



제 428회 국회 (임시회)  
제2차 과학기술정보방송통신위원회

# 데이터 산업진흥 및 이용촉진에 관한 기본법 일부개정법률안

## 검 토 보 고

〈인공지능 학습데이터에 대한 특화체계 마련〉  
■ 김태년의원 대표발의(의안번호 제2210781호)

2025. 8.

과학기술정보방송통신위원회

수석전문위원 이복우

# 목 차

|                                              |   |
|----------------------------------------------|---|
| I . 제안경위 .....                               | 1 |
| II . 제안이유 및 주요내용 .....                       | 1 |
| III . 검토의견 .....                             | 2 |
| 1. 인공지능 학습데이터에 대한 특화체계 마련(안 제14조의2 신설 등) ... | 2 |

## I. 제안경위

가. 발 의 자 : 김태년의원 등 10인

나. 발 의 일 : 2025. 6. 12.

다. 회 부 일 : 2025. 6. 13.

## II. 제안이유 및 주요내용

인공지능(AI) 기술의 발전과 산업 전반에 걸친 데이터 수요의 증가에 따라, 인공지능 학습용 데이터를 안정적이고 효율적으로 유통·거래할 수 있는 기반을 마련하는 것이 국가 데이터산업 경쟁력 강화의 핵심 과제로 부상하고 있음.

그러나 현재는 학습데이터의 경제적 가치에 대한 객관적 평가 기준이 미비하고, 거래 시 활용할 수 있는 표준계약서나 품질인증 체계가 부재하여, 기업 간 거래의 신뢰 부족과 법적 불확실성으로 인해 시장 활성화에 어려움이 있는 실정임.

이에 기존 데이터 가치평가 체계를 인공지능 학습데이터에 특화되도록 보완하고, 반복 활용 및 대규모 가공이 전제된 학습데이터 특성을 반영한 별도의 거래 표준계약서를 마련하며, 개인정보 보호를 위한 익명화·비식별화 기준 및 품질인증 기준을 제도적으로 정비함으로써, 인공지능 학습용 데이터 거래를 촉진하고, 데이터 기반 신산업의 성장을 뒷받침하고자 함(안 제14조의2 신설 등).

### Ⅲ. 검토의견

#### 1. 인공지능 학습데이터에 대한 특화체계 마련(안 제14조의2 신설 등)

##### 가. 개정안 요지

- 과학기술정보통신부장관은 인공지능 학습데이터의 경제적 가치를 객관적으로 평가할 수 있도록 데이터 가치평가 기법 및 체계를 인공지능 학습 목적에 적합하도록 보완·활용할 수 있음(안 제14조의2 제1항).
- 과학기술정보통신부장관은 인공지능 학습데이터의 산업적 활용을 위하여 특화된 거래 표준계약서를 마련·보급할 수 있음(제2항).
- 인공지능 학습데이터의 유통 및 활용 과정에서의 개인정보 보호를 위한 익명화·비식별화 기준은 대통령령으로 정함(제3항).
- 과학기술정보통신부장관은 인공지능 학습데이터의 품질을 확보하기 위하여 별도의 품질인증 기준 및 절차를 마련할 수 있음(안 제20조 제6항 신설).
- 과학기술정보통신부장관은 인공지능 학습데이터의 특성과 반복 활용 가능성을 반영한 별도의 거래 표준계약서를 마련할 수 있으며, 이를 데이터사업자 및 관련 기관에 보급할 수 있음(안 제21조 제3항 신설).

나. 개정안 조문대비표

| 현행                       | 개정안                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><u>&lt;신설&gt;</u></p> | <p><u>제14조의2(인공지능 학습데이터의 가치평가 및 거래활성화)</u></p> <p><u>① 과학기술정보통신부장관은 인공지능 학습데이터의 경제적 가치를 객관적으로 평가할 수 있도록 제14조에 따른 평가기법 및 평가체계를 인공지능 학습 목적에 적합하도록 보완·활용할 수 있다.</u></p> <p><u>② 과학기술정보통신부장관은 인공지능 학습데이터의 산업적 활용을 위하여 제21조에 따른 표준계약서와 별도로 특화된 거래 표준계약서를 마련하고 이를 보급할 수 있다.</u></p> <p><u>③ 인공지능 학습데이터의 유통 및 활용 과정에서 개인정보 보호를 위한 익명화 및 비식별화 기준은 대통령령으로 정한다.</u></p> <p><u>④ 제1항부터 제3항까지에서 규정한 사항 외에 인공지능 학습데이터의 품질인증, 반복 활용, 결합 기준 등에 필요한 사항은 대통령령으로 정한다.</u></p> |

