

2026

**인공지능
영향평가
가이드라인**



과학기술정보통신부

KISDI 정보통신정책연구원

목차

제1장 총론

- 제1절 인공지능 영향평가 개요
- 제2절 추진 근거 및 용어 정의
- 제3절 타 영향평가와의 관계

제2장 인공지능 영향평가 수행 단계별 주요 고려사항

- 제1절 사전 준비 단계
- 제2절 본 평가 수행 단계
- 제3절 사후 단계

부록

1. 인공지능 영향평가서 양식 및 작성 예시
2. 인공지능시스템 기능별 침해 발생 가능 작동원리 예시
3. 인공지능 영향평가 지원 안내
4. 인공지능 영향평가 FAQ

제1장

총론

- 제1절 인공지능 영향평가 개요
- 제2절 추진 근거 및 용어 정의
- 제3절 타 영향평가와의 관계

제1장

인공지능 영향평가 개요

제1절 인공지능 영향평가 개요

1. 개념

- 인공지능 영향평가(이하 “영향평가”)
 - 인공지능사업자가 고영향 인공지능을 이용한 제품 또는 서비스를 제공하는 경우, 사전에 해당 제품 또는 서비스가 사람의 기본권에 미치는 영향을 평가하는 절차



고영향 인공지능 영향평가(법 제35조)

- 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.

2. 목적 및 필요성

- 사업자가 영향평가를 통해 인공지능 제품·서비스가 인간의 존엄성과 기본권에 미치는 잠재적 영향을 사전에 식별·분석하고, 이를 존중·보호하는 방향으로 스스로 개선방안을 마련·이행하도록 유도
- 국민의 기본권을 보호함과 동시에 글로벌 상호운용성 확보
 - 이를 통해 국제사회에서의 신뢰를 강화하고, 역외 규제에 대응하며, 글로벌 인공지능 시장 진출 시 제도적 장벽을 최소화하는 것을 목표로 함
 - ※ 유럽연합(EU)은 「유럽연합 기본권 헌장(Charter of Fundamental Rights of the European Union)」 등 법적 체계에서 보장하는 인간 존엄성, 자유, 평등, 시민의 권리를 포함하는 기본권이 효율적으로 보호될 수 있도록 ‘기본권 영향평가(Fundamental Rights Impact Assessment)’ 제도를 EU 인공지능법(EU AI Act) 제27조에 규정
 - ※ 국내 영향평가는 EU 기본권 영향평가와 상호운용성을 지속적으로 확보해 나갈 예정
- 영향평가는 고영향 인공지능에 국한되지 않고 모든 인공지능 제품·서비스에 대해 자율적으로 수행할 수 있음. 이러한 자율적 영향평가는 인공지능 제품·서비스의 신뢰성을 높이고 글로벌 상호운용성을 확보하여 해외 진출 과정에서 규제 대응력을 강화할 수 있음. 더 나아가 선제적 리스크 관리와 자율적 책임경영을 뒷받침하며, 국제 경쟁력 제고의 핵심 수단이 될 수 있음

3. 평가 대상

- 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 “법”) 제35조에 따라, 고영향 인공지능 제품 또는 서비스를 제공하는 인공지능사업자는 해당 제품·서비스에 대한 영향평가 수행에 노력할 것을 권장
 - **(고영향 인공지능의 확인)** 인공지능사업자는 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우, 해당 인공지능이 고영향 인공지능에 해당하는지 사전에 검토해야 하며, 필요 시 과학기술정보통신부장관에게 해당 여부 확인을 요청할 수 있음(법 제33조제1항)
 - ※ 상세한 사항은 「고영향 인공지능 판단 가이드라인」(과학기술정보통신부·한국지능정보사회진흥원, 2026) 참조
- 법에서는 고영향 인공지능 제품·서비스에 대한 영향평가만을 규정하고 있으나, 인공지능의 책임 있는 활용과 사회적 신뢰 제고를 위해 모든 인공지능사업자가 자율적으로 영향평가를 수행·실천하도록 장려
 - 법적 의무가 없는 영역에서도 인공지능 제품·서비스가 인간의 존엄성과 기본권을 침해할 소지가 있는 영향을 사전에 식별·관리하여, 국제 규범 부합을 통한 시장 확대와 브랜드 가치 제고를 지원할 것으로 기대



고영향 인공지능의 확인(법 제33조)

- ① 인공지능사업자는 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 그 인공지능이 고영향 인공지능에 해당하는지에 대하여 사전에 검토하여야 하며, 필요한 경우 과학기술정보통신부장관에게 고영향 인공지능에 해당하는지 여부의 확인을 요청할 수 있다.
- ② 과학기술정보통신부장관은 제1항에 따른 요청이 있는 경우 고영향 인공지능 해당 여부를 확인하여야 하며, 필요한 경우 전문위원회를 설치하여 관련 자문을 받을 수 있다.
- ③ 과학기술정보통신부장관은 고영향 인공지능의 기준과 예시 등에 관한 가이드라인을 수립하여 보급할 수 있다.
- ④ 그 밖에 제1항에 따른 확인 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.



고영향 인공지능의 확인 절차 등(법 시행령안 제25조)

- ① 인공지능사업자가 법 제33조제1항에 따라 고영향 인공지능에 해당하는지에 대하여 확인을 요청하려는 경우에는 별지 서식의 고영향 인공지능 확인 요청서에 다음 각 호의 서류를 첨부하여 과학기술정보통신부장관에게 제출해야 한다.
 1. 해당 인공지능제품 또는 인공지능서비스의 개요서
 2. 해당 인공지능시스템의 개발 및 학습에 사용된 학습용데이터의 개요
 3. 해당 인공지능시스템을 활용하는 과정 및 결과를 확인할 수 있는 자료
 4. 그 밖에 고영향 인공지능 해당 여부의 확인에 참고가 될 수 있는 서류
- ② 제1항에 따른 요청을 받은 과학기술정보통신부장관은 다음 각 호의 사항을 고려하여 고영향 인공지능 해당 여부를 판단해야 한다.
 1. 인공지능이 법 제2조제4호 각 목의 어느 하나의 영역에서 활용되는지 여부
 2. 사람의 생명, 신체의 안전 및 기본권에 초래할 수 있는 위험의 영향, 중대성, 빈도 및 활용 영역별 특수성
 3. 법 제33조제1항에 따라 인공지능사업자가 사전에 검토한 결과
 4. 법 제33조제2항에 따라 전문위원회(이하 “전문위원회”라 한다)의 자문을 받은 경우에는 그 자문결과
 5. 그 밖에 고영향 인공지능의 해당 여부를 확인하는 데 필요한 사항으로서 과학기술정보통신부장관이 정하여 고시하는 사항
- ③ 과학기술정보통신부장관은 제1항에 따른 요청을 받은 경우에는 요청을 받은 날로부터 30일 이내에 화신해야 한다. 다만, 제품 등의 복잡성 및 중요성 등을 고려하여 30일 이내에 화신할 수 없는 경우에는 30일의 범위에서 한 차례 연장할 수 있으며, 이 경우 확인을 요청한 자에게 연장 사유와 연장 기간을 문서로 알려야 한다.

- ④ 제3항에 따라 회신을 받은 인공지능사업자는 회신 결과에 이의가 있을 때에는 회신을 받은 날로부터 10일 이내에 과학기술정보통신부장관에게 별지 서식의 고영향 인공지능 재확인 요청서에 다음 각 호의 서류를 첨부하여 과학기술정보통신부장관에게 제출해야 한다.
1. 기존 확인 요청에 따라 과학기술정보통신부장관이 회신한 결과
 2. 그 밖에 고영향 인공지능 해당 여부의 재확인에 참고가 될 수 있는 서류
- ⑤ 제4항에 따른 재확인 요청을 받은 과학기술정보통신부장관은 전문위원회의 자문을 받아 고영향 인공지능에 해당하는지를 재확인하고, 재확인 요청을 받은 날부터 30일 이내에 그 결과를 회신해야 한다.
- ⑥ 과학기술정보통신부장관은 제1항에 따른 요청 또는 제4항에 따른 재확인 요청과 관련하여 필요한 경우에는 관계 기관에 의견 제출을 요청할 수 있다.

4. 평가 시기

- 신규 인공지능 제품 또는 서비스 제공 전

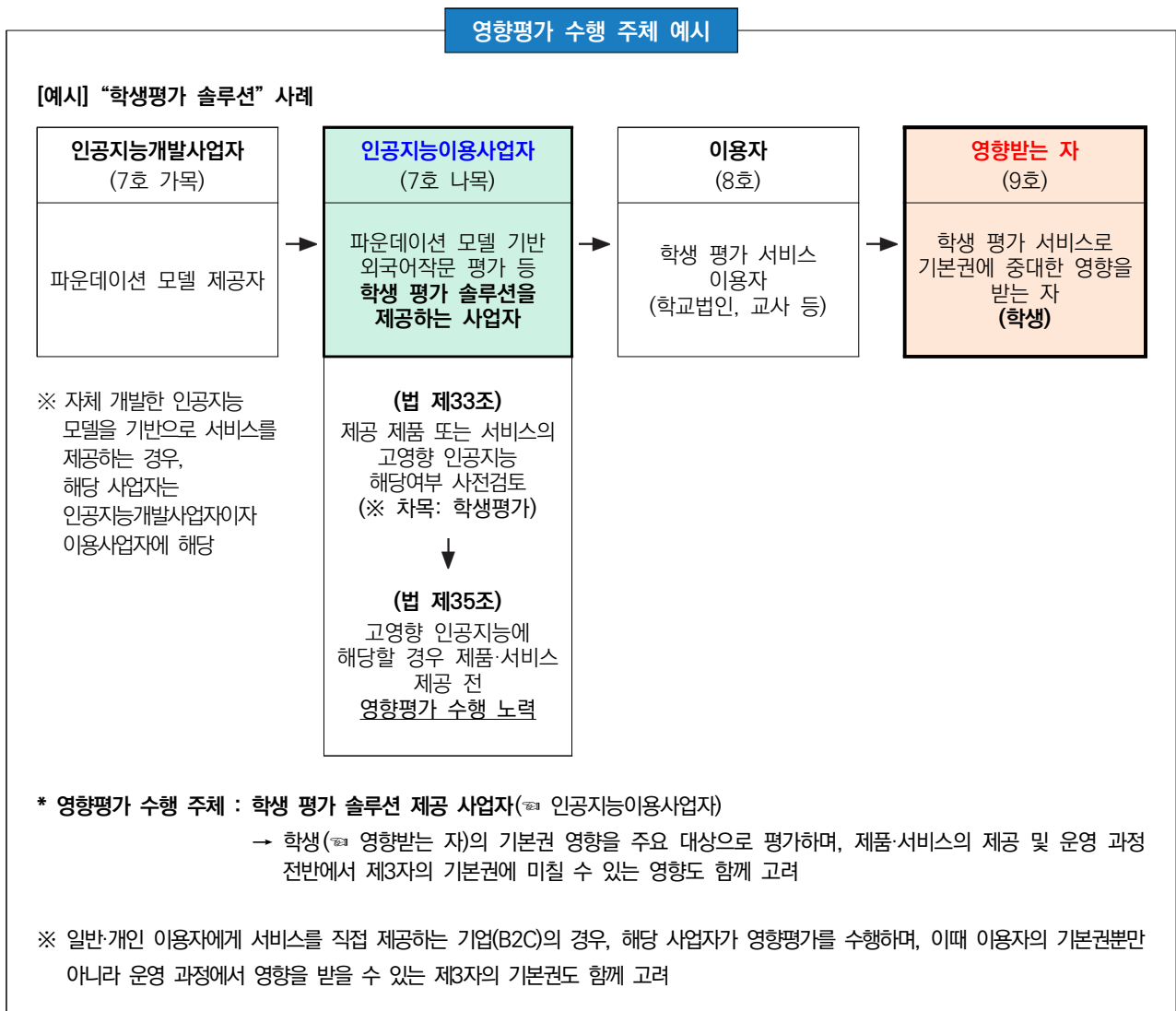
- 사업자가 자율적으로 필요하다고 판단하는 경우 영향평가를 재시행할 수 있으며, 제품·서비스의 핵심 기능이나 영향받는 대상(또는 기본권)에 중대한 변화가 발생한 경우 재시행을 권장함

