

보도 시점 2026. 4. 10.(금) 12:00
(2026. 4. 11.(토) 조간) 배포 2026. 4. 10.(금) 09:00

인공지능 학습용데이터 보유 현황, 전 부처 대상 첫 전수조사 실시

- ‘인공지능 학습용데이터 총조사(센서스)’를 통한 보유 현황 및 인공지능 학습 활용 가능성 파악
- 민간 수요조사 등을 토대로 통합제공 대상 100종 선정 지원

【관련 국정과제】 20. 인공지능 3대 강국 도약을 위한 인공지능 고속도로 구축

과학기술정보통신부(부총리 겸 과기정통부 장관 배경훈, 이하 ‘과기정통부’)와 한국지능정보사회진흥원(원장 김형철)은 각 부처 및 공공기관이 보유한 AI 학습용 데이터의 현황을 체계적으로 파악하고, AI 활용이 가능한 고품질 데이터 확보 기반을 마련하기 위해 범부처 현황조사인 「AI 학습용데이터 현황조사」(이하 ‘AI 학습용데이터 센서스’)를 4.10.(금)부터 실시한다고 밝혔다.

생성형 AI 확산과 함께 산업·공공 전반에서 AI 활용 수요가 급증하고 있으며, 이를 뒷받침할 수 있는 고품질 학습용 데이터 확보가 핵심 경쟁력으로 부상하고 있다. 그러나 현재 공공기관이 보유한 AI 학습용데이터는 기관별로 분산·관리되고 있어 보유 현황과 활용 가능성에 대한 종합적인 파악이 어렵고, AI 기업들이 학습으로 연계·활용에도 한계가 있는 상황이다.

이에 과기정통부는 ‘AI기본법’ 제15조(인공지능 학습용데이터 관련 시책의 수립 등)에 근거, 전 부처를 대상으로 첫 전수 조사를 실시하여 각 기관이 보유한 AI 학습용데이터 현황을 체계적으로 진단하고, AI 학습 활용 가능성이 높은 데이터 100종을 발굴하여 ‘AI 학습용데이터 통합제공체계*’ (이하 ‘통합제공체계’)를 통해 연계·제공 (60억원 규모)할 계획이다.

* ‘AI 기본법’ 제15조에 따라 AI 학습용데이터가 다수 적재된 기존 ‘AI허브(14대 분야 903종 개방)’를 개편하여 공공·민간이 보유한 AI 학습용데이터를 집적하고, 누구나 쉽게 검색·활용할 수 있도록 국가 차원에서 지원하는 플랫폼

■ ‘AI 학습용데이터 센서스’ 주요내용

이번 센서스는 AI 학습용데이터뿐 아니라 향후 가공을 통해 활용 가능한 데이터까지 포함하여 조사하는 것이 특징이며, 단순 현황 파악을 넘어 실제 AI 학습에 활용 가능한 데이터 후보군을 도출하는 데 중점을 두고 추진된다.

조사 항목은 AI 학습용데이터 유형·구조, 구축 목적, 제공 가능 범위 등 AI 학습 활용 가능성과 직결되는 요소 중심으로 구성되며, 표준화된 조사 체계를 통해 조사의 일관성과 신뢰성을 확보할 계획이다.

■ AI 학습 활용 가능 데이터 100종 도출 및 통합제공체계 연계

센서스를 통해 발굴한 각 기관의 AI 학습용데이터 보유 현황을 토대로 기관별·분야별·유형별로 데이터 제공 가능 여부, 가공 수준 등에 대한 종합적인 검토가 이루어지며, 이를 통해 AI 학습 활용 가능성이 높은 데이터 후보군을 도출한다. 이를 바탕으로 전문가 심층 인터뷰, 민간 수요조사 등을 거쳐 통합제공체계를 통해 연계·제공할 AI 학습용데이터 100종을 선정한다.

선정된 100종의 데이터는 품질 보완, 비식별조치 등 AI 학습용데이터 가공 관련 기술적·재정적 지원을 거쳐 제공될 예정이며, 온라인 공개가 어려운 경우 ‘데이터 안심구역*’을 통한 제공을 병행한다.

* ‘데이터산업법 제11조에 따라 기술적·물리적·관리적 보안대책이 갖추어진 시설에서 미개방 데이터를 안전하게 분석·활용할 수 있도록 지정된 구역으로, 현재 11개 기관, 14개 구역 지정 운영 중

한편, 과기정통부는 ‘AI기본법’ 제15조(인공지능 학습용데이터 관련 시책의 수립 등)에 따라 현행 ‘AI허브’를 ‘AI 학습용데이터 통합제공체계’로 고도화 중이며 AI 학습용데이터 거래 활성화 지원 체계를 마련하고, 데이터 안심구역 등을 연계할 계획이다. 이를 통해 AI 개발·활용에 필요한 공공·민간 AI 학습용데이터를 통합적으로 제공·관리하고 AI 학습용데이터 발굴-확보-활용으로 이어지는 선순환 구조를 조성할 예정이다.

김경만 인공지능정책실장은 “AI 성능과 품질의 핵심은 활용 가능한 풍부한 데이터에 있다”며 “이번 조사를 통해 공공부문이 보유하고 있는 AI 학습용데이터 자산을 체계적으로 발굴하고 이를 편리하게 활용할 수 있는 AI 학습용데이터 통합제공 기반을 지속적으로 발전시켜 나가겠다”고 밝혔다.

담당 부서	인공지능정책실 인공지능데이터정책과	책임자	과 장	이소라 (044-202-6580)
		담당자	사무관	문나윤 (044-202-6583)

내일을 만드는 과학기술
내 삶을 채우는 디지털·AI

대한민국
지능책브리핑

