



제 439회 국회 (정기회)
제1차 과학기술정보방송통신위원회

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 일부개정법률안

검 토 보 고

〈인공지능 개발·활용 등 과정에서 성별 편향 방지〉

■ 정춘생의원 대표발의(의안번호 제2218268호)

2026. 9.

과학기술정보방송통신위원회

수석전문위원 이지민

목 차

I . 제안경위	1
II . 제안이유 및 주요내용	1
III . 검토의견	2
1. 인공지능 개발·활용 등 과정에서 성별 편향 방지	2

I. 제안경위

가. 발 의 자 : 정춘생의원 등 10인

나. 발 의 일 : 2026. 4. 13.

다. 회 부 일 : 2026. 4. 14.

II. 제안이유 및 주요내용

2026년 1월 22일부터 시행된 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」은 인공지능 윤리원칙, 고영향 인공지능에 대한 사업자의 책무와 영향평가에 관한 사항을 규정하고 있음.

그러나 인공지능시스템의 개발·활용 과정에서 발생할 수 있는 성별에 따른 편향 및 차별을 방지하기 위한 명시적 규정을 두고 있지 아니하여, 인공지능의 의사결정 과정에서 특정 집단에 대한 차별이 재생산될 우려가 있음.

이에 인공지능의 개발·활용 과정에서 성별 편향 등을 금지하도록 기본원칙에 명시하고, 고영향 인공지능에 대한 영향평가에 성별 편향 등의 발생여부에 대한 평가도 포함하며, 그 결과에 따라 필요한 시정 조치를 마련하도록 함으로써 인공지능으로 인한 차별을 방지하고 국민의 기본권을 실질적으로 보호하려는 것임(안 제3조제3항, 제27조제1항제4호 신설 및 제35조).

Ⅲ. 검토의견

1. 인공지능 개발·활용 등 과정에서 성별 편향 방지

가. 개정안의 주요 내용

- 개정안은 인공지능의 개발, 제공 및 이용 등에 대한 성별 편향 및 차별 방지 노력 의무를 신설하고, 정부가 제정·공표하는 인공지능 윤리 원칙에 포함하는 사항으로 인공지능의 개발·활용 등의 과정에서 성별 편향 등의 방지에 관한 사항을 추가함(안 제3조제3항 및 제27조제1항제4호 신설).
- 개정안은 고영향 인공지능에 대한 영향평가에 성별 편향 등에 관한 사항을 포함하도록 하고, 인공지능사업자가 그 결과에 따른 시정조치를 마련하도록 함(안 제35조제1항 후단 신설).

현행	개정안
제3조(기본원칙 및 국가 등의 책무) ①·② (생략) <u><신설></u>	제3조(기본원칙 및 국가 등의 책무) ①·② (현행과 같음) ③ <u>누구든지 인공지능의 개발, 제공 및 이용 등 모든 과정에서 성별을 이유로 한 편향 및 차별이 발생하지 아니하도록 노력하여야 한다.</u>
③ ~ ⑤ (생략)	④ ~ ⑥ (현행 제3항부터 제5항까지와 같음)

현행	개정안
제27조(인공지능 윤리원칙 등) ① 정부는 인공지능윤리의 확산을 위하여 다음 각 호의 사항을 포함하는 인공지능 윤리원칙(이하 “윤리원칙”이라 한다)을 대통령령으로 정하는 바에 따라 제정·공표할 수 있다.	제27조(인공지능 윤리원칙 등) ① -- ----- ----- ----- ----- ----- -----.
1. ~ 3. (생략) <u><신설></u>	1. ~ 3. (현행과 같음) 4. <u>인공지능의 개발·활용 등의 과정에서 성별 편향 등의 방지에 관한 사항</u>
②·③ (생략)	②·③ (현행과 같음)
법률 제21311호 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 제35조(고영향 인공지능 영향평가)	법률 제21311호 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 제35조(고영향 인공지능 영향평가)
① 인공지능사업자가 고영향 인공지능을 이용한 제품 또는 서비스를 제공하는 경우 사전에 사람의 기본권에 미치는 영향을 평가(이하 “영향평가”라 한다)하기 위하여 노력하여야 한다. 이 경우 영향평가에는 고영향 인공지능을 이용한 제품 또는 서비스의 성격을 고려하여 인공지능취약계층의 특성이 반영될 수 있도록 하여야 한다.	① ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- 하며, 성별 편향 등에 관한 사항을 포함하고 그 결과에 따라 필요한 시정조치를

현행	개정안
②·③ (생략)	<u>마련하여야 한다</u> ②·③ (현행과 같음)
부칙	
이 법은 공포 후 6개월이 경과한 날부터 시행한다.	