※ 「인공지능 기본법」 시행(2026.1.22.) 이전부터 제공되고 있는 인공지능 제품 또는 서비스의 경우, 사업자가 자율적으로 필요성을 판단하여 현행 서비스(또는 제품)를 기준으로 영향평가를 수행하는 것이 바람직함

5. 평가 수행 주체

- 고영향 인공지능 제품 또는 서비스를 제공하는 인공지능사업자는 영향평가를 직접 수행하거나, 전문성을 보유한 제3자에 의뢰하여 수행할 수 있음

※ 「인공지능 기본법 시행령안」 제28조제2항에서 “인공지능사업자는 직접 또는 제3자에 의뢰하여 영향평가를 실시할 수 있다”고 규정



6. 평가 수행 및 지원체계

- 과학기술정보통신부는 영향평가에 관한 법과 시행령 등 제도적 기반을 마련한 이후, 전담기관인 정보통신정책연구원을 통해 가이드라인과 절차 운영에 관한 전문적 지원을 제공하고, 평가 수행에 필요한 실질적 이행을 뒷받침해 나갈 예정

※ 컨설팅 등을 통해 평가 전문성과 결과의 객관성을 높이고, 사회적 신뢰와 인정을 확보할 수 있도록 지원하며, 특히 스타트업·중소기업 등의 원활한 영향평가 수행을 도울 수 있을 것으로 기대

제2절 추진 근거 및 용어 정의

1. 추진 근거

- 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제35조(고영향 인공지능 영향평가)
- 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령안」 제28조(고영향 인공지능 영향평가)

고영향 인공지능 영향평가(법 제35조 및 시행령안 제28조)

「인공지능 기본법」 제35조(고영향 인공지능 영향평가)

- 인공지능사업자가 고영향 인공지능을 이용한 제품 또는 서비스를 제공하는 경우 사전에 사람의 기본권에 미치는 영향을 평가(이하 “영향평가”라 한다)하기 위하여 노력하여야 한다.
- 국가기관등이 고영향 인공지능을 이용한 제품 또는 서비스를 이용하려는 경우에는 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하여야 한다.
- 그 밖에 영향평가의 구체적인 내용·방법 등에 관하여 필요한 사항은 대통령령으로 정한다.

「인공지능 기본법 시행령안」 제28조(고영향 인공지능 영향평가)

- 법 제35조제1항에 따른 영향평가(이하 “영향평가”라 한다)는 다음 각 호의 사항이 포함되어야 한다.
 - 해당 고영향 인공지능을 이용한 제품등에 의하여 사람의 생명, 신체의 안전 및 기본권에 영향 받을 가능성이 있는 개인이나 집단에 대한 식별
 - 해당 고영향 인공지능과 관련하여 영향을 받을 수 있는 기본권 유형의 식별
 - 해당 고영향 인공지능으로 인하여 발생할 수 있는 사람의 기본권에 대한 사회적·경제적 영향의 내용 및 범위
 - 해당 고영향 인공지능의 사용 행태
 - 영향평가에서 활용한 정량적 또는 정성적 평가지표 및 결과산출 방식
 - 해당 고영향 인공지능으로 인한 위험의 예방·완화·손실 복구 등에 관한 사항
 - 영향평가의 결과, 개선이 필요한 경우 그 이행계획에 관한 사항
- 인공지능사업자는 직접 또는 제3자에 의뢰하여 영향평가를 실시할 수 있다.
- 제1항 및 제2항에서 규정한 사항 외에 영향평가에 필요한 사항은 과학기술정보통신부장관이 정하여 고시한다.

2. 용어 정의

인공지능	학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것(인공지능기본법 제2조제1호)
인공지능시스템	다양한 수준의 자율성과 적응성을 가지고 주어진 목표를 위하여 실제 및 가상환경에 영향을 미치는 예측, 추천, 결정 등의 결과물을 추론하는 인공지능 기반 시스템(인공지능기본법 제2조제2호)
인공지능기술	인공지능을 구현하기 위하여 필요한 하드웨어·소프트웨어 기술 또는 그 활용 기술(인공지능기본법 제2조제3호)
기본권	인간 존엄성과 가치, 행복을 추구할 권리 등 국가가 확인·보장할 의무를 지니고 있는, 국민의 불가침한 기본적 권리(대한민국헌법 제10조)
고영향 인공지능	사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템(인공지능기본법 제2조제4호)
인공지능사업자	인공지능산업과 관련된 사업을 하는 자로서 인공지능을 개발하여 제공하는 자(인공지능개발사업자), 인공지능개발사업자가 제공한 인공지능을 이용하여 인공지능제품 또는 인공지능서비스를 제공하는 자(인공지능이용사업자) 중 어느 하나에 해당하는 법인, 단체, 개인 및 국가기관 등(인공지능기본법 제2조제7호)
이용자	인공지능제품 또는 인공지능서비스를 제공받는 자(인공지능기본법 제2조제8호)
영향받는 자	인공지능제품 또는 인공지능서비스에 의하여 자신의 생명, 신체의 안전 및 기본권에 중대한 영향을 받는 자(인공지능기본법 제2조제9호)
인공지능 영향평가 (이하 영향평가)	인공지능사업자가 고영향 인공지능을 이용한 제품 또는 서비스를 제공하는 경우, 사전에 사람의 기본권에 미치는 영향을 평가하기 위한 절차(인공지능기본법 제35조제1항)

제3절 타 영향평가와의 관계

- 다른 영향평가에서 해당 제품 또는 서비스의 기본권 영향을 이미 평가한 경우, 그 결과를 인공지능 영향 평가를 위한 근거로 활용 가능

※ 국내법상 유사 평가제도뿐만 아니라, EU AI Act "기본권 영향평가(Fundamental Rights Impact Assessment)", ISO/IEC 42001, ISO/IEC 42005 등 국제적으로 통용되는 평가제도 및 표준에 따라 수행된 평가 결과도 참고 자료로 활용 가능

※ 이를 통해 중복 평가로 인한 사업자 부담을 완화하고, 국제 기준과의 상호운용성을 확보하여 국내외 사업자의 시장 진출을 지원하고자 함. 다만, 구체적인 인정 요건 및 절차는 제도운영 경과와 국제 논의 동향을 고려하여 단계적으로 검토 예정



국내 유사 평가제도와와의 관계

① 개인정보보호위원회, “개인정보 영향평가” (「개인정보 보호법」 제33조)

- 기본권 범위에 개인정보 관련 정보주체의 권리가 포함되므로, 평가 항목에서 일부 중복 가능성이 존재하나, 양 제도 간 적용 범위(평가주체)가 상이(민간사업자 vs. 공공기관)하여 사업자 대상 직접적인 중복규제 논란은 없을 것으로 판단
- 개인정보 영향평가 적용범위가 민간사업자까지 확대될 경우, ‘인공지능 영향평가’의 개인정보 관련 평가항목은 ‘개인정보 영향평가’ 결과를 인정·보완하는 방식으로 중복성 해소 필요
- ※ EU AI Act에서도 개인정보보호규정(GDPR)에 따른 ‘데이터보호 영향평가’로 ‘인공지능 기본권 영향평가’에서 규정된 의무가 이미 준수된 경우, ‘인공지능 기본권 영향평가’는 ‘데이터보호 영향평가’를 보완하도록(중복이 아닌) 규정

② 식품의약품안전처, “디지털의료제품 영향평가” (「디지털의료제품법」 제37조)

- 식품의약품안전처장은 디지털의료제품의 활용과 확산이 사회·경제·문화 및 국민보건에 미치는 영향을 조사·평가할 수 있으며, 여기에는 국민 건강 및 보건의료 체계에 미치는 영향도 포함
- 양 제도 간 적용 목적(국민 기본권 보호 vs. 의료 안전성 확보)과 평가 주체(사업자 자체 vs. 식약처(인증기관)), 평가내용이 상이하여 중복 규제 논란은 제한적일 것으로 판단
- ※ 디지털의료제품 영향평가는 디지털의료제품의 임상적 안전성 및 유효성, 국민 건강 및 보건의료 체계에 미치는 영향, 사회 및 산업에 미치는 경제적 효과, 정부 정책에 미치는 영향을 평가

제2장

인공지능 영향평가 수행 단계별 주요 고려사항

제1절 사전 준비 단계

제2절 본 평가 수행 단계

제3절 사후 단계

제2장

인공지능 영향평가 수행 단계별 주요 고려사항



- 인공지능 영향평가 제도의 설계와 운영에는 EU 기본권 영향평가(Fundamental Rights Impact Assessment, FRIA) 등 국제 규범과의 상호운용성을 확보하는 것은 매우 중요합니다.
- 그러나 현재 EU 인공지능법(AI Act)에서 규정한 세부 평가 절차와 도구(FRIA Template)는 공식 발표 전 단계에 있어, 이를 직접 적용하는 데 한계 존재합니다.
- 이에 본 지침은 EU 인공지능법, 행동강령(General-Purpose AI Code of Practice), UNESCO 인공지능 윤리영향평가(Ethical Impact Assessment) 등 국제적으로 공신력 있는 평가 프레임워크를 참고하여, 수행 단계별 주요 고려사항 수준으로 정리했습니다.
- 향후 국제 기준 등이 확정·발표되면, 본 가이드라인의 단계별 고려사항을 해당 기준과 정합성 있게 업데이트하고, 필요 시 절차와 세부 항목을 보완해 나갈 예정입니다.

