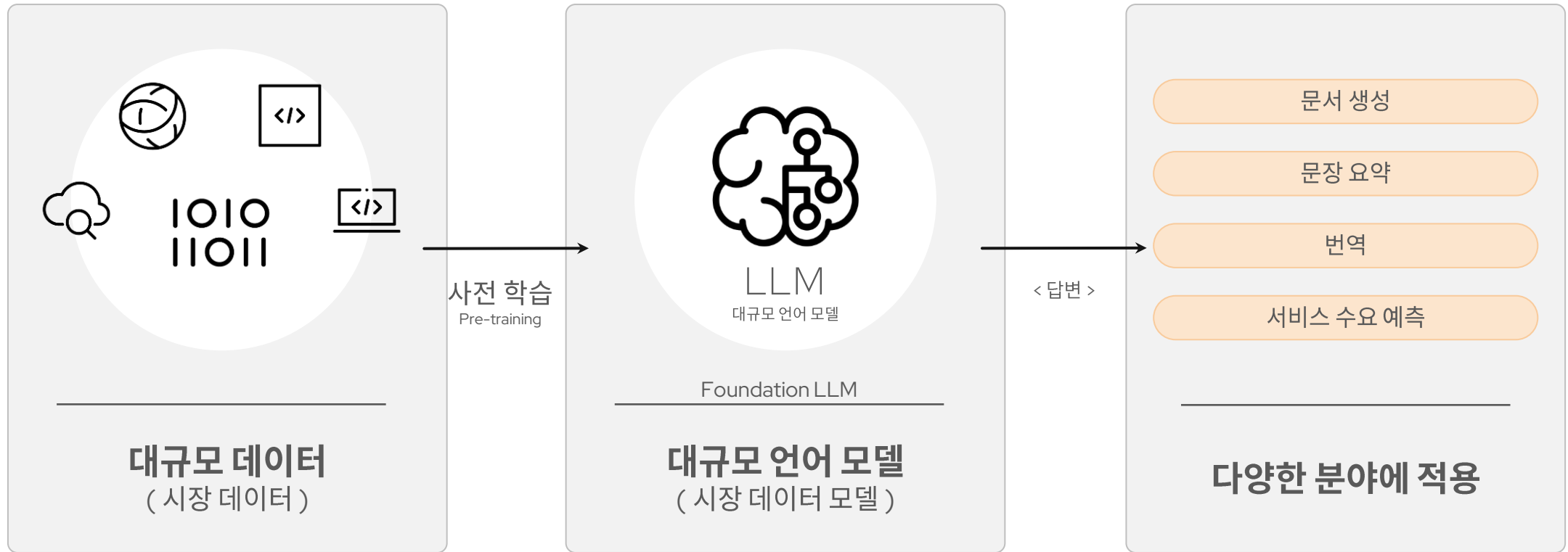
LLM에게 기업의 일을 맡기는 방법,
내부 데이터를 바탕으로

자체 AI를 키우는 것이 가장 큰 가치
로 이어집니다.

금융에 최적화된 AI를 육성하는 방법



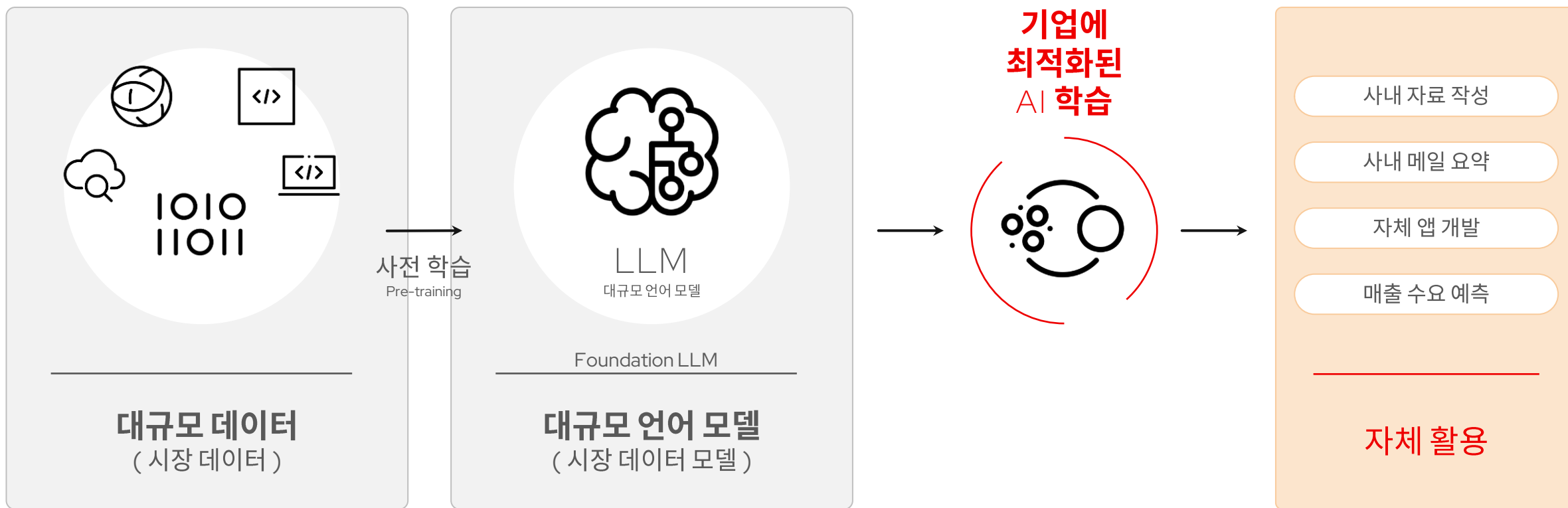
LLM은 어떻게 만들어지는가?



학습 데이터 비공개
사전 학습에 사용된 데이터에 대한 자세한 내용은 비공개

자체 라이선스
라이선스에 따른 이용 범위 내에서만 활용 가능

LLM만으로는 기업의 생산성을 높일 수 없습니다.



기업 데이터를 모델에 통합하면 기업 생산성을 향상시킬 수 있습니다.

