의견서

발행일 2025. 12. 08.

인공지능기본법 고시 및 가이드라인에 대한 시민사회 의견서

일시 | 2025년 12월 8일 (월)

제출 | 디지털정의네트워크·민주사회를 위한 변호사모임 디지털정보위원회·정보인권연구소·참여연대·공공운수노조·무상의료운동본부·문화연대·미디어기독연대·민주주의법학연구회·보건의료단체연합·인권교육온다·전국교직원노동조합·전국금속노동조합·전국민주노동조합총연맹·시민사회단체연대회의·서울YMCA시민중계실·언론개혁시민연대 정책위원회·청소년인권운동연대 지음·표현의 자유와 언론탄압공동대책위원회·한국소비자연맹·한국여성민우회·

한국여성소비자연합

목차

들어가며

지난 2025년 11월 12일, 과학기술정보통신부는 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하, '인공지능기본법')」시행령 제정안(이하 시행령(안))을 입법예고하였다. 그에 앞서 2025년 9월 17일에는 시행령초안을 비롯하여 고시 및 가이드라인 초안을 공개하였다. 우리는 시행령(안)에 대한 의견서와 함께, 아래와 같이 고시 및 가이드라인 초안에 대한 의견서를 제출한다.

인공지능기본법과 시행령(안)의 맥락 하에서 고시 및 가이드라인이 만들어졌기 때문에, 고시 및 가이드라인에 대한 의견도 시행령(안)에 대한 의견과 연결될 수밖에 없고 시행령(안)과 함께 읽어야 한다. 또한, 의견서 제출 양식이 있었음에도 시민사회의 비판과 제안의 맥락을 충분히 설명하기 위해 별도의 문서로 의견서를 작성하였다. (이와 별개로 제출 양식에 따라 제출할 예정이다)

본 의견서에서 우리는 다음과 같은 사항의 개선을 권고하였다.

- 첫째, 안전성 확보 고시(안) 및 AI 안전성 확보 가이드라인(안)과 관련하여, 우선 본 고시와 가이드라인이 모든 인공지능을 대상으로 한 것이 아니므로 최첨단 인공지능 안전성 확보 고시 등으로 명확하게 규정할 것을 제안하였다. 또한 특정 AI 시스템을 대상으로 하고 있어 AI 모델 제공자가 수범 대상이 되는지 모호하다는 점, AI 시스템을 대상으로 할 경우 AI 모델의 고유한 위험성을 통제하기 힘들다는 점, 의무 대상인지 여부를 누가 판단할 것인지가 모호하다는 점을 지적하였다. 또한, 시행령(안)에서도 이미 비판하였지만, 고시와 가이드라인에서도 안전성 확보 의무 대상을 지나치게 협소하게 규정하고 있다는 점을 비판하였다. 이와 함께 과학기술정보통신부는 의무 대상 사업자의 목록을, 의무 대상 사업자는 자신의 위험 관리 체계 및 시스템에 대한 주요 정보를 공개하도록 할 필요가 있음을 제기하였다.

- 둘째, 사업자책무 고시(안) 및 고영향 AI 사업자 책무 가이드라인(안)과 관련하여, 인공지능 개발사업자와 이용사업자별로 책무가 다를 수 있는데 그 구분이 모호하고 자의적이라는 점을 지적하였다. 또한 시행령(안)의 문제와도 연결되는데, AI 시스템을 업무상 이용하는 사업자 역시 이용자로 규정하고 있어서 영향받는 자의 권리 보호와 이용자(또는 이용사업자)의 책무가 규정되지 않은 것은 큰 문제이다. 사업자 책무 가이드라인은 안전성 확보 가이드라인에 비해 그 내용이 구체적이지 않아 매우 부실하게 작성되어 있어, 전반적으로 재검토될 필요가 있다.

- 셋째, AI 투명성 확보 가이드라인(안)과 관련하여, 인공지능기본법은 인공지능개발사업자와 이용사업자에 의무를 부과하며, 단순히 인공지능제품·서비스를 이용한 결과물을 자신의 서비스 등에 활용하는 자는 인공지능기본법상 사업자에 해당하지 않아 투명성 확보 의무가 없다고 명시하고 있음. 이에 따라 YTN와 같은 뉴스사업자나 영화제작자 등 생성 인공지능 기술을 활용해 콘텐츠를 제작한 주제들은 아무런 표시 의무가 없고, 생성형 AI 서비스를 제공하는 사업자는 표시할 수 있는 기능만 제공하면 되기 때문에, 실제 표시 의무를 수행해야 할 주체가 없어지게 됨. 이용자로 명시되어 있는 사업자라고 하더라도 인공지능 개발 및 이를 바탕으로 하는 자신의 고유의 서비스 제공을 위하여 주도권을 가지고 인공지능을 활용하는 경우에는 단순 '이용자'가 아닌

인공지능이용사업자로 규율되어야 할 것임.

- 넷째, 고영향 AI 판단 가이드라인(안)의 경우, 전반적으로 영향 인공지능의 개별 목록의 의미에 대하여 문언과 다르게 임의로 축소 해석하거나, 추가적인 제한 요건을 부가하는 등의 방식으로 고영향 인공지능의 범주를 최소화하려 하고 있는데, 이는 상위법에 위배될 소지가 있고, 기본권 보호의 측면에서도 바람직하지 않다.
- 다섯째, AI 영향평가 가이드라인(안)에 대해서는 기본권 영향평가로 이름을 변경할 것, 평가 수행 주체의 독립성과 전문성 요건을 규정할 것, 국가인권위원회와의 협력을 규정할 것, 사전 준비 단계에서 평가팀의 준비단계도 검토할 것, 이해관계자 및 AI 시스템이 사용되는 사회적 맥락에 대한 파악의 필요성, 영향받는 자와의 협의를 필수적인 절차의 하나로 명시할 것, 영향받는 기본권을 제한하지 말 것 등 다양한 개선을 권고하였다.

안전성 확보 고시(안) 및 AI 안전성 확보 가이드라인(안)

1. 관련 법률 및 시행령(안)

법	시행령(안)
제32조(인공지능 안전성 확보 의무) ① 인공지능사업자는 학습에 사용된 누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템의 안전성을 확보하기 위하여 다음 각 호의 사항을 이행하여야 한다. 1. 인공지능 수명주기 전반에 걸친 위험의 식별·평가 및 완화 2. 인공지능 관련 안전사고를 모니터링하고 대응하는 위험관리체계 구축 ② 인공지능사업자는 제1항 각 호에 따른 사항의 이행 결과를 과학기술정보통신부장관에게 제출하여야 한다. ③ 과학기술정보통신부장관은 제1항 각 호에 따른 사항의 구체적인 이행 방식 및 제2항에 따른 결과 제출 등에 필요한 사항을 정하여 고시하여야 한다.	제23조(인공지능 안전성 확보 의무) ① 법 제32조제1항에서 “대통령령으로 정하는 기준 이상인 인공지능시스템”이란 학습에 사용된 누적 연산량이 10의 26승 부동소수점 연산 이상인 인공지능시스템으로서 과학기술정보통신부장관이 인공지능기술 발전 수준, 위험도 등을 고려하여 고시하는 기준에 해당하는 인공지능시스템을 말한다. ② 제1항에 따른 학습에 사용된 누적 연산량의 구체적인 산정 방식은 과학기술정보통신부장관이 정하여 고시한다.

2. 인공지능 안전성 확보 의무의 대상이 모호함.

(1) 안전성 확보 의무 대상

안전성 확보 의무는 모든 인공지능 시스템을 대상으로 하고 있지 않음. 법 제32조는 안전성 확보 의무의 대상을 ‘학습에 사용된 누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템’으로 정하고 있음. 시행령 제23조는 학습에 사용된 누적 연산량을 10의 26승 부동소수점 연산(이하 FLOPS) 이상으로 규정하면서, 이에 더하여 ‘과학기술정보통신부장관이 인공지능기술 발전 수준, 위험도 등을 고려하여 고시하는 기준에 해당’할 것을 요구하고 있음. 안전성 확보 고시(안)은 제3조 1항은 대상 인공지능시스템의 요건으로 2항 각호의 기준을 모두 만족해야 한다고 규정하고 있음. 이에 따르면, 안전성 확보 의무 대상은 최첨단 인공지능기술이 적용된 인공지능시스템으로서 안전과 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는 경우를 의미함.

법 제32조(인공지능 안전성 확보 의무) ① 인공지능사업자는 학습에 사용된 누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템의 안전성을 확보하기 위하여 다음 각 호의 사항을 이행하여야 한다. 1. 인공지능 수명주기 전반에 걸친 위험의 식별·평가 및 완화 2. 인공지능 관련 안전사고를 모니터링하고 대응하는 위험관리체계 구축 ② 인공지능사업자는 제1항 각 호에 따른 사항의 이행 결과를 과학기술정보통신부장관에게 제출하여야 한다. ③ 과학기술정보통신부장관은 제1항 각 호에 따른 사항의 구체적인 이행 방식 및 제2항에 따른 결과 제출 등에 필요한 사항을 정하여 고시하여야 한다.
시행령(안) 제23조(인공지능 안전성 확보 의무) ① 법 제32조제1항에서 “대통령령으로 정하는 기준 이상인 인공지능시스템”이란 학습에 사용된 누적 연산량이 10의 26승 부동소수점 연산 이상인 인공지능시스템으로서 과학기술정보통신부장관이 인공지능기술 발전 수준, 위험도 등을 고려하여 고시하는 기준에 해당하는 인공지능시스템을 말한다. ② 제1항에 따른 학습에 사용된 누적 연산량의 구체적인 산정 방식은 과학기술정보통신부장관이 정하여 고시한다.
안전성 확보 고시(안)(이하 안전성 고시안) 제3조(적용 대상) ① 이 고시에서 정하는 안전성 확보 조치는 제2항 각 호의 기준을 모두 충족한 인공지능시스템에 적용한다. ② 영 제23조제1항의 인공지능기술 발전 수준, 위험도 등을 고려하여 고시하는 기준이란 다음 각 호와 같다. 1. 학습에 사용된 누적 연산량이 10의 26승 부동소수점연산 이상인 경우 2. 인공지능기술의 발전 수준을 고려할 때 현재 인공지능시스템에 활용되는 인공지능기술 중 최첨단의 인공지능기술을 적용하여 구성·운영되고 있는 경우 3. 인공지능시스템의 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는 경우

(2) 최첨단 인공지능에 대한 유럽 및 미국의 규율

우선 이와 관련된 유럽과 미국의 규율을 간략히 살펴보면 다음과 같음. EU AI Act는 범용 AI 모델에 대한 정의 규정을 두고 이를 5장에서 규율하고 있음.[1] ‘범용 AI 모델’이란, 대규모 자기지도(self-supervision) 학습을 통해 대량의 데이터로 훈련된 경우를 포함하여, 상당한 범용성을 보이고 시장에 출시되는 방식과 관계없이 광범위한 여러 가지 작업을 유능하게 수행할 수 있으며, 다양한 하위(downstream) 시스템 또는 애플리케이션에 통합될 수 있는 AI 모델(Article 3(63))을 의미함. EU의 범용 AI 모델 가이드라인(Guidelines for GPAI Models)[2]에 따르면, 누적연산량이 10의 23승 FLOPS 이상이고 텍스트, 오디오, 이미지, 비디오 등을 생성할 수 있는 모델이 범용 AI 모델로 간주됨. 그러나 10의 23승 FLOPS 이상이더라도 범용성이 없는 전문화된 모델은 범용 AI 모델에서 제외됨. 즉, 단순히

[1] EU AI Act는 2024년 7월 12일에 EU 공식 저널에 출판되었고, 2024년 8월 1일부터 발효되었다. 그러나 AI Act 조항이 실제로 적용되는 것은 단계적으로 이루어진다. 2025년 2월 2일, 금지된 AI 시스템 규정과 AI 리터러시 조항의 적용이 시작되었고, 2025년 8월 2일부터 통지된 기관(Notified bodies :Chapter III, Section 4), 범용 AI 모델(Chapter V), 거버넌스(Chapter VII), 기밀성(Article 78), 벌칙(Articles 99 and 100) 조항의 적용이 시작되었다.

[2] 유럽연합 집행위원회(EC)는 2025년 7월 18일, 범용 AI 모델 가이드라인(Guidelines for GPAI Models) 초안을 발표하였는데, 이 가이드라인은 범용 AI 모델에 적용되는 조항의 해석을 명확하게 하기 위한 목적이다. 이 가이드라인은 범용 AI의 정의 및 범위, 관련 수명 주기 의무, 시스템적 위험의 기준, 제공자의 통지 의무에 대해 어떻게 해석해야 하는지에 대한 지침을 제공한다. https://digital-strategy.ec.europa.eu/en/policies/guidelines-gpai-providers

FLOPS를 기준으로 규율하는 것이 아닌데 이는 '범용 AI 모델'에 고유한 위험성이 있다고 보기 때문임.

AI Act는 51조 1항에 따라 고영향 역량(high-impact capability)을 가진 범용 AI 모델을 '시스템적 위험을 가진 범용 AI 모델(General-Purpose AI Models with Systemic Risk)로 구분하고, 더욱 엄격하게 규율하고 있음. 일반적으로 범용 AI 모델 제공자에 대해서는 기술문서의 작성 및 보관, AI 시스템 제공자에 대한 정보 제공, 저작권 보호 정책의 수립, 훈련 데이터에 대한 충분히 상세한 요약의 공개 등의 의무를 부과하는데(Article 53), 시스템적 위험을 가진 범용 AI 모델 제공자에 대해서는 모델 평가, 시스템적 위험에 대한 평가 및 완화, 심각한 사고 및 조치에 대해 관련 당국에 보고, 사이버 보안 보장 등의 추가적인 의무를 부과함.(Article 55) 훈련에 사용된 누적 연산량이 10의 25승 FLOPS인 범용 AI 모델은 시스템적 위험이 있는 AI 모델로 간주됨.(Article 51(2)) 다만, 해당 시스템이 이러한 기술적 조건을 만족시키지만, 시스템적 위험을 가지지 않는다는 점을 범용 AI 모델 제공자가 입증할 경우 제외가 될 수 있음.(Article 52)

트럼프 2기 정부에서 폐기되었지만, 2023년 10월 30일 바이든 행정부가 발표한 인공지능 행정명령(Executive Order 14110)은 이중용도 파운데이션 모델(dual-use foundation model)에 대한 규율을 포함하고 있음.[3] 여기서 '이중용도 파운데이션 모델'이란 광범위한 데이터로 훈련되고, 일반적으로 자기지도(self-supervision) 학습을 사용하며, 최소 수백억 개 이상의 파라미터를 포함하고, 광범위한 맥락에 적용 가능하며, 안보, 국가 경제 안보, 국가 공중 보건 또는 안전, 혹은 이들의 조합에 심각한 위험을 초래하는 과업에서 높은 수준의 성능을 나타내거나 또는 쉽게 수정하여 그러한 성능을 나타낼 수 있는 AI 모델을 의미함. 행정명령은 이중용도 파운데이션 모델을 개발하는 사업자들은 연방정부에 해당 모델의 훈련, 개발, 생산에 관련된 정보, 소유권, 모델 성능 등에 대한 정보를 제공하도록 하고 있으며, 이러한 모델에 해당하기 위한 기술적 요건을 설정할 것을 명령하면서 잠정적으로 10의 26승 FLOPS 이상의 누적 연산량을 기준으로 설정했음.

(3) 안전성 확보 의무 규정의 문제점

안전성 확보 의무 규정은 해외의 이러한 규정을 참고한 것으로 보이나, 그 의무 대상의 규정에 있어서 다소 모호한 점이 있음. 우선 법령에서 의무 대상의 성격을 규정할 필요가 있음. EU의 경우 (시스템적 위험이 있는) 범용 AI 모델, 미국의 경우에는 '이중용도 파운데이션 모델'로 규제의 대상을 규정하고 있는데, 인공지능 기본법에서는 단지 '학습에 사용된 누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템'으로 규정하고 있음. 조항의 명칭도 '인공지능 안전성 확보 의무'로 되어 있고, 고시의 이름도 '안전성 확보 고시'로 되어 있지만, 사실 내용적으로는 모든 인공지능 시스템의 안전성 의무에 대해 규율하고 있는 것은 아님. 그 취지는 유럽이나 미국과 같이 대규모 범용 AI 모델, 파운데이션 모델, 또는 최첨단 AI 모델을 규율하려는 것으로 보임. 물론 이러한 대규모 AI 모델이 아니더라도 모든 AI 모델 또는 시스템은 일정한 안전성 확보 조치가 필요하지만, 대규모 범용 AI 모델 고유의 위험성이 존재하고 그 영향이 심각하기 때문에 별도의 규율을 하고 있는 것임. 단지 인공지능 안전성 확보 의무, 안전성 확보 고시라고 하는 것은 일반적인 안전성 의무로 오해될 소지가 있기 때문에, 현재 법률을 개정하는 것은 힘들다고 할지라도

---

[3] https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

시행령 및 고시에서는 규율하고자 하는 목적과 대상에 맞게 명확하게 명칭을 규정할 필요가 있음.

둘째, EU와 미국에서는 AI 모델 제공자를 대상으로 하고 있는 반면, 인공지능 기본법에서는 인공지능 시스템을 대상으로 하고 있음. AI 모델은 AI 시스템의 핵심적인 부분일 수 있지만, AI 모델이 시스템이 되기 위해서는 이용자 인터페이스 등 다른 요소들이 필요함.(EU AI Act recital 97) EU AI Act의 경우 AI 시스템에 대한 규율과 범용 AI 모델에 대한 규율을 구분하고 있는데, 국내 인공지능 기본법은 그렇지 않음. 그래서 AI 모델 제공자와 시스템 제공자가 같은 사업자일 경우에는 큰 문제가 아닐 수 있겠지만, 범용 AI 모델이 다른 사업자를 통해 다양한 시스템에 통합될 경우 의무의 주체가 누구인지의 문제가 발생할 수 있음. 우리나라 인공지능 법에 따르면, 이 의무대상 인공지능 시스템의 사업자가 해당 AI 모델 개발자가 아니더라도 이 의무를 이행하는 주체가 될 수 있음.

안전성 고시안 제3조 2항은 의무대상 인공지능 시스템이 다른 사업자에게 제공되는 경우를 다루고 있는데, 애초에 다른 사업자에 제공하는 것을 인공지능 시스템으로 보고 있다는 점에서, AI 모델 제공자와 시스템 제공자가 다를 경우의 문제는 여전히 남아있음.

> 안전성 고시안 제3조(적용 대상)
> ③ 법 제32조제2항에 따라 이행 결과를 제출하여야 하는 인공지능사업자는 다음 각 호와 같다.
> 1. 인공지능시스템이 개발되어 타인에게 제공되는 시점부터 제1항에 따른 인공지능 시스템에 해당하는 경우 : 해당 인공지능시스템을 개발한 인공지능개발사업자
> 2. 제1항에 해당하지 않는 인공지능시스템에 대해 실질적으로 변경을 가하여 제1항 해당하게 한 경우 : 해당 변경을 가한 인공지능사업자

특히, 범용 AI 모델이 오픈소스로 배포되는 경우 문제가 될 수 있음. 즉, 범용 AI 모델이 오픈소스로 공개되고 이에 기반하여 다른 사업자가 범용 AI 시스템을 만들 경우 누가 의무를 이행할 것인가. EU AI Act의 경우 오픈소스로 배포하더라도 시스템적 위험이 있는 범용 AI 모델 제공자의 경우 안전성 조치 의무(EU AI Act Article 55)를 이행해야 함. 그러나 우리 인공지능 기본법은 오픈소스로 공개된 AI 모델에 대한 규정 자체가 없고, 안전성 조치 의무 조항에 따르면 오픈소스로 공개된 범용 AI 모델을 이용하여 AI 시스템을 개발한 사업자가 의무 이행의 주체가 될 수 있음.