제1절 사전 준비 단계

목 표	해당 인공지능 제품·서비스 제공에 앞서 영향평가 필요성 여부를 판단하고, 평가 대상이 되는 제품·서비스 세부 정보를 작성
-----	---

1. 평가 필요성 검토

- 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 “법”) 제35조에 따라, 고영향 인공지능 제품 또는 서비스를 제공하는 인공지능사업자는 해당 제품·서비스에 대한 영향평가 수행에 노력해야 하며, 이를 위해 사전에 해당 여부를 검토
 - 법은 고영향 인공지능 제품·서비스에 대한 영향평가만을 규정하고 있으나, 인공지능의 책임 있는 활용과 사회적 신뢰 제고를 위해 모든 인공지능사업자가 자율적으로 영향평가를 수행·실천하도록 장려
 - ※ 법적 의무가 없는 영역에서도 인공지능 제품·서비스가 인간의 존엄성과 기본권을 침해할 소지가 있는 영향을 사전에 식별·관리하여, 국제 규범 부합을 통한 시장 확대와 브랜드 가치 제고를 지원할 것으로 기대
- 노력 의무 적용 여부를 확인하기 위해, 해당 제품·서비스에 이용된 인공지능이 고영향에 해당하는지 판단
 - **(고영향 인공지능의 확인)** 인공지능사업자는 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우, 해당 인공지능이 고영향 인공지능에 해당하는지 사전에 검토해야 하며, 필요 시 과학기술정보통신부장관에게 해당 여부 확인을 요청할 수 있음(법 제33조제1항)
 - ※ 법 제33조제3항에 따라 별도로 마련된 ‘고영향 인공지능 판단 가이드라인’을 통해 해당 여부를 확인할 수 있음
 - **(고영향 인공지능)** 법 제2조제4호 각 목 중 하나의 영역에서 활용되는 인공지능 시스템
 - 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. 그 밖에 사람의 생명·신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역으로서 대통령령으로 정하는 영역



고영향 인공지능의 확인(법 제33조)

- ① 인공지능사업자는 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 그 인공지능이 고영향 인공지능에 해당하는지에 대하여 사전에 검토하여야 하며, 필요한 경우 과학기술정보통신부장관에게 고영향 인공지능에 해당하는지 여부의 확인을 요청할 수 있다.
- ② 과학기술정보통신부장관은 제1항에 따른 요청이 있는 경우 고영향 인공지능 해당 여부를 확인하여야 하며, 필요한 경우 전문위원회를 설치하여 관련 자문을 받을 수 있다.
- ③ 과학기술정보통신부장관은 고영향 인공지능의 기준과 예시 등에 관한 가이드라인을 수립하여 보급할 수 있다.
- ④ 그 밖에 제1항에 따른 확인 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.



고영향 인공지능의 확인 절차 등(법 시행령안 제25조)

- ① 인공지능사업자가 법 제33조제1항에 따라 고영향 인공지능에 해당하는지에 대하여 확인을 요청하려는 경우에는 별지 서식의 고영향 인공지능 확인 요청서에 다음 각 호의 서류를 첨부하여 과학기술정보통신부장관에게 제출해야 한다.
 1. 해당 인공지능제품 또는 인공지능서비스의 개요서
 2. 해당 인공지능시스템의 개발 및 학습에 사용된 학습용데이터의 개요
 3. 해당 인공지능시스템을 활용하는 과정 및 결과를 확인할 수 있는 자료
 4. 그 밖에 고영향 인공지능 해당 여부를 확인에 참고가 될 수 있는 서류
- ② 제1항에 따른 요청을 받은 과학기술정보통신부장관은 다음 각 호의 사항을 고려하여 고영향 인공지능 해당 여부를 판단해야 한다.
 1. 인공지능이 법 제2조제4호 각 목의 어느 하나의 영역에서 활용되는지 여부
 2. 사람의 생명, 신체의 안전 및 기본권에 초래할 수 있는 위험의 영향, 중대성, 빈도 및 활용 영역별 특수성
 3. 법 제33조제1항에 따라 인공지능사업자가 사전에 검토한 결과
 4. 법 제33조제2항에 따라 전문위원회(이하 “전문위원회”라 한다)의 자문을 받은 경우에는 그 자문결과
 5. 그 밖에 고영향 인공지능의 해당 여부를 확인하는 데 필요한 사항으로서 과학기술정보통신부장관이 정하여 고시하는 사항
- ③ 과학기술정보통신부장관은 제1항에 따른 요청을 받은 경우에는 요청을 받은 날로부터 30일 이내에 회신해야 한다. 다만, 제품 등의 복잡성 및 중요성 등을 고려하여 30일 이내에 회신할 수 없는 경우에는 30일의 범위에서 한 차례 연장할 수 있으며, 이 경우 확인을 요청한 자에게 연장 사유와 연장 기간을 문서로 알려야 한다.
- ④ 제3항에 따라 회신을 받은 인공지능사업자는 회신 결과에 이의가 있을 때에는 회신을 받은 날로부터 10일 이내에 과학기술정보통신부장관에게 별지 서식의 고영향 인공지능 재확인 요청서에 다음 각 호의 서류를 첨부하여 과학기술정보통신부장관에게 제출해야 한다.
 1. 기존 확인 요청에 따라 과학기술정보통신부장관이 회신한 결과
 2. 그 밖에 고영향 인공지능 해당 여부를 재확인에 참고가 될 수 있는 서류
- ⑤ 제4항에 따른 재확인 요청을 받은 과학기술정보통신부장관은 전문위원회의 자문을 받아 고영향 인공지능에 해당하는지를 재확인하고, 재확인 요청을 받은 날부터 30일 이내에 그 결과를 회신해야 한다.
- ⑥ 과학기술정보통신부장관은 제1항에 따른 요청 또는 제4항에 따른 재확인 요청과 관련하여 필요한 경우에는 관계 기관에 의견 제출을 요청할 수 있다.

2. 평가대상 정의

- 본 평가에 앞서 대상이 되는 인공지능 제품·서비스의 특성을 이해하기 위해, 해당 제품·서비스가 의도된 목적에 맞게 사용될 절차 등에 관한 사항을 간략히 기술

평가대상 정의 작성 예시

1. 제품·서비스 개요	평가하려는 인공지능 제품·서비스의 주요 기능과 특징을 간략히 설명
2. 목적 및 필요성	제품·서비스가 제공하려는 핵심 가치나 목적을 작성 (특정 문제 해결이 목표라면, 해결하고자 하는 문제를 구체적 기재)
3. 적용 분야	해당 제품·서비스가 제공·운영될 산업 또는 분야 작성
4. 사용 절차	고객 또는 사용자가 해당 제품·서비스를 이용하는 절차를 설명하고, 운영 단계에서 인공지능이 작동하는 시점과 그 역할을 작성
5. 운영기간 및 빈도	제품·서비스가 운영될 기간과 평균 사용 빈도 작성
6. 이용자·영향받는 자의 범주·특성	현재 제품·서비스를 직접 이용하거나 그 활용으로 인해 직접적 또는 간접적으로 영향을 받을 수 있는 주요 집단과 특성을 작성 (연령대, 숙련도, 전문성 수준, 취약성 여부 등 포함)
7. 이용자 선택권	이용자가 해당 제품·서비스 이용 과정에서 가지는 선택권이나 통제권의 범위 설명 (인공지능 기능 미사용 선택/도출 결과 거부/이의제기, 수정요청 가능/결과 철회 가능)

제2절 본 평가 수행 단계

목 표	인공지능 제품·서비스가 기본권에 미칠 수 있는 영향을 체계적으로 식별·분석
-----	---

1. 피영향자 분석

- 인공지능 제품·서비스의 제공 결과로서, 특정 상황에서 해당 제품·서비스의 사용으로 인해 직접 또는 간접적으로 영향을 받을 수 있는 사람, 집단, 환경을 식별
 - **(직접 피영향자 : 영향받는 자)** 인공지능 제품 또는 서비스에 의해 직접적으로 생명, 신체의 안전 및 기본권에 중대한 영향을 받는 자 (법 제2조제9호)
 - **(간접 피영향자)** 영향받는 자와의 관계, 환경 변화 등으로 인해 간접적이거나 근접하게 영향을 받는 자
 - **(의도치 않은/예상치 못한 피영향자)** 예상치 못하게 영향을 받을 수 있는 사람, 생물, 환경 또는 사물
 - ※ 사전 평가에서 의도하지 않거나 예상치 못한 피영향자를 식별하기 어렵지만, 극단적이거나 가능성이 낮은 시나리오를 상상하며 대비를 강화하기 위한 브레인스토밍의 일환으로 해당 범주 활용

UNESCO, “Ethical Impact Assessment”에서의 피영향자 구분

- 유네스코(UNESCO)는 윤리적 영향평가(Ethical Impact Assessment)에서 인공지능 시스템의 긍정·부정 영향을 받는 대상을 다음의 세 범주로 구분하여 식별 : ① 주요 피영향자(Primary), ② 간접 피영향자(Secondary), ③ 예상치 못한/의도하지 않은 피영향자(Unexpected/Unintended)