기업에 최적화된 AI로 키우기 위한 기법

LLM(생성 AI)에 특정 데이터를 입력시켜 기업이나 개인의 정보에 적합한 특화형 AI를 구축하기 위한 기법

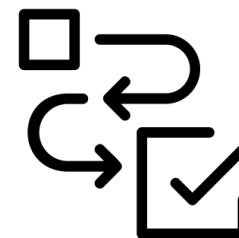
RAG

(Retrieval Augmented Generation)



외부의 지식 기반에서 사실을 검색하여 정확한 정보를 기반으로 LLM이 답변을 생성하도록 하는 기법

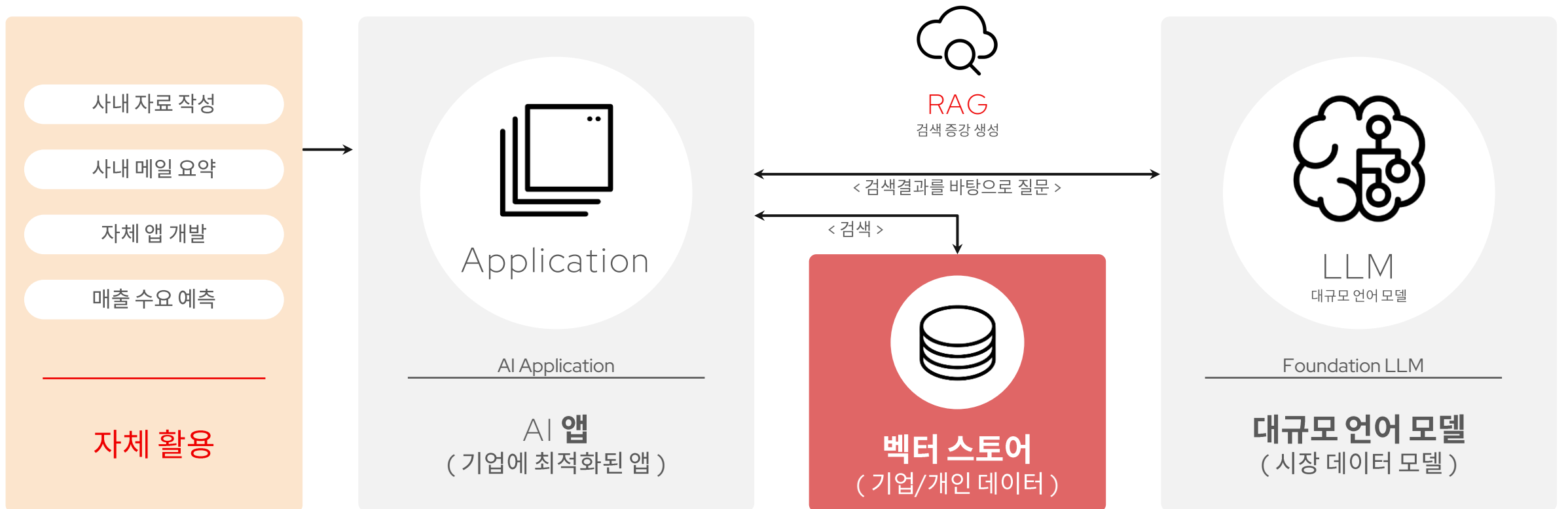
Fine-tuning



LLM의 일부 또는 전체를 다른 데이터 세트를 사용하여 재훈련하여 모델을 미세 조정하는 기법.

RAG

외부 지식 기반에서 사실을 검색하여 정확한 정보를 기반으로 LLM이 답변을 생성하도록 하는 방법



RAG

외부 지식 기반에서 사실을 검색하여 정확한 정보를 기반으로 LLM이 답변을 생성하도록 하는 방법

장점



도입이 용이

- 외부 데이터 업데이트가 용이하다.
- 적은 학습 데이터로도 높은 정확도의 처리가 가능하다.
- 최신 정보와 큰 외부 데이터셋을 활용할 수 있어 유연성이 높다.
- 훈련 데이터에 포함되지 않은 지식도 활용 가능하다.

단점

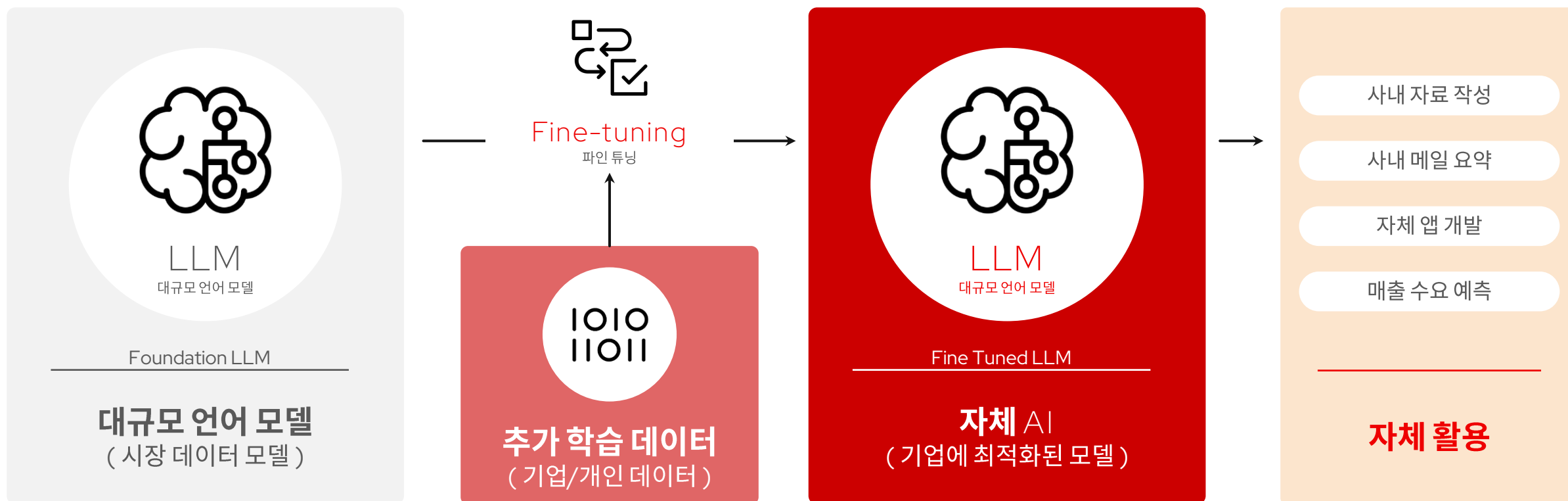


벡터스토어의 조정이 어려움

- 생성되는 콘텐츠의 품질은 검색되는 데이터베이스의 품질에 따라 크게 좌우된다.
- 검색 행위로 인해 답변 작성까지 시간이 소요된다.
- 검색 결과가 부정확하거나 관련성이 낮으면 생성된 답변도 부정확해질 수 있다.

Fine-tuning

LLM의 일부 또는 전체를 다른 데이터 세트로 재훈련하여 모델을 미세 조정하는 방법



Fine-tuning

LLM의 일부 또는 전부를 다른 데이터 세트로 재훈련하여 모델을 미세 조정하는 방법

장점



응답이 빠름

- AI 이용 시 벡터 저장소 등 외부 처리를 끼워 넣을 필요가 없다.
- 많은 학습 데이터를 이용하면 특정 분야에 대한 답변의 정확도가 높아지기 쉽다.

