

인공지능 학습용 데이터 구축사업

똑똑한 인공지능을 만들기 위해서는
고품질, 대규모의 학습용 데이터가 필수입니다.

AI 기술 개발 및 연구에 필수적이지만
중소 · 벤처 · 스타트업 · 연구자들이 구축하기 어려운
대규모 인공지능 학습용 데이터를 지원하는 사업입니다.

2017-2021 191종

- '17년 인공지능 학습용 데이터 구축 시동
- '18년 인공지능 통합 플랫폼 'AI 허브' 런칭
- '17~21년까지 총 191종 데이터 개방
- '20년 클라우드 소싱 도입으로 약 4만여 디지털 일자리 창출
- '21년 안전한 데이터 개방을 위한 안심존 마련

2022 381종

- '21년 구축 인공지능 학습용 데이터 191종 개방(총 381종)
- '22년 인공지능 학습용 데이터 구축사업 추진(310종)
- 이용자 중심의 'AI 허브' 고도화 진행

2023 691종

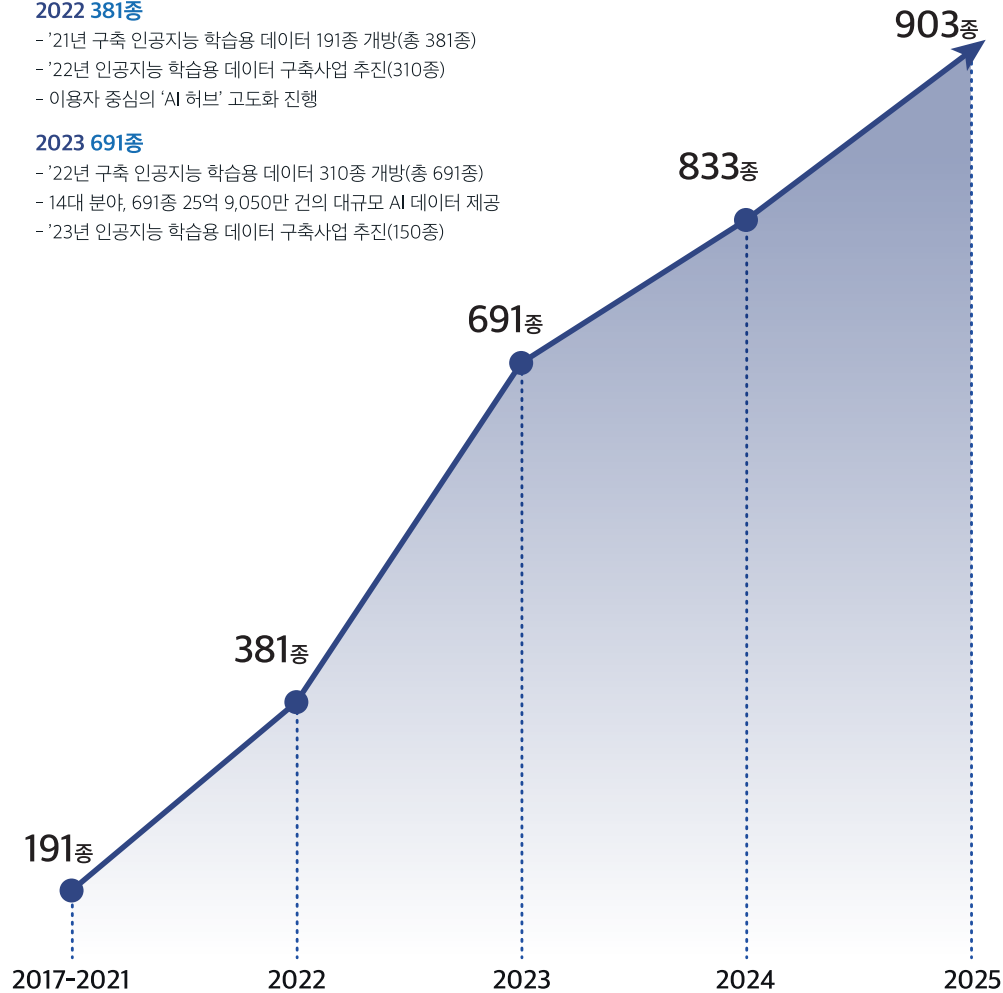
- '22년 구축 인공지능 학습용 데이터 310종 개방(총 691종)
- 14대 분야, 691종 25억 9,050만 건의 대규모 AI 데이터 제공
- '23년 인공지능 학습용 데이터 구축사업 추진(150종)

