

보도시점 2025. 6. 23.(월) 12:00
(2025. 6. 24.(화) 조간)

배포 2025. 6. 23.(월) 09:00

국산 인공지능 기초 모형(AI 파운데이션 모델) 학습데이터, 민·관 협력으로 모은다

- 과기정통부, 데이터 제공 협력기관 공개 모집 시작

과학기술정보통신부(장관 유상임, 이하 ‘과기정통부’)는 세계 최고 수준의 거대 언어모형(LLM) 개발을 목표로 하는 “독자 인공지능 기초 모형 사업(AI 파운데이션 모델 프로젝트)”의 성공적인 추진을 위해, 6월 23일부터 7월 21일까지 공공 및 민간의 데이터 보유 기관을 대상으로 데이터 공급기관을 공개 모집한다고 밝혔다.

이번 프로젝트는 2025년부터 본격 추진되는 정부의 대표 인공지능 전략 과제로, 향후 공모를 통해 선발될 정예 개발팀에 대해 그래픽 처리 장치(GPU), 데이터, 인재 등 필수 자원을 집중 지원함으로써, 독자적인 초거대 인공지능 모형(AI 모델) 개발을 목표로 한다.

이번에 모집하는 데이터 공급기관은 인공지능 개발 정예팀이 인공지능 학습에 활용할 수 있는 유·무료 고품질 학습 데이터를 보유한 기관으로, 각 기관이 보유한 데이터의 최종 활용 여부는 개발팀 선발 이후 실제 수요를 바탕으로 협약을 통해 확정된다. 이 협약은 과기정통부(한국지능정보사회진흥원<NIA>, 한국 데이터산업진흥원<K-Data>)와 데이터 공급기관, 정예 개발팀 간 체결된다.

데이터 제공이 확정된 기관에는 ▲ 데이터 가공·전처리 비용 지원, ▲ 유료 저작물 제공 시 제공 요건에 맞는 비용 보전 등 재정적 지원이 제공될 예정이다. 또한, 해당 사업을 통해 개발되는 ▲ 인공지능 모형(AI모델) 기반의 서비스를 도입할 기회를 우선 제공하거나, ▲ 데이터 이용권(데이터바우처) 지원 사업 등에서 가점 부여를 검토 중이다. 아울러, 공급기관의 희망에 따라 인공지능 거점(AI허브/학습데이터 통합제공 체계<플랫폼>) 등을 통한 데이터 유통·거래도 적극 지원할 계획이다.

< 데이터 공급기관 유인책(인센티브)(예시) >

지원 항목	지원 내용
데이터 가공 및 전처리	원시 데이터에 대한 가공 및 전처리를 통한 품질 향상 지원
데이터 확보 비용	유료 저작물의 경우 데이터 사용 요건(정예팀 한정 사용 등)을 고려하여 적정 비용을 협의하여 지급
정부사업 가점 부여 검토	과기정통부 데이터 관련 사업 참여 시 유인책(인센티브) 가점 검토
데이터 유통·거래 지원	인공지능 거점(AI 허브)·하나 윈도우(원원도우) 등을 통한 데이터 유통 및 거래 지원

과기정통부 김경만 인공지능기반정책관은 “고품질 데이터는 인공지능 성능을 좌우하는 핵심 자산으로, 공공과 민간이 함께 만드는 데이터 협력체계가 세계적 수준의 한국형 거대언어모형(LLM) 개발을 위한 초석이 될 것”이라며, “이번 협력을 계기로 동 프로젝트사업의 성공을 위해 민·관의 데이터 연대가 더욱 확대되기를 기대한다”고 밝혔다.