단점



변경 난이도가 높음

- LLM이나 데이터가 업데이트될 때마다 재 학습이 필요하다.
- 학습(LLM 재작성)을 위해 충분한 양의 양질의 데이터 세트가 필요하다.

대규모 언어 모델
(시장 데이터 모델)

추가 학습 데이터
(기업/개인 데이터)

자사 AI
(기업에 최적화된 모델)

자체 활용

Red Hat이 제안하는 기업에 최적화된 AI를 만드는 방법

LLM에 기업 데이터를 쉽게 도입할 수 있는 방법으로 'LAB (Large-scale Alignment for chatBots)' 프로젝트를 런칭

RAG

(Retrieval Augmented Generation)

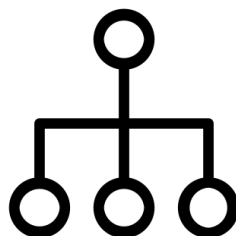


외부의 지식 기반에서 사실을 검색하여 정확한 정보를 기반으로 LLM이 답변을 생성하도록 하는 기법

NEW

LAB

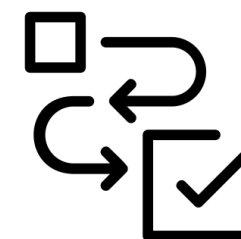
(Large-scale Alignment for chatBots)



LLM의 기존 지식과 새로운 데이터를 계층적으로 매핑하고, 새로운 데이터를 모델에 통합하는 방법론

Fine-tuning

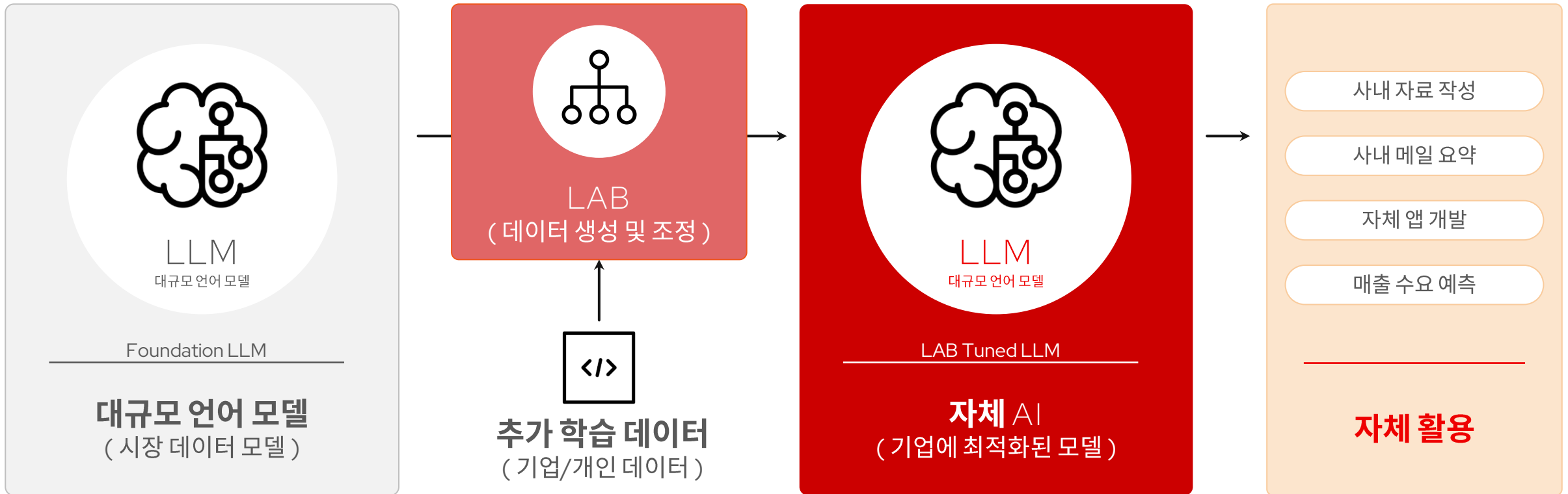
(파인튜닝)



LLM의 일부 또는 전체를 다른 데이터 세트를 사용하여 다시 훈련시켜 모델을 미세 조정하는 기법.

LAB (Large-scale Alignment for chatBots)이란?

LLM의 기존 지식과 새로운 데이터를 계층적으로 매핑하여 모델에 새로운 데이터를 입력하는 방법



LAB의 장점



유연한 데이터 활용

RAG는 검색 단계에서 데이터 활용이 제한적이고, Fine-tuning은 고정된 데이터셋에 의존하지만, LAB는 **실시간으로 데이터를 추가, 제거, 변환**할 수 있습니다.



최신 정보와 실시간 학습

RAG는 최신 정보를 검색할 수 있지만, 모델이 이를 사용하는 방식은 고정적입니다. LAB는 최신 데이터를 동적으로 학습 및 조정하여 **실시간 적응**을 가능하게 합니다.



시간 및 비용 효율성

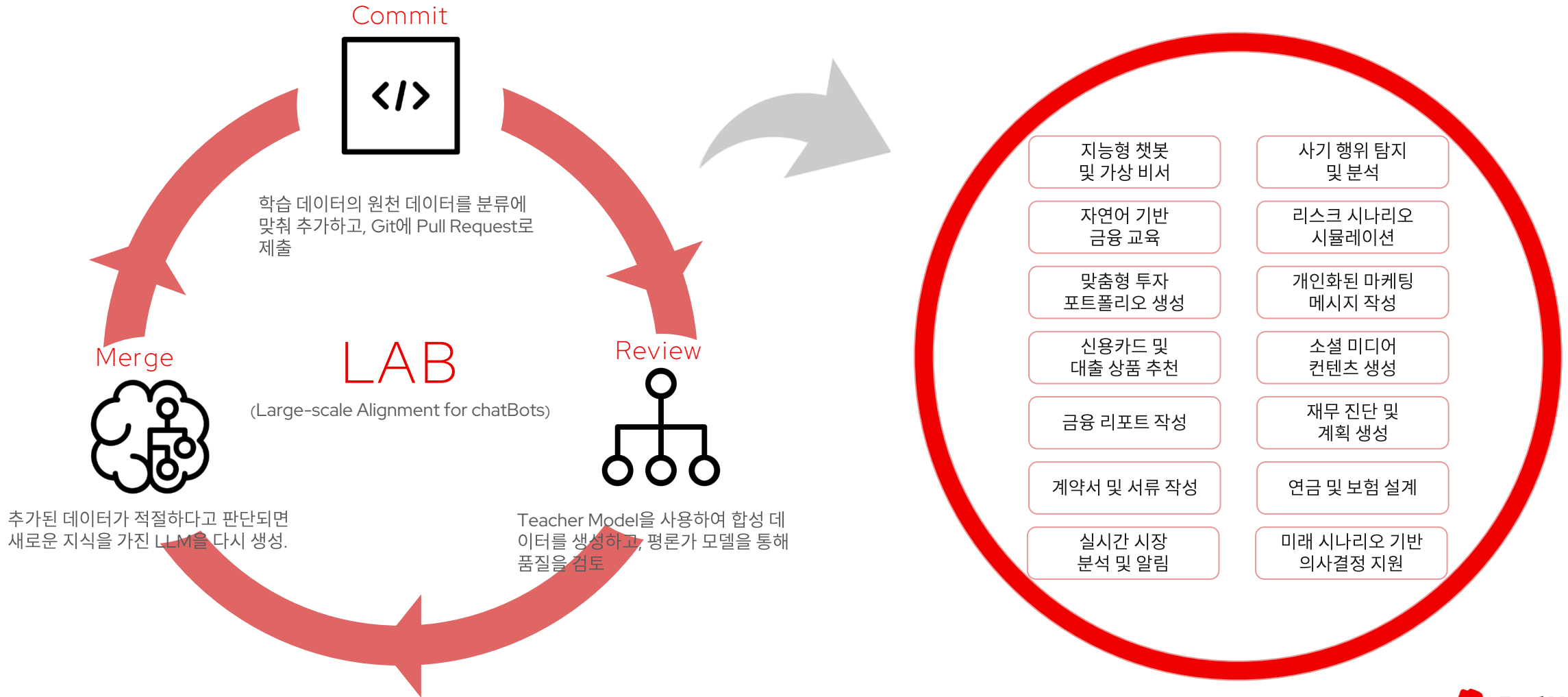
Fine-tuning은 추가 훈련에 높은 비용과 시간이 소요될 수 있지만, LAB는 반복적 실험을 통해 즉각적인 피드백으로 조정을 완료하여 **시간과 비용의 효율성**을 높입니다.



