



제 439회 국회 (정기회)
제1차 과학기술정보방송통신위원회

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 일부개정법률안

검 토 보 고

〈인공지능 학습용데이터 활용 기준 명확화〉

■ 이훈기의원 대표발의(의안번호 제2219563호)

2026. 9.

과학기술정보방송통신위원회

수석전문위원 이지민

목 차

I . 제안경위	1
II . 제안이유 및 주요내용	1
III . 검토의견	3
1. 인공지능 학습용데이터 활용 기준 명확화 등	3

I. 제안경위

가. 발 의 자 : 이훈기의원 등 10인

나. 발 의 일 : 2026. 6. 26.

다. 회 부 일 : 2026. 7. 1.

II. 제안이유 및 주요내용

현행법은 인공지능의 건전한 발전과 신뢰 기반 조성을 위하여 학습용데이터의 생산·수집·관리·유통·활용 촉진 및 품질수준 확보 등에 관한 시책, 학습용데이터 구축사업 및 통합제공시스템 구축·관리 등에 관하여 규정하고 있음.

그러나 현행법은 주로 정부의 학습용데이터 구축·지원사업을 중심으로 규정되어 있어, 민간 인공지능사업자가 인공지능의 개발·학습·검증·고도화 및 활용을 위하여 개인정보에 해당하지 아니하는 학습용데이터를 활용할 수 있다는 기본 원칙과 안전한 활용 기준은 명시되어 있지 않은 실정임.

특히 「개인정보 보호법」은 같은 법 제2조제1호에 따른 개인정보의 처리 및 보호를 중심으로 규율하고 있으므로, 개인정보에 해당하지 아니하는 학습용데이터의 수집·생성·가공·결합·이용 또는 제공에 대해서는 직접적인 규율체계를 두고 있지 않음. 이에 따라 인공지능 산업 현장에서는 학습용데이터 활용에 관한 법적 불확실성이 발생하고 있으며, 동시에 재식별 위험 등 안전성 확보를 위한 기준 마련의

필요성도 제기되고 있음.

이에 인공지능에 관한 기본법으로서 개인정보에 해당하지 아니하는 학습용데이터의 안전한 활용 원칙과 기준을 명시함으로써, 「개인정보 보호법」이 직접 규율하지 않는 영역의 법적 공백을 보완하고자 함.

아울러 학습용데이터 활용 과정에서 개인정보에 해당하게 되거나 재식별 위험이 발생한 경우에는 「개인정보 보호법」에 따르도록 하고, 과학기술정보통신부장관이 개인정보보호위원회 등 관계 중앙행정기관의 장과 협의하여 세부 기준을 마련하도록 함으로써, 인공지능 학습용데이터의 안전하고 원활한 활용을 도모하려는 것임.(안 제5조제1항, 제15조의2 신설).

Ⅲ. 검토의견

1. 인공지능 학습용데이터 활용기준 명확화 등

가. 개정안의 주요 내용

- 개정안은 ‘학습용데이터의 생산·수집·관리·유통·활용’에 대해서 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법을 따르도록 함(안 제5조제1항).
- 개정안은 인공지능사업자가 개인정보에 해당하지 아니하는 학습용데이터를 수집·생성·가공·결합·이용 또는 제공할 수 있도록 명시하되, 학습용데이터를 활용하는 과정에서 해당 학습용데이터가 특정 개인을 알아볼 수 있는 정보에 해당하게 되거나 재식별 위험이 발생한 사실을 알게 된 경우에는 「개인정보 보호법」을 따르도록 함(안 제15조의2제1항 및 제2항).
- 개정안은 과학기술정보통신부장관이 개인정보보호위원회 등 관계 중앙행정기관의 장과 협의하여 학습용데이터의 활용 기준, 재식별 위험관리 및 안전성 확보조치 등에 관한 세부 기준을 고시하도록 함(안 제15조의2제3항).