	설명	예시
Primary	인공지능 시스템의 개발, 배포 및 사용에 직접적으로 관여하는 사람들입니다. 여기에는 개발자, 구매자, 그리고 인공지능 시스템의 최종 사용자 또는 객체가 포함됩니다. 인공지능 시스템에 직접적인 인풋을 제공하고 그 운영에 관여하는 사람, 생물, 환경 또는 사물이 포함됩니다.	특정 알고리즘의 최종 사용자 : 학습 모듈에 사용되는 인공지능 시스템의 경우 학생들, 의료 진단의 효율성을 향상시키기 위해 사용되는 인공지능 시스템의 경우 환자들 등
Secondary	인공지능 시스템의 주요 피영향자(Primary)에게 미치는 영향의 결과로 간접적이거나 근접하게 영향을 받는 사람, 생물, 환경 또는 사물이 포함됩니다.	주된 영향 대상자의 환경, 영향을 받는 대상자에게 의존하는 사람들(예: 보다 효율적인 수확을 위해 인공지능 시스템을 활용하는 농부의 작물 소비자 등)
Unexpected/Unintended	예상치 못하게 영향을 받거나 받을 수 있는 사람, 생물, 환경 또는 사물이 포함됩니다. 사전 평가에서 이를 측정하고 식별하기는 어렵지만, 극단적이거나 또는 가능성이 낮은 시나리오를 상상하며 대비를 강화하기 위한 브레인스토밍의 일환으로 해당 범주가 사용됩니다.	기후/환경, 삶의 질, 지역사회, 일반 대중(예: 누적된 영향으로 인해 영향을 받을 수 있는 경우)

출처 : UNESCO(2023). “Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence”

2. 평가대상 작동원리 분석

- **(제품·서비스 기제 파악)** 인공지능 기반 제품 또는 서비스의 기술·운영상 작동원리를 자세하게 조사하고 체계적으로 기록함으로써 이어지는 기본권 영향 식별 및 분석의 기초 자료로 활용
 - 평가주체는 개발 및 운영 관련 문서와 개발자, 운영자 등 관련 부서 인터뷰 등을 통해 인공지능 제품 또는 서비스의 작동원리를 내부적으로 조사
 - 이 단계에서는 기본권에 미치는 영향과 같이 가치 판단이 필요한 요소는 배제하고, ‘제품 또는 서비스에 이용한 인공지능기술이 작동하는 방식’과 이를 기반으로 제공되는 ‘제품 또는 서비스의 운영 방식’을 상세하게 조사하는 데 집중
 - 작동원리의 체계적 조사를 위하여 인공지능 활용 전 단계를 ‘데이터 수집, 분석·판단, 결정, 피드백’ 등으로 구분하고, 단계별 기술·운영상 작동원리 [부록2 참고] 를 파악
 - 제품 또는 서비스의 기술적·운영상 작동원리가 한눈에 드러나도록 목록화하여, 향후 작동원리별 적절한 보호 조치와 완화 방안 설계 시 참고 자료로 활용 가능

의료 인공지능 영역에서 기술적·운영상 작동원리 (예시)

구분	기능	기술적·운영상 작동원리
데이터 수집	모니터링	IoT 센서를 활용한 환자 생체신호 및 행동 패턴 지속 수집
분석·판단	패턴 인식·분류	딥러닝으로 X-ray, MRI 등 영상에서 병변 자동 탐지
의사결정	의료자원 배분	환자 상태와 자원 현황 고려한 최적 배분 자동 결정
피드백	오류감지·대응	AI 진단 오류, 시스템 장애 실시간 감지 및 대응

3. 기본권 영향 식별 및 분석

3.1 시나리오 기반 위험 분석

- 실제 인공지능 기반 서비스 운영 과정에서 발생할 수 있는 다양한 상황을 설정하고, 앞서 파악한 작동원리가 어떤 방식으로 기본권을 위협할 수 있는지 시나리오를 폭넓게 작성
 - 위험 시나리오 작성은 작동원리와 영향받는 기본권을 연결하는 ‘잠재적 위험’을 도출하는 과정이며, 영향의 현실성은 다음 ‘기본권 영향 판단’ 단계에서 구체적으로 검토되므로, 이 단계에서는 발생 가능성이 다소 낮더라도 최대한 다양한 위험을 폭넓게 제시하는 것이 필요
 - 동일한 작동원리를 가진 제품 또는 서비스라 하더라도 맥락에 따라 전혀 다른 위험을 초래할 수 있으므로, 해당 제품 또는 서비스가 목표로 하는 분야(예: 의료, 금융, 채용 등)와 활용 상황을 기준으로 시나리오를 작성하는 것이 중요

동일한 작동 원리지만, 활용 상황에 따라 주요 위험이 상이한 경우 (예시)

작동원리	패턴 인식 및 예측 모델링	
활용 상황	중환자실 환자의 위험 징후 감지	환자의 질문에 답변하는 챗봇
주요 위험	진단 오류에 따른 직접적 임상적 위험	데이터 보안 실패로 인한 민감 정보의 유출

- 평가주체는 기본권 영향 시나리오 작성 시 ‘작동원리에 내재한 전형적인 위험 요인’을 기초로 삼거나* ‘분야별 주요하게 언급되는 기본권’을 고려하여** 시나리오를 도출할 수 있음
 - * (예시1) 채용 지원자의 식별 정보 및 이력·자격 정보 수집의 경우, 일반적으로 ‘개인정보의 과다 수집’, ‘목적 외 사용’ 또는 ‘부당한 동의’ 등 기본권 위협 요인 고려 가능
 - ** (예시2) 인공지능 기반 채용 서비스의 경우, 평등권이 주로 언급되므로 그 내용을 고려하여 ‘학습데이터의 편향성’, ‘대리변수(Proxy) 기반 차별’ 등 위험을 바탕으로 시나리오 도출
- 평가주체는 개발자, 운영자, 법무팀 등 내부 협업을 통해 시나리오를 작성하고 검토할 수 있으나, 다양한 관점에서 작동원리를 심층적으로 분석하고 극단적인 상황까지 포함하는 시나리오를 확보하기 위해 외부 전문가의 조력을 적극적으로 활용할 수 있음

3.2 관련 기본권의 식별

- 법 제35조제1항은 영향평가의 기준으로 “사람의 기본권”을 규정하고 있으며, 그 범위는 같은 항에 규정된 ‘인공지능’과 ‘제품 또는 서비스’를 근거로 구체화할 수 있음
 - 영향평가의 기준인 ‘기본권’은 헌법에 열거된 기본권은 물론, 열거되지 않았지만, 인간의 존엄과 가치(헌법 제10조 후문)의 실현을 위해 헌법해석을 통해 도출한 기본권(헌법 제37조제1항)을 포함
 - ※ (개념) 기본권이란, 국가 내에서 헌법상 실정화되어 직접 국가권력을 구속하는 근본적인 법규범으로서, 개인이 헌법을 통해 구체적으로 직접 실현할 수 있는 주관적 공권을 의미
 - ※ (예시) 생명권¹⁾, 건강권²⁾, 개인정보자기결정권³⁾, 알권리⁴⁾ 등이 헌법재판소가 헌법해석을 통해 기본권으로 인정한 대표적인 예

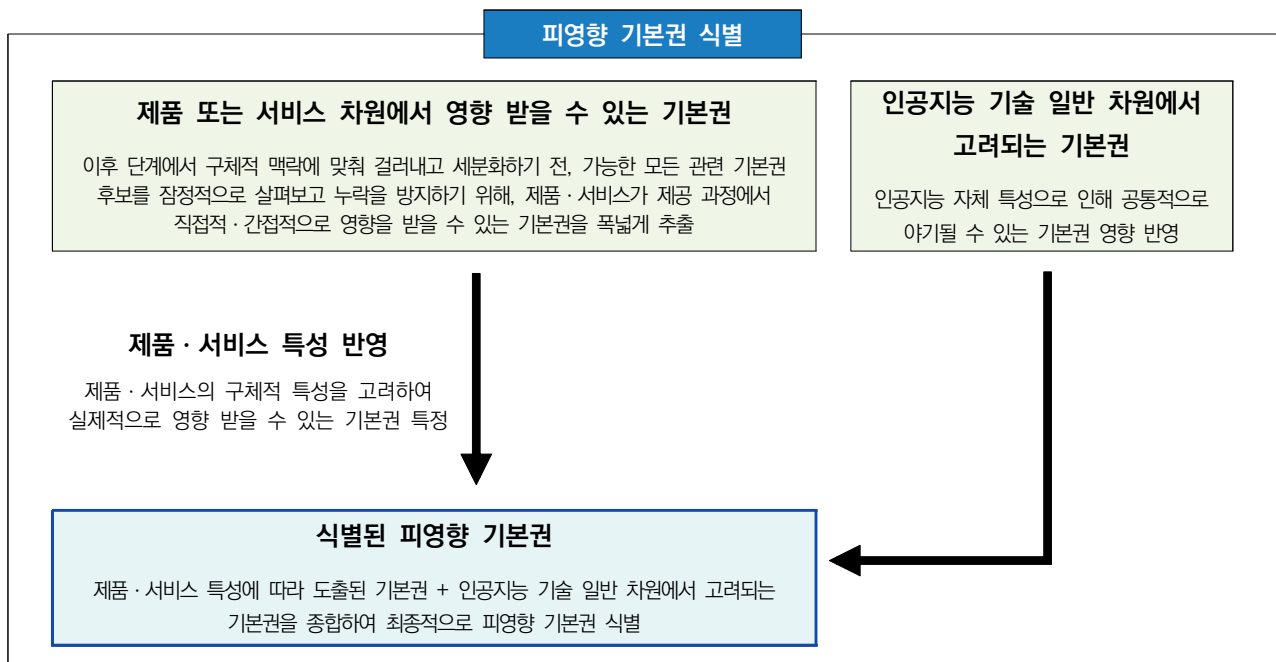
※ (구별) '인권'은 자유롭고 평등한 주체로서 존중받아야 할 모든 개인의 권리이며, 실정법인 「대한민국헌법」에 수용된 기본권 외에 법률 및 국제법상 인정된 가치를 포괄함



인권의 정의(국가인권위원회법 제2조)

1. “인권”이란 「대한민국헌법」 및 법률에서 보장하거나 대한민국이 가입·비준한 국제인권조약 및 국제관습법에서 인정하는 인간으로서의 존엄과 가치 및 자유와 권리를 말한다.

