

# LangChain 기반 LLM framework 를 통한 신소재 합금 설계

이진우<sup>1</sup>, 서기완<sup>2</sup>, 박은수<sup>2</sup>, 이동우<sup>\*1,3</sup>

<sup>1</sup>성균관대학교 반도체융합공학과

<sup>2</sup>서울대학교 재료공학부

<sup>3</sup>성균관대학교 기계공학부

leejw5628@g.skku.edu, kiwan524@snu.ac.kr, espark@snu.ac.kr, \*dongwoolee@g.skku.edu

## New Alloy Design Using a LangChain-Based LLM Framework

Jin-Woo Lee<sup>1</sup>, Ki-Wan Seo<sup>2</sup>, Eun-Soo Park<sup>2</sup>, Dong-Woo Lee<sup>\*1,3</sup>

<sup>1</sup> Department of Semiconductor Convergence Engineering, Sungkyunkwan University, Seobu-ro, Suwon 16419, Republic of Korea

<sup>2</sup> Department of Materials Science and Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 08826 Seoul, Republic of Korea

<sup>3</sup> School of Mechanical Engineering, Sungkyunkwan University, Seobu-ro, Suwon 16419, Republic of Korea.

### 요 약

본 연구에서는 높은 기계적 물성을 가지는 합금 개발을 위해 LangChain 기반의 Large Language Model(LLM) 프레임워크를 구축하였다. LLM 학습을 위한 고품질 데이터 확보를 위해 Human-In-The-Loop (HITL) 방식을 적용하였으며, 모든 LLM 모델에 대해 프롬프트 엔지니어링과 Few-shot 학습 기법을 활용하여 모델 성능을 향상시켰다. 추출된 정보는 Olson Feedback Diagram (OFBD)으로 가공되어, 연구자가 키워드와 메커니즘을 신속하게 파악할 수 있도록 하였다. 최종적으로, OFBD와 LLM 입력 데이터를 활용하여 LLM은 고성능 신소재 합금의 공정 조건을 제안한다.

### I. 서 론

높은 내열성 및 피로 강도 등 우수한 기계적 물성을 보이는 신소재 합금은 항공우주, 의료 분야 등에서 광범위하게 활용되고 있다. 그러나 차세대 산업의 요구를 충족할 수 있는 신 합금 개발은 많은 시간과 인력, 그리고 복잡한 절차를 필요로 한다. 이러한 문제를 해결하고자 밀도 범함수 이론 및 스크리닝 기법이 도입되었으나, 예측 정확도가 낮고 계산 비용이 높다는 한계점이 있다. 최근에는 이런 한계점을 보완한 AI와 LLM을 활용하여 방대한 문헌에서 정보를 추출하고 신 합금 개발을 지원하는 새로운 접근 방식이 주목받고 있다. [1] 본 연구에서는 LLM과 LangChain을 활용하여 고성능 합금 신소재 개발 전략을 구축한다.

### II. 본론

먼저 LLM 훈련에 사용할 고품질 학습 데이터셋 구축을 위해 데이터 추출 과정에 HITL 방식을 적용하였다. [2] 연구자가 재료 과학에 관련된 논문에서 추출된 키워드를 검토함으로써 데이터의 정확성과 신뢰성을 향상시켰다. HITL 방식을 통해 총 2002 편의 논문에서 664 개의 키워드를 추출하였으며, 연구자의 피드백을 통해 grain size, annealing, hardness와 같은 재료 과학에 관련된 중요 정보가 추출된 것을 확인하였다.

HITL을 통해 구축된 데이터셋을 바탕으로 분류 목적의 LLM 훈련 및 모델 선정을 진행하였다. BERT, GPT 등 다양한 모델의 훈련 및 성능 평가를 진행하였으며, Fine-tuned GPT-4o 모델이 F1 score 0.99로 가장 우수한 성능을 보였다. 또한 모호한 단어들이 포함된 데이터 세트에 추가 검증에도 0.90 이상의 성능을 유지하였다. 따라서 해당 모델을 핵심 분류 모델로 채택하였다.

훈련된 LLM을 포함한 LangChain 프레임워크에 대량의 논문을 input으로 사용하여 OFBD를 도출하였다. OFBD는 재료 과학적으로 중요한 공정, 미세구조, 물성에 대한 정보를 담고 있으며, 연관된 메커니즘이 표기되어 실제 신 합금 설계에서 기계 물성 및 공정 및 미세구조의 관계 등의 이해를 돕는다.

또한 프레임워크에서 도출된 정보를 기반으로 우수한 기계적 물성을 가지는 합금 설계 공정을 추천하도록 시스템을 설계한다. 해당 단계는 LLM을 기반으로 진행되며, 공정 변수 및 메커니즘을 표기하여, 공정에 있어서 중요한 요소들을 확인할 수 있도록 구축하였다.

### III. 결론

본 연구에서 구축한 프레임워크는 재료 과학적으로 중요한 정보를 시각화한 OFBD를 추출하였으며, 이를 통해 연구자는 필요한 정보를 신속하게 습득할 수 있다. 또한, 본 프레임워크는 고성능 합금 설계를 위한 공정 조건을 제안할 수 있으며, 향후 특정 합금 시스템(Ti-

6Al-4V 등)을 대상으로 LLM 기반 조건의 실험적 검증은 통해 성능 향상 가능성을 평가할 예정이다. 또한 LLM 기반 자동화시스템 등에 추가하여 다양한 소재 개발에 활용하고자 한다.

## 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 (Grant Number: RS-2024-00511332).

## 참 고 문 헌

[1] Jablonka KM, Ai Q, Al-Feghali A, Badhwar S, Bocarsly JD, Bran AM, et al. 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon. Digital Discovery. 2023;2(5):1233– 50.

[2] Mosqueira-Rey E, et al. Human-in-the-loop machine learning: a state of the art. Artif Intell Rev. 2023 Apr;56(4):3005– 54.