현 행	개 정 안
제5조(다른 법률과의 관계) ① 인공지능, 인공지능기술, 인공지능산업 및 <u>인공지능사회(이하 “인공지능등”이라 한다)</u> 에 관하여 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에서 정하는 바에 따른다.	제5조(다른 법률과의 관계) ① 인공지능, 인공지능기술, 인공지능산업, <u>인공지능사회 및 학습용데이터의 생산·수집·관리·유통·활용(이하 “인공지능등”이라 한다)</u> 에 관하여 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에서 정하는

현행	개정안
<신설>	준, 재식별 위험관리 및 안전성 확보조치 등에 관한 세부 기준을 정하여 고시하여야 한다. ④ 제1항부터 제3항까지에 따른 학습용데이터의 활용 기준, 안전성 확보조치 및 그 밖에 필요한 사항은 대통령령으로 정한다.
부칙	
제1조(시행일) 이 법은 공포 후 6개월이 경과한 날부터 시행한다.	

나. 검토의견

- 현행법은 학습용데이터를 정의하고, 과학기술정보통신부장관이 학습용데이터의 생산·수집·관리·유통 및 활용 촉진과 품질수준 확보 등에 관한 시책을 마련하도록 하는 한편, 학습용데이터 구축사업 및 통합제공시스템 구축·운영 등에 관한 사항을 규정¹⁾하고 있음.

1) 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

12. "학습용데이터"란 인공지능의 개발·활용 등에 사용되는 데이터를 말한다.

제15조(인공지능 학습용데이터 관련 시책의 수립 등) ① 과학기술정보통신부장관은 관계 중앙행정기관의 장과 협의하여 학습용데이터의 생산·수집·관리·유통·활용 촉진 및 품질수준 확보 등을 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 학습용데이터를 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.

⑤ 과학기술정보통신부장관은 통합제공시스템을 이용하는 자에 대하여 비용을 징수할 수 있다.

○ 그런데, 현행법은 정부의 학습용데이터 구축·지원 정책을 중심으로 규정하고 있을 뿐 인공지능사업자의 학습용데이터 활용 기준에 대해서는 별도로 규정하고 있지 아니하고, 현행법 시행 이후 과학기술정보통신부가 마련한 고시나 가이드라인에도 이에 관한 구체적인 기준은 포함되어 있지 않은 것으로 보이며,

현행법 제5조는 인공지능, 인공지능기술, 인공지능산업 및 인공지능사회에 관하여 다른 법률에 특별한 규정이 있는 경우에는 해당 규정을 적용하도록 규정하고 있는바,

과학기술정보통신부에 따르면 학습용데이터의 활용과 관련해서는 「개인정보 보호법」, 「저작권법」 등 관계 법률이 적용되고 있음.

○ 「개인정보 보호법」은 개인정보의 처리 및 보호에 관한 사항을 규정²⁾하는 법률이고, 동법에 따르면 “개인정보”는 살아 있는 개인에 관한 정보로서 개인을 알아볼 수 있는 정보임.

- 「개인정보 보호법」 제2조제1호에 따르면, “개인정보”란 살아 있는 개인에 관한 정보로서 ① 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보, ② 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보, ③ 이들을 가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사

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

2) 「개인정보 보호법」

제1조(목적) 이 법은 개인정보의 처리 및 보호에 관한 사항을 정함으로써 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 한다.

용·결합 없이는 특정 개인을 알아볼 수 없는 정보임.

- 따라서, 특정 학습용데이터가 「개인정보 보호법」 제2조제1호에 따른 개인정보³⁾에 해당한다면 「개인정보 보호법」에 따라 처리해야 하고, 개인정보에 해당하지 않는 정보의 활용에 대해서는 「개인정보 보호법」에서 별도로 규제하지 않고 있음.
- 개정안은 학습용데이터 활용에 대해서도 다른 법률과의 관계를 명시하는 한편, 「개인정보 보호법」의 적용을 받지 않는 학습용데이터에 대해서는 과학기술정보통신부장관이 활용 기준을 마련하도록 하고, 데이터 활용 과정에서 해당 데이터가 개인정보에 해당하게 되거나 재식별 위험이 발생한 경우에는 「개인정보 보호법」을 적용하도록 명확히 하려는 것임.
- 개정안은 인공지능 학습용데이터의 활용에 관한 기준을 명확히 함으로써 현장에서 발생할 수 있는 불확실성을 최소화하고 학습용데이터의 안전하고 원활한 활용을 도모하려는 취지로 보임.
- 다만, 개정안은 다음과 같은 사항을 고려할 필요가 있음.
- 첫째, 안 제15조의2제1항은 개인정보가 아닌 학습용데이터를 인공지