- 평가주체는 사전에 작성한 위험 시나리오를 법률 전문가의 조력을 받아 검토할 수 있으며 또는 아래 제시된 기본권 목록을 참고하여 자체적으로 해당 제품 또는 서비스가 영향을 미칠 수 있는 기본권을 잠정적으로 식별할 수 있음
 - 위험 시나리오별 인공지능이 기본권에 미치는 영향에 대한 심층 평가는 ‘기본권 영향 판단’ 단계에서 본격적으로 수행되므로, 이 단계에서는 시나리오의 내용상 잠재적으로 관련된 모든 기본권을 누락 없이 파악하는 것으로 충분
 - 관련 기본권을 망라하는 작업에 법적 전문성이 요구되기는 하지만, 이 단계의 목적은 최종적 판단이 아니라 제품 또는 서비스의 잠재적 영향 가능성을 폭넓게 탐색하는 것이므로, 제공된 목록을 참조하여 일차적인 식별 작업을 수행하는 것으로도 무방
- 체계적 평가를 위하여 제품 또는 서비스가 영향을 미칠 수 있는 기본권을 ‘인공지능기술 일반 차원’과 ‘제품 또는 서비스의 차원’으로 구분하여 식별

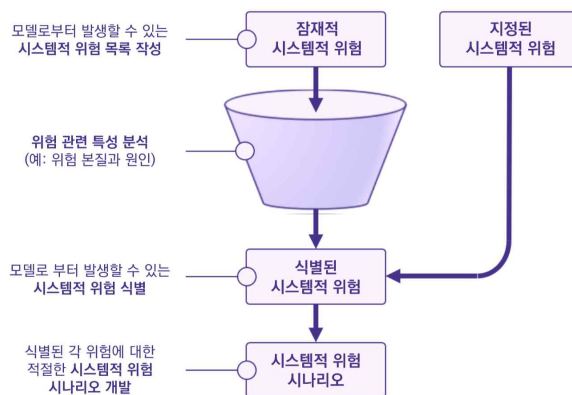


1) 현재 1996.11.28. 선고 95헌바1(사형제도), 판례집 8-2, 537, 546.
 2) 현재 1997.1.16. 선고 90헌마110 등(교통사고처리특례법), 판례집 9-1, 90, 123.
 3) 현재 2005.7.21. 선고 2003헌마282(교육정보시스템), 판례집 17-2, 81, 90.
 4) 현재 1991.5.13. 선고 90헌마133(확정된 형사소송기록 등사신청), 판례집 3, 234, 246.

- 유럽연합(EU)은 인공지능법 제55조1항에 근거하여, 범용 인공지능(GPAI) 모델의 ‘시스템적 위험 영향평가(Systemic Risk Impact Assessment)’ 프레임워크를 담은 행동강령 발표

* 시스템적 위험 : 범용 인공지능 모델이 고영향 역량으로 인해 발생하는 위험으로, 그 영향 범위가 넓어 시장 전반에 중대한 영향을 미치거나 공중 보건, 안전, 공공 보안, 기본권, 사회 전반에 부정적 영향을 줄 수 있으며, 이로 인해 가치사슬 전체에 걸쳐 대규모로 확산될 수 있는 위험

- 이 프레임워크는 시스템적 위험 분석과 수용 여부 결정을 지원하기 위해, 체계적인 절차에 따라 시스템적 위험을 식별하고, 식별된 각 위험에 대해 시스템적 위험 시나리오를 개발하는 과정을 포함



출처 : EU(2025). “Code of Practice for General-Purpose AI Models – Safety and Security Chapter”

- ‘인공지능기술 일반 차원’은 인공지능기술 자체의 작동 방식에 기인한 위험을 다루며, 앞에서 다룬 작동원리의 체계에 따르면 데이터 수집, 분석·판단, 피드백의 단계가 이 차원에 속함
 - 주로 정보의 수집 또는 처리를 다루는 기본권(개인정보자기결정권, 사생활의 비밀)이 평가 기준으로 고려되며, 이 기본권은 인공지능기술을 이용하는 한, 반드시 검토 필요
 - 인공지능기술 일반 차원에서 검토가 필요한 기본권은 아래 표를 참조할 수 있음

인공지능기술 일반 차원에서 고려되는 기본권

잠재적 영향 기본권	기본권 내용 및 침해 시나리오
개인정보자기결정권	개인정보의 공개와 이용에 관하여 스스로 결정할 권리 ⁵⁾ <ul style="list-style-type: none"> - 개인적 대화 내용 학습 후, 다른 사용자에게 대한 답변에 예시로 노출 - AI 예측·추론시스템의 불투명한 작동 방식으로 통제권 형해화 - 이미 모델이 학습한 개인정보에 대한 사후 통제 불가능
사생활의 비밀	의사에 반하여 사생활 영역이 공개되지 않을 권리 ⁶⁾ <ul style="list-style-type: none"> - 음성 AI 비서의 사적 대화 무단 수집·공유 - 헬스케어 AI의 건강 정보를 동의 범위를 넘어 무단 공유 - 사용자의 비공개 행동 패턴(예: 성적 지향, 경제 활동, 일상) 분석 및 공개
평등권	본질적으로 동일한 것의 합리적 이유 없는 차별 금지 ⁷⁾ <ul style="list-style-type: none"> - 편견에 기반한 특정 데이터의 가중치 조정으로 차별적 분류 - 차별을 유발하는 요소와 상관관계가 높은 부차적 변수들을 통한 우회적 차별
재산권	사적 유용성 및 처분권을 지닌 재산적 가치 있는 구체적인 권리의 보호 <ul style="list-style-type: none"> - 생성형 AI가 개인의 창작물을 무단 학습·활용하여 원작자의 경제적 기회 박탈 - 콘텐츠 접근 통제를 위한 기술적 장치를 회피하거나 무력화하여 비즈니스 모델 훼손

5) 현재 2005.7.21. 선고 2003헌마282(교육정보시스템), 판례집 17-2, 81, 90.

6) 현재 2003.10.30. 선고 2002헌마518(좌석안전띠 착용의무), 판례집 15-2하, 185, 206.

7) 현재 1994.2.24. 선고 92헌바43, 판례집 6-1, 72, 75; 현재 1996.8.29. 선고 93헌바57, 판례집 8-2, 46, 56.

- ‘제품 또는 서비스 차원’은 제품 또는 서비스의 활용 맥락에서 발생하는 위험을 다루며, 인간의 삶에 중대한 영향을 미치는 자동화된 의사결정에 초점을 맞춤 → 인공지능의 오류, 편향 등과 자동화에 따른 효율성이 의료 진단, 채용 결정, 신용평가 등 여러 영역에 미치는 영향에 따라 기본권 목록도 유연하게 결정 가능
 - 인공지능의 자동화된 결정으로 영향받는 기본권의 목록은 아래 표를 참조하되, 이는 인공지능의 활용 영역이 확장함에 따라, 인공지능기술의 발전에 따라 정기적인 최신화가 필요함을 유념

※ (주의) 인간의 존엄성, 행복추구권(헌법 제10조)은 다른 기본권과의 관계에서 단지 보충적으로만 기능하는 일반적 자유권이므로, 어떤 기본권에 어떻게 영향을 미치는지 파악하는 영향평가에서는 무의미

제품 또는 서비스 차원에서 영향받을 수 있는 기본권

잠재적 영향 기본권	기본권 내용 및 침해 시나리오
사생활의 자유	외부 개입 없이 사생활을 형성할 자유 ⁸⁾ <ul style="list-style-type: none"> - IoT 기반 스마트홈이 ‘최적화된 생활방식’ 제안 명목으로 개입, - ‘더 나은 인간관계’를 위한 SNS 플랫폼의 개입 - 건강관리 앱이 ‘건강 최적화’를 이유로 생활에 개입 - ‘합리적 소비’를 유도한다는 명목으로 구매 행동에 개입
생명권	생명을 박탈당하지 않을 권리 ⁹⁾ <ul style="list-style-type: none"> - 의료 진단 AI의 치명적 오진(예: 응급상황 미감지, 암 진단 누락 등) - 생명유지 장치나 의료기기 오작동(예: 인공호흡기, 마취 제어 등) - 자율주행 AI의 도로 상황 인식이나 위험 상황 대응 실패 - 인공지능 기반 안전 시스템이나 대응 시스템의 오작동(예: 대피 경로 오안내) - 군사·보안 AI의 위험 상황 오판에 따른 민간인 대상 무력 사용
건강권	신체적 건강을 훼손당하지 않을 권리 ¹⁰⁾ <ul style="list-style-type: none"> - 인공지능 기반 의료 시스템이 분석 오류 또는 획일화로 부적절한 치료 - 인공지능 제어 산업용 로봇의 안전 프로토콜 오류로 신체에 직접적 손상 - 스마트홈 시스템이 가스, 전기 등을 잘못 제어하여 건강에 해를 끼침 - 건강 상태의 과대평가에 따라 과한 운동량 권고로 신체에 해로운 행동을 하도록 유도
평등권	본질적으로 동일한 것의 합리적 이유 없는 차별 금지 <ul style="list-style-type: none"> - 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
표현의 자유	자기 의견을 간섭 없이 표명하고 전파할 자유 <ul style="list-style-type: none"> - 챗봇 서비스가 특정 이념 관련 대화를 ‘부적절한 대화’로 분류하여 대화를 일방적으로 중단 - 인공지능 기반 필터링 시스템이 사회적 소수자 집단의 비판적 의견을 혐오 표현으로 오인하여 자동 삭제
집회의 자유	간섭 없이 집회 참가 여부 및 집회의 목적·시간·장소·방법을 스스로 결정할 권리 <ul style="list-style-type: none"> - 원격 얼굴인식 시스템의 활용으로 위축 효과가 발생해 집회 참여 자체를 주저하게 함
알권리	일반적으로 접근할 수 있는 정보원으로부터 자유롭게 정보를 수집할 자유 <ul style="list-style-type: none"> - 소셜미디어 플랫폼의 AI 시스템이 특정 성향의 의견을 체계적으로 차단 - 검색 결과의 순위를 조작하거나 편향적으로 배열하여 정보 접근 왜곡

8) 현재 2003.10.30. 선고 2002헌마518(좌석안전띠 착용의무), 판례집 15-2하, 185, 206.

9) 현재 1996.11.28. 선고 95헌바1(사형제도), 판례집 8-2, 537, 546.