| 현행                                                                                                  | 개정안                                                                                                                                              |
|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| 제20조(데이터 품질관리 등) ①<br>~ ⑤ (생략)<br><u>&lt;신설&gt;</u>                                                 | 제20조(데이터 품질관리 등) ①<br>~ ⑤ (현행과 같음)<br><u>⑥ 과학기술정보통신부장관은 인공지능 학습데이터의 품질을 확보하기 위하여 제14조의2에 따른 별도의 품질인증 기준 및 절차를 마련할 수 있다.</u>                      |
| 제21조(표준계약서) ①·② (생략)<br><u>&lt;신설&gt;</u>                                                           | 제21조(표준계약서) ①·② (현행과 같음)<br><u>③ 과학기술정보통신부장관은 제14조의2에 따라 인공지능 학습데이터의 특성과 반복 활용가능성을 반영한 별도의 거래 표준계약서를 마련할 수 있으며, 이를 데이터사업자 및 관련 기관에 보급할 수 있다.</u> |
| 부 칙                                                                                                 |                                                                                                                                                  |
| 제1조(시행일) 이 법은 공포 후 6개월이 경과한 날부터 시행한다.                                                               |                                                                                                                                                  |
| 제2조(경과조치) 이 법 시행 당시 과학기술정보통신부장관이 종전의 규정에 따라 추진 중인 인공지능 학습데이터 품질인증 및 유통기준 마련 업무는 이 법에 따라 추진한 것으로 본다. |                                                                                                                                                  |

## 다. 검토사항

- 인공지능기술 및 산업의 발전에 따라 인공지능의 연료(燃料)로 불리는 인공지능 학습데이터의 확보를 지원할 필요성도 커지고 있음.
  - 과학기술정보통신부의 「2024 인공지능산업 실태조사 보고서」에 따르면 조사 대상 인공지능 사업자 중 59.5%가 데이터 확보 및 품질 문제를 사업 운영상 애로사항으로 체감하고 있음<sup>1)</sup>.
- 인공지능 학습데이터는 일반 데이터와 달리 인공지능 학습을 위한 정제 과정이 필요하고, 인공지능 모델 개선을 위해 반복적으로 재사용이 가능하여 데이터 수명주기에 차이가 있으며, 데이터의 가치를 평가함에 있어 편향성, 대표성 등에 대한 판단이 상대적으로 중요한 등 일반 데이터와 구별되는 특성을 지님.
- 이에 개정안은 인공지능 학습데이터에 대한 별도의 가치평가, 표준 계약서, 품질인증 등을 규정함으로써 인공지능 학습데이터의 특성에 부합하는 정책을 수행하도록 하고, 인공지능 개발을 위한 고품질의 데이터 확보 및 활용을 지원하려는 것임.
- 다만 개정안에 대해서는 다음의 사항을 고려할 필요가 있음.
- 첫째, 개정안에 따른 ‘인공지능 학습데이터’는 「인공지능 발전과 신뢰 기반 등에 관한 기본법」(이하 “인공지능기본법”)에 따른 ‘인공지능 학습용 데이터’와 동일한 의미로 보이나,

---

1) 과학기술정보통신부, 「2024 인공지능산업 실태조사」, p. 95

데이터산업법에 인공지능 학습데이터의 정의가 별도로 없다는 점, 인공지능기본법과의 정합성을 확보할 필요가 있다는 점을 고려하여 용어 수정 및 정의가 필요해 보임.

- 참고로, 인공지능기본법에 따른 ‘인공지능 학습용 데이터’는 인공지능의 개발·활용 등에 사용되는 데이터를 의미함<sup>2)</sup>.

○ 둘째, 현행 데이터산업법상 데이터는 인공지능 학습데이터를 포괄하는 개념이고, 현재 데이터 가치평가 기법 및 평가체계, 데이터 품질기준 및 품질관리, 표준계약서의 작성에 대한 세부사항은 동법 시행령 또는 과학기술정보통신부장관에게 위임되어 있으므로, 현행법상으로도 인공지능 학습데이터에 대한 가치평가, 품질관리 및 표준계약서 작성이 가능할 것으로 보임.

- 참고로, 과학기술정보통신부와 한국지능정보사회진흥원(NIA)는 2021년부터 이미 ‘인공지능 학습용 데이터 품질관리 가이드라인’을 마련·제공해 오고 있음.

○ 셋째, 현행 데이터산업법은 데이터 표준계약서 작성 시 공정거래위원회와 협의를 거치고 이해관계자 및 전문가의 의견을 청취하도록 하고 있는데<sup>3)</sup>, 안 제14조의2 제2항 및 제21조 제3항에 따른 인공지능 학습데이터 거래

---

2) 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.

3) 제21조(표준계약서) ① 과학기술정보통신부장관은 데이터의 합리적 유통 및 공정한 거래를 위하여 공정거래위원회와 협의를 거쳐 표준계약서를 마련하고, 데이터사업자에게 그 사용을 권고할 수 있다.

