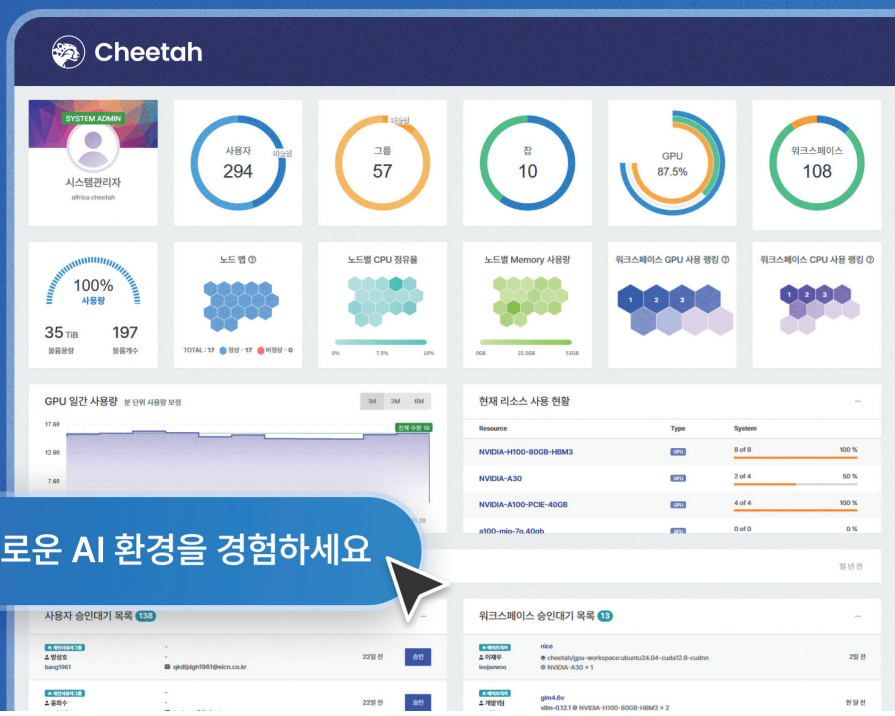
ML & LLMOps 플랫폼

Cheetah

데이터부터 모델 개발 · 배포까지,
AI 전 주기의 효율적인 관리 체계를 제공합니다.



데이터 파이프라인부터 모델 개발, 운영까지

AI를 위한 최적의 프로세스 지금 시작하세요!

AI 전과정을 아우르는 통합 플랫폼

- 데이터부터 모델 학습, 배포, 운영 통합 관리
- ML&LLMOps 통합 환경으로 생산성 극대화
- GitOps 기반 자동화 운영으로 AI 관리 효율화

멀티 엔진 기반 LLM 서빙 최적화

- Triton, vLLM, TensorRT-LLM 등 지원
- 각 엔진 별 성능 메트릭 수집 및 모니터링
- 모델 별 최적화 서빙을 통해 GPU 활용 극대화

AI 서비스 운영 최적화

- AI 서비스 별 리소스 배포 및 운영 관리
- 폐쇄망 환경 지원 및 감사 로그 제공
- SSO 통합 인증 및 권한 관리 기반 접근 통제

엔터프라이즈 GPU 자원 운영

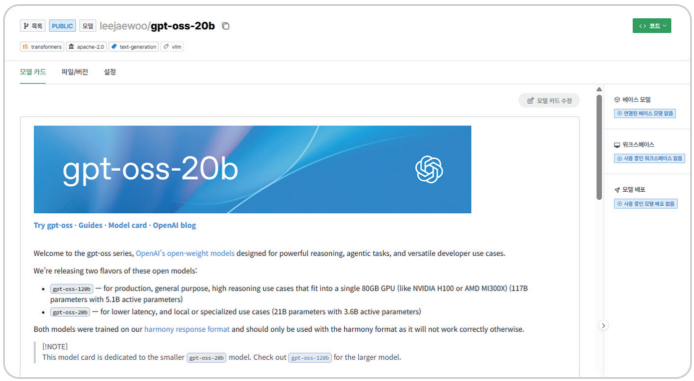
- 팀-사용자 단위 GPU 할당 및 쿼터 관리
- 실시간 GPU·모델·추론 서비스 현황 통합 확인
- AI 워크로드 특성에 맞는 GPU 운영 전략 지원

레퍼런스

DB손해보험	KAI	CELLTRION	경찰청	신한카드
HYUNDAI GLOVIS	DB Inc.	START UP PARK INCHEON	KITECH 한국생산기술연구원	광주과학기술원
공정거래위원회	KEPCO 전력연구원	광주테크노파크 GWANGJU TECHNOPARK	SEOUL BIOHUB 서울바이오허브	농심
KOREN	울산대학교 UNIVERSITY OF ULSAN	전남대학교 CHONNAM NATIONAL UNIVERSITY	국민대학교	한국항공대학교