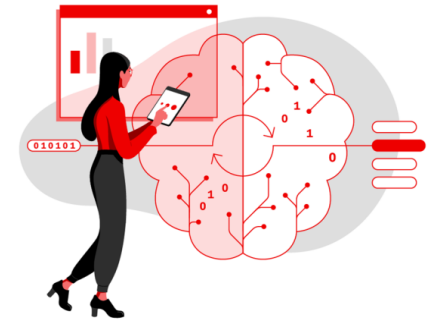
사용자 주도 최적화

LAB는 인터페이스를 통해 **비전문가도 쉽게** 모델을 실험하고 조정할 수 있는 환경을 제공하여 전문성이 낮은 사용자에게도 유리합니다.

금융을 위한 AI 활용



Red Hat의 AI 전략



AI에 대한 Red Hat의 접근 방식

AI + 오픈 소스 커뮤니티



AI 오픈 소스 커뮤니티에 대한 Red Hat의 기여와 커뮤니티 중심의 새로운 LLM 개발 모델 제안

- 1 IBM Research와 협력하여 오픈 소스화한 Granite
- 2 InstructLab의 개방형 LLM 개발 모델

엔터프라이즈 오픈 소스 AI

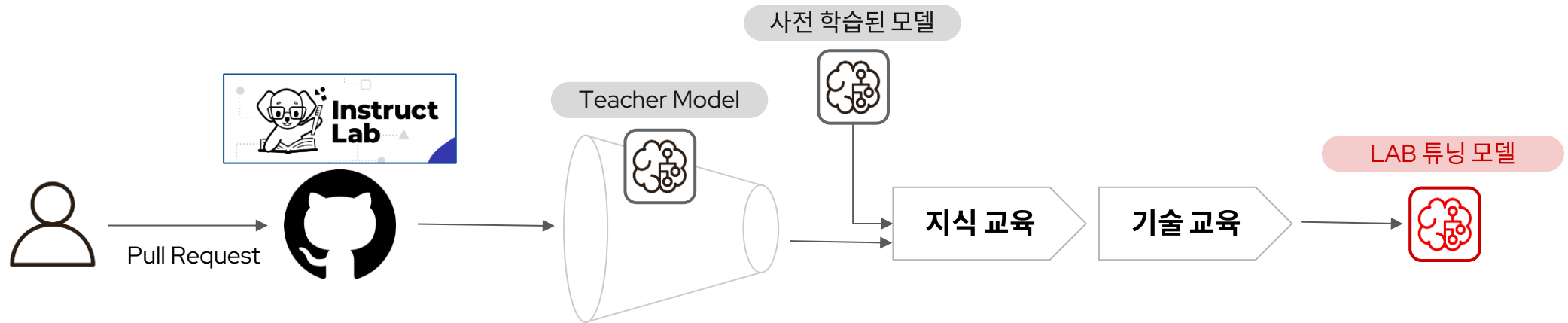


AI 혁신을 비즈니스 가치로 연결하는 Red Hat의 제품 전략

- 1 AI로 핵심 제품 강화
- 2 AI 개발 및 운영을 위한 플랫폼 제공

AI의 개방화를 목표로 하는 InstructLab

AI + 오픈 소스 커뮤니티



1단계 커뮤니티 데이터 수집

기여자는 학습의 기초로 사용될 데이터를 추가하고 GitHub의 InstructLab [리포지토리](#)에 pull request를 제출합니다.

2단계 데이터 증강/필터링

Teacher Model(LLM)은 추가된 데이터를 기반으로 데이터를 확장하고 필터링합니다.

3단계 확장 데이터를 사용한 LLM 추가 학습

2단계에서 생성된 데이터를 사용하여 사전 학습된 LLM에 대해 두 단계의 미세 조정을 수행하여 새로운 지식과 기술을 습득한 LLM을 생성합니다.

Red Hat: 비즈니스에 AI 배포

엔터프라이즈 오픈 소스 AI

1

AI를 활용한 핵심 제품 강화

앤서블 라이트스피드 메커니즘이 RHEL 및 OpenShift에 확장되어 Red Hat 라이트스피드로 발표



Red Hat 라이트스피드



OpenShift **LightSpeed**



Red Hat Enterprise Linux **LightSpeed**



Ansible **LightSpeed**

생성형 AI는 제품 사용에 대한 기술적 장벽을 낮추고 개발자와 운영자의 생산성 향상에 기여합니다.