한편, 안전성 가이드라인에 따르면, 의무 적용 대상 여부는 모델 단위가 아니라 시스템 단위임을 설명하고 있음. 동일 모델이라도 시스템 구성에 따라 다른 인공지능시스템으로 간주되어 적용 대상 여부가 달라질 수 있기 때문임.

> ㅇ [판단 단위] 적용 대상 여부는 모델 단위가 아니라 시스템 단위에서 판단한다.
> - (의미 해석) 동일 모델이라도 시스템 구성에 따라 다른 인공지능시스템으로 간주되어 적용 대상 여부가 달라질 수 있다.
> - (고려 사항) 사업자는 모델 구조뿐만 아니라, 시스템의 목적, 설계, 입출력 방식, 기능적 결합관계 등을 종합적으로 고려하여 새로운 인공지능시스템으로 간주할지 여부를 판단한다.

이처럼 의무 이행 주체를 시스템을 제공하는 사업자로 할 경우, 자신이 직접 개발한 것이 아니라 제공받은 AI 모델에 기반하여 시스템을 개발했을 경우, 해당 사업자가 고시에서

규정한 안전성 의무를 모두 수행할 수 있는지 여부가 문제가 될 수 있음. 경우에 따라서는 AI 모델 제공자의 협조가 필요할 수 있는데, 그렇다면 고시 및 가이드라인에서 이러한 협력 의무에 대해서도 규정할 필요가 있음.

EU가 시스템이 아니라 범용 AI 모델을 규제 대상으로 한 이유는 특정 용도 및 맥락에 따라 위험성의 정도가 달라지는 시스템의 성격(그래서 시스템에 대한 규율도 금지, 고위험 등으로 위험 수준에 따라 규제를 다르게 하고 있다)과 별개로 범용 AI 모델 자체가 갖는 위험성이 있고, 모델 자체의 위험성이 이 모델을 기반으로 만들어지는 수많은 하위 시스템들 전체에 영향을 미친다고 보기 때문임. 또한 시스템 제공자가 API 등을 통해 모델을 활용할 경우 내부 파라미터, 훈련 데이터, 모델 구조 등 위험 평가 및 통제를 위한 요소에 접근하기 힘들 수 있기 때문임. 국내 인공지능 기본법은 범용 AI 모델 자체에 대해서는 규율하지 않고 있는데, 하위 법령에서 이 문제를 근본적으로 해결하는 것은 힘들지라도, 범용 AI 모델에 대한 규율의 부재가 미치는 영향에 대해 검토하고 개선 방안을 마련할 필요가 있음.

셋째, 의무 대상 인공지능 시스템인지 여부를 누가 판단할 것인가의 문제가 모호함. 안전성 고시안 제3조 제2항은 의무 대상 인공지능 시스템의 기준을 정하고 있는데, '인공지능기술의 발전 수준을 고려할 때' '최첨단의 인공지능 기술을 적용하여 구성·운영되고 있는'지 여부나 '위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는'지 여부를 누가 판단할 것인가. 만일 사업자의 판단과 규제기관의 판단이 다를 경우 어떻게 할 것인가.

판단 기준과 관련하여 가이드라인은 다음과 같이 규정하고 있음.

> ② 최첨단의 인공지능기술 판단 요건
> • (판단 시점) 인공지능시스템이 최첨단 인공지능기술을 적용하여 구성·운영되는지 여부는 학습 착수 또는 시스템 설계 단계에서 판단한다.
> • (판단 기준) 사업자는 시스템의 아키텍처, 파라미터 규모, 학습 데이터 특성, 활용목적, 성능 목표, 인공지능 분야의 기술발전수준 등을 종합적으로 고려하여 최첨단 기술 적용 여부를 판단한다.
> • (객관적 근거) 사업자는 공개 벤치마크, 국제 평가 지표, 주요 연구기관·기업의 기술 보고서 등 객관적 자료를 근거로 인공지능시스템의 기술 수준을 검토해야 한다.
> • (요건 충족) 위 검토 결과, 해당 시스템이 최첨단 인공지능기술을 적용하고 있다고 인지되는 것으로 합리적으로 판단되는 경우, 요건을 충족한 것이다.
>
> ③ 중대한 위험성 존재 여부 판단 요건
> • (판단 시점) 인공지능시스템의 위험성은 학습 착수 또는 시스템 설계 단계에서 위험 식별 및 평가 절차를 통해 구조적으로 판단한다.
> • (판단 기준) 사업자는 시스템의 사용 목적, 적용 분야, 오용 가능성(합리적으로 예측가능한 수준을 의미), 타 산업 및 기존 인공지능시스템의 유사 사례, 잠재적 피해 범위 및 심각도 등을 종합하여 생명·신체·기본권에 미칠 영향을 평가한다.
> • (사전 검토) 평가 이전이라도 중대한 위험 가능성이 예상되는 경우, 정부기관 자문 요청 또는 통지 절차를 준비하는 것이 바람직하다.
> • (요건 충족) 위험 평가 결과, 특정 영역에 한하는 것이 아닌, 광범위한 영역에 걸쳐 사람의 생명·신체·기본권에 중대한 영향을 미칠 우려가 있다고 합리적으로 판단되는 경우, 요건을 충족한 것으로 본다.

중대한 위험성 존재 여부 판단 요건과 관련하여 '정부기관 자문 요청 또는 통지 절차를 준비하는 것이 바람직'하다고 하고 있지만, 사업자와 규제 기관의 관점이 다를 경우 어떻게 할 것인지에 대해서는 여전히 모호함. 이와 관련하여 EU AI Act는 범용 AI 모델이 51조(a)의 요건, 즉 10의 25승 FLOPS 이상의 요건을 만족하면 시스템적 위험을 가진 범용 AI 모델로 일단 '간주'되고, 이를 EC에 고지하도록 하고 있음. 다만 제공자가 이러한 요건에도 불구하고 시스템적 위험이 없음을 입증한 경우 의무 적용이 면제될 수 있음. 또한 EC가 직권으로 또는 과학 패널의 권고에 따라 특정한 모델을 시스템적 위험을 가진 범용 AI 모델로 지정할 수도 있음.

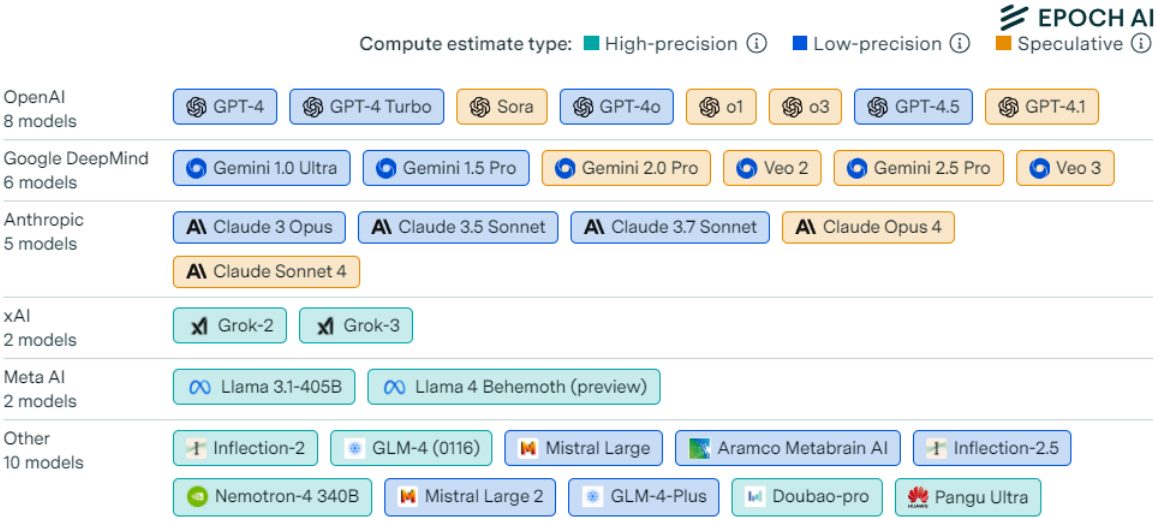
우리도 이와 유사하게 특정한 기준을 넘으면 안전성 조치 의무 대상으로 간주하고, 안전과 인권에 미치는 중대한 위험이 없다는 것을 사업자가 입증할 경우 의무를 면제하며, 사업자가 자발적으로 신고하지 않고 사후적으로 알게된 경우 과기정통부가 의무 대상으로 지정하도록 할 필요가 있음. 이를 가능한 법에서 규정을 하는게 바람직하겠지만, 우선 안전성 고시에라도 포함할 필요가 있을 것임.

3. 인공지능 안전성 확보 의무의 대상이 매우 협소함

시행령(안) 제23조는 안전성 확보 의무 대상을 너무 협소하게 규정하고 있음. 학습에 사용된 누적 연산량이 10의 26승 FLOPS 이상이어야할 뿐만 아니라, 과학기술정보통신부장관이 인공지능기술 발전 수준, 위험도 등을 고려하여 고시하는 기준에 해당할 것을 요구하고 있는데, 고시안 제3조 제2항은 최첨단의 인공지능기술을 적용하여 구성·운영되고, 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는 경우로 제한하고 있음.

우선 '학습에 사용된 누적 연산량이 10의 26승 FLOPS 이상'이라는 규정은 너무 엄격해서 현재 이에 해당하는 AI 시스템이 거의 없을 것으로 보임. 이는 규제의 외양만 갖추었을 뿐 현실적으로 규제의 효과를 거의 가질 수 없음. EU의 경우 10의 25승 FLOPS 이상의 범용 AI 모델을 시스템적 위험이 있는 범용 AI 모델로 간주하는데, 이는 현재 최첨단 AI 모델과 서비스를 대부분 포괄할 수 있는 기준임. 사업자들이 정확한 데이터를 공개하지 않아 추정치이기는 하지만, 2025년 6월 현재 10의 25승 FLOPS 이상의 누적 연산량으로 훈련된 모델이 30 여개 되는 것으로 보이며, 10의 26승 FLOPS 이상의 모델은 몇 개 정도에 불과한 것으로 보임.

출처 : https://epoch.ai/data/ai-models



출처 : https://epoch.ai/data-insights/models-over-1e25-flop

또한, EU의 경우 10의 25승 FLOPS 이상의 기준은 '시스템적 위험을 가진 범용 AI 모델'로 추정하는 기준일 뿐이며, 10의 25승 FLOPS 기준을 만족하지 않는 범용 AI의 모델의 경우에도 EC가 고영향 역량(high-impact capability)을 가졌다고 판단할 경우 '시스템적 위험을 가진 범용 AI 모델'로 지정될 수 있음. 이는 매우 합당한 방식인데, 누적 연산량 자체가 중요한 것이 아니라 해당 시스템이 실제 고영향 역량을 가지고 있는지, 그래서 안전과 인권에 시스템적 영향을 미칠 수 있는지 여부가 중요한 것이기 때문임. 그런데, 우리 고시에서는 10의 26승 FLOPS 기준과 함께 최첨단의 인공지능기술을 적용하여 구성·운영된다는 기준, 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있다는 기준을 '모두' 충족해야만 안전성 확보 조치 의무 대상이 되도록 규정하고 있음. 이는 이 규율의 취지를 망각한 지나치게 엄격한 기준이라고 할 수밖에 없음.

따라서, 안전성 확보 고시 제3조 제1항의 적용대상은 제2호 각 호의 기준을 '하나라도' 충족하는 시스템에 모두 적용할 필요가 있음.

- 2항 1호의 기준인 '학습에 사용된 누적 연산량이 10의 26승 부동소수점연산 이상인 경우'는 신속하고 객관적인 판단을 위한 기준 역할을 할 수 있음. (물론 10의 25승으로 기준 자체는 완화되어야 함.) 다만, 이 기준을 충족하지만 안전과 인권에 미치는 영향이 중대하지 않다는 점을 사업자가 입증할 경우, 적용 대상에서 배제할 수 있을 것임.
- 2항 2호의 기준 '최첨단의 인공지능기술을 적용하여 구성·운영되고 있는 경우'는 최첨단 인공지능기술의 특성상 안전성에 대한 검증이 미흡하고 불확실성이 크다는 점에서 안전성 확보 조치 의무 대상으로 삼을 필요가 있음.
- 2항 3호의 기준 '인공지능시스템의 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는 경우'는 기술적 기준과 무관하게 당연히 안전성 확보 조치 의무 대상이 되어야 할 것임.

4. 인공지능 안전성 확보 의무의 대상의 공개

투명성은 인공지능 사업자, 특히 위험성이 큰 인공지능 시스템을 개발, 제공하는 사업자의 책임성을 높이기 위한 중요한 요건임. 그러나 전반적으로 국내 인공지능 기본법이 사업자에 요구하는 투명성 요구 수준은 매우 미흡함. EU AI Act의 경우 범용 AI 모델 제공자의 범위는 훨씬 넓은데, 기본적으로 누적 연산량이 10의 23승 FLOPS 이상인 모델 제공자에 적용됨. 범용 AI 모델 제공자는 기본적으로 기술문서의 작성 및 보관, AI 시스템 제공자에 대한 정보 제공, 저작권 보호 정책의 수립, 훈련 데이터에 대한 충분히 상세한 요약의 공개 등의 의무를 부과받는데, 시스템적 위험을 가진 범용 AI 모델 제공자도 당연히 이러한 의무의 수범자가 됨. 이들은 이러한 의무에 더하여 안전성 확보 의무를 부담해야 하는 것임. 그러나 국내 인공지능 기본법은 범용 AI 모델 및 시스템에 대한 규정이 없고, 이들의 투명성 의무 역시 규정하지 않고 있음. 그럼에도 불구하고, 안전성 확보 의무 대상으로 지정된 인공지능 시스템의 경우에는 안전과 인권에 중대한 영향을 미치기 때문에, 이에 대한 일정한 정보를 일반 대중에 공개할 필요가 있음.

EU AI Act의 경우 EC로 하여금 시스템적 위험을 가진 범용 AI 모델의 목록을 (지적재산권 및 영업비밀을 보호하면서) 공개하도록 하고 있음. (Article 52(6)) 전술한 바와 같이 시스템적 위험을 가진 범용 AI 모델 제공자는 훈련 데이터의 상세한 요약을 공개해야 함. 시스템적 위험을 가진 범용 AI 모델 제공자가 자발적으로 채택하는 행동 강령(Code of Practice)의 안전 및 보안 챕터(Safety and Security Chapter)는 사업자들의 자신의 안전 및 보안 프레임워크와 모델 보고서의 요약본을 웹사이트 등을 통해 공개하도록 하고 있음.[4]

[4] EU AI 사무국은 2025년 7월 범용 AI 행동강령을 발표하였고, EC와 AI 위원회(AI Board)가 이 행동강령의 적절성을 승인하였다. 이 행동강령은 투명성(Transparency), 저작권(Copyright), 안전 및 보안(Safety & Security) 등 세 개의 챕터로 구성되어 있다. 투명성 챕터는 범용 AI 모델과 관련한 정보의 포괄적인 문서화 및 공개 의무를 다루고 있는데, 모델 문서화 양식, 상세한 라이선스, 기술적 사양, 사용 사례, 데이터셋, 연상 및 에너지 사용량 등의 내용을 포함하고 있으며, 문서는 최소한 10년 이상 보관해야 하고, AI 사무국 및 하위 시스템 사업자에게 제공되어야 하며, 일부 정보는 일반에 공개할 것이 권장된다. 저작권 챕터는 저작권법의 준수를 위한 내부 정책과 책임에 대해 규정하고 있다. 투명성 및 저작권 챕터는 모든 범용 AI 모델 제공자에 적용된다. 반면, 안전 및 보안 챕터는 시스템적 위험을 가진 범용 AI 모델 제공자에 적용되며, 제목 그대로 이들이 취해야 할 안전 및 보안 조치에 대해 다루고 있다.
이러한 행동강령은 시스템적 위험을 가진 범용 AI 모델 제공자가 법적 의무를 충족하기 위해 따를 수 있는 표준 절차나 모범 관행을 규정한 것으로 우리의 '안전성 확보 가이드라인'과 유사한 성격을 가진다. 사업자들은 이 행동강령에 서명하고 자발적으로 준수를 하지만, 행동강령을 따를 경우 감독기관인 AI 사무국(AI Office)가 법을 준수한 것으로 보는 효과를 얻을 수 있다. 물론 이를 따르지 않아도 되지만, 이 경우 문제가 발생했을 때 사업자는 자신이 법을 준수하고 있다는 사실을 다른 방식으로 입증해야 한다.

10.2 조치: 공개 투명성
시스템적 위험을 평가 및/또는 완화하는 데 필요한 경우, 서명자들은 자신들의 프레임워크(약속 1에 따름) 및 모델 보고서(약속 7에 따름)의 요약본을 웹사이트 등을 통해 공개해야 한다. 업데이트된 내용도 포함해야 한다.

공개되는 요약본에는 시스템적 위험 평가 결과에 대한 개략적인 요약과, 안전 및 보안 완화 조치에 대한 개략적인 설명이 포함되어야 한다.

어떤 정보든 안전 및/또는 보안 완화 조치의 효과를 저해하거나 민감한 상업 정보를 위태롭게 할 수 있는 정보는 공개에서 제외하거나 수정(가림 처리)해야 한다.

그러나 우리 시행령(안)이나 안전성 조치 의무에 대한 고시 및 가이드라인에는 이와 같은 의무를 규정하고 있지 않음. 우리도 안전과 인권에 중대한 영향을 미칠 수 있는 최첨단 AI 시스템 제공사업자에 대한 책무성을 높이기 위해서는 이러한 공개 의무를 강화할 필요가 있음. 과학기술정보통신부는 의무 대상 사업자를 공개할 필요가 있으며, 의무 대상 사업자는 자신의 위험 관리 체계 및 시스템에 대한 주요 정보를 공개하도록 할 필요가 있음.

5. 안전성 고시안 및 가이드라인에 대한 기타 의견

국내 안전성 확보 고시 및 가이드라인에서 기타 개선을 검토할만한 점을 살펴보면 다음과 같음.

- 안전성 고시안 제3조 제6항은 “과학기술정보통신부장관은 필요한 경우 인공지능사업자가 제출한 누적 연산량 관련 자료의 검증을 인공지능사업자의 동의를 받아 전문기관에 의뢰할 수 있다.”고 하고 있는데, 인공지능사업자의 동의를 받아 전문기관에 의뢰할 수 있다고 규정하는 것은 적절하지 않음. 동의하지 않으면 전문기관 의뢰를 통해 검증할 수 없는 것인가? 인공지능사업자는 감독기관의 조사에 협력할 의무가 있음.

- 안전성 고시안 제5조 제2항은 “② 인공지능사업자는 위험을 평가하기 위하여 위험평가조직을 구성하여야 하며, 필요한 경우 외부 기관 또는 전문가가 참여하게 할 수 있다.”고 하고 있는데, 외부 기관 또는 전문가에 포함될 수도 있겠지만, 해당 인공지능 시스템으로부터 영향을 받는 사람이나 이들을 대리할 수 있는 단체가 위험평가에 참여할 수 있도록 보다 명시적으로 규정할 필요가 있음. 안전 및 인권에 중대한 영향을 미치는 인공지능 시스템의 위험을 평가하는 것인만큼 해당 시스템의 영향을 받는 사람의 관점이 고려될 필요가 있기 때문임.

- 안전성 고시안 제9조는 매 3년마다 본 고시의 타당성을 검토하도록 하고 있는데, 통상적으로는 3년이 적정할 수 있지만 인공지능 기술의 급속한 발전을 고려할 때, 당분간 매년 고시의 타당성을 검토하고 개선하는 것을 고려할 필요가 있음.

