

멀티 프레임워크 기반 AI 워크플로우를 지원하는 AIMI 플랫폼 설계

Design of the AIMI Platform for Multi-Framework-Agnostic AI Workflows

Chan Jung^{1, 2}, Won-Yong Shin^{†1, 2}, Jun-Chae Na², Sung-Il Yang², Young-Jin Yu^{1, 2},
Ju-Hye Lee², Min-Hee Lee^{1, 2}, Yong-eun Cho^{*1, 2}, Seung-Jun Han^{2, 3}

¹School of Mathematics and Computing (CSE), Yonsei University, Seoul 03722, Republic of Korea

²KAILOS LAB Co. Ltd., Seoul 06349, Republic of Korea

³Department of Mechanical Engineering (ME), Kyung Hee University, Yongin 17104, Republic of Korea

Email: lucy.yu@kailoslab.com, wy.shin@yonsei.ac.kr, david@kailoslab.com, simon.yang@kailoslab.com, may.lee@kailoslab.com, chloe@kailoslab.com, kevin@kailoslab.com, chann.jung@kailoslab.com, hansj2k@khu.ac.kr

Abstract

인공지능 (AI) 기반 신소재 개발 플랫폼인 AIMI (AI for Material Innovation)는 실험 데이터를 Lakehouse 구조로 통합하고, MLOps 기반 모델 자동화, Agent 기반 RAG-LLM 시스템을 결합하여 신소재 R&D 전주기를 연결하는 클라우드 네이티브 스택을 제공한다. 이러한 멀티 프레임워크 기반 AI 워크플로우를 통해 연구 데이터 관리 비용을 절감하고, 모델 전달 및 의사결정 속도를 비약적으로 향상시킬 수 있다.

I. 서론

신소재 연구는 다양한 실험 장비에서 생성되는 이질적이고 대용량의 데이터를 처리해야 하며, 복잡한 시뮬레이션 결과와의 통합도 요구된다. 그러나, 여전히 많은 연구 현장에서는 데이터 저장소, 실험 제어 시스템, AI 분석 환경이 분리되어 있어, 연구자들이 데이터 정제와 반복 실험에 많은 시간과 자원을 소모하고 있다. 이는 데이터 형식 불일치, 버전 관리 부재, 실험 재현성 부족 등 문제로 이어져 연구의 효율성과 정확성을 저해한다 [1]. AI 기반 신소재 개발 플랫폼인 AIMI는 이러한 문제를 해결하고자, Data Lakehouse, MLOps 자동화, Agentic RAG-LLM 기술을 통합한 일관된 AI 워크플로우를 제공한다. 이러한 멀티 프레임워크 기반 AI 워크플로우를 통해 실험 데이터의 표준화, 모델 학습 및 배포의 자동화, 지식 기반 의사결정이 하나의 흐름으로 연결되어 R&D 전체 주기를 최적화할 수 있다.

II. 본론

AIMI 플랫폼은 다음 세가지 모듈로 구분 할 수 있다.

- 1) **데이터 Lakehouse 모듈**: 실험 데이터를 Raw, Structured, AIMI Schema의 3단계로 관리한다. 실험 장비로부터 수집된 원시 데이터 (Raw)는 메타데이터 기반 카탈로그를 통해 구조화되고, AIMI 표준 스키마에 따라 정규화된다. 오브젝트 스토리지 기반 설계로 비용 및 확장성을 확보하며, 데이터 계보 추적 및 역할 기반 접근 제어 (RBAC)를 통해 데이터 거버넌스를 강화하고, 외부 소재 데이터베이스와의 연동도 지원한다.
- 2) **MLOps 모듈**: (i) 조성 및 공정을 입력으로 물성을 예측하는 순방향 모델, (ii) 목표 물성을 바탕으로 최적 조성 및 공정을 제안하는 역방향 모델, (iii) 실험 설계를 자동 최적화하는 디자인 모델 등 세 가지 유형의 모델을 통합적으로 관리한다. 학습-평가-배포 전 과정을 자동화하는 파이프라인과 소재 특화된 Feature Engineering 인터페이스를 제공함으로써 반복 작업을 줄이고 모델 전달 시간을 단축한다.
- 3) **AIMI Agent 모듈 (RAG 기반 LLM)**: Data Lakehouse, 사내·외 AI 모델 및 LLM을 연결하여 RAG 기반의 지식 탐색 및 실험 설계를 지원한다. Vector ETL 파이프라인은 데이터를 주기적으로 임베딩해 VectorDB를 갱신하며, MCP (Model Context Protocol) 서버는 MLOps에서 관리되는 모델 및 외부 데이터와의 연동 인터페이스로 기능한다. 사용자는 대화형 Agent 인터페이스를 통해 데이터 질의, 시뮬레이션, 보고서 생성 등을 수행할 수 있어, 데이터 기반 의사결정의 속도와 정확도를 크게 향상시킬 수 있다.

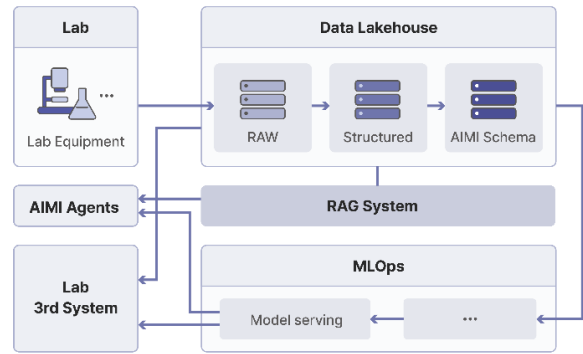


그림 1. 멀티 프레임워크 기반 AIMI 플랫폼 개요

III. 결론

AIMI 플랫폼은 Lakehouse 기반 데이터 관리, MLOps 기반 모델 자동화, Agentic RAG-LLM 기반 지식 탐색 기능을 하나의 통합된 워크플로우로 연결함으로써, 실험 데이터 수집부터 실험 설계·검증에 이르는 전 과정을 자동화하고 재현성 있게 수행할 수 있도록 한다. 플랫폼의 모듈형 설계는 화학, 생명과학 등 다양한 도메인으로 확장 가능하며, 표준화된 API와 MCP 기반 거버넌스를 통해 보안성과 협업 효율성도 함께 확보할 수 있다. 향후에는 실험 장비 표준 인터페이스 고도화, 데이터 품질 추적 자동화, 대규모 멀티모달 모델 연동을 통해 완전 자율형 디지털 연구실로의 전환을 추진할 계획이다.

ACKNOWLEDGMENT

This research was supported by SMEs Technology Innovation Development Program through the Technology Innovation and Promotion Agency (TIPA), funded by Ministry of SMEs and Startups (RS-2024-00511332) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2021-NR059723, No. RS-2023-00220762).

참고 문헌

- [1] B. Blaiszik *et al.*, "CRADLE: A distributed, data-centric infrastructure for heterogeneous materials data," *MRS Bulletin*, 2024.