



제 429회 국회 (정기회)
제16차 과학기술정보방송통신위원회

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 일부개정법률안

검 토 보 고

<인공지능 학습을 위한 공공데이터 활용 지원 등>
■ 최민희의원 대표발의(의안번호 제2212675호)

2025. 11.

과학기술정보방송통신위원회

수석전문위원 이복우

목 차

I . 제안경위	1
II . 제안이유 및 주요내용	1
III . 검토의견	2
1. 인공지능 학습을 위한 공공데이터 활용 지원 등(안 제2조 제12호 신설 등)	2

I. 제안경위

가. 발 의 자 : 최민희의원 등 10인

나. 발 의 일 : 2025. 9. 3.

다. 회 부 일 : 2025. 9. 4.

II. 제안이유 및 주요내용

현행법은 인공지능 학습용 데이터의 수집·활용 등에 관한 시책 추진 등을 규정하고 있으나 국가기관과 지방자치단체 등이 보유·관리하는 공공데이터를 이용한 인공지능 학습용 데이터에 대하여 따로 규정하고 있지 않음.

인공지능 산업의 발전을 위하여 양질의 공공데이터를 산업 생태계에 원활하게 공급하고 이를 적절하게 활용할 수 있는 기반 마련이 시급함. 국가기관과 지방자치단체 등이 보유·관리하는 공공데이터를 인공지능의 개발과 활용에 사용할 수 있도록 가공하여 제공할 수 있는 법적 근거가 필요함.

이에 공공데이터를 이용한 인공지능 학습용데이터의 생성 및 제공에 관한 사항 등을 추가함(안 제2조, 제6조제2항, 제8조제1항, 제15조제1항 및 제16조제1항).

Ⅲ. 검토의견

1. 인공지능 학습을 위한 공공데이터 활용 지원 등(안 제2조 제12호 신설 등)

가. 개정안 요지

- ‘인공지능의 개발·활용 등에 사용되는 데이터’를 “학습용 데이터”로 정의함
(안 제2조 제12호 신설).
- 인공지능 기본계획에 포함되어야 할 사항에 “「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따른 공공데이터를 학습용데이터로 제공하기 위한 우선 생성·제공 범위 및 기준, 생성·제공 활성화에 관한 사항”을
추가함(안 제6조 제2항 제4호의2 신설).
- 국가인공지능위원회의 심의·의결 사항에 “학습용데이터 수집·관리 및
활용 촉진에 관한 사항”을 추가함(안 제8조 제1항 제6호의2 신설).
- 과학기술정보통신부장관은 학습용데이터의 생산·수집·관리·유통·
활용 촉진 및 품질수준 확보 등을 위하여 필요한 시책을 추진하여야
함(안 제15조 제1항).
- 국가 및 지방자치단체는 기업 및 공공기관의 인공지능기술 도입 촉진
및 활용 확산을 위하여 필요한 경우에 “공공기관이 보유·관리하고 있
는 데이터를 학습용데이터로 생성·제공하고 적절한 품질수준을 확보
하는데 필요한 지원”을 할 수 있음(안 제16조 제1항 제2호의2 신설).

나. 개정안 조문대비표

현행	개정안
법률 제20676호 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다. 1. ~ 11. (생략) <u><신설></u>	법률 제20676호 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 제2조(정의) ----- -----. 1. ~ 11. (현행과 같음) 12. <u>“학습용데이터”란 인공지능의 개발·활용 등에 사용되는 데이터를 말한다.</u>
제6조(인공지능 기본계획의 수립) ① (생략) ② 기본계획에는 다음 각 호의 사항이 포함되어야 한다. 1. ~ 4. (생략) <u><신설></u>	제6조(인공지능 기본계획의 수립) ① (현행과 같음) ② ----- -----. 1. ~ 4. (현행과 같음) 4의2. <u>「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따른 공공데이터를 학습용데이터로 제공하기 위한 우선 생성·제공 범위 및 기준, 생성·제공 활성화에 관한 사항</u>
5. ~ 7. (생략) ③ ~ ⑦ (생략) 제8조(위원회의 기능) ① 위원회는 다음 각 호의 사항을 심의·의결한다.	5. ~ 7. (현행과 같음) ③ ~ ⑦ (현행과 같음) 제8조(위원회의 기능) ① ----- ----- -----.

현행	개정안
1. ~ 6. (생략) <u><신설></u> 7. ~ 13. (생략) ②·③ (생략) 제15조(인공지능 학습용데이터 관련 시책의 수립 등) ① 과학 기술정보통신부장관은 관계 중 앙행정기관의 장과 협의하여 <u>인공지능의 개발·활용 등에</u> <u>사용되는 데이터(이하 “학습용</u> <u>데이터”라 한다)의 생산·수집</u> <u>·관리·유통 및 활용 등을 촉</u> <u>진하기</u> 위하여 필요한 시책을 추진하여야 한다. ② ~ ⑥ (생략) 제16조(인공지능기술 도입·활용 지원) ① 국가 및 지방자치단 체는 기업 및 공공기관의 인공 지능기술 도입 촉진 및 활용 확산을 위하여 필요한 경우에 는 다음 각 호의 지원을 할 수 있다. 1.·2. (생략) <u><신설></u>	1. ~ 6. (현행과 같음) <u>6의2. 학습용데이터 수집·관리</u> <u>및 활용 촉진에 관한 사항</u> 7. ~ 13. (현행과 같음) ②·③ (현행과 같음) 제15조(인공지능 학습용데이터 관련 시책의 수립 등) ① ----- ----- ----- <u>학습용데이터의 생산·수집·</u> <u>관리·유통·활용 촉진 및 품</u> <u>질수준 확보 등을</u>----- -----. ② ~ ⑥ (현행과 같음) 제16조(인공지능기술 도입·활용 지원) ① ----- ----- ----- ----- ----- ---. 1.·2. (현행과 같음) <u>2의2. 공공기관이 보유·관리하</u>