재산권	<p>사적 유용성 및 처분권을 지닌 재산적 가치 있는 구체적인 권리의 보호</p> <ul style="list-style-type: none"> - 인공지능 신용평가 시스템이 분석 오류로 실제 신용능력과 무관하게 신용등급 조작 - 인공지능이 일시적 시장 변동을 오판하여 임의로 자산의 대량 매도 결정
직업의 자유	<p>(직업 선택) 원하는 직업을 타인의 간섭 없이 선택할 권리</p> <ul style="list-style-type: none"> - 편향된 채용 시스템이 특정 배경을 가진 지원자의 진입을 체계적으로 차단 - 교육 분야 AI가 적성 분석을 근거로 진로를 제한하거나 특정 방향으로 강요 - 플랫폼 경제 AI가 과거 평점이나 수행 이력을 근거로 특정 노동자를 부적격으로 분류 - 인공지능이 사회신용점수를 산출하고 점수가 낮은 개인을 신뢰가 필요한 직업군에서 배제 <p>(직업 수행) 선택한 직업을 자유롭게 수행할 권리</p> <ul style="list-style-type: none"> - 인공지능 진단·치료시스템이 의료진의 진단 결과에 과도하게 간섭하여 진료 방해 - 플랫폼 AI가 성공한 창작자의 스타일을 최적 모델로 정의하고, 이를 따르도록 종용 - 인공지능 법률 분석 시스템이 변론 전략을 과도하게 제한하여 업무 수행 방해
예술의 자유	<p>간섭 없이 예술 작품을 창작하고 전시·공연·보급할 자유¹¹⁾</p> <ul style="list-style-type: none"> - 인공지능 기반 저작권 검증 시스템이 과도한 기준을 적용하여 패러디, 오마주 등을 차단 - 기존 범주에 속하지 않는 실험적 작품을 격하하여 전시 기회 박탈 - 특정 문화권의 상징을 다양한 문화적 배경을 무시하고 획일적으로 검열
학문의 자유	<p>간섭 없이 연구의 전 과정을 결정하고 그 결과를 발표할 자유</p> <ul style="list-style-type: none"> - 연구 주제나 방법론 선택 시 AI가 효율성만을 기준으로 제한 - 인공지능 논문 심사 시스템이 편향된 기준으로 논문을 사전 선별 - 연구 성과 평가 시 AI가 획일적이고 단기적인 지표만 중시하여 장기적 연구 저해

3.3 기본권 영향 판단

- 단순한 연상 또는 직관에 대한 의존을 피해 위험 시나리오를 바탕으로 ‘작동원리-위험’과 ‘잠재적 피영향 기본권’의 연결 관계와 현실성을 평가
 - 이 단계에서는 ‘시나리오 기반 위험 분석’과 ‘관련 기본권의 식별’ 단계에서 느슨한 기준에 따라 파악한 요소들을 종합하여 해당 제품 또는 서비스가 기본권에 미치는 영향을 구체적으로 판단
 - 시나리오에서 확인한 ‘위험’이 잠정적으로 파악한 ‘관련 기본권’의 보호영역을 제약하는지 판단이 주요 내용이며, 이를 위해 기본권 내용을 바탕으로 구체적 위험, 즉 ‘영향 내용’을 작성
 - 평가주체는 영향 내용을 바탕으로 시나리오에 제시된 위험의 ‘발생 가능성’을 자체적으로 평가하여 해당 제품 또는 서비스와 관련성이 높은 시나리오 및 기본권을 선별하고, 이를 ‘사후 단계’에 활용
 - 위 과정은 양식 [부록1] 에 해당 내용을 작성하는 방식으로 진행하며, 위험 시나리오별로 하나의 양식을 각각 작성

10) 현재 1997.1.16. 선고 90헌마110 등(교통사고처리특례법), 판례집 9-1, 90, 123.

11) 현재 1993.5.13. 선고 91헌바17(음반제작자 등록제), 판례집 5-1, 275, 283.

인공지능 영향평가 작성 예시

“환자상태를 고려한 최적 배분 자동 결정”이라는 작동원리가 “자동화된 차별”이라는 위험을 통해 “생명권”과 “개인 정보자기결정권”에 미치는 영향의 분석:

위험 시나리오								
예) 편향된 의료 데이터를 학습한 모델이 ‘노숙인’ 신분을 응급도 평가 변수로 활용하여, 사회적 편견에 기반한 특정 서비스(예: 알코올 중독 프로그램)를 추천								
피영향자	<div>직접 피영향자(영향받는 자)</div> <div><div>환자 개인</div><div>(이유) 인공지능 모델에 의해 편향된 응급도 평가 → 치료 지연 → 생명·건강 위험 초래</div></div>							
	<div>간접 피영향자</div> <div><div>동일 시스템에 적용받는 다른 취약계층 환자</div><div>(이유) 시스템 편향이 다른 취약계층에도 동일하게 적용 → 유사한 불이익 발생</div><div>응급실 의료진</div><div>(이유) 의료진도 잘못된 추천으로 인해 의학적 판단에 제약</div><div>보호자·가족</div><div>(이유) 가족은 치료 지연으로 정신적·경제적 피해</div></div>							
	<div>의도치 않은/예상치 못한 피영향자</div> <div><div>의료 데이터 제공자·수집기관 (데이터 품질 문제로 비난·소송 위험)</div><div>병원 및 지역사회 (서비스 신뢰 하락, 의료 이용 기피)</div><div>공중보건 당국 (편향된 데이터가 장기적으로 정책 의사결정 왜곡)</div><div>(이유) 원래는 위험 대상으로 고려되지 않았으나, 신뢰 저하·법적 분쟁·사회적 비용 증가로 영향을 받음</div></div>							
관련 기본권	영향 내용	영향수준						이행 중인 관리방안
		영향 규모		영향 기간		발생 가능성		
생명권	편향된 응급도 평가를 내려 뇌출혈 환자의 치료가 지연되고, 적시 치료를 받지 못해 생명에 치명적 결과(사망 등)가 초래	보통/경미	<input type="checkbox"/>	단기	<input type="checkbox"/>	낮음	<input type="checkbox"/>	· 모델 편향 검증 및 개선 · 휴먼 인 더 루프 · 실시간 모니터링 및 피드백
		중간	<input type="checkbox"/>	중기	<input type="checkbox"/>	중간	<input checked="" type="checkbox"/>	
		높음	<input type="checkbox"/>	장기	<input type="checkbox"/>	높음	<input type="checkbox"/>	
		매우 높음	<input checked="" type="checkbox"/>	세대 간	<input checked="" type="checkbox"/>	매우 높음	<input type="checkbox"/>	
개인정보 자기결정권	불투명성과 설명 부재로 인해 환자가 잘못된 라벨링(알코올 중독 위험군)을 인지·정정할 기회를 보장받지 못해, 개인정보 통제권과 자기결정권이 침해됨	보통/경미	<input type="checkbox"/>	단기	<input type="checkbox"/>	낮음	<input type="checkbox"/>	· 설명가능성 제고(XAI 등) · 정정·이의제기 절차 마련
		중간	<input type="checkbox"/>	중기	<input type="checkbox"/>	중간	<input type="checkbox"/>	
		높음	<input checked="" type="checkbox"/>	장기	<input checked="" type="checkbox"/>	높음	<input checked="" type="checkbox"/>	
		매우 높음	<input type="checkbox"/>	세대 간	<input type="checkbox"/>	매우 높음	<input type="checkbox"/>	
⋮	⋮	⋮		⋮		⋮		⋮

[참고] 해외 유사 제도 사례



EU, “Fundamental Rights Impact Assessment” 평가요소 (EU 인공지능법 제27조)

- EU 인공지능법(EU AI Act) 제27조에서 평가 대상이 되는 고위험 인공지능 시스템 배포자는 다음 사항을 포함한 기본권 영향 평가를 수행할 것을 규정
 - (a) 대상 AI 시스템이 의도된 목적에 맞게 사용될 절차 설명
 - (b) 대상 AI 시스템의 사용 기간 및 빈도 설명
 - (c) 특정 상황에서의 사용으로 인해 영향을 받을 가능성이 있는 개인(자연인) 또는 집단의 범주
 - (d) 식별된 대상에 영향을 미칠 가능성이 있는 구체적인 위험 요소
 - (e) 사용 지침에 따른 인적 감독 조치 이행에 대한 설명
 - (f) 위험이 현실화될 경우 취할 조치(내부 거버넌스 및 이의 처리 체계를 포함)

출처 : EU(2025). “Code of Practice for General-Purpose AI Models – Safety and Security Chapter”



UNESCO, “Ethical Impact Assessment”에서의 윤리영향 식별·완화 측정 문항

- 유네스코(UNESCO)는 윤리적 영향평가(Ethical Impact Assessment)에서 발생 가능한 긍정적·부정적 영향을 윤리 원칙별(① 안전 및 보안, ② 공정성, 차별금지, 다양성, ③ 지속가능성, ④ 프라이버시, 정보 보호, ⑤ 인간 감독 및 결정, ⑥ 투명성 및 설명 가능성, ⑦ 인식 및 리터러시) 식별하고, 규모, 범위, 발생 가능성 및 해결 가능성 등을 평가하도록 평가도구 제시

긍정적 영향				
윤리 가치에 대한 인공지능 시스템의 긍정적 영향				
개방형 (정성) 작성 예: 인공지능 시스템은 개인에 대한 악성 소프트웨어(Malware) 및 해킹 위험을 식별하는 데 활용될 수 있음				
예상되는 긍정적 결과의 규모 평가	예상되는 긍정적 결과의 범위 평가 · 영향을 받는 개인/그룹/단체에 대한 설명(모든 살아있는 유기체가 포함 가능) · 영향의 기간		예상되는 긍정적 결과가 발생할 가능성 평가	
유의미한 수준 · 매우 높음 · 높음 · 중간 · 보통/ 경미한	영향의 범위 영향을 받는 대상 · 직접적/ 간접적/ 예기치 않은 의도하지 않은 영향 기간 · 단기/ 중기/ 장기/ 세대 간		발생 가능성 · 낮음 · 중간 · 높음 · 매우 높음	
부정적 영향				
윤리 가치에 대한 인공지능 시스템의 부정적 영향				
개방형 (정성) 작성 예: 인공지능 시스템은 사용자의 심리적 안녕을 위협할 수 있는 허위 정보나 해악 또는 남용을 확산시킬 가능성				
윤리가치에 대한 시스템의 부정적 영향	예상되는 부정적 결과의 범위 평가 · 영향을 받는 개인/그룹/단체에 대한 설명(모든 살아있는 유기체가 포함 가능) · 영향의 기간	예상되는 부정적 결과가 해결될 가능성 평가	예상되는 부정적 결과가 발생할 가능성 평가	절차적 안전조치의 부정적 영향 완화 정도 추가적 완화 및 구제 전략 시행 필요성 · 시스템을 즉시 중지해야 할 극단적 경우 설명 · 조사 및 구제가 필요한 경우, 시기와 그 이유 설명 · 잠재적 피해가 반복되지 않기 위한 보장 방안
중대한 수준 · 재앙적 · 치명적 · 심각한 · 보통/ 경미한	영향의 범위 영향을 받는 대상 · 직접적/ 간접적/ 예기치 않은 의도하지 않은 영향 기간 · 단기/ 중기/ 장기/ 세대 간	해결가능 정도 · 매우 낮음 · 낮음 · 중간 · 높음	발생 가능성 · 낮음 · 중간 · 높음 · 매우 높음	개방형 (정성)