2024 833종

- '24년 인공지능 학습용 데이터 142종 개방(총 833종)
- '24년 인공지능 학습용 데이터 구축사업 추진(70종)

2025 903종

- '24년 인공지능 학습용 데이터 개방(903종) 예정
- '25년 초거대 인공지능 학습용 데이터 추가 구축 예정



www.aihub.or.kr

AI Hub



과학기술정보통신부

NIA 한국지능정보원

인공지능 통합 플랫폼 AI-Hub 활용 가이드

AI의 시작은 DATA,
DATA의 시작은 AI-Hub

AI Hub

AI Hub DATA

AI 통합 플랫폼 AI Hub

AI 허브는 대규모 인공지능 학습용 데이터와 AI 기술 및 서비스 개발에 필요한
다양한 AI 인프라를 지원하는 AI 통합 플랫폼입니다.

누구나 인공지능 서비스를 개발할 수 있도록
AI로 우리 삶을 풍요롭게 만들 수 있도록
대한민국이 인공지능 강국으로 나아갈 수 있도록

국민이 기획, 구축, 활용하는 인공지능 학습용 데이터로
국가 디지털 대전환을 주도합니다.

LOCALIZATION DATA

한국 특화 데이터

한국어, 한국인 얼굴, 도로, 농축수산물 등
한국 특화 데이터 제공

고품질 데이터

체계적인 품질관리, 실사용자 의견 반영
등을 통한 품질 고도화

HIGH QUALITY DATA

LARGE SCALE DATA

대규모 데이터

14대분야, 833종의 대규모 인공지능
학습용 데이터를 제공

AI 허브 오픈 API 'aihubshell'

Linux 등 다양한 개발환경에 최적화된
AI 데이터 다운로드 제공

AI-Hub Open API 'aihubshell'

Open Ko-LLM LeaderBoard

Open Ko-LLM 리더보드

'한국어 초거대 언어 모델(LLM)'의
성능평가와 순위를 경쟁하는 리더보드

데이터 개방 플랫폼 “AI 허브”

실사용자 중심의 편리한 데이터 활용 환경 조성을 위해
유형별/분야별 데이터셋 검색 기능 등 다양한 서비스를 제공하고 있습니다.

데이터셋 검색

데이터셋 단위 상세 검색 기능 제공
- 분야별, 유형별, 구축년도별 검색 지원

AI 허브 오픈 API ('24.1~)

이용자 개발환경에서 데이터를 손쉽게
다운로드 할 수 있도록 AI허브 다운로드 API 제공

한국어 LLM 리더보드 시즌2 ('24.8~)

한국어 거대 언어 모델(LLM)의 성능을 평가하고,
순위를 경쟁할 수 있는 참여형 리더보드
(NIA, 업스테이지 공동 주최)

글로벌 최대 규모의 인공지능 학습용 데이터

총 833종(*2024년 개방 데이터 기준)



데이터 주권의 핵심 한국어 데이터

다양한 분야, 상황에 적용 가능한 음성·대화·질의응답·번역 등의
학습용 데이터로 한국어 AI 개발을 지원합니다.



인공지능의 눈 영상이미지 데이터

인공지능이 사물·상황 등을 감지·인식·분석할 수 있도록 다양한
종류·유형의 영상·이미지 데이터를 제공합니다.



건강도 안심, 보안도 안심 헬스케어 데이터

한국인의 문화적·유전적 특성을 반영한 헬스케어 데이터를
개방하고, 데이터 안심존을 통해 데이터의 반출 없이 AI 모델을
개발할 수 있는 안전한 개발환경을 지원합니다.



미래 스마트팜의 씨앗 농축수산 데이터

농·축·수산업의 생산성 향상 및 스마트팜 조성을 위해 사육 및
질병 진단·처방 정보 등에 대한 데이터를 개방합니다.



스마트시티를 달리는 교통물류 데이터

스마트시티의 핵심 기술인 자율주행, 스마트물류 등에 필요한
데이터를 개방합니다.



국민의 안전을 책임지는 재난안전환경 데이터

홍수·태풍 등 자연재난 및 환경오염, 화재 등 사회재난에 대한
데이터를 개방하여 재난 예측·대응 시스템 구현을 지원합니다.