- 가이드라인은 위험 평가 기준으로 심각성(severity), 중대성(criticality) 등을 고려하도록 하고 있는데, 심각성은 피해가 심각한 정도, 중대성은 부정적 영향을 받은 범위를 의미하는 것으로 보임. 그런데 중대성은 심각성과 유사한 의미로 보이므로 적절한 다른 이름으로 변경하는 것이 바람직할 것임.

사업자책무 고시(안) 및 고영향 AI 사업자 책무 가이드라인(안)

1. 관련 법률 및 시행령(안)

법	시행령(안)
제34조(고영향 인공지능과 관련한 사업자의 책무) ① 인공지능사업자는 고영향 인공지능 또는 이를 이용한 제품・서비스를 제공하는 경우 고영향 인공지능의 안전성・신뢰성을 확보하기 위하여 다음 각 호의 내용을 포함하는 조치를 대통령령으로 정하는 바에 따라 이행하여야 한다. 1. 위험관리방안의 수립・운영 2. 기술적으로 가능한 범위에서의 인공지능이 도출한 최종결과, 인공지능의 최종결과 도출에 활용된 주요 기준, 인공지능의 개발・활용에 사용된 학습용데이터의 개요 등에 대한 설명 방안의 수립・시행 3. 이용자 보호 방안의 수립・운영 4. 고영향 인공지능에 대한 사람의 관리・감독 5. 안전성・신뢰성 확보를 위한 조치의 내용을 확인할 수 있는 문서의 작성과 보관 6. 그 밖에 고영향 인공지능의 안전성・신뢰성 확보를 위하여 위원회에서 심의・의결된 사항 ② 과학기술정보통신부장관은 제1항각 호에 따른 조치의 구체적인 사항을 정하여 고시하고, 인공지능사업자에게 이를 준수하도록 권고할 수 있다. ③ 인공지능사업자가 다른 법령에 따라 제1항 각 호에 준하는 조치를 대통령령으로 정하는 바에 따라 이행한 경우에는 제1항에 따른 조치를 이행한 것으로 본다.	제26조(고영향 인공지능과 관련한 사업자의 책무) ① 인공지능사업자는 법 제34조제1항 각 호의 조치를 이행하고 그 근거를 문서로 5년간 보관해야 한다. ② 인공지능사업자는 법 제34조제1항 각 호의 조치 중에서 다음 각 호에 해당하는 내용을 사업장 또는 인공지능사업자의 홈페이지 등에 게시하여야 한다. 다만, 「부정경쟁방지 및 영업비밀 보호에 관한 법률」 제2조제2호에 따른 영업비밀에 해당하는 사항은 제외할 수 있다. 1. 위험관리정책 및 조직체계 등 법 제34조제1항제1호에 따른 위험관리방안의 주요 내용 2. 법 제34조제1항제2호에 따른 설명 방안의 주요 내용 3. 이용자 보호 방안 4. 해당 고영향 인공지능을 관리・감독하는 사람의 성명 및 연락처 ③ 법 제34조제1항제1호부터 제3호까지의 사항의 조치를 모두 또는 일부 이행한 인공지능시스템을 제공받은 인공지능이용사업자가 인공지능시스템의 중대한 기능 변경을 초래하지 않은 경우에는 법 제34조제1항에 따른 조치를 이행한 것으로 본다. ④ 인공지능이용사업자는 인공지능개발사업자에게 법 제34조제1항에 따른 책무를 이행하기 위하여 필요한 자료의 제공을 요청할 수 있고, 인공지능개발사업자는 이에 협력하도록 노력하여야 한다. ⑤ 법 제34조제3항에서 “대통령령으로 정하는 바에 따라 이행한 경우”는 별표 1에 따른 조치를 해당 법령에 따라 이행한 경우를 말한다.
[별표 1] 이행조치 인정 기준과 절차(제26조 관련) 1. 법 제34조제3항에 따라 법 제34조제1항제1호의 조치를 이행한 경우란 다음과 같다. 가. 법 제2조제4호라목의 영역에서 활용되는 고영향 인공지능에 대하여 「디지털의료제품법」 제8조제4항 또는 같은 법 제12조제3항에 따라 품질관리체계를 갖추고, 같은 법 제24조제2항에 따라 품질관리기준에 적합하다는 판정을 받은 경우	

나. 법 제2조제4호마목의 영역에서 활용되는 고영향 인공지능에 대하여 「원자력안전법」 제26조에 따른 안전조치를 취한 경우
다. 법 제2조제4호아목의 영역에서 활용되는 고영향 인공지능에 대하여 「자율주행자동차 상용화 촉진 및 지원에 관한 법률」 제43조의 책무 및 「자동차관리법」 제30조의9 제1항·제2항 및 같은 법 제34조의5에 따른 책무를 모두 이행한 경우
라. 법 제2조제4호아목의 영역에서 활용되는 고영향 인공지능에 대하여 「자율운항선박 개발 및 상용화 촉진에 관한 법률」 제19조제1항에 따른 안전성 평가와 같은 법 제20조제1항에 따른 승인을 받은 경우

2. 법 제34조제3항에 따라 법 제34조제1항제2호의 조치를 이행한 경우란 다음과 같다.
가. 법 제2조제4호라목의 영역에서 활용되는 고영향 인공지능에 대하여 「디지털의료제품법」 제8조제3항에 따라 제조허가, 제조인증 또는 제조신고를 하고, 같은 법 제22조에 따라 디지털의료기기소프트웨어의 정보를 표시 또는 첨부한 경우
나. 법 제2조제4호라목의 영역에서 활용되는 고영향 인공지능에 대하여 「디지털의료제품법」 제12조제2항에 따라 수입허가, 수입인증 또는 수입신고를 하고, 같은 법 제22조에 따라 디지털의료기기소프트웨어의 정보를 표시 또는 첨부한 경우
다. 법 제2조제4호마목의 영역에서 활용되는 고영향 인공지능에 대하여 「원전감독법」 제8조에 따른 원자력발전시설의 관리 및 같은 법 제13조에 따른 윤리행동강령에 관련 내용이 포함된 경우
라. 법 제2조제4호사목의 영역에서 활용되는 고영향 인공지능에 대하여 「신용정보의 이용 및 보호에 관한 법률」제35조의2에 따른 설명의무를 준수하고, 같은 법 제36조의2에 따른 설명의무 이행을 위한 절차를 갖춘 경우
마. 법 제2조제4호아목의 영역에서 활용되는 고영향 인공지능에 대하여 「자율운항선박 개발 및 상용화 촉진에 관한 법률」 제19조제1항 따른 안전성 평가와 같은 법 제20조제1항에 따른 승인을 받은 경우

3. 법 제34조제3항에 따라 법 제34조제1항제3호의 조치를 이행한 경우란 다음과 같다.
가. 법 제2조제4호라목의 영역에서 활용되는 고영향 인공지능에 대하여 「디지털의료제품법」 제13조에 따른 준수사항을 이행한 경우
나. 법 제2조제4호사목의 영역에서 활용되는 고영향 인공지능에 대하여 「금융소비자보호법」 제10조에 따른 금융상품판매업자등의 책무를 이행한 경우
다. 법 제2조제4호아목의 영역에서 활용되는 고영향 인공지능에 대하여 「자율운항선박 개발 및 상용화 촉진에 관한 법률」 제19조제1항 따른 안전성 평가와 같은 법 제20조제1항에 따른 승인을 받은 경우

4. 법 제34조제3항에 따라 법 제34조제1항제4호의 조치를 이행한 경우란 다음과 같다.
가. 법 제2조제4호라목의 영역에서 활용되는 고영향 인공지능에 대하여 「디지털의료제품법」 제8조제7항(같은 법 제12조제4항에 따라 준용되는 경우를 포함한다)에 따라 품질책임자를 두는 경우
나. 법 제2조제4호마목의 영역에서 활용되는 고영향 인공지능에 대하여 「원전감독법」 제8조에 따른 원자력발전시설의 관리 및 같은 법 제13조에 따른 윤리행동강령에 관련 내용이 포함된 경우
다. 법 제2조제4호아목의 영역에서 활용되는 고영향 인공지능에 대하여 「자율운항선박 개발 및 상용화 촉진에 관한 법률」 제19조제1항 따른 안전성 평가와 같은 법 제20조제1항에 따른 승인을 받은 경우

5. 법 제34조제3항에 따라 법 제34조제1항제5호의 조치를 이행한 경우란 다음과 같다.
가. 인공지능사업자가 법 제2조제4호라목의 영역에서 활용되는 고영향 인공지능에 대하여 다음의 사항을 모두 이행한 경우
1)「디지털의료제품법」 제8조제4항 또는 같은 법 제12조제3항에 따라 품질관리체계를 갖출 것
2)「디지털의료제품법」 제24조제2항에 따라 품질관리기준에 적합하다는 판정을 받을 것

3)「디지털의료제품법」 제5조에 따라 준용되는 「의료기기법」 제13조에 따라 품질관리체계를 유지할 것
나. 법 제2조제4호아목의 영역에서 활용되는 고영향 인공지능에 대하여「자율운항선박 개발 및 상용화 촉진에 관한 법률」 제19조제1항 따른 안전성 평가와 같은 법 제20조제1항에 따른 승인을 받은 경우

6. 인공지능사업자가 「지능정보화기본법」 제60조제1항에 따른 안전성 보호조치를 이행한 경우에는 법 제34조제1항의 제1호・제2호 및 제4호를 이행한 것으로 본다. 다만, 법 제34조제1항의 조치 중 일부를 이행한 경우에는 해당 조치만을 이행한 것으로 본다.

7. 인공지능사업자가 「개인정보 보호법」 제4장 개인정보의 안전한 관리 및 제5장 정보주체의 권리 보장에서 규정하는 조치·의무를 이행한 경우, 개인정보 처리 및 보호에 대하여는 법 제34조제1항 각호를 이행한 것으로 본다. 다만, 법 제34조제1항의 조치 중 일부를 이행한 경우에는 해당 조치만을 이행한 것으로 본다.

2. 인공지능 개발/이용사업자별 책무가 모호하고 자의적임

EU AI Act의 경우 고위험 인공지능 시스템과 관련된 규정을 다루고 있는 3장(Chapter III) 2절(Section 2)에서 우선 고위험 인공지능 시스템이 준수해야 할 요건을 규정한 후, 3절(Section 3)에서 고위험 인공지능 시스템 제공자(provider), 수입업자(importer), 유통업자(distributor), 배치자(deployer) 등의 의무를 주체별로 각각 규정하고 있음. 반면, 우리 인공지능 기본법에서는 법과 시행령을 봐서는 해당 책무의 주체가 누구인지 명확하지 않음.

법 제34조 제1항에서 각 호의 조치를 이행해야 하는 주체는 인공지능사업자임. 법 제2조(정의)에서 인공지능사업자는 인공지능개발사업자와 인공지능이용사업자로 구분되지만, 법 제34조는 이를 구분하지 않고 책무를 규정하고 있음. 시행령(안)에서 조치 이행의 방법을 규정한 제1항 역시 조치의 내용에 따른 주체의 구분없이 인공지능사업자로만 규정하고 있음. 시행령 제26조 제2항은 “법 제34조제1항제1호부터 제3호까지의 사항의 조치를 모두 또는 일부 이행한 인공지능시스템을 제공받은 인공지능이용사업자가 인공지능시스템의 중대한 기능 변경을 초래하지 않은 경우에는 법 제34조제1항에 따른 조치를 이행한 것으로 본다.”라고 정하고 있음. 원칙적으로 법 제34조 제1항 각호의 의무는 모든 인공지능사업자에 대한 것인데, 마치 개발사업자와 이용자업자의 책무가 나누어진 것처럼 표시하고 있음. 제3항에서는 책무 이행을 위해 인공지능이용사업자가 인공지능개발사업자에게 자료 제공을 요청할 수 있도록 하고 있음. 따라서 법과 시행령의 내용만 보면 인공지능사업자 중 누가 해당 책무를 해야하는지 개념상 명확하지 않음.

사업자 책무 고시(안) 역시 대부분 주체를 구분하지 않고 모든 인공지능사업자에 조치의 이행을 요구하고 있음. 다만, 제5조(설명방안의 수립・시행) 제3항은 인공지능이용사업자를 특정하여 이용자에게 설명방안을 제공할 의무를 부여하고 있을 뿐임. 제6조는 이용자 보호방안의 수립・운영에 대한 내용이지만, 모든 사업자를 의무 주체로 상정하고 있음.

가이드라인은 소주제마다 인공지능개발사업자 또는 이용사업자 중 누구에게 해당되는 의무인지 표시하고 있는데, 대부분은 모든 사업자에 해당되는 것으로 표시하고 있으며, 개발사업자 또는 이용사업자에게만 해당하는 의무는 아래와 같음.

- 개발사업자에게만 해당하는 의무
2-1-1. 투명성 및 설명가능성 확보
2-2-1. 학습용데이터 개요
3-1-2. 알고리즘 설계 및 모델 개발

- 이용사업자에게만 해당하는 의무
2-3-1. 설명 방안의 주요 내용 수립
2-3-2. 설명 방안의 시행
3-2-3. 이용자 권리 보장

이처럼 가이드라인까지 살펴보면 어떤 책무를 누가 이행해야 하는지 구분되어 있지만, 법과 시행령만 보아서는 개발사업자와 이용사업자 중 누가 어떠한 책무를 이행해야 하는지 명확하지 않음. 심지어 고시까지 보아도 명확하게 구분되어 있지 않음. 따라서 주체들의 책무를 명확히 하고, 그 책무와 연결되어 있는 권리 주체들의 권리 보호를 위해서라도 법과 시행령에서 보다 구체적으로 인공지능개발사업자, 인공지능이용사업자 나누어 책무를 명확히 규율하여야 함

2. 이용자(또는 이용사업자)의 책무와 영향받는 자 권리 보호의 공백

법과 시행령에서 누가 어떠한 책무를 이행해야하는지 명확하게 규정되어 있지 않다보니, 실제로 책무로 규정되어야 할 사항과 고영향 인공지능의 부정적 영향으로부터 보호되어야 할 사람들의 권리 역시 제대로 규정되어 있지 않음. 이 문제는 비단 법 제34조만의 문제가 아니라, 인공지능 기본법에서 정의하고 있는 개념의 모호함과도 연결되어 있음.

(1) 인공지능 개발사업자, 이용사업자, 이용자, 영향받는 자 사이의 관계
과기정통부가 발표한 <AI기본법 하위법령 제정방향>(2025.9) 문서에 따르면, 인공지능개발사업자는 EU AI Act의 제공자(Provider)에 대응하는 개념으로, 인공지능이용사업자와 이용자를 배치자(Deployer)에 대응하는 개념으로 설명하고 있음.[5]

한국(인공지능기본법)	EU(AI ACT)
인공지능개발사업자	Provider[1)]
인공지능개발사업자 : 인공지능을 개발하여 제공하는 자 ※ 인공지능개발사업자가 개발한 인공지능을 제품 또는 서비스 형태로 이용자에게 직접 제공한 경우 인공지능개발사업자이자 인공지능이용사업자에 해당	"공급자"란 AI 시스템 또는 범용 AI 모델을 개발하거나 타인이 개발하도록 하여 자신의 명의·상호로 시장에 출시하거나 서비스를 공급(유·무료의 경우를 모두 포함)하는 자연인, 법인, 정부·공공기관, 기타 기관·단체 등을 의미한다.

[5] <AI기본법 하위법령 제정방향>(2025.9) 문서에서는 Provider를 공급자로, Deployer를 배포자로 번역하였으나, 이 의견서에서는 provider를 제공자로, Deployer를 배치자로 번역하였다.

한국(인공지능기본법)	EU(AI ACT)
인공지능이용사업자 : 인공지능개발사업자가 제공한 인공지능을 이용하여 인공지능제품·서비스를 제공하는 자 이용자 : 인공지능제품·서비스를 제공받는 자	Deployer[2)] "배포자"란 자신의 권한에 따라 AI 시스템을 이용하는 자연인, 법인, 정부·공공기관, 기타 기관·단체 등을 의미한다. 단, 직업적 활동이 아닌 개인적 활동 과정에서 AI 시스템을 이용하는 경우는 제외된다.

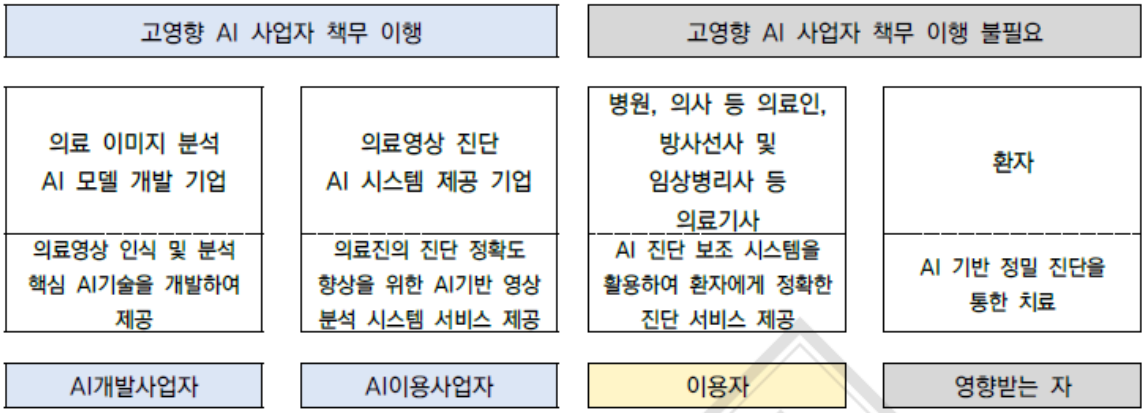
EU AI Act에서 제공자(provider)를 "AI 시스템 또는 범용 AI 모델을 개발하거나 타인이 개발하도록 하여 자신의 명의나 상표로 시장에 출시하거나 서비스를 공급하는 자연인, 법인, 정부·공공기관, 기타 기관·단체 등"으로 규정하고 있고, 타인이 개발한 AI를 이용하여 자신의 명의로 서비스나 제품을 출시하는 자도 제공자로 규정하고 있다는 점에서, 우리 인공지능 기본법 상 인공지능이용사업자 역시 EU AI Act 의 제공자(provider)라고 해석할 수도 있음.

그리고 인공지능 제품·서비스를 제공받는 자로 규정한 이용자는 업무를 목적으로 자신의 권한에 따라 AI 시스템을 이용하는 배치자(Deployer)와 개인 소비자로서의 AI 서비스 이용자를 포괄하는 개념으로 해석할 수 있음. 이처럼 인공지능 기본법의 정의와 해석 상의 모호함에도 불구하고, 본 의견서는 일단 과기정통부가 해석한대로 인공지능개발사업자는 EU 법 상의 제공자(Provider)로, 인공지능이용사업자와 이용자는 배치자(Deployer)로 간주함.