출처 : UNESCO(2023). “Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence”

제3절 사후 단계

1. 후속 조치

- 영향평가 완료 후에는 다음 사항을 고려하여 후속 조치를 자율적으로 시행하는 것을 권장함
 - 식별된 영향요소 중 부정적 영향을 완화·제거하기 위한 개선·보완 계획 수립
 - 영향평가 과정에서 수집된 피드백(내부·외부 전문가, 이해관계자 의견 포함)을 반영하여 인공지능 제품·서비스 운영 정책 및 절차 업데이트
 - 부정적 영향이 현실화될 경우 취할 조치를 사전에 마련(내부 거버넌스 절차, 피해 최소화 방안, 이해관계자와 이용자의 이익 제기·처리 체계를 포함)

2. 문서의 보관 및 결과의 공개

- 문서의 보관
 - ① 인공지능 영향평가의 결과와 ② 영향평가 수행 과정에서 수집·작성한 근거자료는 사업자가 자율적으로 보관·관리하는 것을 권장함
 - 영향평가를 수행하였음을 입증하고, 그 결과와 과정이 신뢰성 있게 관리되고 있음을 보여주기 위해, 사업자가 해당 문서와 근거자료를 자율적으로 보관하도록 권장함. 이는 법 제35조제2항과 관련하여 국가기관 등이 고영향 인공지능 제품 또는 서비스를 도입하는 과정에서 영향평가 결과 제출을 요청하는 경우, 사업자가 보관해 둔 문서를 활용할 수 있도록 하기 위함임. 이는 법적 의무가 아니며, 자율적 보관을 권장하는 데 목적이 있음
- 영향평가 결과의 공개
 - 영향평가 결과는 사업자의 판단에 따라 자율적으로 공개 여부를 결정할 수 있음
 - 인공지능 제품·서비스에 대한 이해 증진 및 사회적 신뢰 확보 차원에서, 영향평가 결과의 공개를 권장

부록

1. 인공지능 영향평가서 양식 및 작성 예시
2. 인공지능시스템 기능별 침해 발생 가능 작동원리 예시
3. 인공지능 영향평가 지원 안내
4. 인공지능 영향평가 관련 FAQ

부록1

인공지능 영향평가서 양식 및 작성 예시

〈제품·서비스명〉

1. 제품·서비스 개요	평가하려는 인공지능 제품·서비스의 주요 기능과 특징을 간략히 설명
2. 목적 및 필요성	제품·서비스가 제공하려는 핵심 가치나 목적을 작성 (특정 문제 해결이 목표라면, 해결하고자 하는 문제를 구체적 기재)
3. 적용 분야	해당 제품·서비스가 제공·운영될 산업 또는 분야 작성
4. 사용 절차	고객 또는 사용자가 해당 제품·서비스를 이용하는 절차를 설명하고, 운영 단계에서 인공지능이 작동하는 시점과 그 역할을 작성
5. 운영기간 및 빈도	제품·서비스가 운영될 기간과 평균 사용 빈도 작성
6. 이용자·영향받는 자의 범주·특성	현재 제품·서비스를 직접 이용하거나 그 활용으로 인해 직접적 또는 간접적으로 영향을 받을 수 있는 주요 집단과 특성을 작성 (연령대, 숙련도, 전문성 수준, 취약성 여부 등 포함)
7. 이용자 선택권	이용자가 해당 제품·서비스 이용 과정에서 가지는 선택권이나 통제권의 범위 설명 (인공지능 기능 미사용 선택/도출 결과 거부/이의제기, 수정요청 가능/결과 철회 가능)

8. 위험 시나리오								
피영향자	<div>직접 피영향자(영향받는 자)</div> <div><div></div><div>(이유)</div></div>							
	<div>간접 피영향자</div> <div><div></div><div>(이유)</div></div>							
	<div>의도치 않은/예상치 못한 피영향자</div> <div><div></div><div>(이유)</div></div>							
관련 기본권	영향 내용	영향수준						이행 중인 관리방안
		영향 규모		영향 기간		발생 가능성		
		보통/경미	<input type="checkbox"/>	단기	<input type="checkbox"/>	낮음	<input type="checkbox"/>	
		중간	<input type="checkbox"/>	중기	<input type="checkbox"/>	중간	<input type="checkbox"/>	
		높음	<input type="checkbox"/>	장기	<input type="checkbox"/>	높음	<input type="checkbox"/>	
		매우 높음	<input type="checkbox"/>	세대 간	<input type="checkbox"/>	매우 높음	<input type="checkbox"/>	
⋮	⋮	⋮		⋮		⋮		⋮

작성 예시1

스마트채용(AI기반 채용 후보자 평가·적합도 판정 서비스)

※ 본 예시 영향평가는 “채용 후보자의 각종 정보(후보자가 제출하지 않은 정보 중 사회관계망 등에 공개된 정보까지 포함)를 수집·분석하여 채용 적격 여부를 판정하는 AI 기반의 채용 시스템”을 가상의 서비스로 설정하여 수행한 것으로, 실제 존재하는 서비스가 아니라는 점을 유의하여 주시기 바랍니다.

1. 제품·서비스 개요	<ul style="list-style-type: none"> 채용 후보자가 제출한 이력서·경력사항 등의 공식 자료뿐만 아니라 논문·프로젝트 이력, 공개된 사회관계망(SNS) 정보, 인터뷰 결과 등을 자동으로 수집·분석하여 후보자의 직무 적합성과 경쟁력을 점수화하고 평가 리포트를 제공하는 AI 기반 서비스
2. 목적 및 필요성	<ul style="list-style-type: none"> 기업이 다수의 채용 후보자를 효율적으로 비교·평가하고, 직무 요구사항에 부합하는 인재를 객관적으로 선발할 수 있도록 지원하는 것이 주요 목적 이를 통해 채용 과정의 속도·정확성을 높이고, 의사결정의 편향을 줄이며 후보자와 기업 모두에게 보다 투명한 채용 경험을 제공
3. 적용 분야	<ul style="list-style-type: none"> 일반 기업 및 공공기관의 HR·채용 업무 신규·경력직 채용, 정기·수시 채용, 내부 인사 이동 및 승진 심사 지원
4. 사용 절차	<ol style="list-style-type: none"> 기업 HR 담당자가 채용공고·직무 요구 역량 입력 후보자 데이터(제출자료 + 공개 정보) 자동 수집 AI 모델이 데이터 분석 → 직무 적합도·경쟁력 점수 산출 평가 리포트(점수·분석 근거 포함) 자동 생성 기업이 결과 리포트를 검토하고 자체 심사 프로세스를 통해 최종 결정 <p>▶ 단계 2~3에서 데이터 수집·분석 및 점수 산출 수행</p>
5. 운영기간 및 빈도	<ul style="list-style-type: none"> 운영기간 : SaaS 형태로 상시 제공 사용 빈도 : 채용 공고별로 1회 이상 분석 수행 (정기 또는 수시)
6. 이용자·영향받는 자의 범주·특성	<p>HR 담당자는 시스템 사용 능력은 있으나 AI 모델의 내부 로직 이해도는 제한적일 가능성 있음</p> <ul style="list-style-type: none"> (직접 이용자) 기업 HR 담당자, 인사팀 관리자, 시스템 관리자 (간접 영향자) 채용 후보자, 추천인, 채용 결과에 영향을 받는 내부 직원·팀
7. 이용자 선택권	<ul style="list-style-type: none"> HR 담당자(이용자): 시스템 분석 결과를 수용·재검토·무시 여부 선택 가능 채용 후보자: 사전 동의 절차를 통해 시스템 평가 참여 여부 선택 가능 (동의 거부 시 대체 평가 절차 제공 권고) 결과 접근권: 기업·후보자 모두 결과 리포트 확인 가능 (기업 정책에 따라 범위 조정) 이의제기: 후보자가 평가 결과에 대해 설명 요청·정정 요구할 수 있는 절차 제공 가능

8. 위험 시나리오(복수 작성 가능)

※ 필요 시 복수의 시나리오를 추가 기재하고, 시나리오별로 피영향자, 관련 기본권, 영향수준, 이행 중인 관리 방안 등 작성

<div> <div>시나리오(1)</div> </div>	<p>채용 후보자와 관련한 다양한 데이터를 광범위하게 수집·분석하는 과정에서, 비공개적이거나 사적인 영역의 정보까지 활용되어 사생활의 자유가 침해될 수 있음. 또한 부실하거나 편향된 데이터가 포함될 경우 차별적·불공정한 평가 결과가 산출되어 평등권 및 직업 선택의 자유가 침해될 위험이 있음. 더 나아가 AI 시스템의 분석 과정이 불투명하여, 후보자가 본인의 평가 결과 산출 근거를 확인하거나 이의제기하기 어려워 알 권리가 제한될 가능성이 있음</p>
<div> <div>피영향자</div> </div>	<div> <div>직접 피영향자(영향받는 자)</div> <ul style="list-style-type: none"> <div> <div>채용 후보자</div> <div>(이유) 후보자는 정보 수집의 대상이 될 뿐만 아니라, AI 시스템의 결정 결과에 따라 직업 기회가 제한되거나 영향을 받는 당사자</div> </div> </div>