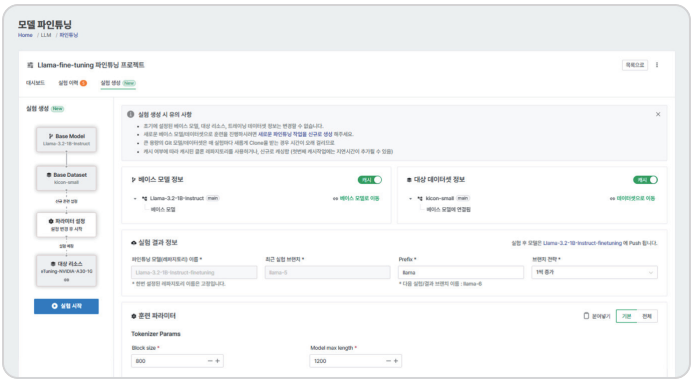
주요 기능

데이터셋/모델 관리



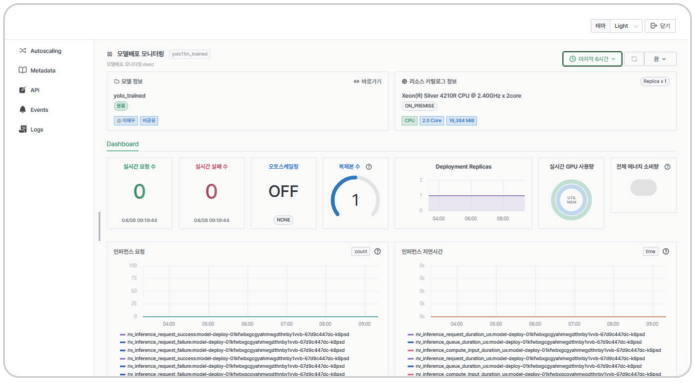
- Airflow 기반의 데이터 파이프라인 자동화
- Git/오브젝트 스토리지 기반의 대규모 데이터/모델 관리

모델 개발/파인튜닝



- 40여 종의 빌트인 이미지를 통해 개발환경 1분 내 구성
- GUI 방식의 쉽고 빠른 모델 파인튜닝 및 모델평가

모델 배포/모니터링

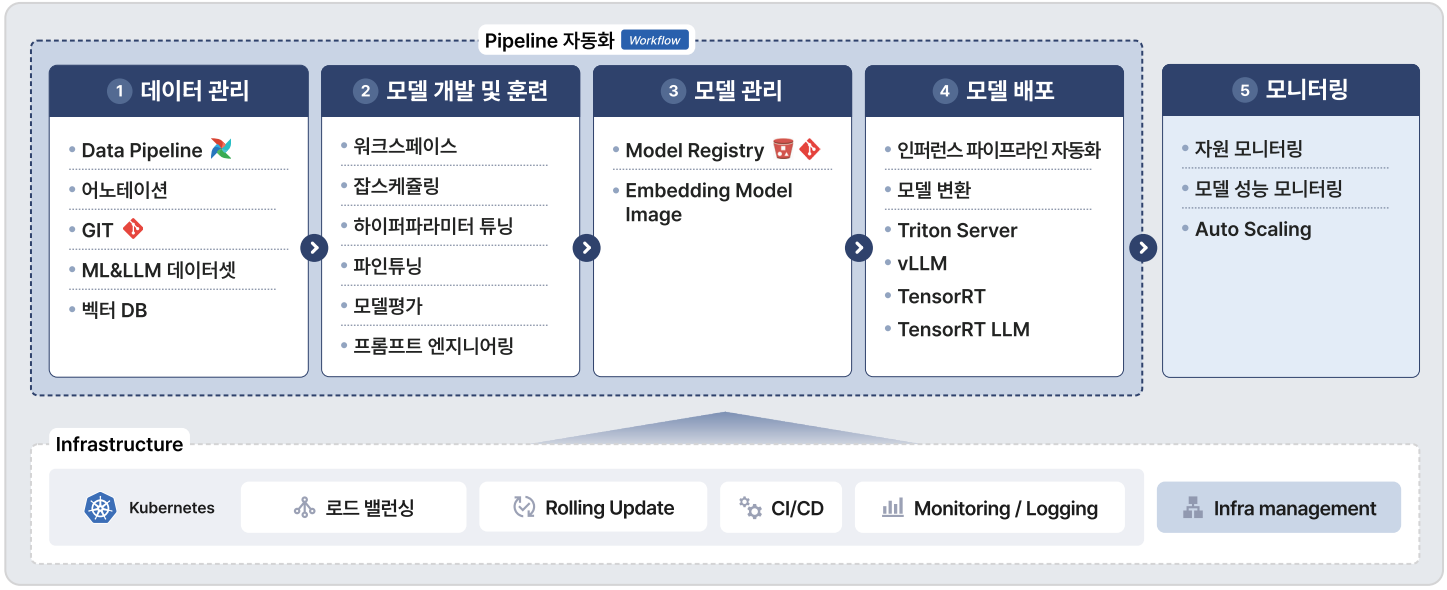


- Triton / vLLM / TensorRT-LLM 등 멀티 서빙 엔진 지원
- 배포 시 API 자동 생성 및 성능 모니터링 대시보드 제공

사용자/자원 관리

- 사용자, 그룹별 체계적인 자원 할당 및 회수 관리
- AI 인프라 자원에 대한 통합 모니터링 및 대시보드

아키텍처



AIFRICA
(주)에이프리카
T 1599-4855 | H www.aifrica.co.kr | E master@aifrica.co.kr

aifrica.co.kr에서 더 자세한 내용을 확인해보세요