

# MLOps 지원을 위한 AI 워크벤치 고도화

김성찬<sup>1,2</sup>, 최명석<sup>1</sup>

<sup>1</sup>한국과학기술정보연구원 인공지능데이터연구팀, <sup>2</sup>과학기술연합대학원대학교 응용AI  
sckim@kisti.re.kr, mschoi@kisti.re.kr

## Enhancement of AI Workbench for MLOps Support

Seongchan Kim<sup>1,2</sup>, Myung-Seok Choi<sup>1</sup>

<sup>1</sup>AI Data Research Team, Korea Institute of Science and Technology Information <sup>2</sup>Applied AI,  
UST-KISTI

### 요약

본 논문에서는 MLOps(Machine Learning Operations) 전주기를 지원하는 문제 해결 플랫폼인 AI(Artificial Intelligence) 워크벤치를 소개한다. AI 워크벤치는 MLOps 라이프사이클의 전 과정을 원활하게 지원하도록 설계되었으며, 단계별로 필요한 기능을 갖추고 있다. 이 기능에는 데이터 수집, 구축 및 큐레이션, 맞춤형 학습 데이터 생성, AI 모델 학습 및 서빙, 모델 스토어 관리, OpenAPI 등록, 그리고 모델 데이터 공유 시스템인 AIDA(AI Data Archive)와의 연계가 포함된다. AI 워크벤치를 활용함으로써 데이터 기반의 해결이 필요한 과학적 및 사회적 문제에 대해 다양한 연구자들이 DATA/AI 기반의 협업을 통해 효율적으로 문제를 해결할 수 있을 것이다.

### 1. 서론

최근 데이터 및 AI 기반의 문제해결이 주목을 받는 가운데 다양한 과학 사회 분야에서 기계학습 모델 개발과 서비스 제공이 활발하게 이루어지고 있다. 이에 발맞추어 모델을 개발하여 학습하고 추론과 서비스를 제공할 수 있는 환경과 인프라 및 모델에 대한 관리 플랫폼의 수요도 늘어나고 있다.

이를위해 한국과학기술정보연구원은 과학기술 기계학습데이터 처리·분석을 지원하는 플랫폼인 AI 워크벤치를 구축하여 고도화하고 있다[1]. AI 워크벤치의 개발 목표는 개발과 운영을 따로 나누지 않고 개발의 생산성과 운영의 안정성을 최적화하기 위한 문화이자 방법론인 DevOps를 ML

시스템에 적용하는 것이다. 데이터를 수집하고 분석하는 단계, ML 모델을 학습하고 배포하는 단계까지의 AI Lifecycle 전 과정을 대상으로 한다. 그림 2에서 보는 바와 같이 AI 워크벤치의 주요 서비스로 데이터분석, 모델학습, 모델서빙, API 활용 제공하고 있으며 AIDA와의 연계, 맞춤형 학습데이터 생성 시스템, 모델학습/서빙 시스템, 연산인프라 관리시스템(BackendAI), 학습데이터 레이블링 시스템이 결합되어 있다. IRods 저장소와 Lustre, NAS와 같은 저장장치 및 CPU/GPU 시스템의 연산장치를 활용한다. AI 워크벤치는 기계학습 데이터를 공개하는 KISTI 기계학습 데이터 공유·활용 서비스 AIDA(AI Data Archive)[2]의 하위 시스템이며 연동되어 서비스를 제공하고 있다.