이번 “독자 인공지능 기초 모형 사업(독자 AI 파운데이션 모델 프로젝트)” 참가 개발팀에 데이터 공급을 희망하는 기관은 붙임 안내문에 따라 신청서를 작성하여 7월 21일까지 **온라인(ai_data@kdata.ork.kr)**으로 제출하면 된다. (문의처 : T.02-3708-5445)

붙임 : 독자 인공지능 기초 모형 사업(독자 AI 파운데이션 모델 프로젝트) 데이터 공급 기관 모집 상세 안내문

담당 부서	과학기술정보통신부 데이터진흥과	책임자	과장	김수정 (044-202-6290)
		담당자	사무관	심지섭 (044-202-6293)
관련 기관	한국지능정보사회진흥원 AI데이터정책팀	책임자	본부장	신신애 (053-230-4200)
		담당자	팀장	심호찬 (053-230-4201)
	한국데이터산업진흥원 데이터활용본부	책임자	본부장	조수완 (02-3708-5400)
		담당자	팀장	홍상현 (02-3708-5440)

국민 누구나 AI를 활용하고 혜택을 누릴 수 있는 생태계 실현을 위해, 다양한 고품질 데이터를 확보하여 세계적 수준의 독자 AI 파운데이션 모델 개발을 지원하기 위해 아래와 같이 데이터 공급기관을 모집합니다.

< 독자 AI 파운데이션 모델 프로젝트 개요 >

- (개요) 모든 국민이 우리가 개발한 세계적 수준의 AI를 활용할 수 있도록 역량있는 AI 기업들을 선발하여 GPU, 데이터, 인재 등 모델 개발에 필요한 핵심 인프라를 집중 지원
- (주무부처/전담기관) 과학기술정보통신부/정보통신산업진흥원(총괄) 및 분야별 전담기관 등
- (지원대상) 국내 AI 기업(주관) + 국·내외 산·학·연 컨소시엄(참여, 필요시)
- (지원내용) GPU(임대), 데이터(구매·구축), 인재(유치) 등

□ 데이터 공급기관 모집 개요

- (목적) ‘독자 AI 파운데이션 모델 프로젝트’ 정예 개발팀의 AI 파운데이션 모델 개발에 필요한 데이터 제공 기관 모집
- (주무부처/전담기관) 과기정통부, 한국지능정보사회진흥원(NIA)
 - ※ 협력기관 : (데이터 공동구매) 한국데이터산업진흥원(K-Data)
- (신청대상) AI 학습용데이터(도서, 문제집 등) 제공 가능 기관·기업*
 - * AI학습용데이터 제공을 위한 적법한 권원(저작권, 소유권 등)을 보유한 기관 또는 기업

□ 추진 절차

6~7월	7~8월	8월~	9월~
공급기관 Pool 모집	독자 AI모델 프로젝트 정예팀 공모	데이터 공급기관 확정 (정예팀 수요 기반)	데이터 활용 및 AI 모델 개발
NIA (K-Data)	NIPA(총괄), NIA-IITP	NIA (K-Data)	정예팀

- 데이터 공급 참여기관(pool) 모집(6~7월)
 - AI모델 개발에 활용될 데이터인 점 등을 고려하여 각 기관에서 제공할 수 있는 데이터 및 활용 요건을 제출
 - ※ (예) 동 프로젝트에 한해 非상업용으로만 활용, 추후 상용모델화 하는 경우 별도 협의
- 독자 AI 파운데이션 모델 개발 정예팀 선발 후 데이터 공급기관 확정(8월)
 - 정예팀, 데이터 공급기관, 사업관리기관(NIA, K-Data) 간 데이터 이용 협약 체결
 - ※ 협약 조건에 데이터 이용기간, 이용범위, 사용기준, 책임관계 등 명시 가능