현행	개정안
3. ~ 5. (생략) ② (생략)	<u>고 있는 데이터를 학습용데이터로 생성·제공하고 적절한 품질수준을 확보하는데 필요한 지원</u> 3. ~ 5. (현행과 같음) ② (현행과 같음)

다. 검토사항

- 인공지능 시대에는 고성능 GPU와 같은 물리적 연산 인프라의 확보도 필수적이지만, 실제로 인공지능 모델의 성능을 결정짓는 핵심 요소는 다양한 영역과 형태의 데이터를 충분히 확보하고 이를 학습에 활용할 수 있는 환경을 조성하는 것이며, 데이터의 양과 질이 모델의 정확도·활용성을 좌우한다는 점에서 풍부한 학습용 데이터의 제공이 중요함.
- 이를 반영하여 현행 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 “인공지능 기본법”)은 국가가 인공지능 학습용 데이터의 구축·제공·품질관리를 지원할 수 있는 근거규정¹⁾을 마련하고 있고,

1) 제15조(인공지능 학습용데이터 관련 시책의 수립 등) ① 과학기술정보통신부장관은 관계 중앙행정기관의 장과 협의하여 인공지능의 개발·활용 등에 사용되는 데이터(이하 “학습용데이터”라 한다)의 생산·수집·관리·유통 및 활용 등을 촉진하기 위하여 필요한 시책을 추진하여야 한다.
② 정부는 학습용데이터의 생산·수집·관리·유통 및 활용 등에 관한 시책을 효율적으로 추진하기 위하여 지원대상사업을 선정하고 예산의 범위에서 지원할 수 있다.
③ 정부는 학습용데이터의 생산·수집·관리·유통 및 활용의 활성화 등을 위하여 다양한 학습용 데이터를 제작·생산하여 제공하는 사업(이하 “학습용데이터 구축사업”이라 한다)을 시행할 수 있다.
④ 과학기술정보통신부장관은 학습용데이터 구축사업의 효율적 수행을 위하여 학습용데이터를 통합적으로 제공·관리할 수 있는 시스템(이하 “통합제공시스템”이라 한다)을 구축·관리하고 민간이 자유롭게 이용할 수 있도록 제공하여야 한다.

과학기술정보통신부는 AI 허브(AI-Hub) 운영을 통해 다수의 공공·민간 데이터를 개방하는 등 정부 예산으로 데이터 확보 사업을 추진하면서 인공지능 생태계 조성을 위해 노력하고 있음.

- 특히, 정부·지자체 및 공공기관이 축적해온 공공데이터는 구축 기간이 길고 국민 전반을 대상으로 수집된 경우가 많아 데이터의 규모와 다양성 측면에서 매우 풍부하며, 이는 기업의 영업비밀이 얹혀 있는 민간 데이터에 비해 비교적 자유롭게 활용할 수 있다는 점에서 인공지능 학습용 데이터로서의 가치가 높다고 평가됨.
- 그러나 현재 공공데이터는 주로 단순한 일반 데이터 형태로 공공데이터 포털 등을 통해 제공되는 수준에 머물고 있으며, 학습용 데이터로 적합하게 구조화·정제·라벨링되어 제공되는 사례는 제한적이어서, 인공지능 연구자와 개발자가 직접 활용하기에는 여전히 품질과 형식 면에서 제약이 많은 상황임.
- 개정안은 국가가 공공데이터의 인공지능 학습용 데이터로써 공공데이터 활용을 지원할 수 있도록 정책적·제도적 기반을 마련함으로써, 공공 부문이 보유한 데이터가 체계적으로 정비·가공하여 인공지능 모델 학습에 적극적으로 제공되도록 하려는 것임.
- 다만, 공공데이터에도 국민의 개인정보가 포함되거나 저작권이 복잡하게 얹혀 있을 수 있으므로, 이러한 법적·윤리적 쟁점을 해소하기 위한

⑤ 과학기술정보통신부장관은 통합제공시스템을 이용하는 자에 대하여 비용을 징수할 수 있다.

⑥ 그 밖에 제2항에 따른 지원대상사업의 선정 및 지원, 학습용데이터 구축사업의 시행, 통합제공시스템의 구축·관리 및 제5항에 따른 비용의 징수 등에 필요한 사항은 대통령령으로 정한다.

세부 지침과 가이드라인이 병행되어야 실제로 공공데이터의 인공지능 학습용 전환이 활성화될 수 있을 것으로 보이며, 향후 이를 구체화하는 후속 제도 설계가 함께 이루어질 필요가 있음.

- 과학기술정보통신부는 인공지능 기술·산업 진흥의 맥락에서 공공기관의 학습용 데이터 업무 지원을 규정함으로써 인공지능 개발·활용과 연계한 정책 추진에 기여할 수 있을 것으로 보인다는 입장임.
- 행정안전부는 인공지능 기본계획 내용 중 공공데이터를 학습용데이터로 제공하는 것에 관한 사항은 공공·민간 데이터를 포괄하는 형태로 수정이 필요하다는 의견을 제출함.

문 의 처
02)6788-5230