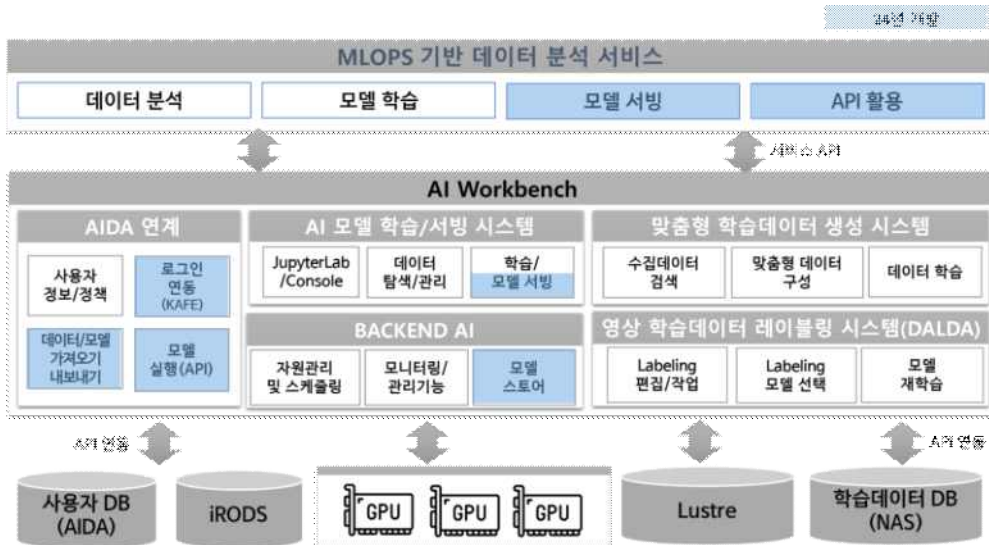


그림 1 MLOps 전주기 지원 AI 워크벤치

AI 워크벤치는 최근 다음과 같은 고도화를 진행하였다. KAFE(Korean Access Federation, 연구교육인증연합)[3]로 로그인을 연동하여 기관 이메일과 패스워드로 로그인을 간편하게 진행할 수 있다. 또한 https 인증서를 적용하여 보안을 강화하였다. 모델스토어를 도입하여 만들어진 AI 모델의 공유 활용성을 높였으며, 모델 API의 출판도 빠르게 지원한다. 마지막으로 AIDA의 모델을 가져오고 내보낼 수 있는 연계성도 강화하였다.

## II. AI 워크벤치 주요 기능

AI 워크벤치는 다음의 주요 서비스와 기능을 제공한다. 이 주요 기능들은 그림 2와 같이 MLOps 파이프라인의 단계에 맞는 서비스를 제공한다.

- 학습데이터 레이블링 저작도구
- AI 모델생성 및 관리
- 모델 스토어
- OpenAPI 등록 및 관리
- AIDA 연계

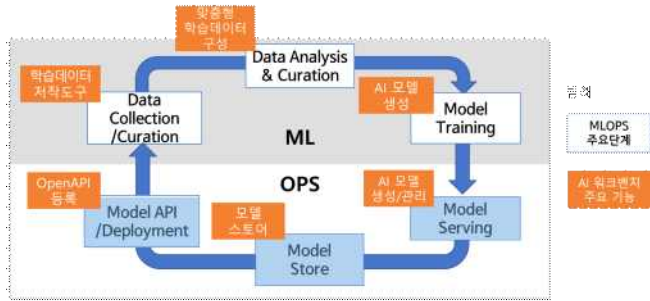


그림 2 MLOps파이프라인 및 AI워크벤치 기능

### 1) 학습데이터 레이블링 저작도구

본 기능은 맞춤형 학습 데이터 수집 및 구축하기 위한 것으로 국내외 공개 데이터셋의 수집 및 관리 시스템을 제공한다. 컨텍스트 추출 모델을 기반으로 한 이미지 데이터의 전처리 기술을 포함하며, SQL 기반 학습 데이터베이스(DB) 구축 및 iRODS를 활용한 학습 데이터셋 제공 환경을 구축하였다. 또한, 객체 인식 자동 레이블링 기능과 민감 정보 비식별화 기능을 제공하여 데이터 처리의 효율성을 높이고 있다.