2

AI 개발 및 운영을 위한 플랫폼 제공

RHEL AI 및 Podman AI Lab과 같은 신제품과 기존 OpenShift AI의 새로운 기능 및 파트너십 확장

Red Hat AI 플랫폼



Red Hat Enterprise Linux AI

로컬 환경에서 AI 개발 기능 제공

- OSS Granite 제공
- InstructLab CLI
- RHEL 이미지 모드

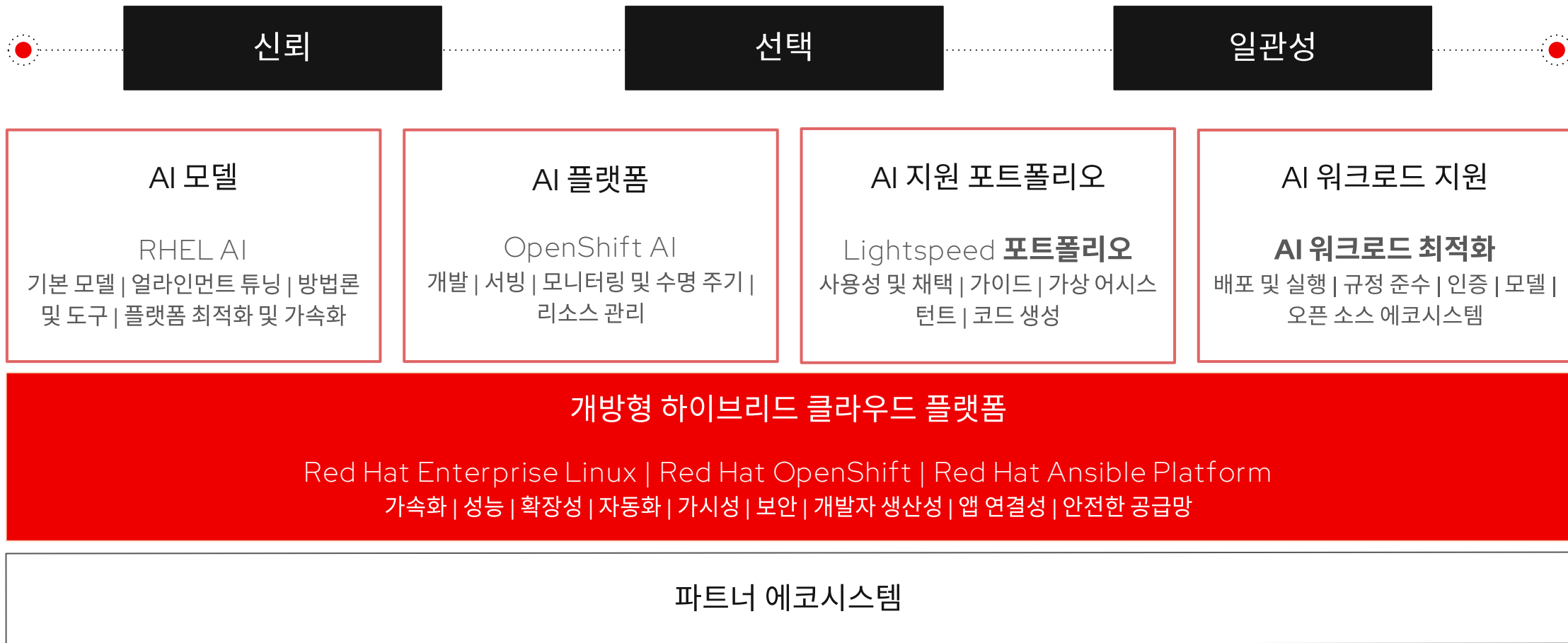


Red Hat OpenShift AI

통합 MLOps 플랫폼 제공

- 엣지에 모델 배포
- 분산 교육
- vLLM 지원
- 파트너십 발표

Red Hat의 AI 포트폴리오 전략



AI 플랫폼 포트폴리오

엔드 투 엔드(End to End)로 기업의 AI 육성 지원



노트북 / 데스크탑
(InstructLab)

싱글 서버
(RHEL AI)

멀티 클러스터
(OpenShift AI)