3) 「개인정보 보호법」

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.

가. 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보

나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보. 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다.

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)

능사업자가 활용할 수 있도록 규정하려는 것인데, 현행 법체계에서 개인정보에 해당하지 않는 데이터를 인공지능사업자가 학습용데이터로 활용하는 것에 대하여 별다른 규정을 두고 있지 않아 이미 규제 없이 활용할 수 있으므로, 이에 대한 법적 근거를 명시적으로 규정할 실익이 크지 않은 것으로 보임.

- 학습용데이터가 개인정보에 해당한다면 인공지능사업자가 학습용데이터를 수집·생성·가공·결합·이용 또는 제공할 때 「개인정보 보호법」을 적용받지만 그렇지 않은 정보에 대해서는 「저작권법」 등 다른 개별법을 준수하는 한 자유롭게 활용이 가능함.

○ 둘째, 안 제15조의2제2항은 인공지능사업자가 개인정보가 아닌 학습용데이터를 활용하다가 재식별⁴⁾이 발생한 경우 「개인정보 보호법」을 따르도록 규정하려는 것인데, 현재도 개인정보가 아닌 정보를 활용하다가 개인정보로 재식별되면 「개인정보 보호법」의 적용을 받게 되므로 개정의 실익이 크지 않은 것으로 보임.

○ 또한, 개정안은 재식별이 발생하지 않았더라도 그 위험이 발생한 사실을 알게 되면 「개인정보 보호법」을 따르도록 하고 있는데, 「개인정보 보호법」은 개인정보에 관하여 규정하고 있으므로 재식별이 아직 발생하지 않은 “재식별 위험”인 상태에서 개인정보가 아닌 정보에 대하여 「개인정보 보호법」을 어떻게 적용할 것인지가 불분명

4) 개정안은 ‘재식별’의 정확한 의미에 대해서 규정하지 않고 있고, 「개인정보 보호법」에서도 재식별에 대한 정의를 별도로 명시하고 있지는 아니하지만, 재식별이란 가명정보가 특정 개인을 알아볼 수 있는 정보(「개인정보 보호법」 제2조제1호 가목 및 나목의 개인정보)가 되는 것을 의미하는 것으로 이해됨.

한 것으로 보임.

- 셋째, 안 제15조의2제1항은 「개인정보 보호법」상 개인정보만 아니라면 인공지능사업자가 학습용데이터로 활용할 때 다른 개별법에 따른 규제 사항을 적용받지 않는 것이 허용되는 것으로 해석될 수 있어 이에 대한 검토가 필요함.
- 학습용데이터의 활용을 규제하기 위해 적용되는 개별법은 「개인정보 보호법」 외에도 제3자의 권리 보호와 관련된 「저작권법」, 「부정경쟁방지 및 영업비밀보호에 관한 법률」 등이 있고,
현행법 제5조는 다른 법률에서 특별한 규정이 있는 경우 해당 법률을 적용하도록 하고 있으므로, 개인정보에 해당하지 않는 학습용데이터를 활용하는 경우에도 「저작권법」 등 다른 개별 법률은 당연히 적용되는 것으로 해석하는 것이 타당함.
 - 예를 들어, 개인정보에 해당하지 않는 학습용데이터는 「개인정보 보호법」의 적용 대상이 아니지만, 저작권이 있는 저작물을 포함하는 경우에는 「저작권법」을 준수하여 활용하여야 하는 등 다른 법률에 따른 규율은 여전히 적용됨.
- 그런데, 안 제15조의2제1항과 같이 「개인정보 보호법」의 적용에 대해서만 별도로 규정하면 학습용데이터에 대해서 적용되어야 하는 다른 법률들의 적용에 대해서 수범자나 이해관계자의 해석상 혼란이 발생할 수 있음.

