

NIA 한국지능정보원	보 도 자 료		다시 도약하는 대한민국 함께 잘사는 국민의 나라
보도 일시	2024. 08. 12.(월) (온라인 배포시점)부터 보도하여 주시기 바랍니다.		
배포 일시	담당팀	AI데이터활용팀	
2024. 8. 12.(월) 09:00	담당자	신선영 팀장(053-230-4281) 박수영 수석(053-230-4214)	

「Open Ko-LLM 리더보드」 평가 지표 전면 업그레이드로 시즌2 시작

- 추론 능력, 감성 이해 능력, 무해성 능력 등으로 평가 지표 전면 개편 -

- 한국지능정보사회진흥원(원장 황종성, 이하 NIA)과 업스테이지(대표 김성훈(성킴))가 공동 운영하는 ‘Open Ko-LLM 리더보드’가 시즌2로 전면 개편된다.
- ‘Open Ko-LLM 리더보드’는 ’23년 9월 민관협력을 통해 개설된 국내 최대 개방형 한국어 LLM 평가 체계로, 산·학·연 각계 분야에서 7월 말 기준 1,700개가 넘는 LLM 모델이 제출되어 경쟁하는 등 많은 관심과 참여가 지속되고 있다.
 - 하지만 LLM 기술이 빠르게 발전함에 따라 1년 전에 만들어진 리더보드 벤치마크(평가지표)의 개선 필요성이 대두되었고
 - 글로벌 ‘Open LLM 리더보드’를 운영 중인 허깅페이스 또한 지난 6월에 벤치마크를 새롭게 개편한 리더보드 시즌2를 공개하였다.
- 이에 따라 NIA와 업스테이지는 LLM 기술 변화와 허깅페이스의 리더보드 개편을 반영하여 8월 12일부터 새로운 성능 평가지표를 적용한 ‘Open

Ko-LLM 리더보드 시즌2' 운영을 시작한다.

○ 시즌2에서는 기존에 운영되던 평가지표(벤치마크)들을 전면 폐지하고 AI 데이터 전문 기업인 '플리토'와 '셀렉트스타', 'KAIST AI 대학원'이 참여하여 추론능력 지표, 감성 지표, 무해성 지표 등 9개의 새로운 벤치마크를 구축하였고 이를 본격 적용한다.

- 이에 따라 시즌1에 제출되었던 모든 모델을 포함하여 향후 제출되는 모든 LLM은 신규 벤치마크를 통해 평가된다.

□ NIA 황종성 원장은 “리더보드는 민관 협업을 통해 한국어 LLM 생태계 구성에 크게 기여해 오고 있으며 시즌2 개편을 통해 한국어 LLM 기술이 세계적 수준에 한발 더 올라설 수 있는 계기가 되기를 기대한다”고 밝혔다.

○ 또한 공동 운영기관인 업스테이지의 박찬준 리더는 “많은 전문 기업과 학계의 참여로 시즌2를 빠르게 준비할 수 있었으며 시즌2에서도 지속적인 관심과 참여를 부탁한다”고 전했다.

□ Open Ko-LLM 리더보드 시즌2는 AI허브(<https://www.aihub.or.kr>) 사이트의 참여하기-리더보드 메뉴에서 확인 할 수 있다.

붙임1. Open Ko-LLM 리더보드 시즌2 성능 평가지표

붙임2. Open Ko-LLM 리더보드 시즌2 오픈 배너. <끝>.

성능지표	주요내용	제공기관
추론능력	· (Winogrande) 글 속 대명사에 대한 독해 평가를 통해 추론능력 측정	플리토
산술추론능력	· (GSM8K, Grade School Math 8K) 대규모 초등학교 수학 문제를 바탕으로 산술 연산 분야 추론능력 평가	
정보추출능력	· (Ko-GPQA) 대학원 수준의 Google 검증 Q&A 벤치마크	
지시 이행 능력	· (Ko-IFEval) 주어진 지시를 얼마나 잘 이해하고 따르는지를 평가	
감성 평가	· (Ko-EQ-Bench) 대화 맥락에서 다양한 감정과 사회적 상호작용을 이해하는 능력 검증	
정보정확성	· (KorNAT-Knowledge) 국가 구성원들의 공통 지식에 대해 LLM이 얼마나 갖추고 있는지 평가	셀렉트스타, KAIST SI대학원
소셜 얼라인먼트	· (KorNAT-Social-Value) 국가 구성원들의 사회적 가치관에 대해 LLM이 얼마나 일치하는지 평가	
무해성	· (Ko-Harmlessness) 사회적으로 유해할 수 있는 영역에 대해 LLM이 얼마나 무해한지 평가	
도움 적정성	· (Ko-Helpfulness) LLM이 사용자 의도에 따라 쿼리의 유용성을 얼마나 잘 판단할 수 있는지 평가	