InstructLab

1단계

소규모 데이터 세트에 대한 제한된 데스크톱 규모 교육 방법을 통해 학습 및 실험



Red Hat
Enterprise Linux AI

2단계

완전한 합성 데이터 생성, 교사 및 비평가 모델을 사용한 프로덕션급 모델 교육. 스크립트 가능한 기본 요소에 중점을 둔 툴링

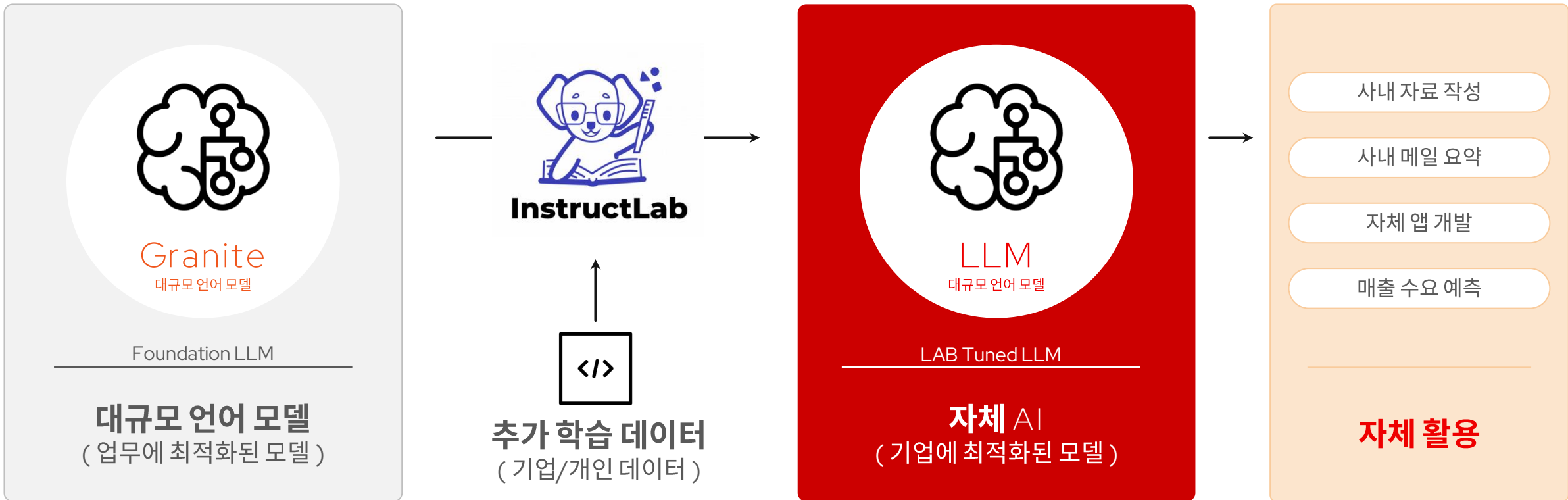


Red Hat
OpenShift AI

3단계

Kubernetes 확장, 자동화 및 MLOps 서비스의 모든 기능을 사용하여 RHEL AI에 서와 같이 프로덕션급 모델 학습

누구나 모델을 최적화할 수 있는 플랫폼



자신의 지식을 활용하여 AI를 육성하는 「InstructLab」



InstructLab

AI의 전문적 지식은 물론 랩에 필요한 스킬과 지식을 추가할 수 있는 툴

질문과 답변 형식으로
지식 공유

```

version: 2
task_description: '주요 금융규제'
created_by: gildong
domain: fsi

seed_examples:
- question: 자산관리 회사가 따라야 할 주요 규제는 무엇인가요?
  answer: |
    자산관리 회사는 '자산운용법' 및 '금융투자업법'을 준수해야 하며, 특히
    고객 자산의 투명한 관리와 보고 의무, 내부통제 시스템 구축, 그리고 고객
    의 투자성향에 맞는 상품 추천 의무를 따라야 합니다.

- question: 금융 상품의 세금은 어떻게 부과되나요?
  answer: |
    금융 상품에 대한 세금은 투자 상품의 종류와 수익에 따라 달라집니다.
    예를 들어, 주식의 매매 차익에 대해서는 '양도소득세'가 부과되며, 채권 이
    자에 대해서는 '이자소득세'가 부과됩니다. 또한, 연금 상품에 대해서는 '연
    금소득세'가 적용될 수 있습니다.
...

document:
repo: https://github.com/gildong/fsi_knowledge.git
commit: e1744af
patterns:
- fsi_results.md

```

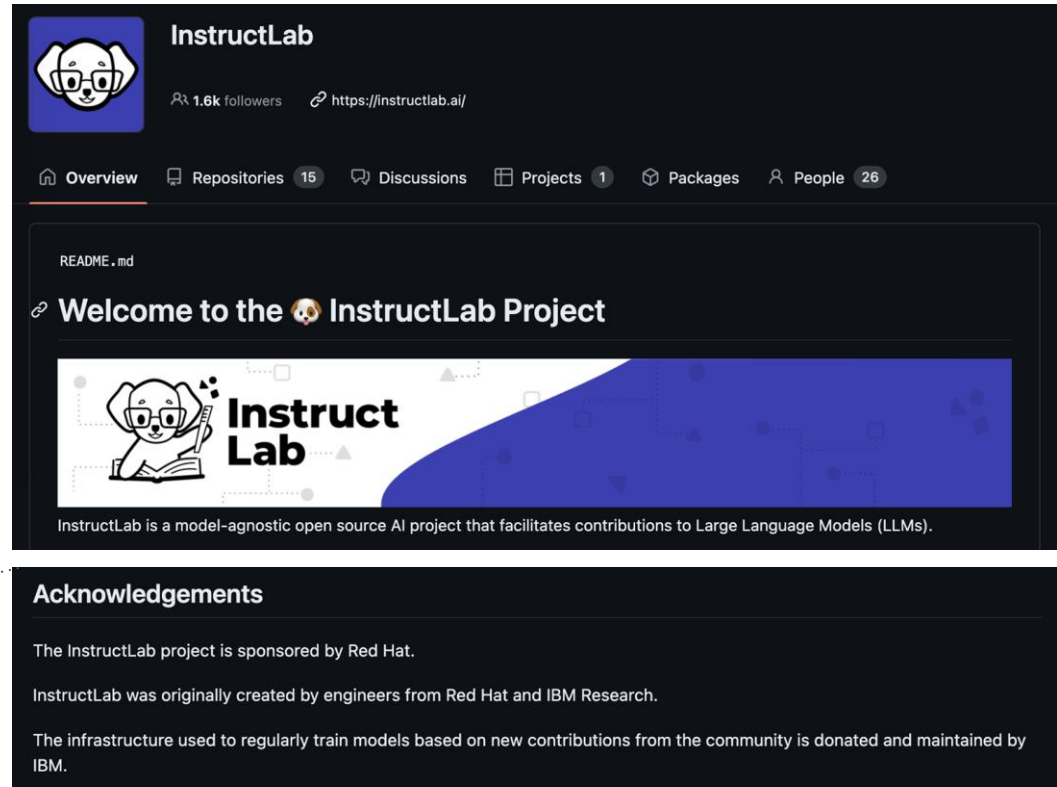
InstructLab 프로젝트는 Red Hat이 후원합니다.

<https://github.com/instructlab>



InstructLab

진정한 오픈 소스 LLM을 구축하기 위한 새로운 커뮤니티 기반 접근 방식



InstructLab 프로젝트는 Red Hat이 후원합니다.
InstructLab은 원래 Red Hat과 IBM Research의 엔지니어들에 의해 만들어졌습니다.
커뮤니티의 새로운 기여를 기반으로 모델을 정기적으로 훈련하는 데 사용되는 인프라는 IBM에서 기부하고 유지 관리합니다.