다. 검토의견

- 최근 개발·활용되고 있는 인공지능 시스템은 대량의 데이터를 기반으로 스스로 학습하고 결과를 도출하는 특성이 있어 개발자나 이용자의 의도와 무관하게 기존 사회에 존재하던 성별 고정관념이나 편향된 데이터가 반영될 가능성이 있음.
 - 특히, 인공지능 제품·서비스가 국민의 일상생활과 밀접한 영역에서 활용되고 있는 현실을 고려하면 인공지능시스템의 성별 편향성으로 인한 차별은 국민의 권익 침해와 사회적 불평등 심화로 이어질 수 있다는 점에서 이에 대한 관리 필요성이 있다고 볼 수 있음.
 - 이에, 개정안은 인공지능의 개발·활용 등의 과정에서 성별 편향으로 인한 사회적 문제가 발생하는 것을 예방하기 위하여 기본원칙 및 인공지능 윤리원칙에 성별 편향 방지에 관한 사항을 반영하고, 고영향 인공지능에 대한 영향평가 시에도 성별 편향 발생 여부를 평가하도록 하려는 것으로,
- 인공지능 기술의 확산 과정에서 발생할 수 있는 차별적 요소를 사전

에 관리할 수 있는 제도적 기반을 마련한다는 점에서 의의가 있는 것으로 보임.

- 다만, 개정안은 다음과 같은 사항을 고려할 필요가 있음.
- 첫째, 현행법에 따라 제정·공표된 인공지능 윤리원칙에는 이미 성평등에 관한 원칙이 포함되어 있고, 정부가 작성한 「인공지능 영향평가 가이드라인」에서도 영향평가 시 고려해야 할 기본권 중 ‘평등권’ 사항에 성별에 따른 차별적 요소를 배제하도록 이미 명시하고 있음.

<인공지능 윤리원칙>

< ③공정성·포용성(Fairness and Inclusiveness) >

- 인공지능으로 인한 편익과 부담이 특정 개인이나 집단에 편중되지 않아야 하며, 인공지능은 성별·연령·장애·언어·지역·종교·활용역량 등 개인이나 집단의 특성을 이유로 부당한 차별이나 불이익을 초래하지 않도록 개발·제공·이용되어야 한다. **[차별 금지 및 공정한 배분]**
- 인공지능에 활용되는 데이터와 알고리즘에 내재된 편향이 사회적 불평등이나 부당한 불이익을 재생산하지 않도록, 인공지능의 개발·제공·이용 전 과정에서 데이터·알고리즘 및 산출물에 대한 지속적인 점검과 개선이 이루어져야 한다. **[편향 최소화]**
- 이용자의 디지털 역량이나 환경적 제약과 관계없이 누구나 인공지능의 혜택에 동등하게 접근할 수 있어야 하며, 기술적 소외가 사회적 배제로 이어지지 않도록 포용적인 이용 환경이 보장되어야 한다. **[접근성 및 포용성 확보]**

공정성·포용성, 무엇을 고려해야 하나?

인공지능의 혜택과 부담은 모두에게 똑같이 나타나지 않을 수 있습니다. 과거의 차별이나 불평등이 데이터에 남아 있으면, 인공지능의 학습·활용 과정에서 이러한 경향이 반영되어 특정 개인이나 집단에 부당한 불이익으로 이어질 수 있습니다. 또한 디지털 기기나 서비스 이용이 어려운 집단·개인에게는 인공지능의 혜택이 충분히 닿지 않을 수 있습니다.