- 협약 내용에 따라 해당 데이터를 전용 공간(AI허브 전용게시판*) 또는 당사자 간 직접 데이터를 제공

* AI 허브 내 전용 공간에서 데이터를 안전하게 저장할 수 있는 별도 스토리지 제공

□ 데이터 협조 기관에 대한 인센티브

- 가용 예산 범위 내에서 공급기관별 이용 협의 및 계약 조건에 따라 데이터 제공 방식(무상 또는 유상)별 인센티브 지급

< 데이터 제공에 따른 인센티브 지급(안) >

▶ (무상 제공 시) 데이터 수집·가공 등에 소요된 기초 지원* 및 실비** 지급

* (예) 원시데이터에 대한 OCR, 데이터 완전 중복 제거 등의 기본적인 지원

** (예) 데이터 제공 관련 법률 검토, 데이터 전처리 및 가공, 전송 등에 필요한 금액

▶ (유상 제공 시) 유료 판매 저작물의 제공요건에 맞는 비용 보전 등 재정적 지원

- 지속적 협력을 위해 ① AI 허브를 통한 데이터 공유 및 활용 활성화, ② 과기정통부 데이터 관련 사업(예_데이터 바우처 등) 가점, ③ AI 모델 기반 서비스 도입 지원 등 추가 인센티브 제공 검토

□ 유의사항

- 타 정부지원 사업* 등을 통해 구축·재활용한 데이터는 제외

* 초거대AI 확산 생태계 조성 사업(舊 AI학습용 데이터 구축 사업, 한국지능정보사회진흥원) 등

- 계약 체결 후, 공급기관이 제공한 데이터 관련 법률적 분쟁 또는 이슈가 발생할 경우, 계약서 등에 명시된 절차와 기준 등에 따라 처리
- 이용협의, 계약체결 등 과정에서 필요 서류는 각 단계별로 별도 안내 예정

□ 신청방법

- (접수기간) 2025년 6월 23일 ~ 2025년 7월 21일
- (문의처) 한국데이터산업진흥원 활용기반팀(02-3708-5445/5442)
- (접수방법) 이메일을 통한 온라인 접수(ai_data@kdata.or.kr)
- (제출서류) 데이터 공급기관 등록 신청서(서식1), 데이터 제공 계획서(서식2), 사업자등록증(또는 고유번호증)

서식1 데이터 공급기관 등록 신청서

데이터 공급기관 등록 신청서				
신청기관	기관명			
	기관유형	☐ 중앙부처 ☐ 지자체 ☐ 공공기관 ☐ 협회 ☐ 기업 ☐ 기타()		
	대표자명			
	주 소	(-)		
	사업자등록번호 (고유번호)			
담당자	소 속		직위	
	성 명		Tel	
	E-mail		Fax	
보유 데이터	데이터명	// 제공할 데이터의 명칭		
	데이터 종류	//전문서적, 교과서, 학습교재 등		
	제공 조건	☐ 무상(기여) ☐ 유상		
	저작권 범위	☐ 전체 ☐ 일부		
위 기관은 선발된 복수의 AI 기업(컨소시엄)의 'AI 파운데이션 모델 개발'에 필요한 데이터 제공 등을 위한 데이터 공급기관으로 등록 신청함에 있어 신청 내용이 사실임을 확인하고, 당사는 데이터를 제공함에 있어 상호 협의를 통해 적극적으로 협력할 것을 약속합니다.				
별첨 1. 데이터 제공 계획서 1부 2. 사업자등록증(고유번호증) 사본 1부				
2025년 월 일				
담당자 : (인)				
신청기관의 장 : (인)				

서식2 데이터 제공 계획서

1. 기관 개요

기관명		사업 분야	//현재 영위 중인 주요 사업 (전자책 유통, 학습콘텐츠 제작, SW개발 등)
-----	--	-------	--

2. 보유데이터 현황(※ 제공 가능한 데이터별로 작성)

<작성 요령>

- 신청기관이 적절한 권원(저작권, 소유권 등)을 보유한 데이터를 기준으로 작성
- AI 학습 등의 활용을 위해 신청기관이 제공하는 데이터를 명확하게 이해할 수 있도록 작성
※ 신청서 접수 시 가능한 경우 데이터 샘플도 함께 첨부