② 과학기술정보통신부장관은 제1항에 따른 표준계약서를 제정 또는 개정하는 경우에는 관련사업자단체 등 이해관계자와 전문가의 의견을 들어야 한다.

표준계약서의 경우 해당 절차를 포함하지 않고 있으므로 이를 추가할 필요가 있다고 보임.

○ 넷째, 데이터산업법과 「개인정보 보호법」의 관계 및 현행 제도 운영 상황을 고려할 때 안 제14조의2 제3항에 따른 익명화·비식별화 기준에 대해서는 「개인정보 보호법」 등 관계 법령으로 정하는 것이 타당하다고 보임.

- 현행 데이터산업법은 개인정보에 관한 사항의 경우 동법이 아닌 「개인정보 보호법」에서 정하는 바에 따르도록 하고 있으므로 동법에 개인정보에 관한 사항을 규정하는 것은 법률의 취지 및 법체계에 부합하지 않는 측면이 있음<sup>4)</sup>.

- 개인정보의 익명화 및 비식별화에 대해서는 현재 「개인정보 보호법」에서 가명정보 처리 및 조치의무에 관한 사항을 규정하고 있으며<sup>5)</sup>, 개인정보보호위원회는 이에 따라 ‘가명정보 처리 가이드라인’, ‘보건의료 데이터 활용 가이드라인’, ‘교육분야 가명·익명정보 처리 가이드라인’

---

4) 제7조(다른 법률과의 관계) ② 개인정보, 저작권 및 공공데이터에 관하여는 각각 「개인정보 보호법」, 「저작권법」, 「공공데이터의 제공 및 이용 활성화에 관한 법률」 등 다른 법률에서 정하는 바에 따른다.

5) 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

1. “개인정보”란 살아 있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보를 말한다.

다. 가목 또는 나목을 제1호의2에 따라 가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보(이하 “가명정보”라 한다)

제28조의2(가명정보의 처리 등) ① 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있다.

② 개인정보처리자는 제1항에 따라 가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함해서는 아니 된다.

제28조의4(가명정보에 대한 안전조치의무 등) ① 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 경우에는 원래의 상태로 복원하기 위한 추가 정보를 별도로 분리하여 보관·관리하는 등 해당 정보가 분실·도난·유출·위조·변조 또는 훼손되지 않도록 대통령령으로 정하는 바에 따라 안전성 확보에 필요한 기술적·관리적 및 물리적 조치를 하여야 한다.

등을 마련·제공하고 있음.

- 다섯째, 안 제14조의2 제2항과 제21조 제3항은 모두 인공지능 학습데이터에 대한 거래 표준계약을 마련·보급하도록 하는 것으로서 그 내용이 동일하므로 중복적으로 규정할 필요성이 낮을 것으로 보임.
- 과학기술정보통신부는 인공지능 학습데이터의 특성을 반영하여 기존 제도를 보완할 수 있는 법적 근거를 마련하려는 취지에 공감하여 개정안에 동의하는 의견임.

다만, 법체계의 정합성을 고려하여 인공지능 학습데이터에 대한 정의 신설, 조문 위치 이동 및 문구 수정 등 일부 정비가 필요하고,

안 제14조의2 제3항의 학습데이터 익명화 및 비식별화의 경우 「개인정보 보호법」상 가명정보 규정의 특례에 해당하므로, 개정안의 규정이 「개인정보 보호법」보다 우선 적용될 수 있도록 현행 제7조(다른 법률과의 관계)를 함께 개정할 필요가 있다는 입장임.

- 공정거래위원회는 안 제21조 제3항에 따른 표준계약서 마련 시에도 공정거래위원회와 협의를 거치도록 하여 현행 조문과의 통일성을 유지할 필요가 있다는 의견을 제출함.
- 한국방송협회는 인공지능 학습데이터 중 ‘저작물’은 「저작권법」상 보호되는 창작물로서 데이터와 구분되는 개념으로, 저작물에 대해서는 데이터 산업법이 아닌 「저작권법」에서 다루는 것이 바람직하고, 인공지능 및 인공지능 학습데이터의 정의가 없는 상황에서 개정안에 따른 인공지능 학습데이터에 저작물이 포함되는 것으로 해석되는 경우

문화산업 등 관련 업계를 불안정하게 만들 우려가 있으며,  
인공지능 학습데이터에 대한 거래가 활성화되지 않은 상황에서 데이터  
가치평가 기관이 학습데이터의 가치를 평가하는 것이 타당하다고 보기  
어렵다는 의견을 제출함.

|              |
|--------------|
| 문 의 처        |
| 02)6788-5230 |