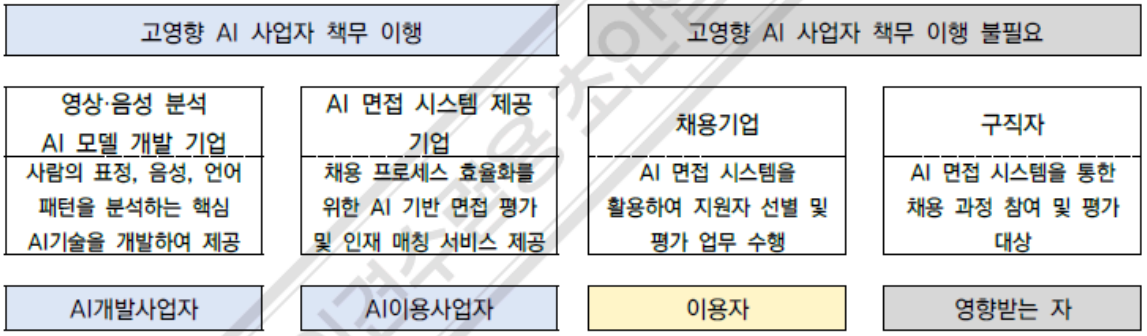
과기정통부의 해석에 따르면, 병원, 채용회사, 금융기관 등 업무상 목적으로 인공지능을 이용하는 사업자 역시 이용자로 규정이 되며, 환자, 구직자, 대출 신청자 등은 '영향받는 자'의 위치에 있게 됨.[6]

[6] <AI기본법 하위법령 제정방향>(2025.9) p169

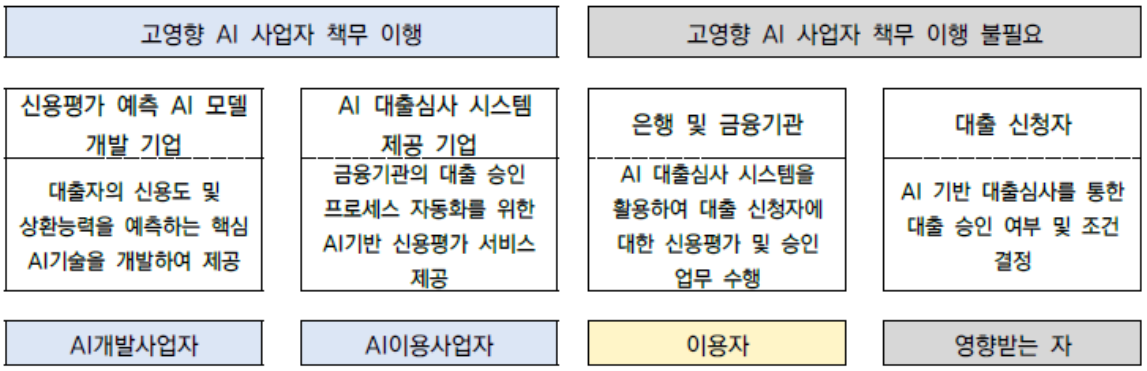
보건의료 산업



채용분야



대출심사 분야



(2) 이용자의 책무와 영향받는 자 권리의 부재

문제는 인공지능 기본법과 하위법령들이 개발사업자의 이용사업자에 대한 책무, 인공지능사업자의 이용자에 대한 책무는 규정을 해놓았지만, 배치자(Deployer)인 이용자의 영향받는 자에 대한 책무는 규정하고 있지 않다는 점임.

우선 법 제34조 제1항에서는 인공지능사업자의 책무로 '이용자 보호 방안의 수립·운영'만을 포함하고 있을 뿐, 영향받는 자에 대해서는 언급하지 않고 있음. 시행령(안) 제26조에서도 제1항에서 '이용자 보호 방안'만을 홈페이지에 게시하도록 하고 있을 뿐임.

고시에서도 제5조(설명방안의 수립・시행) 제3항에서 인공지능이용사업자로 하여금 단지 이용자에 대해서만 설명방안을 제공할 의무를 부여하고 있음.

> 제5조(설명방안의 수립・시행)
> ③ 인공지능이용사업자는 홈페이지에 게시, 해당 문서(전자문서를 포함한다. 이하 같다)의 제공 등의 방법으로 설명방안을 이용자가 제공받을 수 있도록 하여야 한다.

제6조는 이용자 보호방안의 수립・운영에 대해 다음과 같이 규정하고 있음. 제1항 제1호는 '안전하고 적법한 데이터 수집 및 관리' 조치를 정하고 있는데, 인공지능사업자가 개인정보를 비롯한 데이터를 수집할 때 이용자의 개인정보만 처리하는지 의문임. 오히려 환자, 구직자, 대출 신청자 등 영향받는 자의 개인정보를 처리하게 될 가능성이 크지만, 영향받는 자 개인정보의 안전하고 적법한 수집 및 관리 조치에 대해서는 언급하고 있지 않음. '안전한 알고리즘 설계 및 모델 개발'이나 '시험 및 평가 수행'과 같은 시스템 보안을 위한 여러 조치들도 비단 이용자만을 위한 조치는 아님. 그런데 제6조 2항에서도 '이용자' 만의 의견을 수렴하고 이에 기반하여 지속적으로 개선할 것, '이용자' 만의 권리를 보장하고 피해 발생 시 이를 보상할 수 있는 방안을 수립할 것을 규정하고 있음.

> 제6조(이용자 보호방안의 수립・운영) ① 사업자는 고영향 인공지능 개발 과정에서 이용자를 보호하기 위한 방안을 수립하는 경우 다음 각 호의 조치를 포함하여야 한다.
> 1. 안전하고 적법한 데이터 수집 및 관리
> 2. 적대적 공격 등에 대응하기 위한 안전한 알고리즘 설계 및 모델 개발
> 3. 다양하고 예외적인 상황을 고려한 시험 및 평가 수행
> ② 사업자는 고영향 인공지능 운영 과정에서 이용자를 보호하기 위한 방안을 수립하는 경우 다음 각 호의 조치를 포함하여야 한다.
> 1. 문제를 실시간으로 탐지하기 위한 모니터링 및 대응방안 수립
> 2. 이용자의 의견을 수렴하고 이에 기반하여 지속적으로 개선
> 3. 이용자의 권리를 보장하고 피해 발생 시 이를 보상할 수 있는 방안 수립

가이드라인 3-2-3 에서도 이용자 권리 보장만을 언급하고 있음. 그러나 이용 중단, 이의제기, 설명 요구 권리는 오히려 영향받는 자에게 주어져야 할 권리임. 일부 이용자(특히 최종 소비자로서의 이용자)는 영향받는 자일 수 있는데, 아마도 이 가이드라인은 영향받는 자로서의 이용자를 염두에 두고 작성되었을 수 있음. 하지만, 그렇다고 하더라도 대상을 '영향받는 자' 또는 '이용자 및 영향받는 자'로 규정했다면, 권리를 보장받아야 할 주체를 모두 포괄할 수 있었을 것임.

> 3-2-3. 이용자 권리 보장 [이용사업자]
> - 목표 : 이용자 권리 보장을 위한 체계 및 보호 정책을 수립하고 이를 이용자에게 안내해야 함.
> - 설명 :
>   - 이용사업자는 이용자에게 개인정보 활용 내역 등을 사전에 명확히 고지하고, 필요한 경우 이용 중단, 이의제기, 설명 요구 권리를 부여해야 함.
>   - 이용사업자는 안전한 사용을 위한 정책을 마련하여 이용자가 오남용하지

않도록 안내하고, 계약 조건을 명확히 규정하여 이용자의 권리 관계나 책임 소지에 대해 인지할 수 있도록 노력해야 함.
- 이용사업자는 운영 지침에 피해 발생 시의 보상 체계 등을 포함하여 이용자의 권리 보장이 이루어질 수 있는 체계를 도입하기 위해 노력해야 함.

이처럼, 오히려 병원, 채용회사, 금융기관 등 사실상 사업자에 해당하는 이용자의 권리는 보호하면서도, 환자, 구직자, 대출 신청자 등 인공지능 시스템에 의해 실제 '영향받는 자'의 권리에 대한 보호와 이들에 대한 인공지능사업자 및 이용자의 책무를 규정하지 않은 것은 커다란 문제임.

이러한 문제를 해결하기 위해서는 병원, 채용회사, 금융기관 등 업무를 목적으로 인공지능 서비스를 이용하는 사업자들도 인공지능이용사업자로 규정을 하던가, 또는 업무를 목적으로 인공지능 시스템을 이용하는 이용자 역시 영향받는 자에 대해 일정한 책무(예를 들어, 위험관리, 설명책임, 사람의 관리・감독 등)를 지도록 규정할 필요가 있음.

이는 매우 중요하고 본질적인 한계임. 법 제34조 제1항 제3호의 '이용자 보호 방안'은 '이용자'뿐만 아니라 '영향을 받는 자'에 대한 보호방안 수립운영에 더 의미가 있음. 그런데 법, 시행령, 고시의 3체계 하에서 '영향을 받는 자'에 대한 내용이 구체화되지 않아 사실상 보호방안과 구제수단이 부재한 상황임.

3. 사업자 책무 가이드라인의 내용이 부실함

다른 가이드라인, 특히 인공지능 안전성 확보 가이드라인(안)에 비하여 사업자 책무 가이드라인(안)의 내용이 매우 부실함.

예를 들어, 안전성 확보 가이드라인(안)과 사업자 책무 가이드라인(안)은 공통적으로 위험 식별, 평가, 처리(제거·완화) 등과 관련한 내용을 포함하고 있는데, 가이드라인의 상세한 정도에서 매우 큰 차이를 보이고 있음. 예컨대, 사업자 책무 가이드라인 1-1-2.는 위험 식별에 관한 챕터이고, 안전성 확보 가이드라인 3.1 역시 위험 식별을 다루고 있음. 사업자 책무 가이드라인 1-1-2.는 한 페이지 정도로 간략하게 다루고 있으나 (이 챕터 뿐만 아니라 전반적으로 그러함) 안전성 확보 가이드라인 3.1 은 7 페이지에 걸쳐서 상세하게 설명하고 있음. 이는 안전성 확보 가이드라인이 최첨단 인공지능의 안전성을 다루고, 사업자 책무 가이드라인은 고영향 인공지능 시스템의 안전성을 다루는 등 대상의 차이에 기인한 것으로 보이지 않음.

가이드라인이 전체적으로 체계적이기 위해서는 고영향 인공지능 시스템과 최첨단 인공지능 시스템의 위험 식별, 평가, 처리(제거·완화)가 어떠한 차이가 있는지, 공통점과 차이점은 무엇인지, 이 외에 인공지능 기본권영향평가도 있는데, 기본권영향평가에서의 위험 식별, 평가, 처리와는 어떻게 다른지, 이를 위한 위험평가체제는 각각 구성이 되어야 하는지 등이 체계적으로 설명될 필요가 있음.

AI 투명성 확보 가이드라인(안)

1. 관련 법률 및 시행령(안)

법	시행령(안)
제31조(인공지능 투명성 확보 의무) ① 인공지능사업자는 고영향 인공지능이나 생성형 인공지능을 이용한 제품 또는 서비스를 제공하려는 경우 제품 또는 서비스가 해당 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지하여야 한다. ② 인공지능사업자는 생성형 인공지능 또는 이를 이용한 제품 또는 서비스를 제공하는 경우 그 결과물이 생성형 인공지능에 의하여 생성되었다는 사실을 표시하여야 한다. ③ 인공지능사업자는 인공지능시스템을 이용하여 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등의 결과물을 제공하는 경우 해당 결과물이 인공지능시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하여야 한다. 이 경우 해당 결과물이 예술적・창의적 표현물에 해당하거나 그 일부를 구성하는 경우에는 전시 또는 향유 등을 저해하지 아니하는 방식으로 고지 또는 표시할 수 있다. ④ 그 밖에 제1항에 따른 사전고지, 제2항에 따른 표시, 제3항에 따른 고지 또는 표시의 방법 및 그 예외 등에 관하여 필요한 사항은 대통령령으로 정한다.	제22조(인공지능 투명성 확보 의무) ① 인공지능사업자는 고영향인공지능이나 생성형 인공지능을 이용한 제품 또는서비스(이하 “제품등”이라 한다)를 제공하기 전에 다음 각 호의 어느 하나에 해당하는 방법으로 법 제31조 제1항에 따른 고지를 하여야 한다. 1. 제품 등에 직접 기재하거나, 계약서, 사용설명서, 이용약관 등에 기재 2. 이용자의 화면 또는 단말기 등에 표시 3. 제품 등을 제공하는 장소(해당 장소와 합리적으로 관련된 범위의 장소를 포함한다)에 인식하기 쉬운 방법으로 게시 4. 그 밖에 제품 등의 특성을 고려하여 과학기술정보통신부장관이 인정하는 방법 ② 인공지능사업자가 법 제31조제2항에 따른 표시(법 제31조제3항에 따른 실제와 구분하기 어려운 결과물을 제공하는 경우로서 해당 결과물이 인공지능시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하는 경우에는 적용하지 아니한다)를 할 때에는 다음 각호 중 하나의 방법으로 할 수있다. 1. 사람이 인식할 수 있는 방법 2. 기계가 판독할 수 있는 방법. 다만, 이 경우에는 생성형 인공지능에 의하여 생성되었다는 사실을 1회 이상 안내문구·음성 등으로 제공하여야 한다. ③ 법 제31조제3항에 따른 고지 또는 표시는 인공지능사업자가 다음 각 호의 사항을 고려하여 이용자가 명확하게 인식할 수 있는 방식으로 해야 한다. 1. 이용자가 시각, 청각 등을 통하거나 소프트웨어 등을 이용하여 쉽게 내용을 확인할 수 있는 방법으로 고지 또는 표시할 것 2. 주된 이용자의 연령, 신체적・사회적 조건 등을 고려하여 고지 또는 표시할 것 ④ 법제31조제1항부터 제3항까지는 다음 각 호에 해당하는 경우에는적용하지 아니한다. 다만, 제3호의 경우에는 법 제31조제1항부터 제3항까지 중 전부 또는 일부를 적용하지 아니할 수 있다. 1. 제품・서비스명, 이용자 화면이나 제품

	겉면 및 결과물에 표시된 문구 등을 고려할때 고영향 인공지능 또는 생성형인공지능을 활용한 사실이 명백한 경우 2. 인공지능사업자의 내부 업무 용도로만 사용되는 경우 3. 그 외 제품 등의 유형・특성이나 결과물의 내용, 이용형태 및 기술수준 등을 고려하여 법제31조제1항부터 제3항까지 중전부 또는 일부에 대한 적용 예외가 필요한 사항으로 과학기술정보통신부장관이 정하여 고시하는 경우

2. 인공지능 투명성 확보 의무 주체 축소해석

EU AI Act의 경우 고위험 인공지능 시스템과 관련된 규정을 다루고 있는 3장(Chapter III) 2절(Section 2)에서 우선 고위험 인공지능 시스템이 준수해야 할 요건을 규정한 후, 3절(Section 3)에서 고위험 인공지능 시스템 제공자(provider), 수입업자(importer), 유통업자(distributor), 배치자(deployer) 등의 의무를 주체별로 각각 규정하고 있음. 반면, 우리 인공지능 기본법에서는 법과 시행령을 봐서는 해당 책무의 주체가 누구인지 명확하지 않음.

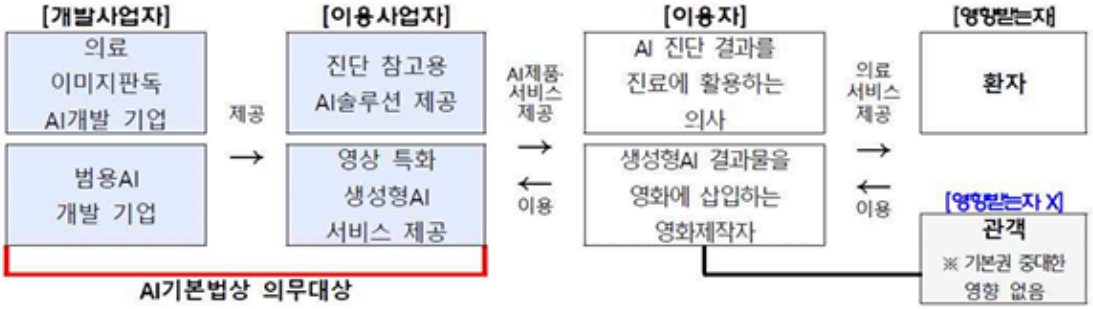
현행 인공지는기본법은 생성 결과물 유통과정에서의 투명성 확보 의무를 별도로 규정하고 있지 않음. 따라서, 인공지능기본법 상 투명성 확보에 따른 이용자와 영향 받는 자를 보호하기 위하여 '인공지능이용사업자'에 대한 해석을 명확하게 할 필요가 있음.

인공지능기본법 상 '인공지능이용사업자'는 "인공지능개발사업자가 제공한 인공지능을 이용하여 인공지능제품 또는 인공지능서비스를 제공하는 자"로 규정되어 있는데, '인공지능서비스'란 인공지능제품과 관련한 서비스를 포함하는 개념으로 가이드라인에서는 사업자가 주도적으로 인공지능을 구성하고 활용하여 고유(행정업무, 교육, 언론, 영화·문화산업 등)의 업무서비스를 제공하는 경우에도 단순 '이용자'가 아닌 인공지능이용사업자로 해석되어야 할 것임

그럼에도 불구하고, 가이드라인(안)에서는 아래와 같이 인공지능기본법은 인공지능개발사업자와 이용사업자에 의무를 부과하며, 단순히 인공지능제품·서비스를 이용한 결과물을 자신의 서비스 등에 활용하는 자는 인공지능기본법상 사업자에 해당하지 않아 투명성 확보 의무가 없다고 명시하고 있음

○ **(투명성 확보 대상 예시)** AI기본법은 AI개발사업자와 이용사업자에 의무를 부과하며, 단순히 AI제품·서비스를 이용한 결과물을 자신의 서비스 등에 활용하는 자는 AI기본법상 사업자에 미해당하여 투명성 확보 의무가 없음

- AI를 이용하여 CG를 생성한 후 영화에 삽입한 영화제작자는 단순히 AI 서비스를 이용한 결과물을 자신의 콘텐츠에 활용한 것 이므로 AI기본법 상 사업자가 아니며 이용자에 해당



초기 가이드라인(안)에 의하면 오디오 콘텐츠 투명성 확보 방안의 예시로 제시된 '클로바 더빙' 및 'YTN 뉴스' 예시를 명시하여 해당 사업자에게도 투명성 확보 의무가 있는 것으로 명시되었으나, 최종 가이드라인(안)에서는 해당 사업자가 단지 이용자가 생성 인공지능 기술을 활용해 콘텐츠를 제작한 사례로 보아 단순히 이용자에 해당한다는 이유로 해당 예시를 삭제하여 책무의 주체를 오히려 축소함

이용자로 명시되어 있는 사업자라고 하더라도 인공지능 개발 및 이를 바탕으로 하는 자신의 고유의 서비스 제공을 위하여 주도권을 가지고 인공지능을 활용하는 경우에는 단순 '이용자'가 아닌 인공지능이용사업자로 규율되어야 할 것임

고영향 AI 판단 가이드라인(안)

1. 관련 법률 및 시행령(안)

법	시행령(안)
제2조(정의) 4. “고영향 인공지능”이란 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템으로서 다음 각 목의 어느 하나의 영역에서 활용되는 것을 말한다. 가.「에너지법」 제2조제1호에 따른 에너지의 공급 나.「먹는물관리법」 제3조제1호에 따른 먹는물의 생산 공정 다.「보건의료기본법」 제3조제1호에 따른 보건의료의 제공 및 이용체계의 구축・운영 라.「의료기기법」 제2조제1항에 따른 의료기기 및「디지털의료제품법」 제2조제2호에 따른 디지털의료기기의 개발 및 이용 마.「원자력시설 등의 방호 및 방사능 방재 대책법」 제2조제1항제1호에 따른 핵물질과 같은 항 제2호에 따른 원자력시설의 안전한 관리 및 운영 바. 범죄 수사나 체포 업무를 위한 생체인식정보(얼굴・지문・홍채 및 손바닥 정맥 등 개인을 식별할 수 있는 신체적・생리적・행동적 특징에 관한 개인정보를 말한다)의 분석・활용 사. 채용, 대출 심사 등 개인의 권리・의무 관계에 중대한 영향을 미치는 판단 또는 평가 아.「교통안전법」 제2조제1호부터 제3호까지에 따른 교통수단, 교통시설, 교통체계의 주요한 작동 및 운영 자. 공공서비스 제공에 필요한 자격 확인 및 결정 또는 비용징수 등 국민에게 영향을 미치는 국가, 지방자치단체,「공공기관의 운영에 관한 법률」 제4조에 따른 공공기관 등(이하 “국가기관등”이라 한다)의 의사결정 차.「교육기본법」 제9조제1항에 따른 유아교육・초등교육 및 중등교육에서의 학생 평가 카. <u>그 밖에 사람의 생명・신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역으로서 대통령령으로 정하는 영역</u>	없음

2. 총설

인공지능기본법 제2조 4호에서 정하고 있는 고영향 인공지능 목록은 인공지능시스템이 활용되는 경우 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 개별 영역을 정한 것으로서 각 영역에서 활용되고, 기본권에 중대한 영향 또는 위험을 야기할 가능성이 있는 일체의 인공지능 시스템은 원칙적으로 고영향 인공지능에 포함된다고 보아야 할 것임.