LLM에 기업 데이터를 입력하는 환경

엔터프라이즈 애플리케이션을 위한 Granite 제품군 LLM을 원활하게 개발, 테스트 및 실행



Granite 제품군 모델

Apache 2.0 라이선스에 따라 제공되는 LLM
투명한 트레이닝 데이터 세트



InstructLab 모델 정렬 도구

누구나 모델 개발을 할 수 있는 도구 도입



Granite
대규모 언어 모델



InstructLab



최적화된 부팅 가능한 모델 런타임 인스턴스

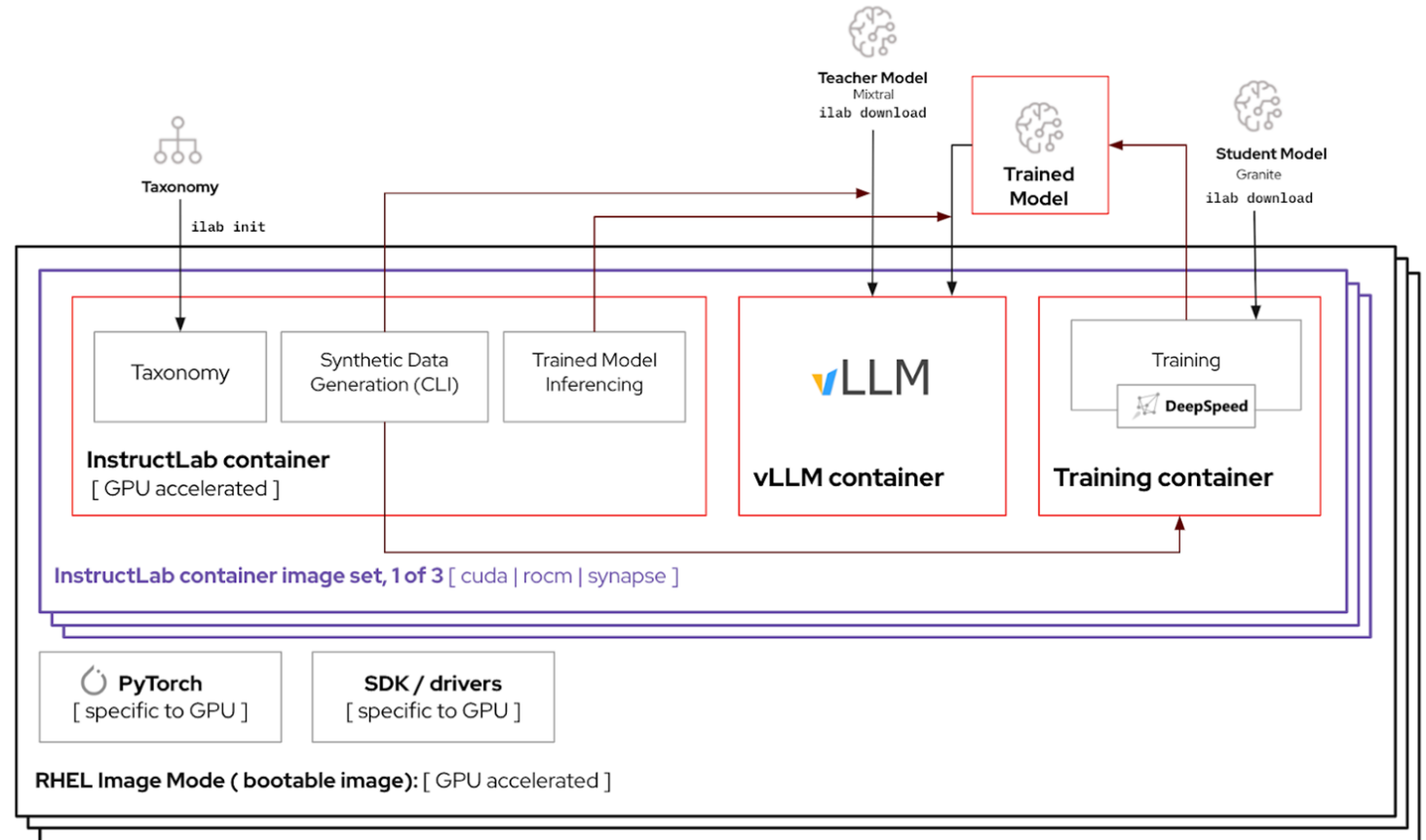
(Optimized bootable model runtime instances)
런타임 라이브러리(+Pytorch)와 하드웨어 최적화(NVIDIA/
Intel/AMD)를 포함한 부팅 가능한 OS 이미지 패키징

RHEL AI 컴포넌트



Red Hat Enterprise Linux AI

- [Press Release](#)
- [RHEL AI blog](#)
- [InstructLab Community Page](#)
- [RHEL AI developer preview](#)
- [LAB: Large-Scale Alignment for ChatBots](#)
- [IBM research blog on Synthetic training for LLMs](#)

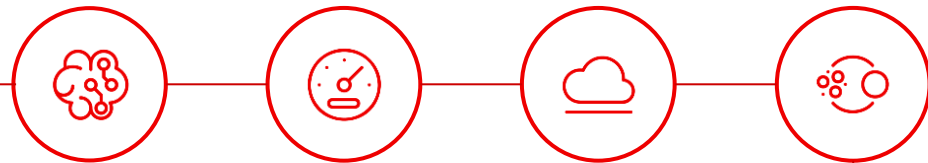


All logos and marks are the property of their respective owners. Details in the appendix.

개방형 하이브리드 클라우드를 위한 AI 엔터프라이즈급 오픈 소스 하이브리드 AI 및 MLOps 플랫폼



실험에서 생산에 이르기까지 AI/ML 모델 및 애플리케이션의 수명 주기를 개발, 서비스, 모니터링 및 관리.



Model tuning & alignment

PyTorch와 InstructLab을 비롯한 핵심 AI/ML 라이브러리와 프레임워크에 접근하여 JupyterLab에서 탐색적 실험을 수행해 보세요. 노트북 이미지나 직접 만든 이미지를 사용할 수 있습니다.



Model serving & monitoring

모든 클라우드, 완전 관리형 (fully managed), 자체 관리형 (self-managed) OpenShift 환경에 모델을 배포하고 확장하며, 모델의 성능을 중앙에서 모니터링합니다.



라이프사이클 관리

기업 전체에 걸쳐 모델의 튜닝, Alignment (Manual 혹은 InstructLab), 검증 및 전달을 위한 모델 파이프라인을 만듭니다.

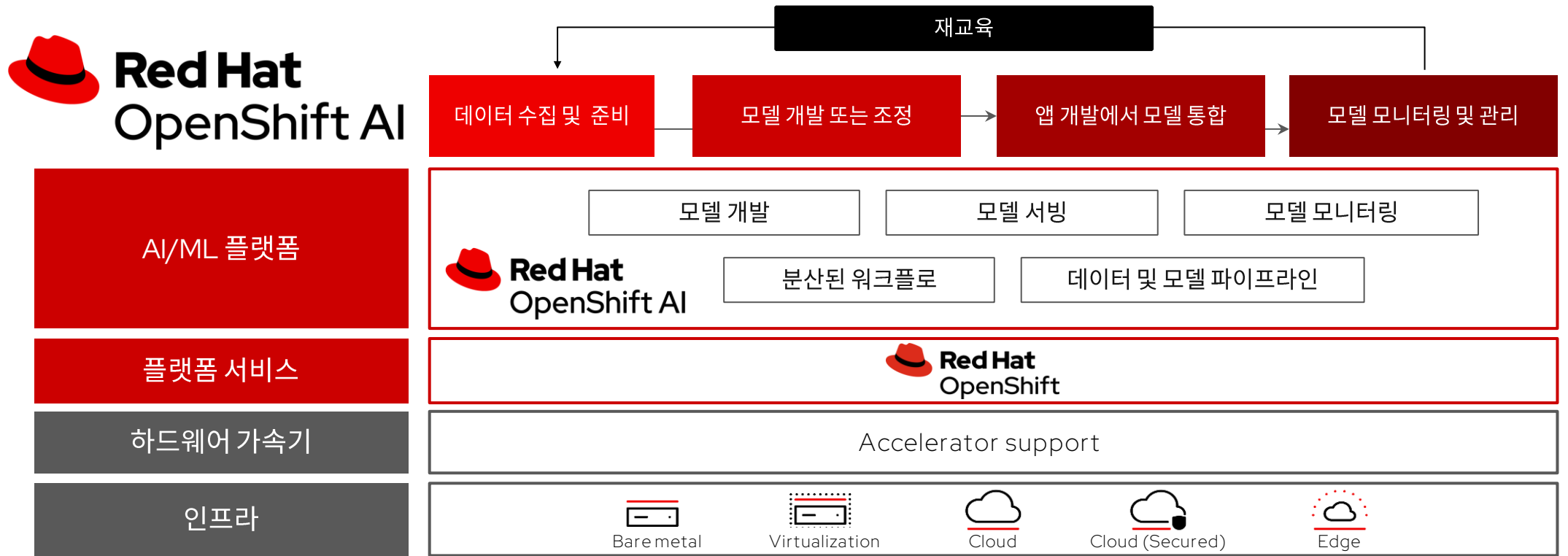


리소스 관리 및 최적화

튜닝 및 추론에 사용할 리소스를 최적화합니다. 사용자 간 리소스, 프로젝트 및 모델을 공유합니다.