- 예컨대, 문화체육관광부와 한국방송협회는 안 제15조의2제1항에서 개인정보에 해당하지 않는 학습용데이터는 인공지능사업자가 활용할 수 있도록 규정하고 있는 것이 「저작권법」 및 「부정경쟁방지 및 영업비밀보호에 관한 법률」의 적용을 받지 않는 활용까지 허용하는 것으로 오인될 우려가 있다는 의견을 제시하였음.
- 넷째, 안 제15조의2제3항은 과학기술정보통신부가 개인정보가 아닌 학습용데이터의 활용 기준을 정하도록 하고 있는데, 오히려 이로 인해 인공지능사업자의 데이터 활용이 현재보다 제한될 가능성이 있음.
- 현재는 인공지능사업자가 개인정보가 아닌 학습용데이터를 활용하는 경우 「저작권법」, 「부정경쟁방지 및 영업비밀보호에 관한 법률」 등 개별법을 준수하면 별도의 기준 없이 데이터를 처리할 수 있음.
- 개정안과 같이 과학기술정보통신부장관이 학습용데이터의 활용 기준을 마련할 경우 인공지능사업자에게 데이터 활용에 관한 구체적이고 명확한 기준을 제시하여 불확실성을 해소하는데 기여하는 측면은 있을 것으로 보이지만,

현재 별도의 규제 없이 관계 법률에 따라 활용할 수 있는 데이터에 대하여 새로운 활용 기준을 마련할 경우 인공지능사업자의 데이터 활용에 추가적인 제한이 발생할 가능성도 있으므로, 이를 고려할 필요가 있는 것으로 보임.
- 다섯째, 현행법 제5조제1항에서 규정하는 “인공지능, 인공지능기술,

인공지능산업 및 인공지능사회”는 모두 분야 또는 영역을 의미하는 반면, 개정안에서 추가하려는 “학습용데이터의 생산·수집·관리·유통·활용”은 행위를 의미하므로 규정 대상의 성격이 서로 상이한바,

“학습용데이터의 생산·수집·관리·유통·활용”을 “학습용데이터”로 수정하는 방안을 검토할 필요가 있음.

- 여섯째, 안 제15조의2제1항은 인공지능사업자가 개인정보에 해당하지 아니하는 학습용데이터를 “수집·생성·가공·결합·이용 또는 제공”할 수 있도록 규정하고 있으나,

안 제5조제1항과 현행법 제15조 등에서는 “생산·수집·관리·유통 및 활용”이라는 용어를 사용하고 있어 학습용데이터를 대상으로 하는 주요 행위에 대한 용어를 서로 다르게 사용하게 됨.

- 동일한 개념을 규율하면서 용어를 달리 사용할 경우 해석상 혼란이 발생할 수 있으므로 조문에서 사용하는 용어는 서로 통일하는 것이 적절해 보임.

- 광의의 데이터 활용에는 데이터를 목적에 맞게 이용하기 위한 데이터의 가공·결합이 포함될 수 있으므로 현행법 조문에서 명시하는 “생산·수집·관리·유통 및 활용”으로 통일하는 방안을 검토할 필요가 있음.

다. 정부 및 관계 단체 의견

1) 과학기술정보통신부: 신중검토

- 민간 인공지능사업자의 법적 불확실성을 해소하려는 취지에는 일부 공감하나, 개인정보가 아닌 학습용데이터까지 재식별 위험 관리 기준, 활용 기준 등을 별도로 정하도록 법률에서 규정하면 민간부문의 학습용데이터 활용을 위축할 우려가 있어 신중한 검토가 필요함.