데이터명	
데이터 설명(요약)	
데이터 유형	☐ 텍스트 ☐ 이미지 ☐ 음성 ☐ 영상 ☐ 기타()
데이터 종류	// 전문서적, 교과서, 학습교재, 뉴스(방송, 신문) 등
데이터 규모	// 텍스트 00GB, 전자책 00권 분량, 뉴스 000건 등
데이터 확보 경로	☐ 직접 생성 ☐ 신탁관리 ☐ 업무협력/제휴(//원본데이터 보유기관명 기재) ☐ 기타 ()
데이터 형식	☐ txt ☐ pdf (☐ OCR 가능 ☐ OCR 불가) ☐ csv ☐ json ☐ xml ☐ Parquet ☐ 기타 (//이미지 등)
데이터 상태	☐ 원본 상태 (정제·전처리 필요) ☐ 부분 가공됨 (일부 정제/파싱 완료) ☐ 구조화 완료 (csv, json, xml 등)
개인정보 포함 여부	☐ 포함 ☐ 해당없음
권원 보유 범위 (저작권, 소유권 등)	☐ 전체 ☐ 일부 (예시) 문제집 내 지문 등 제외)

3. 데이터 제공 조건(※ 2. 보유데이터 현황에 기재한 데이터별로 작성)

<작성 요령>

- 신청기관의 데이터 제공 조건(무상/유상)에 대한 세부 항목 체크 및 작성
 - ※ (무상 제공 시) 데이터 제공에 필요한 기초 지원 및 실비 지급, (유상 제공 시) 데이터 사용 요건을 고려하여 적정 비용 협의 지급
- 정예팀(선발 후)의 공동활용 데이터 수요를 기반으로 이용협의, 계약 등을 통해 공급기관 최종 확정

구 분		상세 제공 조건(해당 항목 체크 및 작성)			
활용 요건		☐ 제한없음(상업/비상업 모두 사용) ☐ 비상업용 ☐ 기타()			
제공 형태		☐ 원시 데이터 ☐ 원천 데이터 ☐ 학습용 데이터			
이용 기간		☐ 영구 ☐ 기간(//1년, 3년 등) ☐ 기타()			
제공 방식	☐ 무상 제공	지원 필요 항목	☐ 기초 지원(원시데이터에 대한 OCR, 완전중복 제거 등) ☐ 실비 지원(법률 검토, 전처리 및 가공 등에 필요한 비용) ☐ 기타() ☐ 해당없음		
		필요 예산 (실비 지원 선택 시)	총액(//000백만원) <산출내역 작성> 필요한 실비 지원 예산에 대한 구체적인 산출내역 작성 예시) 데이터 준비를 위한 담당자 투입 공수 : m/m 000원		
			AI 허브 개방 희망 여부		
		☐ 유 ☐ 무			
	☐ 유상 제공	거래 이력	☐ 유 ☐ 무	거래 이력 증빙 자료	☐ 유(//견적서, 계약서 등) ☐ 무
	이용 방식·요건 (이용료 산정 기준)	// 정예팀에 한정하여 사업 목적 내에서만 사용			
	희망 이용료	총액(000백만원) <산출내역 작성> ※ 비용 보전 요청 금액에 대한 구체적인 산출내역 작성 예시: 000 데이터 건 당 00원 x 000원 = 000원 ※ 본 금액은 독자 AI 파운데이션 모델 프로젝트의 사업범위에서만 이용되는 데이터를 제공하는 기관에 대해 데이터 구축 등 해당 기관의 노력을 일부 보전하기 위한 취지의 비용으로 데이터의 시장 가격과는 무관함			

4. 추가 인센티브 희망사항

※ 추가 인센티브 제공은 확정된 사항은 아니며, 지속적 협력기반 마련을 위해 검토 예정

구 분	필요 항목 (해당 항목 체크)
① 제공 데이터를 데이터 마켓(AI 허브 등)에 유통할 수 있도록 상품등록·홍보 등 지원	
② 기관의 AI-데이터 활용 역량에 대한 대외적 홍보(언론보도, 성과보고회 등) 지원	
③ 우수 공급기관 추가 인센티브(포상, 우수사례 콘텐츠 제작 등) 제공	
④ 과기정통부 및 산하기관이 주관하는 데이터 관련 사업 참여 시 가점(인센티브) 부여	