인공지능기본법은 인공지능 분야의 기본법으로, 고영향 인공지능시스템 전반에 적용될 최소한의 공통된 규율과 책무를 규정하고 있으므로, 인공지능 기본법 하위법령에서 임의의 기준에 따라 고영향 인공지능의 범주에서 배제하는 것은 인공지능 기본법의 취지에 반함.

그러나 가이드라인(안)에서는 인공지능기본법 제2조 4호 각목에서 규정하고 있는 고영향 인공지능의 개별 목록의 의미에 대하여 문언과 다르게 임의로 축소 해석하거나, 추가적인 제한 요건을 부가하는 등의 방식으로 고영향 인공지능의 범주를 최소화하려 하고 있음. 이는 상위법에 위배될 소지가 있고, 기본권 보호의 측면에서도 바람직하지 않다 할 것임.

3. 분야별 고영향에 대한 검토

(1) 에너지

가. '에너지 공급'의 법정 영역을 전력망 중심으로 축소하고 있음.

가이드라인(안)에서는 적용범위를 발전·송전·변전·배전·소비로 이어지는 전력 공급의 관리·운영에 중점을 둔다고 설명하면서 상위법인 인공지능기본법의 "에너지 공급"을 사실상 "전력 공급"으로 축소하고 있음. 이는 법률이 정한 영역을 기술 분야 중심으로 재구성하는 것이기 때문에, 상위법의 문언과 취지에 부합하지 않음. 이로 인해 상위법상으로는 "에너지 공급" 전체가 영역으로 포함되지만 가이드라인(안)은 전력망 중심으로 해석해 적용 범위가 축소되며 고영향 인공지능의 판단이 누락될 위험이 있음.

나. 고영향 판단 기준에 상위법에 없는 기술 요건을 추가해 범위가 협소해지고 있음.

인공지능기본법 제2조제4호는 "사람의 생명·신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템"이라고 명시하고 있으나 가이드라인(안)은 고영향 여부 판단에 기능 중요도·잠재적 위험성·시스템 신뢰성·데이터 처리 능력·자율성·의사결정 능력 등을 제시하고 있음. 이러한 조건들은 법에 존재하지 않는 추가적인 기술 요건이며 이를 통해 고영향 인공지능의 범위가 좁아질 위험이 있음. 예를 들어, 기술적 자율성이 낮거나 인력의 보조를 담당하는 인공지능이어도 에너지 공급 공정에서 오류가 발생하면 충분히 사람의 생명·신체의 안전에 위험을 발생시킬 수 있음. 고영향의 본질적인 판단 기준은 생명과 신체, 기본권 위협 여부임을 명시하고 해당 기술적 조건들은 참고 기준으로만 두어야 함.

(2) 먹는물

인공지능기본법 제2조제4호나목은 먹는물관리법 제3조제1호에 따른 먹는물 생산 공정을 고영향 영역으로 규정하고 있음에도 가이드라인(안)은 먹는물 인공지능을 '기초 모니터링 → 분석·예측 → 자동 조정 → 완전자율 운영'이라는 단계 구조로 제시하며 단계별 위험 판단 기준을 추가하고 있음. 이는 상위법이 요구하지 않은 기술 단계 요건을 사실상 고영향 판단 기준처럼 구성하는 것으로, 단순 모니터링 시스템의 오작동만으로도 수질 사고가 발생할 수 있는 먹는물 생산 공정의 특성을 충분히 반영하지 못함. 기술적 발달 단계가 낮다는 이유로 고영향 판단에서 제외되는 해석이 현장에서 발생한다면 생명과 신체, 기본권에 위험을 끼칠 수 있음.

수질관리 단계별 도출 판단기준

단계	내용	도출 판단기준
1단계: 기초 모니터링 및 데이터 수집	• 인공지능은 먹는물과 관련된 기본적인 데이터 수집 및 모니터 담당 • 예를 들어, 수질 센서 데이터의 실시간 수집이나 간단한 통계적 분석의 수행 • 위험도는 낮으나, 데이터의 정확성과 신뢰성이 중요한 기초가 됨	데이터 수집의 정확성 데이터 처리 및 저장의 신뢰성
2단계: 진단 및 예측	• 인공지능은 수집된 데이터를 바탕으로 수질의 변화를 예측, 잠재적 문제를 조기에 진단 • 특정 오염물질의 수치가 기준치에 근접하거나 초과할 가능성을 예측하여 조치를 취함 • 이때 인공지능에 의한 결정은 반드시 인간의 검토를 거쳐야 함	데이터 분석 및 패턴 인식 예측의 정확성
3단계: 자동 조정 및 제어	• 먹는물의 수질을 자동으로 조정하고 관리하는 시스템 • 예를 들어, 수질이 기준치를 벗어날 때 자동으로 처리설비를 조정하거나 특정 화학물질을 투입하는 결정 • 이 단계는 높은 수준의 신뢰성과 정확성을 요구하며, 잘못된 조정은 직접적으로 인간 건강에 영향을 미침	수질 자동 조정의 정확성 실시간 피드백 능력
4단계: 완전 자동화 및 독립적 운영	• 최종 단계의 인공지능은 거의 모든 운영을 독립적으로 수행하며, 인간 최소한의 간섭으로 관리 • 이 단계의 인공지능에는 매우 높은 기술 수준이 요구되며, 가능한 모든 시나리오에서 안전성과 효율성을 입증해야 함 • 오류에 대한 높은 회복성과 자가 진단 능력이 필수적	자율 운영 시스템의 안정성 자가 진단 및 자율 복구 능력

또한 가이드라인(안)에 포함된 "하수처리는 고영향 분야에 해당하지 않는다"는 단정적 문구 역시 문제임. 인공지능기본법 제2조제4호나목은 먹는물 생산 공정을 영역으로 지정할 뿐 하수처리에 대한 배제 규정을 두지 않음. 하수처리는 먹는물과 직접 연결된 수질 관리 체계에서 중요한 요소이므로, 하수처리 인공지능의 오작동이 먹는물 안전에 영향을 미칠 수 있음. 그럼에도 불구하고 가이드라인(안)이 '고영향이 아니다'라고 명시하면 법적 근거 없이 예외를 신설하는 결과가 되고, 현장에서 위험이 과소평가될 가능성이 높음.
가이드라인(안)은 기술적 요소와 단계 구분을 설명적·보조적 기준으로 재배치하고, 고영향 여부는 반드시 인공지능기본법 제2조제4호 본문이 정한 위험 기준을 중심으로 판단하도록 명확히 해야 함.

(3) 보건, 의료

가. 의료기기 중 4등급으로 한정하여 고영향으로 포함시킬 수 있는 근거가 없음.

가이드라인(안)에서는 의료기기법 시행규칙 별표1상 식품의약품안전처장에 의한 분류기준표상 4등급에 해당하는 의료기기만 고영향으로 파악하여야한다고 설명하고 있으나, 중증도 이상 잠재적 위해성을 갖는 3등급 의료기기에는 치과용 임플란트, 인공호흡기, 복막투석장치, 엑스선투시진단장치 등이 포함되어 있는데, 이들 의료기기는 인체와의 접촉 정도가 상당하고, 환자의 신체 등에 미치는 영향이 중대하므로 고영향으로 분류되어야 함.
아울러 잠재적 위해성이 거의 없는 경우 부여되는 1등급(수동식 부항기, 진료의자, 진료대, 수동식 휠체어), 잠재적 위해성이 낮은 경우 부여되는 2등급(초음파 진단기기, 전자 혈압계, 전자체온계) 의료기기 또한 부정확하거나 오용될 경우 의사의 진단이나 판단을 좌우할 수 있어 환자의 생명권, 신체의 자유 등에 중대한 영향을 미칠 가능성이 있다는 점에서 고영향 인공지능에 포함되어야 마땅함. 그러나 가이드라인(안)은 1, 2, 3등급 의료기기의 경우 임의로 고영향의 범주에서 제외될 가능성이 높은 것처럼 설명하고 있다는 점에서 적절치 못함.

나. 보건의료 이용체계에 대한 고영향 판단기준에 인공지능에 의해 의사결정이 좌우되어야 한다는 별도 요건을 부가하고 있음.

인공지능기본법 제2조 제4호 다목에서는 보건의료기본법 제3조 제1호에서 정한 ‘보건의료’의 제공 및 이용체계의 구축·운영에 활용되는 인공지능시스템을 고영향 인공지능으로 분류하고 있으나, 가이드라인(안)에서는 보건의료 이용체계에 대하여 ‘인공지능을 통해 의사결정이 좌우되어 서비스 대상자의 생명·신체·정신에 중대한 피해가 발생하는 경우’ 고영향에 해당한다고 설명하고 있어 인공지능에 의해 의사결정이 좌우되어야 한다는 별도의 요건을 부가하고 있음.
인공지능 기본법은 기본권에 대한 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템을 고영향 인공지능으로 정의하고 있는바, 인공지능에 의해 의사결정이 좌우된다는 요건은 인공지능기본법에서 전혀 예정하고 있지 않은 요건을 추가하는 것이어서 적절치 못함.

(4) 원자력

가이드라인(안)에서는 인공지능이 원자력위원회 고시(원자로시설의 안전등급과 등급별 규격에 관한 규정)상 1, 2, 3 등급의 안전기능을 수행하는 경우에 원칙적으로 고영향 인공지능에 해당한다고 보면서, 비안전등급으로 분류된 경우에도 고시 제8조 제2항 각호의 기능을 수행하는 경우에는 원칙적으로 고영향에 해당한다고 하고 있어, 이러한 설명은 합당해 보임.

<원자로시설의 안전등급과 등급별 규격에 관한 규정 제8조 제2항 각호>

1. 방사성폐기물을 처리・추출・포장 또는 저장하는 기능
2. 정상 운전 중에 원자로냉각재계통 또는 핵연료저장 냉각계통으로부터 방사성물질을 정화하는 기능
3. 조사된 중성자흡수재의 재사용을 위하여 방사성물질을 추출・저장 또는 운반하는 기능(예:붕산화합물)
4. 방사성물질의 방출률 및 전체 방출량이 정상 운전 및 과도상태에서 설정된 제한치 이하로 유지되도록 방사성유출물을 감시하는 기능
5. 안전등급 1, 2 또는 3의 설비가 안전기능을 수행하는 데 지장을 초래할 수 있는 고장을 방지하는 기능
6. 원자력발전소 내 인원 및 안전등급 1, 2 또는 3의 설비를 보호하기 위한 영구적인 차폐기능
7. 방사성물질과 관련된 운전, 보수 또는 사고후 복구할 때에 공중의 보건 및 안전에 과도한 위험을 유발하지 아니 하도록 하는 기능
8. 비상시 원자로제어실로부터 대피한 후 원자로를 안전정지상태로 도달・유지시키는 데 필요한 안전등급 1, 2 또는 3의 설비에 적절한 환경을 제공하는 기능
9. 핵연료 손상시 다량의 방사성물질이 방출될 우려가 있는 사용후핵연료의 취급에 관한 기능
10. 저장중인 핵연료의 반응도를 제어하는 기능
11. 화재발생시 원자로를 안전정지시키거나 유지하는 데 필요한 안전등급 2 또는 3의 설비를 보호하는 기능
12. 다음 각 목의 사항을 이행하기 위한 변수를 감시하는 기능
가. 원자력발전소의 운전조건이 운영기술지침서의 제한치 이하임을 확인(예: 핵연료 재장전 저장탱크 수위, 안전관련 냉각수 온도)
나. 보호계통의 운전에 의하여 자동적으로 제거되지 않은 보호계통 우회상태 지시
다. 안전등급 1, 2 또는 3의 설비상태 지시
라. 사고발생 후 사건의 원인을 조사하고 결과를 판정하기 위한 지원
13. 제1호부터 제12호의 어느 하나에 해당하는 기능을 수행하는 비안전등급 설비를 구조적으로 지지하거나 보호하는 기능

그러나 ① 인공지능이 설비나 기능을 직접 운전하지 않는 경우, ② 기능이 수행되기 전에 사람이 cross-check를 수행하는 경우, ③ 사람의 판단 및 기능 수행에 인공지능이 보조적인 자료를 제공하는 경우 ④ 기존에 이미 안전기능이 있는 상태에서 추가적인 안전기능을 제공하는 경우를 나열하면서, 위 각 사항 가운데 한가지에만 해당하여도 고영향에 해당하지 않는다고 설명하고 있으나, 이러한 설명에는 아무런 근거가 없음.
위 4가지 사항 가운데 1가지를 충족한다고 하여도, 나머지 3가지 사항을 충족하는 경우 여전히 기본권에 중대한 영향을 미치거나 부정적인 영향을 미칠 가능성이 존재한다고 볼 수 있으므로 이러한 판단 방법은 타당성이 없으며, 아무런 법상의 근거도 없음. 이는 자의적으로 고영향의 범주를 축소시키는 판단 기준으로 불합리하다는 평가를 면하기 어려움.

(5) 범죄 수사 및 체포

가. 범죄 수사 및 체포업무를 위한 목적으로 범죄예방, 예측 치안 업무에 활용하는 인공지능 시스템을 임의로 배제시킴.

가이드라인(안)에서는 '범죄의 예방', '범죄의 예측', '위험도 분석' 등은 '범죄 수사나 체포 업무'의 범위에 포함되지 않고, 인공지능기본법 제2조 제4호 바목에서 설명하는 '고영향 인공지능'의 범주에 포함될 수 없다고 설명하고 있으나 인공지능기본법 제2조 제4호 바목의 문언에서 범죄수사와 체포 관련 고영향 인공지능의 영역을 정의하면서 다른 제한적, 한정적 문구를 사용하지 않고 범죄 수사나 체포업무를 '위한'이라고 정한 취지는 '범죄 수사나

체포’를 목적으로 예방적 치안 활동에 활용되는 인공지능 시스템 또한 고영향 범주에 포함시키려는 의도로 이해할 수 있음.

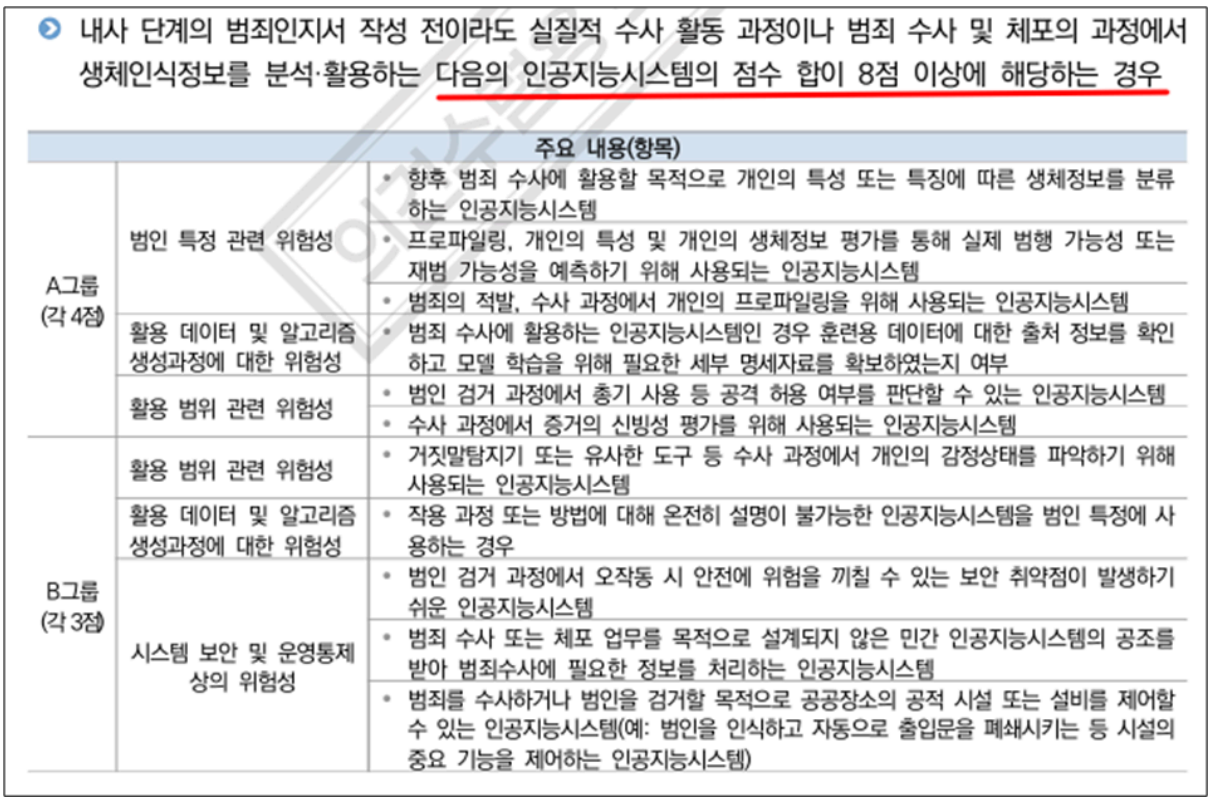
예측 치안 또는 범죄 예방 영역에서 활용되는 인공지능 시스템이 부정확하고, 부당한 방식으로 작동하는 경우 기본권에 중대한 영향을 미치거나, 차별적인 결과를 초래할 수 있고, 인공지능 시스템의 신뢰성이 보장되지 않고, 관련 문서가 작성, 보관되어 있지 않은 경우 공정한 재판을 받을 권리 등 절차적 기본권의 행사에도 중대한 영향을 미칠 가능성이 존재함.

EU AI Act는 수사기관을 포함한 법집행 기관 및 관련 기관이 사용하는 피해자 위험 평가, 증거의 신뢰성 평가, 범죄성 및 재범 가능성의 평가에 활용되는 인공지능 시스템 뿐만 아니라[부속서 III 제6조], 수사나 체포 목적으로 한정하지 않고 생체인식분류에 사용되도록 의도된 인공지능시스템은 모두 고위험 인공지능으로 분류하고 있으며 [부속서 III 제1조], 생체인식 정보를 활용하여 인종, 정치적 의견 등의 추론에까지 나아가는 경우 금지되는 인공지능으로 분류하여 개발 및 활용 자체를 금지하고 있음[EU AI Act 제5조 제1항(g)].

그러나 가이드라인(안)에서는 ‘범죄수사 및 체포업무’라는 문언만을 강조하며, ‘범죄의 예방’, ‘범죄의 예측’, ‘위험도 분석’ 등 영역을 모두 고영향 인공지능에 포함되지 않는다고 설명하고 있는데, 이렇게 해석하는 경우 범죄 패턴 예측, 잠재적 범죄자 식별, 범죄발생 가능성이 높은 지역과 시간 추적, 예방적 순찰, 사전 감시 등 기본권 침해, 편향, 차별의 발생 가능성이 높은 영역에서 아무런 안전성 또는 신뢰성 확보조치 없이 인공지능 시스템이 도입되고, 활용하는 결과가 됨. 이러한 해석은 부당하고 임의적인 축소 해석이라 할 것임.