공정성·포용성 원칙은 인공지능으로 인한 편익과 부담이 특정 개인이나 집단에 편중되지 않도록 하고, 데이터·알고리즘·산출물의 편향을 지속적으로 점검·개선하도록 하는 기준입니다. 이를 통해 누구나 인공지능의 혜택에 동등하게 접근하고, 기술적 소외가 사회적 배제로 이어지지 않도록 하는 데 의미가 있습니다.

자료 : 과학기술정보통신부

<인공지능 영향평가 가이드라인

- 제품 또는 서비스 차원에서 영향받을 수 있는 기본권>

잠재적 영향 기본권	기본권 내용 및 침해 시나리오
평등권	본질적으로 동일한 것의 합리적 이유 없는 차별 금지 - 성별과 관련된 편향으로 인해 특정 성(性)에 불리한 결과 발생 - 인공지능 기반 신용평가 시스템이 대출 신청자를 거주 지역을 이유로 차등 평가 - 인공지능 기반 안면인식 출입 시스템이 특정 인종의 인식률이 낮아 추가 인증 요구 - 복지 부정수급 예측 시스템의 설계 및 운영상 편향으로 저소득층만 집중적으로 감시

자료 : 과학기술정보통신부, 2026.5.

- 즉, 현행법과 이에 따른 정부 지침으로도 인공지능 제품·서비스의 성별 편향을 방지하도록 노력하는 제도적 근거가 이미 일정 부분 마련되어 있는 상황으로 개정안을 통한 입법 실익이 크지 않을 것으로 보임.
 - 둘째, 인공지능 제품·서비스의 편향성으로 인한 사회적 위험이 발생할 수 있는 요소는 성별뿐만 아니라 장애인, 외국인, 취약계층 등 다수 있음에도 성별 편향에 관한 것만 법률에 조문으로 명시하는 것이 적절한지 검토가 필요해 보임.
 - 셋째, 현행법을 제정할 때 인공지능시스템으로 인한 사회적 위험을 관리하기 위하여 고영향 인공지능의 개념과 이에 대한 영향평가 제도를 도입하였으나,
- 국내 인공지능 산업 발전을 저해할 우려가 있어 영향평가를 사업자의 의무로 규정하지 않고 국가기관등이 영향평가를 받은 제품·서비스를

우선 고려하도록 하여 유인책을 마련하는 수준으로 정리하였음.

- 그런데, 개정안은 사업자가 고영향 인공지능 영향평가를 실시하는 경우 성별편향 등에 관한 사항을 포함하고 그 결과에 따라 필요한 시정 조치를 마련하도록 의무를 부과하려는 것인바, 영향평가를 인공지능사업자의 의무로 규정하지 아니한 현행법의 취지에 부합하지 않는 측면이 있음.

다. 정부 및 관계 단체 의견

1) 과학기술정보통신부: 수정동의

- 이미 현행 인공지능기본법에 따른 영향평가 사항 중 “사람의 기본권에 미치는 영향”에 성별 편향에 관한 사항을 포함하고 있더라도 인공지능 제품·서비스의 성별 편향으로 인한 사회적 문제를 예방할 중요성을 고려할 때 성별 편향에 관한 사항을 관련 규정에 명시적으로 추가할 필요가 있으나,

현행 인공지능기본법은 영향평가를 노력 의무로 규정하고 사업자의 자율성을 전제하고 있어 성별 편향에 대해 시정조치를 의무로 규정하는 내용은 제외하는 것이 바람직함.

2) 한국인터넷기업협회: 신중검토

- 개정안에 따른 성별 편향은 현재 법체계상 정의가 불명확하고 사회

적·기술적으로도 그 판단기준이 불분명하여 이에 대한 사업자 의무
부과 규정은 명확성 원칙에 위배될 우려가 높으므로,

성별편향 방지를 위한 조치는 사업자의 자율적 영역으로 남겨두는 것이
바람직하며 기본원칙·윤리원칙과 영향평가의 새로운 고려 요소로 법제
화하는 방향은 불명확성을 증가시킬 우려가 있어 재검토가 필요함.

문 의 처
02)6788-5230