

보도시점 2025. 7. 17.(목) 12:00
(2025. 7. 18.(금) 조간)

배포 2025. 7. 17.(목) 09:00

독자 인공지능 기초모형(AI 파운데이션 모델) 확보를 위한 성능평가 데이터 모음(데이터셋) 구축 추진

- 과기정통부, 인공지능 모형(AI모형) 성능 평가 데이터 모음(데이터셋) 구축 관련 공개모집

과학기술정보통신부(장관 배경훈, 이하 ‘과기정통부’)는 한국지능정보사회진흥원(원장 황종성, 이하 ‘NIA’)과 「독자 인공지능 기초 모형(독자 AI 파운데이션 모델)」 사업에 참여할 정예팀이 개발한 인공지능 모형의 성능을 평가하기 위해 「성능 평가 데이터 모음(데이터셋) 구축 사업」 수행기관을 7월 17일부터 8월 7일까지 공개 모집한다고 밝혔다.

동 사업은 「독자 인공지능 기초 모형(독자 AI 파운데이션 모델)」 사업의 일환으로, 국내 인공지능 모형에 대한 성능평가에 활용할 데이터 모음(데이터셋)을 구축하는 것을 목표로 한다.

그간 세계적인 정보통신기술 대기업(글로벌 빅테크 기업)을 필두로 다양한 생형 인공지능 서비스가 등장하였으나, 대부분의 성능평가에서는 영어권의 평가지표가 활용되며, 국내의 서비스 이용환경 등을 충분히 반영하지 못한다는 한계가 있었다.

이에, 우리나라의 문화, 사회적 가치 등을 반영하고, 국내외 다양한 인공지능 모형의 성능을 객관적으로 진단하기 위해 총 24억 원(총 3개 과제, 과제별 8억원)을 투입하여 고품질 성능 평가 데이터 모음(데이터셋) 구축을 추진한다.

이를 위해, 인공지능 모형 개발 및 평가 방법 분야의 국내외 산학 전문가 의견 수렴 및 검토를 거쳐 성능평가 데이터 모음(데이터셋) 구축

분야를 발굴하였다. 올해에는 우선적으로 ①거대언어모형(LLM)의 수학 풀이 능력 평가 데이터(수학 분야), ②한국형 지식을 평가하기 위한 주제별 질의-정답 및 추론 데이터(지식 분야), ③장문 문맥에서 다양한 과업에 대한 성능평가 데이터(장문이해 분야) 등 3개 분야의 데이터 모음(데이터셋)을 구축하는 한편, 향후에는 다중모형(멀티모달), 대리인(Agent) 등 생성형 인공지능 모형의 다양한 영역을 평가하기 위한 데이터 모음(데이터셋)을 구축해 나갈 계획이다.

한편, 사업 참여를 희망하는 기업·기관은 초거대 인공지능 또는 자연어 처리, 다중형태(멀티모달) 인공지능 등 대규모 데이터 모음(데이터셋) 기반의 인공지능 개발 역량을 갖춘 기업·기관이 1개 이상 수행기관 연합체(컨소시엄)에 포함되어야 한다.

과기정통부 김경만 인공지능기반정책관은 “국민이 체감할 수 있는 고성능 국내 독자 인공지능 기초 모형(독자 AI 파운데이션 모델)을 확보하기 위해서는 그 성능을 평가하는 데이터 모음(데이터셋)도 국내 사회·문화 환경 등이 반영되어야 한다”라며, “이번 사업을 통해 구축한 성능 평가 데이터 모음(데이터셋)은 정예팀 뿐만 아니라 향후 국내 인공지능 개발기관이 활용 할 수 있도록 공개할 계획이며, 국내 인공지능 생태계 경쟁력 강화에 기여할 것으로 기대한다”고 밝혔다.

붙임 : 2025년도 성능 평가 데이터 모음(데이터셋) 구축 사업 모집 상세 안내문

담당 부서	인공지능기반정책관 데이터진흥과	책임자	과 장	이소라 (044-202-6290)
		담당자	사무관	위재승 (044-202-6293)
관련 기관	한국지능정보사회진흥원 AI데이터확산팀	책임자	본부장	신신애 (053-230-4200)
		담당자	팀 장	신선영 (053-230-4281)

내일을 만드는 과학기술
내일을 채우는 디지털·AI

대한민국
지책브리핑



AI 모델 성능을 정량·정성적으로 평가할 수 있는 **고품질 평가검증 데이터셋 구축**으로 표준화된 **모델 평가체계 마련**과 **국내 AI 모델의 글로벌 경쟁력 강화**를 지원하기 위해 아래와 같이 **성능평가 데이터셋 구축기관을 모집**합니다.