2) 개인정보보호위원회: 신중검토

- 현행 「개인정보 보호법」은 익명정보를 적용대상에서 제외하여 활용 기준, 안전조치 없이 자유롭게 활용할 수 있도록 규정 중인데, 개정안은 개인정보가 아닌 학습용데이터의 재식별 위험관리 기준을 정하도록 하여, 익명정보 활용을 위축시킬 우려가 있음.
- 또한, 학습용데이터가 특정 개인을 알아볼 수 있는 정보가 된 경우 개인정보로서 「개인정보 보호법」이 적용되므로 개정안과 같이 별도로 규정할 실익이 크지 않음.
- 아울러, 인공지능 개발·활용 과정에서의 개인정보 보호 및 재식별 위험 관리에 관한 안전장치는 「개인정보 보호법」 체계 내로 일원화되고 있는 상황으로, 「개인정보 보호법」 이외의 법률에서 인공지능 개발·활용과 관련하여 익명정보 및 재식별 위험 관리에 관한 별도 기준을 신설하는 것은 개인정보 보호 및 안전관리 체계의 파편화를 유발하고 법률 간 정합성을 저해할 우려가 있음.

- 참고로, 최근 법제사법위원회를 통과한(2026.7.29.) 「개인정보 보호법 일부개정법률안」(대안)⁵⁾은 익명·가명 처리로는 인공지능 기술 개발이 곤란한 경우 적법하게 수집한 개인정보의 원본 활용을 허용하는 ‘개인정보 처리 특례(안 제28조의12부터 제28조의15까지)’를 규정하여, 인공지능 학습을 위한 개인정보 이용 요건 및 위험요인 평가, 개인정보보호위원회의 주기적 관리·감독과 특례 위반 시 개인정보의 처리제한 등 데이터 활용과 안전 관리에 관한 통합적 규율체계를 구축하고 있음.

3) 문화체육관광부: 신중검토

- 안 제15조의2제1항은 「저작권법」 상 보호받는 저작물일지라도 개인정보가 아니라면, 인공지능사업자가 학습용데이터로 저작물을 활용하면서 저작권자의 허락없이 이용 가능하다는 오해의 소지가 있어 신중한 검토가 필요함.
- 안 제5조제1항의 경우, 안 제15조의2와 연계되어 학습용데이터의 생산·수집·관리·유통·활용 과정에서 다른 법률에 특별한 규정이 있는 경우를 제외하고 이 법에서 정한 바에 따를 경우 다른 법률에서 권리 보호 규정이 있음에도 해당 규정을 받지 않고 학습용데이터를 수집·생성·가공·결합·이용 또는 제공할 수 있는 것으로 오해 소지가 있어 신중한 검토가 필요함.

5) 의안번호 제2220246호

4) 한국방송협회: 신중검토

- 개정안과 같이 개인정보 해당 여부만을 기준으로 인공지능사업자의 학습용데이터 활용 가능성을 선언하는 입법 방식은 부적절함.
- 안 제15조의2제1항은 개인정보가 아니라는 이유만으로 저작권이 있는 데이터를 학습 데이터 생성을 위해 자유롭게 이용할 수 있다는 취지로 오독될 수 있고, 결과적으로 「개인정보 보호법」이 아닌 다른 법률이 보호하고 있는 권리를 실질적으로 침해하는 결과를 초래할 수 있으므로, 삭제하는 것이 타당함.
- 만약 인공지능 산업 진흥을 위해 해당 조항의 신설이 불가피하다면, 저작권 등 권리 침해에 대한 인공지능사업자의 면책 규정으로 악용되지 않도록 단서 규정⁶⁾이 반드시 포함되어야 함.

문 의 처
02)6788-5230

6) 다만, 이 경우에도 저작권법에 따라 보호받는 저작물 또는 부정경쟁방지 및 영업비밀보호에 관한 법률에 따른 영업비밀 제2조 제1항 파목의 '성과'에 해당하는 경우 인공지능사업자는 권리자에게 사전에 이용허락을 받아야 한다.