Red Hat OpenShift AI

예측 및 차세대 AI 애플리케이션을 위한 Red Hat의 AI/ML 플랫폼



Red Hat OpenShift AI 소프트웨어 스택



MLOps on Open Hybrid Cloud

모델 학습

모델 추론

실험적 학습 환경

- JupyterLab
- PyTorch/TensorFlow



학습 작업 관리

- Kueue



분산된 학습 환경

- Ray
- CodeFlare



파이프라인

- KubeFlow Pipelines



리소스 관리*

- Nvidia GPU Operator
- Intel Habana AI Operator
- AMD GPU Operator



추론 API 서빙

- KServe
- OVMS



LLM 서비스

- TGIS Serving
- vLLM

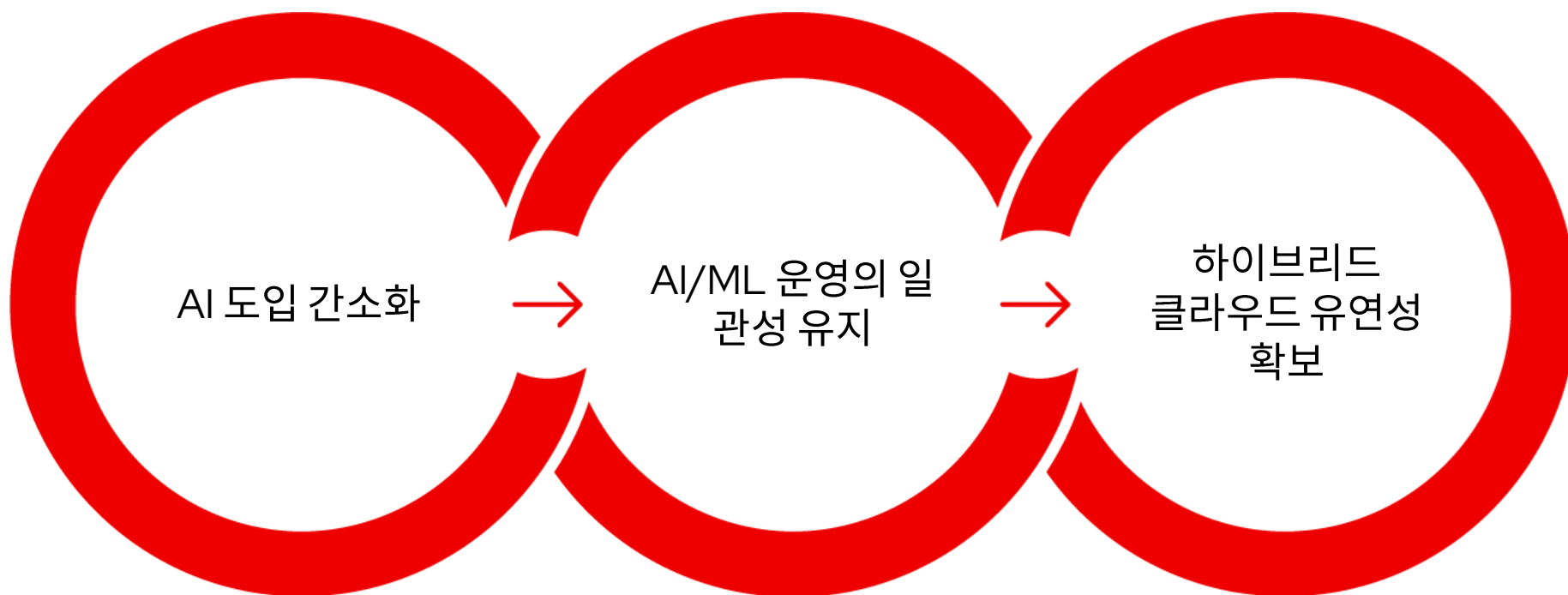


AI 모델 모니터링

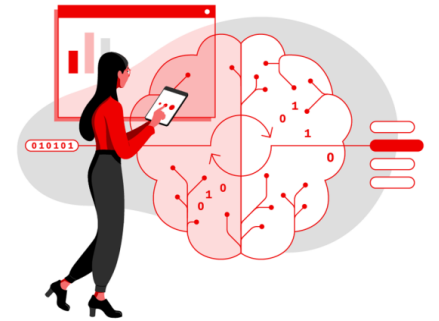
- TrustyAI
- Prometheus



Red Hat OpenShift AI를 통한 혁신



요약



AI 통합 플랫폼을 통한 금융 혁신



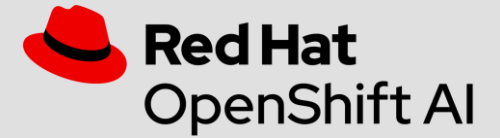
1단계

소규모 데이터 세트에 대한 제한된 데스크톱 규모 교육 방법을 통해 학습 및 실험



2단계

완전한 합성 데이터 생성, 교사 및 비평가 모델을 사용한 프로덕션급 모델 교육. 스크립트 가능한 기본 요소에 중점을 둔 툴



3단계

Kubernetes 확장, 자동화 및 MLOps 서비스의 모든 기능을 사용하여 RHEL AI에서와 같이 프로덕션급 모델 학습

지능형 챗봇 및 가상 비서

자연어 기반 금융 교육

리스크 시나리오 시뮬레이션

신용카드 및 대출 상품 추천

소셜 미디어 콘텐츠 생성

계약서 및 서류 작성

연금 및 보험 설계

사기 행위 탐지 및 분석

맞춤형 투자 포트폴리오 생성

개인화된 마케팅 메시지 작성

금융 리포트 작성

재무 진단 및 계획 생성

실시간 시장 분석 및 알림

미래 시나리오 기반 의사결정 지원

금융 AI 구축을 위한 차별화된 전략



오픈 소스

오픈 소스 AI/ML 툴링의 변경 사항과 수정 사항을 추적하고 고객이 업스트림 혁신에 액세스할 수 있도록 지원



손쉬운 관리

안전하고 검증된 플랫폼에서 간단한 구성으로 적은 노력으로 확장 또는 축소 가능



협업

확장 가능한 공통 플랫폼에서 협업하여 IT, 데이터 사이언티스트 및 애플리케이션 개발 팀의 협업 증대



하이브리드 클라우드

온프레미스 또는 클라우드에서 인텔리전트 앱을 위한 컨테이너화된 형식으로 모델 배포

감사합니다.

김종규 jonggkim@redhat.com
Red Hat Korea



linkedin.com/company/red-hat



youtube.com/user/RedHatVideos



facebook.com/redhatinc



twitter.com/RedHat