나. 고영향 인공지능을 판별하기 위한 항목을 제시하며, 합계 점수 8점 이상인 경우 고영향에 해당한다는 설명은 자의적인 기준에 의한 것으로서 객관성이 없음.

가이드라인(안)에서는 내사 단계라 하더라도 실질적 수사활동 과정이나 범죄 수사 및 체포 과정에서 생체인식정보를 활용하는 인공지능 시스템 가운데, 합계 점수가 8점 이상에 해당하는 경우 고영향 인공지능에 해당한다고 하며, 아래와 같이 항목별 점수 기준을 제시하고 있음.



내사 단계의 범죄인지서 작성 전이라도 실질적 수사 활동 과정이나 범죄 수사 및 체포의 과정에서 생체인식정보를 분석·활용하는 다음의 인공지능시스템의 점수 합이 8점 이상에 해당하는 경우

		주요 내용(항목)
A그룹 (각 4점)	범인 특정 관련 위험성	• 향후 범죄 수사에 활용할 목적으로 개인의 특성 또는 특징에 따른 생체정보를 분류하는 인공지능시스템
		• 프로파일링, 개인의 특성 및 개인의 생체정보 평가를 통해 실제 범행 가능성 또는 재범 가능성을 예측하기 위해 사용되는 인공지능시스템
		• 범죄의 적발, 수사 과정에서 개인의 프로파일링을 위해 사용되는 인공지능시스템
	활용 데이터 및 알고리즘 생성과정에 대한 위험성	• 범죄 수사에 활용하는 인공지능시스템인 경우 훈련용 데이터에 대한 출처 정보를 확인하고 모델 학습을 위해 필요한 세부 명세자료를 확보하였는지 여부
	활용 범위 관련 위험성	• 범인 검거 과정에서 총기 사용 등 공격 허용 여부를 판단할 수 있는 인공지능시스템
		• 수사 과정에서 증거의 신빙성 평가를 위해 사용되는 인공지능시스템
B그룹 (각 3점)	활용 범위 관련 위험성	• 거짓말탐지기 또는 유사한 도구 등 수사 과정에서 개인의 감정상태를 파악하기 위해 사용되는 인공지능시스템
	활용 데이터 및 알고리즘 생성과정에 대한 위험성	• 작용 과정 또는 방법에 대해 온전히 설명이 불가능한 인공지능시스템을 범인 특정에 사용하는 경우
	시스템 보안 및 운영통제상의 위험성	• 범인 검거 과정에서 오작동 시 안전에 위험을 끼칠 수 있는 보안 취약점이 발생하기 쉬운 인공지능시스템
		• 범죄 수사 또는 체포 업무를 목적으로 설계되지 않은 민간 인공지능시스템의 공조를 받아 범죄수사에 필요한 정보를 처리하는 인공지능시스템
		• 범죄를 수사하거나 범인을 검거할 목적으로 공공장소의 공적 시설 또는 설비를 제어할 수 있는 인공지능시스템(예: 범인을 인식하고 자동으로 출입문을 폐쇄시키는 등 시설의 중요 기능을 제어하는 인공지능시스템)

가이드라인(안)에 의하면, 기준상 대상 항목들이 A그룹과 B그룹으로 구분되고, A그룹에 대해서 각 4점, B그룹에 대해서 각 3점이 부여되며, 합산 점수 8점을 초과해야 고영향 인공지능에 해당한다는 것이나, A그룹과 B그룹이 구분되어 배점을 달리하는 이유에 대해서 아무런 설명이 제시되어 있지 않고 합산 점수 8점을 초과해야 고영향으로 볼 수 있다는 판단 또한 자의적이라는 평가를 면하기 어려움.

가이드라인(안)에 제시되어 있는 각 개별 항목 중 '향후 범죄 수사에 활용할 목적으로 개인의 특성 또는 특징에 따른 생체정보를 분류하는 인공지능 시스템', ' 범죄의 적발, 수사과정에서 개인의 프로파일링을 위해 사용되는 인공지능시스템', '범인 검거 과정에서 총기 사용 등 공격 허용여부를 판단할 수 있는 인공지능 시스템'은 그 자체로 기본권 침해 및 차별 위험이 높다고 할 수 있고 실제 EU AI Act에서 금지하거나 고위험 인공지능 시스템으로 분류하고 있는 영역에 해당하는 것으로 보임. 그러나 여러 항목에 중첩적으로 해당하여야(즉, 합산 점수 8점 이상에 해당하여야) 고영향 인공지능으로 분류될 수 있다는 설명 또한 임의적인 해석에 따른 것으로 보여짐.
따라서 가이드라인(안)에서 제시하고 있는 기준은 객관적 근거가 결여되어 있을 뿐 아니라, 부당하게 고영향 인공지능의 범주를 축소시킬 우려가 있어 인권보호의 측면에서도 바람직하지 않다 할 것임.

(6) 채용

가. 임의로 설정한 채용 과정 인공지능시스템 사용 '목적'을 한정하고 중대한 영향'의 범위를 제한하여 입법 취지를 몰각함

「고영향 인공지능 판단 가이드라인(안)」중 6. 채용 부분은 고영향 인공지능시스템의 적용범위를 "채용 전(全) 과정(모집, 서류심사, 필기평가, 면접평가, 실기평가)에서 공정성과

효율성을 높이기 위하여 사용되는 인공지능시스템을 의미한다"라고 인공지능시스템의 사용 목적을 '공정성','효율성'으로 한정하고 있음
그러나 인공지능시스템의 사용 목적이 정당하더라도 ▲그 절차와 과정이 투명하게 공개되지 않거나 ▲결과에 부당한 요소 개입이 존재한다면 개인의 권리·의무에 회복할 수 없는 피해를 줄 위험이 있기에, 위와 같이 사용 목적을 한정하는 것은 무의미(**無意味**)하게 함. 오히려 사용 목적을 한정함으로 인해, 인공지능시스템을 개발·활용하여 채용 과정에서 발생할 위험을 예방하기 위한 ▲사전평가와 ▲관리·감독을 우회하는 길을 열어줄 수 있음.

또한 「고영향 인공지능 판단 가이드라인(안)」중 6. 채용 부분은 고영향 인공지능시스템의 '고영향'의 의미를 "채용 과정에서 활용되는 인공지능시스템이 헌법상 보장된 구직자의 직업선택의 자유 및 평등권에 중대한 영향을 미칠 우려가 있는 판단 또는 평가를 하는 경우 '고영향 인공지능'에 해당한다"고 제시하여, 고영향 인공지능시스템이 적용되는 범위를 제한하고 그 결과 입법 목적에 반할 우려가 있음.
법 제2조 제4호 사.목은 열거 조항으로 채용은 "개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가"의 예시를 든 것에 불과함.
채용 관련 인공지능시스템 개발·활용 과정에서 '구직자의 직업 선택의 자유', '평등권'뿐 아니라 '인격권''사생활의 비밀과 자유''개인정보자기결정권','알권리''공무담임권' 등 다양한 인권 및 기본권이 침해될 수 있음. 그렇기에 법 제2조 제4호 사.목 법문에서도 "등 개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가""영역에서 활용되는 것"이라 명시하고 있는 것임. 「고영향 인공지능 판단 가이드라인(안)」중 6. 채용 부분은 '고영향'의 의미를 입법 취지에 반하여 제한적으로 제시함으로써 고영향 인공지능시스템의 위험을 증가시키고 이로 인해 인권침해가 발생한 경우 적절하게 권리구제를 받을 수 없게 함.

나. '고영향' 확인 기준에 아무 근거 없이 '실질적이고 의미 있는' '인적 개입이 있는지'를 추가함

「고영향 인공지능 판단 가이드라인(안)」중 6. 채용 부분은 '고영향' 확인 기준을 어떠한 근거 없이 임의로 '실질적이고 의미있는 인적 개입이 있는지 여부'로 제시하여 인권·기본권의 제한 내지 침해를 야기할 위험이 있음.
법 제2조 제4호 사.목은 "채용" 두 글자만을 규정하고 있을 뿐 실질적이고 의미있는 인적 개입이라는 기준을 정하고 있지 않습니다. 그런데 위 가이드라인은 아무런 법적 위임이나 합리적인 설명 없이 채용에 활용되는 인공지능시스템이 '고영향'인지를 확인하는 기준을 임의로 설정하여 제시하였음. 이는 헌법 제37조, 제75조 행정기본법 제8조에 명백히 위반되는 행정작용임.
이러한 위헌·위법적인 가이드라인(내부지침, 행정작용)은 헌법상 법치주의·법적 안정성 원칙에 위배되어 인권과 기본권을 침해할 우려가 큼. 이러한 내용을 가이드라인에 그대로 둔다면 국민의 기본권을 심각히 침해할 뿐 아니라 향후 법적 분쟁을 다수 유발할 것으로 사료됨.

다. 단편적으로 서술된 '사례로 알아보는 고영향 확인 기준'은 인권 침해 예방 및 권리 구제에 방해가 됨

「고영향 인공지능 판단 가이드라인(안)」중 6. 채용 부분 '사례로 알아보는 고영향 확인 기준'은 위헌·위법적인 가이드라인(내부지침, 행정작용)인 "실질적이고 의미있는 인적 개입이 있는지 여부"를 기초로 작성된 것으로 보임. 위 나.에서 살펴본 바와 같이 이는 헌법상 법치주의·법적 안정성 원칙에 위배되어 인권과 기본권을 침해할 우려가 매우 큼.

이러한 내용을 가이드라인에 그대로 둔다면 헌법상 위헌의 문제와 더불어 법적 분쟁이 야기될 것임.

(7) 대출심사

가. 대출 심사 등 금융분야에서 고영향 인공지능시스템의 적용 범위 정의를 협소하고 자의적으로 해석함

「고영향 인공지능 판단 가이드라인(안)」중 7. 대출 심사 부분은 법 제2조 제4호 사.목을 문언 자체에 집착하여 고영향인공지능시스템이 적용되는 범위를 '대출심사''판단 또는 평가'에 한정하여 협소하게 제시하고 있음. 구체적으로는 대출심사를 "금융회사가 신용정보 등에 근거하여 개인의 신용을 평가하고 여신의 가부와 범위를 결정하는 데 직접적으로 관련되는 업무만이 대출심사에 포함된다"고 하면서, 대출 과정 관련 업무인 '대출 상담' , '이상거래탐지서비스', '금융회사 업무 효율화에 인공지능시스템을 활용하는 경우' 등을 제외하고 있음. 또 판단 또는 평가에 대하여 "대출 심사 결정 과정에서 ...단순 참고하는 경우에는 (중대한 영향을 미치는) '판단 평가 목적의 인공지능시스템'에 해당한다고 볼 수 없다"고 정하고 있음.

그러나 법 제2조 제4호 사.목은 열거 조항으로 대출 심사는 "개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가"의 예시를 든 것에 불과하므로 고영향인공지능시스템 적용 범위를 "대출심사" 자체에만 한정할 이유가 없음. 실제 대출심사는 대출상담-신용정보확인-이상거래탐지-대출승인 등과 같은 일련의 유기적인 관계를 거쳐 이루어지는 것이고, 각 단계에서 인공지능시스템을 활용하여 개인의 권리·의무 관계에 중대한 영향을 미치는 판단·평가를 한다면 이는 고영향인공지능 적용범위 내에 포섭된다 할 것임.

법 제2조 제4호 법문을 체계적·종합적으로 해석하더라도 마찬가지임. 법 제2조 제4호는 "고영향 인공지능"에 대하여 정의하면서 "다음 각 목의 어느 하나의 영역에서 활용되는 것을 말한다"고 정하고 있음. 특정 행위를 위해 인공지능시스템을 활용하는 것이 아니라, 당해 분야에 인공지능시스템이 활용된다면 고영향 인공지능에 해당한다고 보아야 함. 이는 법 제33조 내지 제34조에서 고영향 인공지능시스템과 관련하여 특별히 엄격한 적합성 평가와 투명성 책무를 부여한 입법 취지에도 부합함.

나. 고영향 인공지능 확인 기준이 아무 근거 없이 임의적으로 정해짐

「고영향 인공지능 판단 가이드라인(안)」중 7. 대출 심사 부분 "2.7.3. '고영향' 확인 기준"은 "❶대출 심사 중간 단계에서 최종 결정에 상당한 영향을 미치는 결정을 하거나 대출 심사 개별 단계를 종합하여 최종 결정을 하는 인공지능시스템이나 ❷대출 심사에 차별적 데이터 또는 민감한 데이터를 사용하는 인공지능시스템의 경우 '고영향 인공지능'에 해당한다"고 제시함. 심지어 "A그룹 2개 항목 이상에 해당하거나 A그룹 1개 및 B그룹 2개에 해당하는 경우 '고영향 인공지능'에 해당한다"고 하는데, 그 근거가 무엇인지 전혀 설명하지 않고 있음.

그런데 ❶ 내지 ❷의 '중간단계','상당한','최종결정'이라는 기준은 추상적이고 불명확하여 어떠한 기준점이 될 수 없음. 더욱이 위 기준은 기준을 설정한 근거나 맥락을 전혀 제시하지 않고 있음. EU AI Act에서 '신용평가시스템'을 고위험 인공지능 시스템(High-Risk AI

System)으로 분류한 이유는 해당 시스템이 ▲부당한 대출 거부 ▲사회적 차별 ▲기본적 생계 위협 등을 가능하게 하고 있기 때문임. 위와 같은 위험은 대출 심사 전반 과정에서 일어날 수 있으며, 잘못된 결과가 개인의 삶에 회복 불가한 피해를 주기 때문에 사전에 위험 평가와 감독·관리가 필요하므로, 이러한 점을 고려하더라도 고영향 인공지능 확인 기준을 아무 근거 없이 임의로 정하는 것은 부당하다고 사료됨.

(8) 교통 : 차

가. 부분 자율주행시스템, 인공지능 시스템이 보조수단으로 적용되는 경우를 '고영향'에서 제외시킬 수 있는 근거가 없음.

가이드라인(안)에서는 조건부 완전자율주행시스템 이상의 자율주행시스템과 관련한 인공지능 시스템을 고영향 인공지능으로 분류하면서, 이보다 낮은 수준의 '부분 자율주행시스템'에 대하여 ① 인공지능이 도로주행 제어에 관여하지 않고, ② 관련 법령상 안전기준을 준수하게 되면 피해정도를 현저히 낮출 수 있는 경우라는 두가지 요건을 충족하는 경우 고영향 인공지능에 해당하지 않는다고 설명하고 있으나, 제시된 두가지 요건은 고영향의 예외를 광범위하게 설정하면서도 그 자체로 의미가 명확하지 않으며 이를 고영향에서 제외시킬 법상의 근거 또한 부재함.
부분 자율주행시스템(레벨3)이라 하더라도 작동한계상황 등이 아닌 일반적인 경우 운전자의 개입없이 자동차를 운행할 것을 가정하고 있을 뿐만 아니라, 인공지능이 도로주행의 제어에 관여하지 않더라도 인공지능이 인지 및 판단에 개입하도록 되어 있는 경우 오류나 부정확성으로 인한 사고 발생 등 안전에 중대한 영향을 미칠 가능성이 있다는 점에서 고영향으로 분류되어야 함.
한편, 가이드라인(안)에서는 인공지능기술이 도로주행에 관여하는 경우에도, 자율주행시스템이 미적용된 차량의 경우 고영향에서 제외된다고 설명하고 있으나, 인공지능기술이 차량 운행의 주요한 기능인 도로 주행에 관여하는 이상 그 정도가 제한적이라 하더라도 신체의 안전과 위험 등 영향을 미칠 가능성이 상당하다는 점에서 고영향에 포함되어야 마땅함. 가이드라인(안)에서 이를 제외시키는 것은 상위 법령에 반하고, 인권보호의 측면에서도 바람직하지 못함.

나. 교통시설, 교통체계에 관한 인공지능시스템 중 일부를 제외시킬 수 있는 근거가 부재함.

가이드라인(안)에서는 교통시설, 교통체계에 적용된 인공지능시스템에 대해서도, 법률로 정한 성능평가, 안전기준이 존재하고 이를 충족하면 사람의 생명, 신체의 안전에 중대한 영향을 미칠 가능성이 극히 낮다고 평가될 수 있는 경우 고영향 인공지능에 해당하지 않는다고 설명하고 있으나, 이러한 설명은 고영향의 예외를 설정하면서도 그 자체로 의미가 명확하지 않으며 이를 고영향에서 제외시킬 법상의 근거 또한 부재하다는 점에서 적절한 설명으로 볼 수 없음.

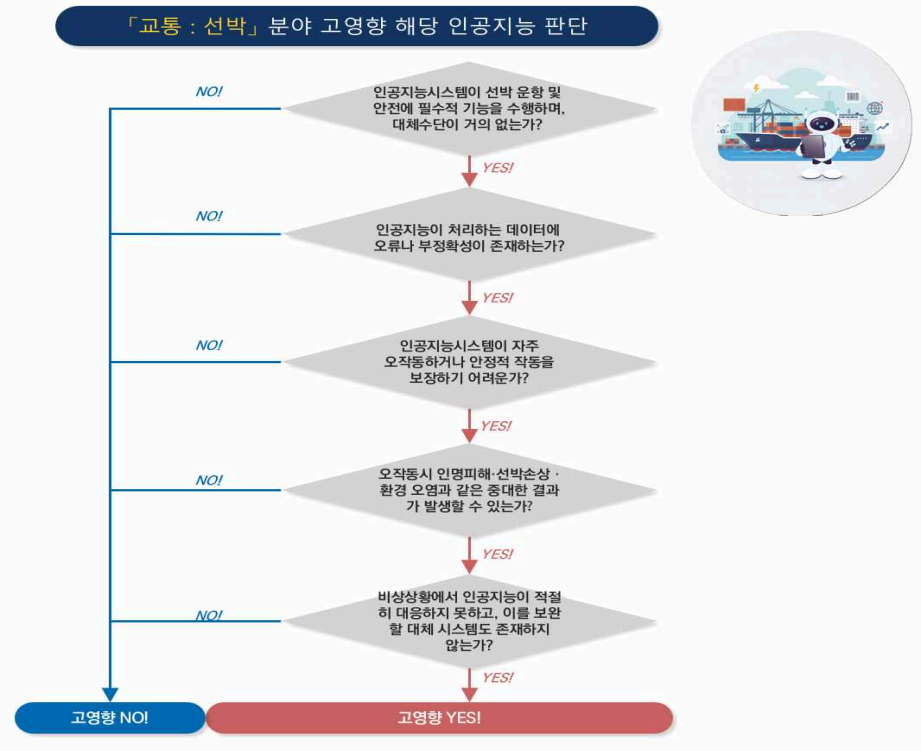
(9) 교통 : 선박

법제33조에 제1항에 따라 인공지능사업자가 과기부 장관에게 고영향 인공지능인지 여부의 확인을 요청하면 과기부장관은 고영향 인공지능 해당 여부를 확인하여야 함. 이 경우 전문위원회를 설치하여 자문을 받을 수 있고 과기부장관은 고영향 인공지능의 기준과 예시 등에 관한 가이드라인을 수립하여 보급할 수 있다(제3항)고 규정하고 있어 가이드라인 수립의 법적 근거가 제시되고 있음.