□ 성능평가 데이터셋 구축기관 모집 개요

- (목적) 민관 협력 AI 파운데이션 모델 개발 추진으로, 모델 성능을 객관적으로 검증 가능한 데이터셋 구축 및 모델 평가 추진
- (주무부처/전담기관) 과기정통부, 한국지능정보사회진흥원(NIA)
- (지원대상) 기업, 대학, 공공기관, 정부·지자체, 협회 등
 - (지원조건) 대규모 데이터셋 기반의 인공지능 개발 역량을 갖춘 1개 이상의 기업·기관이 컨소시엄 형태로 수행기관 구성 필수
 - ※ 2025년 「독자 AI 파운데이션 모델 프로젝트」에 수행기관으로 최종 선정된 기관은 본 사업 참여 불가, 지원 조건에 관한 세부 사항은 신청 유의사항 참고
- (지원규모) 총 2,400백만원(3종 데이터)
 - (세부 지원규모) 정부지원금 800백만원(1종당 정부지원금)
 - (자유공모) 3개 분야 3개 수행기관 선정 및 지원
 - ※ 정부 지원금은 과제조정위원회 심의·조정에 따라 일부 조정될 수 있음
 - (공모분야) 수학(Math), 지식(Knowledge), 장문 이해(Long context)

연번	분야(과제)	데이터(종)	수행기관	예산(억원)
1	수학(math)	1	1	8
2	지식(knowledge)	1	1	8
3	장문이해(Long context)	1	1	8

□ 주요 추진일정

7~8월	8월	11월	12월	26.1월~
본공고 및 과제제안	과제평가 및 협약체결	과제별 중간 점검	데이터 품질검증 및 최종 평가	데이터 품질 및 산출물 보완

※ 추진일정은 일부 변경될 수 있음

□ 공모 과제 분야별 세부 요구사항

- 언어모델의 추론(Reasoning) 성능을 객관적으로 측정할 수 있도록 분야별 요구사항과 평가 방식 등이 명확한 데이터 구축

<공모 분야별 세부 요구사항>

공모분야	내용	종수 (예산)
수학 (math)	○ (설명) LLM의 수학 풀이 능력 등을 평가하기 위한 데이터셋 ○ 데이터 구성 유의사항 - (평가데이터 구성) 추론형 문제-정답 쌍 데이터 구성 - (언어별 구성) 한국어-영어 1:1 병렬로 구성 ※ (참고) MATH Dataset(https://arxiv.org/abs/2103.03874), GSM8K, OlymMATH 이상 수준을 요함	1종 (8억원)
지식 (knowledge)	○ (설명) 기초 지식(역사, 문화 등) 평가 데이터셋 ○ 데이터 구성 유의사항 - (평가데이터 구성) 지식 평가 가능한 추론형 문제-정답 쌍 구성 - (주제 구성) 한국형 지식을 평가하기 위한 주제(category)를 제안하고 주제별 지식 평가가 가능한 질의-정답 및 추론형 데이터(CoT)*로 데이터 구성 * CoT 데이터는 단순 나열 형태가 아닌 순서, 인과 등을 확인 가능하도록 구성 - (글로벌 공통) 국가, 문화에 구애받지 않는 공통 지식에 대한 질의응답 구성 ※ (참고) SimpleQA, Humanity's Last Exam(Text only)	1종 (8억원)
장문 이해 (Long context)	○ (설명) 장문, 긴 문맥에서의 추론을 요하는 업무수행 능력 평가 ○ 데이터 구성 유의사항 - (평가데이터 구성) 장문이해, 문서 수준 길이에서의 업무수행 능력을 평가하기 위한 데이터 구성 * 객관적인 단위를 기반으로 한 구축량 제시 필수(ex. 32K 이상 토큰 00개, 128K 이상 토큰 00개 등) - (태스크별 구성) 긴 문맥에서의 다양한 업무 수행을 평가할 수 있는 태스크별 성능평가 데이터 구성(문맥 결속력, 논리 흐름, 참/거짓 판단, 인과 등) - (난이도 구성) 높은 난이도의 실제 전문시험 수준 포함한 문항 구성 ※ (참고) LongLeader, Michelangelo	1종 (8억원)

※ AI허브 내 데이터와 중복성 검토, 저작권 이슈 해결방안, 데이터 구성 및 구축량 제안 필수

□ 신청방법

- (접수기간) 2025년 7월 17일 ~ 2025년 8월 7일
- (문의처) 한국지능정보사회진흥원 AI데이터확산팀(053-230-4216/4225)
- (접수방법) 디지털제안서 통합관리시스템(propose.nia.or.kr)을 통한 온라인 접수
- (평가일정) 기관 개별 통보(제안서 접수시스템 내 일정 개별 확인)