한국의 스마트 관광가이드 문화관광 데이터

대한민국의 아름다운 관광자원과 전통문화 등의 학습용
데이터로 한국만의 스마트 관광 서비스 개발을 지원합니다.



과학과 스포츠를 접목시키는 스포츠 데이터

선수 경기력 향상, 팀 전술 능력 향상을 위한 과학적 분석, 스포츠
저변 확대를 위한 서비스 구현에 필요한 데이터를 개방합니다.



에듀테크의 기본 교과서 교육 데이터

학생의 능력·수준에 따른 맞춤형 AI 교육 서비스 등에 필요한
데이터를 개방하여 에듀테크 서비스 개발을 지원합니다.



로봇기술 고도화의 동력 로보틱스 데이터

로봇 기술 수준 향상을 위한 인지·모션 등의 데이터를 개방하여
글로벌 로봇산업 경쟁력 강화에 필요한 기술개발을 지원합니다.



제조산업의 혁신을 위한 제조 데이터

제조산업 특화 데이터를 개방하여 국내 제조산업의 품질·생산성
향상에 기여할 수 있는 서비스 개발을 지원합니다.



새로운 기술 발견의 시작 지식재산 데이터

특허·기술·분쟁·거래 정보 등이 담긴 지식재산 데이터를
개방하여 특허상당, 유망기술 예측 기술 개발을 지원합니다.



스마트 법률서비스의 기초 법률 데이터

법률 지식과 접근성 증진, 높은 수준의 법률 서비스(리걸테크)
개발을 위한 법령·용어·판례 데이터를 개방합니다.



스마트 금융 서비스시대의 밑거름 금융 데이터

다양한 출처로부터 투자정보를 생산·제공하여 정보비대칭을 해소하거나,
서비스 개발·고도화를 지원할 수 있는 데이터를 개방합니다.

데이터 안심존 활용 가이드

AI 허브는 안심존을 통해 헬스케어 데이터 등
민감정보가 포함된 데이터를 안전하게 개방합니다.

안심존 활용 단계별 프로세스

AI 허브 - AI 허브 가입 및 로그인을 통해 헬스케어 카테고리 내 헬스케어 데이터 탐색 - AI 허브와 안심존 포털 간 회원 연동(SSO)	안심존 포털 접속 AI 허브 헬스케어 데이터 이용신청 버튼을 통해 안심존 전용 포털로 이동	신청서 작성 및 제출 - 안심존 포털 내 이용 신청서 작성 - 기관생명윤리위원회 심의 결과 통지서, 연구계획서 등 제시된 필수 서류 제출
안심존 자원 및 데이터셋 준비 승인된 사용 신청 건 관련 데이터셋·안심존 자원 준비 및 접속 정보 사용자 전달	안심존 가상화 서버 접속 부여된 접속 정보 기반 안심존 서버 접속 (데이터셋 접근권한 부여 전 사용자 환경구성 기간 최대 5일 제공)	의료데이터 활용 및 모델반출 - 데이터 분석·연구 - 사용자 : 학습 모델 반출 신청 - NIA : 학습 모델 반출 검토 및 파일 전달



안심존 사용 필수서류

- 기관생명윤리위원회(IRB) 심의 결과 통지서
 - 기관생명윤리위원회(IRB) 승인된 연구계획서
 - 신청자 소속 증빙 서류 (재직증명서, 재학증명서, 근로계약서 등 택 1)
 - 보안서약서
 - 안심존 이용 신청서
- ※ 교육·연구기관에 소속되지 않은 개인의 경우
공용기관생명윤리위원회를 통해 IRB 심의를 받으실 수 있습니다.

온라인 안심존



※ PC 최적화

사용자 매뉴얼



AI 허브 프로그램 참가하기

AI 허브의 인공지능 학습용 데이터는 여러분과 함께 만들고 성장합니다.
데이터를 경험하며 얻은 인사이트를 AI 허브에 공유해주세요.

신규 데이터 제안

신기술 트렌드와 현장 수요를 반영한
신규 데이터 및 기구축 데이터 고도화 제안

데이터 품질개선 의견 제안

인공지능 학습용 데이터 활용에 있어
개선이 필요한 내용을 국민 누구나 자유롭게 제안