또한 과기부가 2025.9.17. 공개한 고영향 인공지능 판단 가이드라인(159쪽)에서 “본 가이드라인은 인공지능기본법의 입법목적에 따라 사업자가 책무 유무를 스스로 판단할 수 있도록 고영향 인공지능의 기준과 예시 등을 마련하여 사업자의 예측 가능성을 제고하고 산업현장에서의 혼란을 방지하기 위한 목적으로 마련되었다”고 밝히고 있음.
목표 설정은 법과 시행령의 규정에 따라 고영향 인공지능의 확인에 있어서 기준과 예시 등에 대해 가이드라인을 수립, 보급한다는 취지를 밝히고 있음.
위임법령에 따른다면 과기부장관이 사업자가 스스로 판단할 수 있도록 하는 기준, 방법을 마련하여 보급할 수 있고, 이때 사업자의 예측가능성을 높이기 위해서는 정확하고 명료한 기준제시가 있어야 할 것임. 특히 선박,항공, 철도 분야의 경우, 그 영역 특성상 과기부장관이 전문성이 있다고 보기 어려운 분야이기 때문에 관련부처-해양수산부, 국토교통부 및 관련 분야의 전문성과 경험을 보유한 전문가와의 협력이 필요함. 따라서 가이드라인 및 취지에 이와 같은 부처 및 전문위원회와의 협력 등에 대한 내용도 포함되는 것이 필요해 보임.

나. 선박 영역 가이드라인(안)주요 내용 및 문제점

가) 고영향 판단기준으로 제시한 항목의 내용이 명확하지 않고 추상적, 포괄적임



고영향 확인 기준으로 사업자가 자율적으로 체크하는 사항에는 5가지 항목을 제시하고 있음. 그런데 선박법에 따라 선박은 설계, 건조, 검사 및 운항과 관련된 최초검사 및 정기검사를 받도록 하고 있음. 따라서 가이드라인은 이와 같은 선박의 최초 및 정기검사에 준하는 항목별로 세부적 고영향인공지능 판단 기준을 제시해 주어야 한다고 보여짐. 각 항목의 내용도 추상적이고 포괄적임.
특히, 선박 고영향 확인 기준의 예시로 자동보고기술을 사용한 인공지능시스템을 통한 원격관제는 고영향인공지능이 아니라고 하고 있음. 이유는, 해당 시스템의 주요 역할은 관제 지원으로, 선박 운항과 안전에 필수적인 자율 제어 기능을 직접 수행하지 않아 오작동 시 위험이 발생할 가능성은 있으나, 비상시 원격운항자가 직접 대응할 수 있다는 것임.

그러나 선박 원격관제시스템은 자율운항선박의 핵심 기술이라고 알려져 있고, 관제는 해양 사고 예방, 해상 사고 및 선박사고를 실시간 모니터링해 인명사고를 예방하거나 피해를 최소화하는 한편 해상에서의 교통 흐름을 원만하게 지원하는 역할 및 기상 상황, 주변 교통 상황, 위험 정보 등 항해자가 안전한 의사 결정을 내리는 데 필요한 중요 정보를 적시에 제공함. 또한 원격관제시스템은 자율운항선박에서 필수적인 요소임에 따라 관제를 지원한다는 이유로 인공지능에 의해 운영되는 관제시스템을 고영향이 아니라고 판단하는 예시는 임의로 선박 영역에서의 고영향 인공지능 범위를 축소하는 것임. 자율운항뿐 아니라 관제 영역에 운용되는 인공시스템도 고영향인공지능으로 판단할 수 있도록 정교한 체크항목을 제시하는 것이 필요함.
예컨대 순서도에 따르면 첫번째 항목인 인공지능시스템이 선박 운항 및 안전에 필수적 기능을 수행하며, 대체수단이 거의 없는가라고 묻고 아니면 고영향이 아닌 것으로 판단하도록 하고 있음. '거의 없는가'라는 질문은 사업자의 자체적 판단에 따라 달라질 수 있는 것임. 또 그 다음 항목은 인공지능이 처리하는 데이터에 오류나 부정확성이 존재하는가인데, 만약 아니라고 기업이 판단하면 고영향이 아닌 것으로 판단하도록 함. 그러나 오류나 부정확성이 존재하는지 아닌지 판단을 사업자가 하도록 하면 대부분의 사업자는 아니라고 할 것임. 인공지능의 오류가능성 부정확성은 개발자조차 예측하기 어려운 측면이 있기 때문에 법에서 고영향을 정의한 취지가 사람의 생명, 신체안전, 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능은 특별히 관리하여야 한다는 취지임을 고려할 때 이는 부적절함.

나) 고영향이 아니라고 하는 판단의 법적 근거가 없고 자의적으로 고영향 인공지능의 범위를 축소하고 있음

고영향인지 아닌지를 판단하는 항목이 5가지로만 제시되어 있고 각각의 항목 자체만으로도 고영향이라고 판단해야 하는 내용임. 또한 이 다섯가지 항목만으로 고영향 인공지능을 평가하는 것이 적절한지 의문이고 이 외의 경우의 수가 있을 수 있는데 이렇게 함으로써 고영향의 범위를 축소하는 것은 근거가 없다고 보여짐.
고영향은 법2조 "사람의 생명, 신체 안전, 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 AI"라고 하였는데, 기본권에 대한 중대한 영향, 그로 인해 초래할 위험에 대해 사업자 등이 판단할 수 있으려면 중대한 영향에 대한 평가기준, 위험에 대한 평가기준이 있어야 할 것임.
가이드라인은 법적 구속력은 갖지 않지만 상위법에서 위임한 이상 기업들에게 법규적 효력을 발생할 수 있고, 무엇보다 그동안 기업들에게 가이드라인은 법적 책임의 면책 근거가 되어 온 것도 사실임. 따라서 기업들이 자율적으로 판단하는 근거와 기준을 제시하겠다는 취지로 만들어졌으나 사실상 법의 취지를 따르지 않거나 임의로 판단근거를 제시하는 것은 위험함. 무엇보다 가이드라인은 고영향 판단여부를 과기부장관에게 요청하기 전 사업자 스스로 판단하는 단계이고 만약 기업 스스로 가이드라인에 따라 고영향이 아니라고 판단했는데 실제 국민의 생명, 안전, 기본권에 중대한 영향을 끼쳤을 때 그 책임소재, 구제 등을 어떻게 할 것이냐도 중요한 문제임.

세번째 항목인 '자주 오작동하거나 안정적 작동보장이 어려운데, 오작동해도 인명피해나 선박손상 환경오염과 같은 중대한 결과가 발생할 수 없다면 고영향이 아니'라고 한다면, 선박 분야 자체가 인명사고로 이어지거나 환경오염의 위험성이 있는 영역이라 고영향으로 정의되었음에도 고영향이 아니라고 판단하는 것은 법취지를 형해화 시키는 것이며 임의로 고영향 인공지능의 범위를 축소하는 것임

네번째 항목인 ‘오작동시 인명피해 선박손상 환경오염과 같은 중대한 결과가 발생할 수 있는 인공지능이면서 비상상황에서 인공지능이 적절히 대응하고 이를 보완할 대체 시스템이 존재한다고 하면 고영향이 아니’라고 판단하도록 한 것은 사업자의 자의적 판단에 따라 자칫 고영향임에도 적절한 책임과 의무를 다하지 않을 가능성을 열어두는 것이라 문제가 있음

(10) 교통 : 항공

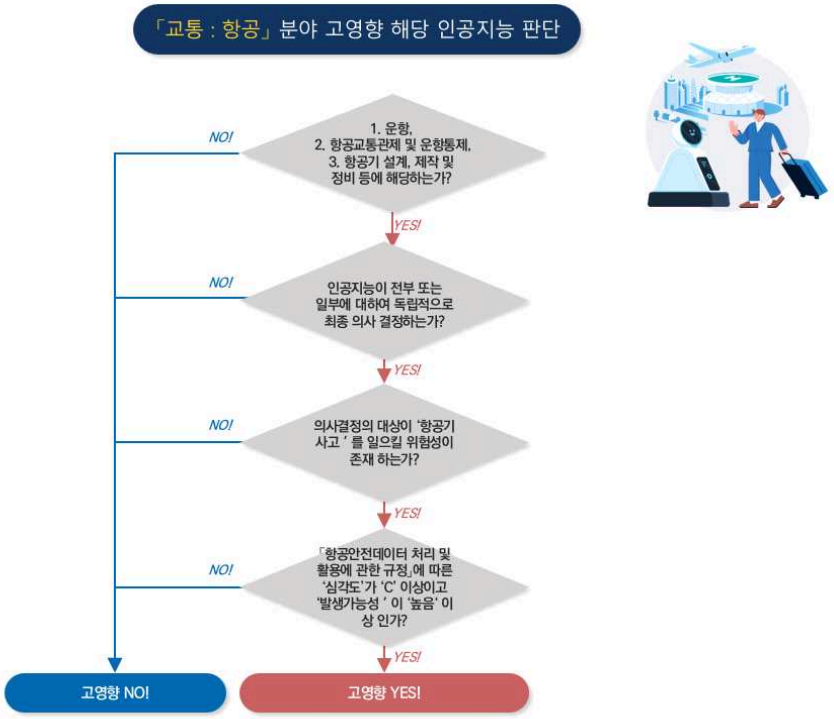
항공교통 분야에서 항공기의 직접적인 운항, 안전시설 및 장비 운용에 사용되는 인공지능시스템이고, 항공교통 분야는 교통수단의 특성상 인공지능시스템의 오류 또는 잘못된 판단으로 인해 다수의 인명 피해를 수반하는 심각한 사고가 발생할 가능성이 크므로, “항공교통 분야에 사용되는 인공지능은 모두 ‘고영향 인공지능’에 해당한다고도 볼 수 있을 것이라고 하면서, 다만, 항공교통 분야에서 사용되는 현행 항공교통 관련 법령에서 정하고 있는 위험도 기준을 적절하게 반영하여 ‘고영향 인공지능’ 해당여부를 판단한다”고 하고 있음.

가. 고영향 인공지능의 기준 모호, 사업자의 자의적 판단에 따라 제외될 가능성이 큼

항공교통 분야의 특수성에 따라 항공교통 분야에 사용되는 모든 인공지능을 고영향인공지능으로 보는 것은 적절하면서도, 현행 항공교통 관련 법령에서 정하고 있는 위험도 기준을 적절하게 반영하여 ‘고영향 인공지능’ 해당여부를 판단한다고 함으로써 “적절하게” 반영하여 판단하라는 것은 다소 모호함. 기본전제가 이 분야는 다 고영향이다라고 하면서, 다시 위험도 기준을 반영해서 판단하라는 것은 위험도가 낮으면 고영향이 아니라는 것인지 모호함. 사업자가 제대로 명확하게 판단할 수 있도록 명확한 기준을 제시해야 함.

인공지능의 속성상 오작동의 가능성, 예측 불허성이 존재하는데, 순서도 항목 중 “인공지능이 일부 또는 전부에 대하여 최종 의사결정 하지 않으면 고영향인공지능이 아닌 것”으로 판단하라는 것인데 이때 일부의 기준이 무엇인지 모호한 기준으로 보여짐. 의사결정의 대상이 항공기 사고를 일으키지 않을 것이란 판단도 사업자가 자체적으로

판단해서 고영향인지 아닌지 판단하는 것도 자의적 판단에 따라 위험성을 낮게 판단할 수



있음.

나. 고영향 인공지능의 예외가 자의적

순서도에서 심각도가 C이상이고 발생빈도가 높음 이상일 경우 고영향으로 판단하도록 하고 있음. 위에서 언급했듯이 항공교통은 조금만 위험요인이 있어도 큰 사고로 이어질 수 있으므로 최대한 보수적으로 판단하여 발생빈도를 보통이상으로는 해야 한다고 보여짐. 역시 임의로 고영향을 축소하는 것으로 법적 근거가 없어 문제가 있음.

〈표 5. 위해요인 별 발생결과의 빈도〉

구분	발생 가능성	정량적 판정 기준	
		1회 이상 발생하는 비행시간	2백만 비행시간으로 환산
5	매우 높음 (Frequent)	~1만 시간 미만	일 단위
4	높음 (Occasional)	1만 시간 이상~10만 시간 미만	2주 단위
3	보통 (Remote)	10만 시간 이상~100만 시간 미만	월~분기 단위
2	낮음 (Improbable)	100만 시간 이상~1,000만 시간 미만	반기~연간 단위
1	매우 낮음 (Extremely Improbable)	1,000만 시간 이상~	5년 이상에 해당하는 단위

- 비해당 사례

C 항공사는 D 인공지능시스템을 통해 운항 중인 객실 내에서 승객의 행동을 분석하여 비상 상황에 대비하여 신속한 대처방안을 승객에게 제시하고 있다.

(검토) D 인공지능시스템은 승무원의 업무를 지원하고, 비상 상황에서 빠르고 정확한 대응을 할
수 있도록 하여 객실 내 승객들의 안전을 보장하는 시스템이다.

인공지능시스템은 기내 CCTV를 통해 승객들을 감시하여 승객의 건강과 위험한 행동을 사전에 발견하고, 객실 내의 온도, 습도, 압력 등을 모니터링 하여 객실 내에서 발생할 수 있는 비상 상황을 감지하고 승무원에게 경고를 보내거나 승객들에게 비상 안내를 할 수 있다.

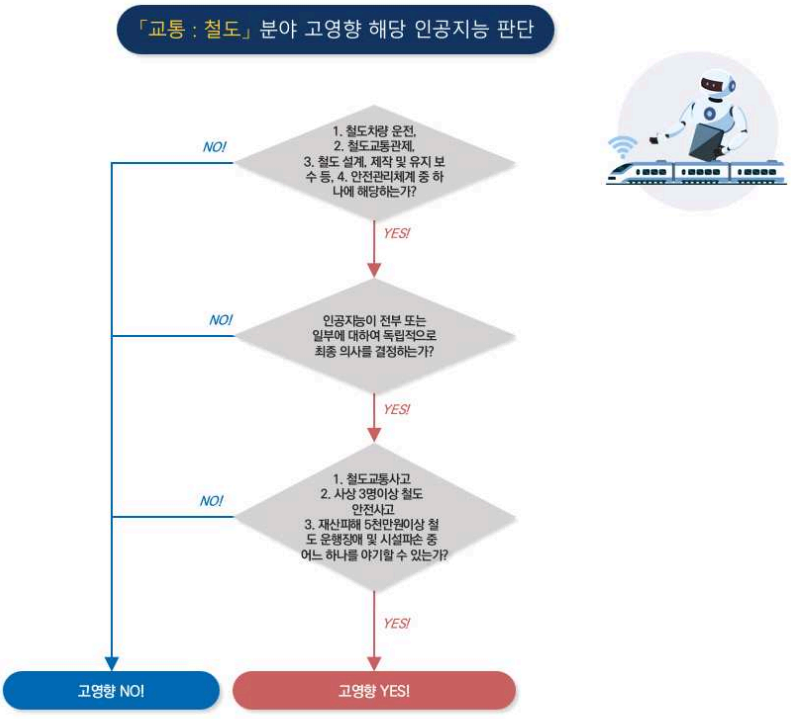
(결론) 따라서 항공교통의 위험에 영향을 미치는 ①운항 ②항공교통관제 및 운항통제 ③항공기
설계·제작 및 정비 등 분야에서 사용되는 인공지능시스템에 해당하지는 않으므로, D 인공지능시스템은 고영향 인공지능에 해당하지 않는다.

위 예시에서 제시한 사례는 비록 인공지능시스템이 항공 운항과 관련된 것이 아니라 기내 승무원 업무 지원이고, 비상시 승객안정 보장 시스템이긴 하지만, 기내 승객 감시 및 비상상황 감지하고 승무원에게 경고를 보내고 비상안내를 할 수 있도록 하는 것인데, 만약 오작동하여 비상상황임에도 경고를 하지 않는다면 승객들의 안전에 위험이 될 수 있다면 상식적으로 고영향으로 판단하는 것이 적절해 보임. 특히 “항공교통 분야에 사용되는 인공지능은 모두 ‘고영향 인공지능’에 해당한다고도 볼 수 있을 것이다” 라고 밝히고 있는 점과도 모순됨. 따라서 근거없이 고영향에서 제외하는 것은 법의 취지에도 맞지 않음.

(11) 교통 : 철도

철도분야 전반에 사용되는 인공지능은 오류발생시 철도사고를 유발할 수 있으므로 고영향 인공지능에 해당할 가능성이 높다고 하고 있음. 즉, “인공지능시스템은 철도 분야에서의 위험을 낮추는 목적으로 도입되고 있지만 비상대응에 있어 인공지능시스템의 도입이 오히려 기존 비상대응 시스템을 무력화함으로써 위험을 증가시킬 여지가 있어 ‘고위험 인공지능’에 해당한다고 서술하고 있음.

가. 고영향 인공지능 판단 기준이 협소함



철도사고는 법이 일정한 의무를 부여하는 것이 정당화될 만큼의 위험이 인정된다고 할 수 있으므로 이와 관련된 사고가 인공지능시스템에 기인한다면 '고영향 인공지능'에 해당한다고 하고 있음. 따라서 철도사고 또는 유사한 기준에 따라 고영향을 판단하도록 하고 있으면서 철도교통의 위험에 영향을 미치는 ① 철도운행관리 및 유지보수, ② 철도교통관제, ③ 철도시설의 설치·운영 및 유지보수, ④ 철도안전관리 분야에서 사용되는 인공지능시스템 중에서도, 인공지능이 전부 또는 일부에 대하여 독립적으로 최종 의사결정을 하며, 인공지능의 의사결정으로 '화재·충돌·탈선사고' 등의 철도교통사고를 일으킬 위험성이 존재하는 경우에 또는 3명 이상의 사상자가 발생한 철도안전사고를 야기할 수 있는 경우, 또는 재산피해 5천만원 이상의 철도 운행장애 및 시설파손을 유발할 수 있는 경우에는 고영향 인공지능에 해당한다고 함. 이외의 경우는 고영향이 아니라는 것이라면 이는 지나치게 고영향 인공지능을 협소하게 규정하는 것임. 사고는 결과일 뿐이고 안전을 위해 전단계에서 인공지능이 관여하는 것이라면 고영향으로 보는 것이 적절해 보임.

나. 고영향 인공지능의 예외를 넓게 열어둠

철도 분야 전반에 사용되는 인공지능은 오류발생시 철도사고 유발 가능성이 있으므로 전반적으로 고영향 인공지능이라고 하는 것이 적절할 것임. 따라서 순서도에서 제시한 항목 중 ① 철도운행관리 및 유지보수, ② 철도교통관제, ③ 철도시설의 설치·운영 및 유지보수, ④ 철도안전관리 분야에 전부 또는 일부에 대해 독립적 최종의사결정을 하는 경우는 이하 항목을 만족하든 아니든 고영향으로 하는 것이 적절해 보임. 그 아래 항목에서 열거한 대로 철도사고로 이어질 가능성이 있어야만 고영향으로 평가하도록 하는 것은 예외를 설정하는 것으로 법의 취지를 몰각한 것이며, 무엇보다 그렇게 할 정당한 근거가 없음.

순서도의 두번째 항목은, 독립적인 최종의사결정을 인공지능이 하지 않는다고 하더라도 의사결정에 주요한 근거를 제시한다면 이를 토대로 철도운행, 시설운영, 교통관제 등을 한다면 이는 철도사고로 이어질 수 있으니 이는 고영향 인공지능으로 보아야 함에도 이 항목에 해당하지 않는다는 이유만으로 고영향이 아닌 것으로 판단하는 것은 적절하지 않고 근거도 없어 보임. 따라서 아래 비해당 사례로 제시한 사례의 경우 "이는 열차운행에 방해되는 지장물을 발견하면 영상과 알람을 작업자에게 실시간 전송하게 되는데, 적절한 선로유지보수가 이루어지지 않았다면 궤도를 이탈하는 탈선사고를 유발할 수 있다. 그러나, 해당 기술은 인공지능이 열차운행에 방해되는 지장물을 알려주는 역할을 할 뿐 독립적인 최종 의사결정을 하지 않는다"며 고영향이 아니라고 한 예시는 정정되어야 할 것임.