### 2) AI 모델 학습

GUI 기반의 맞춤형 인공지능 모델 학습을 위한 플랫폼을 구축하였다. 이 플랫폼은 사전 학습 모델의 선택 및 옵션 설정을 통해 효율적인 학습을 가능하게 하며, MLOps 기반의 Backend.AI 플랫폼과 AI 분석 워크벤치를 연계하여 통합적인 환경을 제공한다. 또한, 온프레미스 클러스터 환경에서 GPU 자원의 분할 가상화 및 효율적인 관리를 통해 웹 기반으로 온디맨드 가상 연산 환경을 제공한다. 또한 JupyterLab 및 웹 terminal 서비스 기반의 사용자 인터페이스 제공으로 사용자의 프로그래밍 환경 및 시스템 명령의 편의성을 높여 모델개발의 생산성을 높였다 [4].

### 3) 배포 및 모델 서빙

모델스토어 기반의 AI 모델 배포 및 서빙 시스템을 개발하였다. 이 시스템은 AI 모델의 서빙 및 패키징 기능을 통해 사용자들이 동일한 플랫폼 내에서 모델을 개발하고 서비스할 수 있도록 지원한다. 또한, 프로덕션 환경에서 ML 모델의 지속적이고 안정적인 배포를 위한 유지 관리 및 모니터링 기능을 제공하며, 모델스토어 운영을 통해 공유 활용 환경을 조성한다.

### 4) OpenAPI 관리 및 서비스

다양한 연구자들이 개발된 과학기술 AI 분석 모델을 쉽게 활용할 수 있도록

OpenAPI를 통해 서비스하는 방안을 제안한다. 이를 통해 논문 문장 의미 태깅 API, 논문 주제 분류 API, 상세 개체 인식 API, 영상 데이터 상황 수준 인식 API, 영상 데이터 객체 수준 인식 API, 영상 데이터 중복성 제거 API 등 다양한 기능을 제공하여 연구자들의 접근성을 높이고 있다.

## III. 결론

본 연구에서 제안한 AI 워크벤치는 기계학습 데이터의 수집, 처리, 분석 및 모델 개발을 통합적으로 지원하는 플랫폼으로, 다양한 과학사회 분야에서 AI 활용을 촉진하고자 한다. MLOps 파이프라인에 따라 구성된 주요 기능은 데이터 레이블링, AI 모델 학습, 모델 서빙, OpenAPI 관리 등에서 효율성과 안정성을 극대화하도록 설계되었으며, KISTI의 AIDA와 연계하여 공개된 기계 학습 데이터를 쉽게 접근하고 활용할 수 있는 환경을 제공한다.

이는 사용자가 연구를 보다 효과적으로 수행하도록 돕고, 다양한 연구자들이 AI 모델을 개발하고 공유할 수 있는 플랫폼으로 자리매김할 것이다. 지속적인 업데이트와 기능 고도화를 통해 발전할 것으로 기대되며, 데이터 기반 문제 해결을 위한 핵심 인프라로 자리잡아 연구 및 산업적 응용에 기여하고, 기계 학습 기술의 확산과 발전에 중요한 역할을 할 것으로 판단된다. 이를 통해 연구자들과 기업들이 더욱 혁신적이고 효율적인 AI 솔루션을 개발하는 데 필요한 토대를 마련할 것이다.

## ACKNOWLEDGMENT

이 연구는 2024년도 한국과학기술정보연구원의 기본사업 “디지털 기술 기반 국가·사회 현안 문제 해결(K24ILAM2C6-01)”의 지원으로 수행되었습니다.

## 참고 문헌

- [1] KISTI 인공지능 데이터 공유·활용 서비스 AI 워크벤치 (<https://aida.kisti.re.kr/workbench>)
- [2] KISTI 인공지능 데이터 공유·활용 서비스 (<https://aida.kisti.re.kr/>)
- [3] Korean Access Federation, 연구교육인증연합 (<https://www.kafe.or.kr/>)
- [4] 김성찬, 장래영, 최명석, “효율적 기계학습 모델 구축지원을 위한 JupyterLab 기반의 AI 워크벤치 개발”, 2022, 한국통신학회 추계종합 학술대회발표회 논문집 pp. 810-811