- 비해당 사례

B 교통공사는 LTE 통신망과 카메라, 라이다(Lidar) 센서를 장착한 철도시설물 '자율주행 점검 로봇'을 선로유지보수 점검 서비스에 활용하고 있다. 이 로봇은 지정된 장소까지 선로를 자율주행하며, 열차 운행에 방해되는 지장물을 발견하면 영상 및 알람을 작업자에게 실시간으로 전송한다.

(12) 공공서비스

가. 행정기본법 제20조에 따라 이루어지는 자동적 처분 일체를 고영향 범주에서 제외시킬 근거가 없음.

가이드라인(안)에서는 '국가기관 등이 사용하는 인공지능 중 공공서비스 제공에 필요한 자격확인, 결정 또는 비용징수 중 어느 하나에 해당하는 인공지능시스템'은 고영향

인공지능에 포함된다고 설명하면서, ‘활용’의 의미와 관련하여 「행정기본법」 제20조의 자동적 처분 등 기속행위를 제외한 나머지 공공서비스 영역에서 사용되는 인공지능의 경우에만 고영향 인공지능에 해당할 가능성이 있는 것과 같이 설명하고 있음.
그러나 「행정기본법」 제20조는 개별 공무원의 의사가 개입될 여지가 없는 기속행위의 경우 인공지능 시스템을 활용하여 자동적 처분을 할 수 있도록 허용하고자 마련된 규정일 뿐, 기속적 행정행위가 이루어지는 영역 일체에 대해서 아무런 사전, 사후 규제없이 인공지능 시스템을 도입할 수 있다는 취지의 규정이 아님. 즉, 사람의 재량이나 판단여지가 존재하지 않는 경우(완전 자동화된 의사결정이 가능한 경우)에만 인공지능 시스템에 의한 자동화된 처분이 가능하도록 하자는 취지에서 마련된 「행정기본법」 제20조를 근거로, 공공서비스와 관련한 기속행위 영역 일체에 대해서 고영향 인공지능에 해당할 가능성이 없는 것처럼 설명할 수 없음.
예를 들어, 법령에 따른 자격요건을 갖춘 신청자에게 반드시 지급하도록 규정되어 있는 사회복지급여의 지급 결정은 기속행위에 해당할 수 있는데, 해당 지급결정이 인공지능시스템에 의해 자동적 처분 형태로 이루어지는 경우, 해당 인공지능시스템은 대상자의 권리, 의무에 직접적인 영향을 미칠 수 있기 때문에 고영향 인공지능으로 분류되어야 마땅함. 즉, 자동적 처분이 이루어지는 기속행위라고 하여 고영향 인공지능에서 제외된다는 설명은 상위법인 인공지능법 제2조 제4호 자목의 문언에 반하는 자의적인 해석이므로, 「행정기본법」 제20조의 자동적 처분 등 기속행위로 이루어지는 공공서비스 영역에서 활용되는 인공지능 시스템이 고영향 인공지능에 해당될 가능성이 없다는 취지의 설명은 가이드라인에서 삭제되어야 함.

나. 자격확인, 비용징수와 관련하여 고영향의 범주를 제한하는 취지의 설명에는 아무런 근거가 없음.

가이드라인(안)에서는 국가기관등에 의해 이루어지는 공공서비스 제공에 필요한 자격확인, 비용징수는 원칙적으로 고영향에 해당한다고 하면서도,
거주지·국가유공자여부·실업급여수급 등의 수급자격 확인 등의 경우 해당 자격·사실이 객관적이고 단순하여 국가기관 등의 판단 재량 여지가 없다고 하며 고영향에 해당하지 않는다고 설명하고 있음. 또 비용징수와 관련하여서도, 비용이 ‘정액화’되어 있어 국민이 징수액을 쉽게 예상할 수 있고, 결정 과정이 투명한 경우 위험성이 낮다고 하며 고영향에 해당하지 않는다고 설명하고 있음.
그러나 이러한 설명에는 아무런 객관적인 근거가 없고, 상위법인 인공지능 기본법의 문언에도 반함. 인공지능 기본법은 공공서비스의 제공에 필요한 자격확인, 비용징수에 활용되는 인공지능 시스템에 대하여 고영향에 해당한다는 점을 분명히 하고 있음에도 불구하고, 가이드라인(안)에서 그 범주를 임의로 제한하는 것은 상위법에 위배되는 해석으로 볼 소지가 다분함.
가이드라인(안)에서는 자격확인과 관련하여 근거가 되는 자격·사실이 객관적이고, 단순한 경우 위험성이 적다고 하였으나, 객관적이고 단순하다는 개념은 다분히 주관적이고 모호하며, 이 경우 위험성이 적다는 설명 또한 일반화하기 어려움. 자격확인 여하에 따라 공공서비스 제공 대상자에게 미칠 부정적 영향, 차별 발생 가능성 등은 자격 요건의 객관성, 단순성과는 직접적인 관련이 존재하지 않음. 따라서 자격 요건의 객관성, 단순성은 기껏해야 위험성을 판단하는 요소 중 일부로 볼 수 있을 뿐, 위험성 정도를 결정하는 충분조건이 될 수 없음.
가이드라인(안)에서는 비용징수와 관련하여, 비용이 ‘정액화’되어 있는 경우 위험성이 낮다고 하나, ‘정액화’라는 개념은 그 자체로 모호하며, 비용이 정액화되어 있다고 하여 그로 인한 위험성이 낮다고 설명하고 있으나, 이러한 설명 또한 근거가 없음. 비용 징수의 사유,

비용 징수 금액의 다소, 대상자의 범위 등에 따라 위험성은 달라질 수 있으므로, ‘정액화’라는 단편적인 기준에 따라 위험성을 일률적으로 구분할 수 없다 할 것임.

(13) 교육

가. “교육평가”의 의미를 자의적으로 축소 해석함

교육기본법 제9조 제1항은 “유아교육・초등교육・중등교육 및 고등교육을 하기 위하여 학교를 둔다.”고 정하고 있음. 이에 인공지능기본법 제2조 제4호 차항「교육기본법」제9조 제1항에 따른 유아교육·초등교육 및 중등교육에서의 학생 평가 의미는 각 교육과정에서 이루어지는 평가 일련의 과정 및 그 결과를 포괄하여 의미할 것임.
그러나 인공지능법 시행령 및 가이드라인은 인공지능기본법 제2조 제4호 차항을 “인공지능시스템으로 시행한 학생평가의 결과가 향후 입시·취업 등에 활용되는지”를 기준으로 각 유아교육・초등교육・중등교육 및 고등교육 과정에서 마다 다르게 판단하려 함. 예컨대 유아교육과정에서의 평가에 인공지능시스템을 활용할 경우 고영향인공지능에 해당할 가능성이 있으나, 학생평가 결과가 향후 입시·취업에 활용될 가능성이 낮아 고영향인공지능이 아니라고 판단함.
EU AI Act에서는 교육기관의 감정인식 AI를 법률로 금지하는데, 학교 현장에 도입되는 인공지능 도구와 그 대상이 되는 학생의 권리 역시 이러한 규범으로부터 영향을 받을 수밖에 없기 때문이라고 보기 때문임. 같은 맥락에서 학교에서 학생을 평가하거나 모니터링하는 인공지능시스템 역시 고위험 인공지능으로 분류함.

나. 인간의 개입 없이 자동화된 의사결정으로 평가에 이른 경우에만 고영향 인공지능으로 봄

인공지능법 시행령 및 가이드라인은 인공지능기본법 제2조 제4호 차항과 관련한 사례로, 교사의 별도 개입없이 인공지능 자체의 자동화결정에 따라 평가 결과를 그대로 활용하게 하는 경우에 고영향 인공지능으로 판단하고, 교사의 개입을 통해 오류 발생을 확인, 수정할 수 있어 학생의 입시·취업 등에 활용되지 않으면 고영향 인공지능이 아니라고 판단함.
인공지능시스템이 학생 평가에 활용될 경우 단순히 결과 산출의 정확성 문제뿐 아니라, 인공지능시스템의 편향성, 불투명성 등으로 학생들의 학습권이라는 기본권에 미치는 영향이 더 막대할 것이라 사료됨.
EU AI Act는 이러한 취지에서 학교 교육과정에서 사용하는 인공지능시스템의 경우, 해당 인공지능 시스템을 개발하거나 배치하는 학교 역시 설명, 인간의 관리감독, 문서화 등 일정한 책무를 이행하고 그 영향을 받는 학생의 권리를 보장하여야 한다고 정하고 있음.
하지만 현행 인공지능법 시행령 및 가이드라인은 그에 미치지 못하는 바, 향후 인공지능으로부터 영향받는 사람인 학생의 학습권 등 기본권에 미칠 위험을 예방할 수 없는 한계가 있음.

AI 영향평가 가이드라인(안)

1. 관련 법률 및 시행령(안)

법	시행령(안)
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. ② 국가기관등이 고영향 인공지능을 이용한 제품 또는 서비스를 이용하려는 경우에는 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하여야 한다. ③ 그 밖에 영향평가의 구체적인 내용・방법 등에 관하여 필요한 사항은 대통령령으로 정한다.	제27조(고영향 인공지능 영향평가) ① 법 제35조제3항에 따른 영향평가(이하 “영향평가”라 한다)는 다음 각 호의 사항이 포함되어야 한다. 1. 해당 고영향 인공지능을 이용한 제품 또는 서비스에 의하여 사람의 생명, 신체의 안전 및 기본권에 영향 받을 수 있는 가능성이 있는 개인이나 집단에 대한 식별 2. 해당 고영향 인공지능과 관련하여 영향을 받을 수 있는 기본권 유형의 식별 3. 해당 고영향 인공지능으로 인하여 발생할 수 있는 사람의 기본권에 대한 사회적·경제적 영향의 내용 및 범위 4. 해당 고영향 인공지능의 사용 행태 5. 영향평가에서 활용한 정량적 또는 정성적 평가지표 및 결과산출 방식 6. 해당 고영향 인공지능으로 인한 위험의 예방·완화·손실 복구 등에 관한 사항 7. 영향평가의 결과, 개선이 필요한 경우 그 이행 계획에 관한 사항 ② 인공지능사업자는 직접 또는 제3자에 의뢰하여 영향평가를 실시할 수 있다. ③ 제1항 및 제2항에서 규정한 사항 외에 영향평가에 필요한 사항은 과학기술정보통신부장관이 정하여 보급할 수 있다.

2. AI 영향평가 가이드라인의 이름 변경

AI 영향평가는 다양한 종류가 있을 수 있음. 예를 들어, 위험에 대한 영향평가를 할 수도 있고, 지능정보화 기본법은 ‘사회적 영향평가’를 규정하고 있음. 그런데 인공지능 기본법 제35조 제1항은 고영향 인공지능 사업자에 대해 ‘사람의 기본권에 미치는 영향을 평가’하도록 하고 있음.
이에 가이드라인 이름도 명확하게 ‘AI 기본권 영향평가 가이드라인’으로 하는게 그 취지를 명확하게 드러낼 수 있을 것으로 보임. EU AI Act도 ‘기본권 영향평가(Fundamental Rights Impact Assessment)’로 지칭하고 있고, 가이드라인에서도 글로벌 상호운용성 확보의 중요성을 언급하고 있음.

3. 평가 수행 주체

가이드라인은 "인공지능사업자는 영향평가를 직접 수행하거나, 전문성을 보유한 제3자에 의뢰하여 수행할 수 있음"이라고 규정하고 있음.(p386)

그런데 인공지능사업자가 영향평가를 직접 수행할 수 있다고 하더라도, 이 경우 독립적인 평가가 어려울 수 있음. 따라서 직접 수행하는 경우에는 평가 수행 주체의 독립성을 보장할 것을 권고할 필요가 있음.
전문성을 보유한 제3자에 의뢰하여 수행하는 경우에는 어떠한 전문성을 보유한 기관에 의뢰해야 하는지 명시해줄 필요가 있음. 기본권 영향평가를 위해서는 비단 AI에 대한 전문성 뿐만 아니라, 해당 도메인 분야 전문성 및 인권 전문성이 요구되므로, 수행 기관에 이러한 전문성을 갖출 것을 요구할 필요가 있음.

4. 지원 체계

가이드라인은 "전담기관인 정보통신정책연구원(인공지능 영향평가 센터(가칭))을 통해 가이드라인과 절차 운영에 관한 전문적 지원을 제공"할 것이라고 규정하고 있음. 그런데 인권전문기구인 국가인권위원회는 이미 '인공지능 인권영향평가 도구'를 발표한 바 있음. 인공지능 영향평가 센터(가칭)가 인권 원칙에 기반한 전문적 지원을 충실히 하기 위해 국가인권위원회와 협력하도록 규정할 필요가 있음.

5. 사전 준비 단계

가이드라인은 영향평가의 사전 준비 단계로 평가 필요성 검토와 평가대상 정의만을 포함하고 있는데(p393), 사전에 영향평가를 수행하는 조직의 정책 및 영향평가 수행주체(평가팀)에 대해 검토하고 문서로 정리해놓을 필요가 있음. 즉, 조직에 대해서는 해당 조직이 기본권 영향평가와 관련된 정책을 수립하고 있는지, 영향평가의 결과가 어떠한 절차를 거쳐 조직의 사업에 반영되는지 확인할 필요가 있음. 수행주체(평가팀)와 관련해서는 수행주체가 적절한 역량을 갖추었는지, 적절한 권한과 지원을 확보하고 있는지 등에 대해 준비 단계에서 검토하도록 할 필요가 있음.

6. 평가대상 정의

가이드라인은 평가대상 정의 작성 예시(p394)에서 '6. 이용자 범주・특성'의 설명으로 "현재 제품・서비스를 직접 이용하거나 간접적으로 영향을 받을 주요 이용자 집단과 특성을 작성(연령대, 숙련도, 전문성 수준 등 포함)"으로 서술하고 있음. 그런데 해당 AI 제품이나 서비스의 직접 이용자, 간접적으로 영향을 받을 이용자 집단이 무엇을 의미하는지 모호함. 만일 해당 AI 제품이나 서비스를 직접 운영하는 이용자가 아니라, 그 결과물의 영향을 받는 사람들을 의미한다면, 평가 수행 단계에서 언급하고 있는 '피영향자' 또는 '영향받는 자'로 규정하는 것이 타당함.

7. 이해관계자 및 AI 시스템이 사용되는 사회적 맥락에 대한 파악 필요

가이드라인은 제2절 본 평가 수행 단계에서 '피영향자 분석'과 '평가대상 작동원리 분석'을 언급하고 있으나, 이에 더하여 해당 AI 시스템과 관련된 이해관계자 및 법적, 사회적 맥락에 관련한 정보 역시 파악할 필요가 있음. 영향평가를 수행하는 인공지능사업자가 개발사업자인지 또는 이용사업자인지에 따라 접근할 수 있는 정보에 한계가 있을 수 있고, AI 시스템의 일부는 외주 업체가 관련될 수도 있음. 따라서 영향평가 대상이 되는 AI 시스템과 관련하여 영향평가에 필요한 이해관계자가 누구인지 파악할 필요가 있으며, 영향평가 과정에서 그들에게 자료나 의견을 요청해야할 수도 있음.

또한 평가대상의 작동원리 뿐만 아니라 해당 AI 시스템이 사용되는 법적, 사회적 맥락 등에 대한 정보도 파악할 필요가 있음. 똑같은 시스템이더라도 특수한 사회적, 문화적, 제도적 맥락에 따라 인권에 미치는 영향이 달라질 수 있음. (p397에서도 '동일한 작동원리를 가진 제품 또는 서비스라 하더라도 맥락에 따라 전혀 다른 위험을 초래할 수 있으므로'라고 하고 있음)

8. 영향받는 자와의 협의를 필수적인 절차의 하나로 명시

유엔 인권영향평가(HRIA)의 토대가 되는 '유엔 기업과 인권 이행원칙(UN Guiding Principles on Business and Human Rights)'은 영향받는 자와의 의미 있는 협의(meaningful consultation)를 기업의 인권 실사(Human Rights Due Diligence) 핵심 요소로 강력하게 권고하고 있음. 영향평가는 인권 실사의 핵심적인 수단이며, AI 기본권영향평가 역시 일반적 인권영향평가 원칙을 존중할 필요가 있음. 따라서 '제2절 본 평가 수행 단계'에서 필수적인 절차의 하나로 영향받는 자 또는 이를 대변하는 인권단체와의 협의를 포함할 필요가 있음.

참고로 p404에서 "영향평가 과정에서 수집된 피드백(내부・외부 전문가, 이해관계자 의견 포함)을 반영하여"라고 하고 있는 바, 이러한 피드백 과정을 제2절에서 절차로 명확하게 규정할 필요가 있음.

9. 시나리오 기반 위협 분석

시나리오 기반 위협 분석(p397)에서 해당 시스템이 의도된 목적에 따라 작동하더라도 위험이 발생할 수 있을 뿐만 아니라, 기능상 결합으로 오작동하는 경우, 또는 이용자의 실수나 악의에 의해 의도된 목적과 달리 사용되는 경우에도 위험이 발생할 수 있으므로, 이러한 경우에 발생할 수 있는 위험 시나리오를 모두 포괄하도록 제안할 필요가 있음.

10. 관련 기본권의 식별

가이드라인은 '3.2 관련 기본권의 식별'에서 "평가주체는 제품 또는 서비스를 직접 설계하며 데이터의 수집부터 활용에 이르는 전 과정을 사전에 면밀히 검토하므로 인공지능기술의 작동으로 영향받는 기본권을 어렵지 않게 식별 가능"(p399)이라고 언급하고 있으나, 영향평가 주체는 인공지능 사업자이고 여기에는 인공지능 개발사업자 뿐만 아니라 이용사업자 역시 포함할 뿐 아니라, 해당 AI 시스템의 일부는 외주를 줄 수 있으므로, 평가주체가 항상 제품 또는 서비스를 직접 설계하는 주체라고 볼 수는 없음.

11. 영향받을 수 있는 기본권을 자유권에 한정한 문구 삭제

가이드라인은 "※ (주의)「인공지능기본법」 제35조 제1항은 "기본권"으로 규정하고 있지만, 국가의 일정한 행위나 급부에 대한 적극적 청구권인 사회권적 기본권의 포함 여부는 심층적인 논의가 필요하므로, 현 단계에서는 자유권적 기본권에 한정"이라고 언급하고 있는데, 이는 가이드라인에서 법률의 위임을 벗어나서 규정하는 것일 뿐만 아니라, 기본권 영향평가를 자유권적 기본권에 한정할 이유도 없음. 이렇게 될 경우 노동권, 환경권 등 AI의 영향을 받을 중대한 기본권을 배제하게 될 우려가 있음. 따라서 이 문구는 삭제되어야 함.

12. 제품 또는 서비스 차원에서 영향 받을 수 있는 기본권

가이드라인은 p400 표에서 AI 시스템에 의해 영향 받을 수 있는 기본권 사례를 예시로 들고 있는데, 사회보장 시스템의 편향으로 인한 '차별받지 않을 권리'의 침해, 원격 얼굴인식 시스템 등의 활용에 의한 '집회 시위의 자유' 침해 등도 포함할 것을 제안함.

13. 인공지능 영향평가 작성 예시

예시이기는 하지만 '인공지능 영향평가 작성 예시(p402)'에서 '영향 규모' 개념에 중대성(영향이 얼마나 심각한지)과 영향의 범위(해당 시스템이 미치는 영향의 범위가 얼마나 광범위한가)가 혼재되어 있음.

또한, 국내외 AI 인권영향평가 사례가 많지 않으므로, 예시로 한국 국가인권위원회의 인권영향평가도구도 소개할 필요가 있음.