	<div>간접 피영향자</div> <div><div>▪ 채용 기업 (이유) AI 시스템의 작동 상태(정상·비정상)에 따라 채용 결과가 영향을 받을 수 있으며, 장기적으로는 기업의 채용 심사 기준이 AI 시스템의 평가 기준을 반영하는 방향으로 변화할 가능성</div><div>▪ 향후 잠재 채용 후보자군 (이유) AI 시스템이 채용 과정에서 일반화될 경우, 지원자는 시스템의 평가 기준을 고려해 학력, 경력, 온라인 활동 등을 준비해야 할 필요 발생</div></div>							
	<div>의도치 않은/예상치 못한 피영향자</div> <div><div>▪ 직업교육 종사자 (이유) 채용 평가가 AI 중심으로 전환되면서 기존 교육·훈련 프로그램이 인사 담당자의 주관적 평가를 전제로 설계된 경우 방향 전환이 요구됨</div></div>							
관련 기본권	영향 내용	영향수준						이행 중인 관리방안
		영향 규모		영향 기간		발생 가능성		
직업의 자유	<div>· 데이터에 내재된 사회적·관습적 편향이 시스템 설계 단계에서 교정되지 않으면 불공정한 채용 결과를 초래할 가능성이 있음</div> <div>· 기업 평가 기준이 사회적 기준과 다르면, 채용 결과가 불공정하게 인식될 수 있음</div>	보통/경미 중간 높음 매우 높음	<div><input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/></div>	단기 중기 장기 세대 간	<div><input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/></div>	낮음 중간 높음 매우 높음	<div><input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div>	<div>· 객관적· 다양한 데이터를 활용하고, 소수자 배려 요소를 모델 학습 데이터· 알고리즘에 반영</div> <div>· AI 채용시스템의 평가 결과와 근거를 HR 담당자가 최종 확인·보완하여 결과의 공정성 확보</div>
사생활의 자유	<div>· 채용 시스템의 공정성과 정확성을 도모하기 위해 후보자에 대한 광범위한 정보 분석이 이루어질 수 있으며, 이 과정에서 불필요한 정보 수집 위험이 존재</div> <div>· 분석 이후 적절한 절차에 따른 개인정보 파기가 이루어지지 않을 경우, 인사 과정 종료 후에도 해당 정보가 잔존하거나 유출될 가능성 있음</div> <div>· 한편, 시스템 보안을 강화하고 접근 통제를 적용하면 사생활 정보에 대한 불필요한 접근을 차단하여 프라이버시 보호에 기여할 수 있음</div>	보통/경미 중간 높음 매우 높음	<div><input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div>	단기 중기 장기 세대 간	<div><input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div>	낮음 중간 높음 매우 높음	<div><input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div>	<div>· 당사자의 사전 동의를 받은 정보 또는 합리적으로 공개된 정보만 분석 대상으로 활용</div> <div>· 채용 과정에서 수집·저장되는 정보를 암호화 접근통제하여 인사 절차 종료 후 즉시 파기</div>

평등권	<ul style="list-style-type: none"> · 데이터가 양적으로 충분하고 사회적 관습을 그대로 반영하고 있더라도, 역사적으로 발전해 온 평등·차별금지 가치가 별도로 고려되지 않는다면 특정 계층에 대한 취업차별로 이어질 수 있음 · 탈락 사유에 대한 설명이 부족할 경우, 실제로는 객관적 요인에 따른 결과라도 소수자 차별로 인식될 가능성이 있음 	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<ul style="list-style-type: none"> · 성별·지역·인종·종교·학벌 등 차별적 요소가 될 수 있는 변수를 평가 기준에서 배제 · 시스템 운영 과정에서 평가 결과를 주기적으로 점검하고, 소수자에 불리한 편향이 발견되면 알고리즘 개선
알권리	<ul style="list-style-type: none"> · AI 채용 시스템은 평가 과정의 불투명성으로 인해 채용 자원자의 알 권리가 제한될 수 있다는 논의가 있음 · 다만, 기업 채용 절차의 의사결정 과정은 일반적으로 알 권리의 직접적 보호 범위에 포함되지 않아 실질적 침해 가능성은 제한적 이라고 볼 수도 있음 	보통/경미 중간 높음 매우 높음	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<ul style="list-style-type: none"> · 시스템이 수집·분석하는 정보의 범위와 주요 평가 기준을 사전에 안내 · 후보자 요청 시 평가 결과의 주요 근거와 판단 요소를 설명할 수 있는 절차 제공
⋮	⋮	⋮		⋮		⋮		⋮

작성 예시2

Face Swap Studio(딥페이크 얼굴 합성 영상 제작 서비스)

※ 본 예시 영향평가는 “사용자가 업로드한 동영상·이미지를 합성하여 결과 영상을 생성·공유할 수 있는 AI 기반 얼굴 합성 영상 서비스”를 가상으로 설정하여 수행한 것으로, 실제 존재하는 특정 기업·서비스를 지칭하지 않습니다.

1. 제품·서비스 개요	<ul style="list-style-type: none"> · 사용자가 업로드한 동영상·이미지와 선택한 타인의 얼굴 데이터를 AI 모델이 합성하여 자연스러운 얼굴 변환 영상을 생성하는 서비스 - 얼굴 인식·추출, - GAN·Transformer 기반 합성 알고리즘으로 얼굴 변환 - 결과 영상 미리보기 및 간단한 편집 - 결과물 다운로드·공유 기능 제공
2. 목적 및 필요성	<ul style="list-style-type: none"> · 엔터테인먼트·창작 지원으로, 개인이 영화 장면 속 캐릭터 얼굴을 자신의 얼굴로 교체하거나 패러디·유머 콘텐츠를 제작할 수 있도록 지원함. 이를 통해 영상 제작 비용과 시간을 절감하고, 비전문가도 창의적 콘텐츠 제작에 참여할 수 있는 접근성을 높임
3. 적용 분야	<ul style="list-style-type: none"> · 엔터테인먼트 : 패러디, 밈, 팬 콘텐츠 제작 · 마케팅·광고 : 맞춤형 캠페인, SNS 홍보 영상 제작 · 교육·연구 : 영상합성 기술 교육, 연구용 합성 데이터 생성
4. 사용 절차	<ol style="list-style-type: none"> 1. 원본 영상 업로드 또는 템플릿 선택 2. 교체할 얼굴 이미지 업로드 → AI가 얼굴 인식·특장점 추출 3. 합성 모델 작동 → 얼굴 합성 결과 생성 4. 결과 영상 미리보기·편집 5. 결과물 다운로드·SNS 공유 <p>▶ 단계 3~4에서 얼굴 인식·합성 모델 작동</p>
5. 운영기간 및 빈도	<ul style="list-style-type: none"> · 운영기간 : 상시 운영 · 사용 빈도 : 1인당 평균 월 3~5회
6. 이용자·영향받는 자의 범주·특성	<p>대체로 비전문가, 영상 제작 경험이 있으나 AI 기술 수준은 중간 이하</p> <ul style="list-style-type: none"> · (직접 이용자) 10~40대 일반 사용자, 크리에이터, 광고·마케팅 담당자 · (간접 영향자) 합성에 사용된 인물, 원본 영상·사진의 저작권자, 합성영상 시청자 등
7. 이용자 선택권	<ul style="list-style-type: none"> · AI 기능 비사용 선택 : 불가(서비스 본질 기능) · 결과 거부·재생성 : 합성 결과 미리보기 후 재합성 가능 · 이의제기·수정 요청 : 초상권·저작권 침해 신고 접수 채널 운영 · 결과 철회 : 플랫폼 내 게시물 삭제 가능(외부 공유 이후 통제 어려움 사전 고지) · 사전 동의 절차 : 타인 얼굴 업로드 시 동의 여부 확인 체크

8. 위험 시나리오(복수 작성 가능)

※ 필요 시 복수의 시나리오를 추가 기재하고, 시나리오별로 피영향자, 관련 기본권, 영향수준, 이행 중인 관리 방안 등 작성

시나리오(1)	<p>사용자가 유명인의 얼굴 사진을 인터넷에서 무단으로 수집해 서비스에 업로드하고, 서비스는 이를 합성해 음란물 형태의 패러디 영상을 생성 → 사용자가 해당 영상을 유포하고, 이를 본 불특정 다수가 실제 사건으로 오인해 피해자의 평판이 실추되고 정신적 피해를 입음</p>				
피영향자	<table border="1"> <tr> <th colspan="2">직접 피영향자(영향받는 자)</th></tr> <tr> <td>▪ 합성 대상이 되는 개인(유명인, 일반인 포함)</td><td>(이유) 원치 않는 합성으로 인한 이미지 훼손, 사생활 침해 가능</td></tr> </table>	직접 피영향자(영향받는 자)		▪ 합성 대상이 되는 개인(유명인, 일반인 포함)	(이유) 원치 않는 합성으로 인한 이미지 훼손, 사생활 침해 가능
직접 피영향자(영향받는 자)					
▪ 합성 대상이 되는 개인(유명인, 일반인 포함)	(이유) 원치 않는 합성으로 인한 이미지 훼손, 사생활 침해 가능				

	간접 피영향자							
	<div>합성 영상 시청자, 일반 대중 (이유) 허위 정보 영상 접촉으로 인해 사실 오인, 사회적 불신 증가</div>							
	의도치 않은/예상치 못한 피영향자							
<div>동일 · 유사 얼굴 특징을 가진 제 3자 (이유) 합성 결과가 제3자와 유사해 오인 · 괴롭힘 등 2차 피해 발생 가능</div>								
관련 기본권	영향 내용	영향수준						이행 중인 관리방안
		영향 규모		영향 기간		발생 가능성		
개인정보 자기결정권	본인 동의 없이 얼굴 이미지가 수집 · 합성되어 통제권 상실	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	· 업로드 시 동의 여부 확인 체크 · 권리자 요청 시 신속 삭제 · 차단
사생활의 비밀	사적 이미지가 음란물 등으로 변형 · 유포되어 정신적 피해 발생	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/>	단기 중기 장기 세대 간	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	· AI 유해물 탐지 · 자동 차단 · 신고 접수 후 모니터링 · 삭제
평등권	특정 집단(성별 · 인종)이 합성물 타깃이 될 경우 차별 조장	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	· 약관에 혐오 · 차별 목적 합성 금지 명시 · 반복 위반 계정 차단
⋮	⋮	⋮		⋮		⋮		⋮

작성 예시3

Edu Score(AI기반 학생 평가 서비스)

※ 본 예시 영향평가는 “학생이 제출한 시험 답안·과제·에세이 등 학습 데이터를 수집·분석하여 평가 점수를 산출하고 제한적 피드백을 제공하는 “AI 기반 학생평가 서비스”를 가상으로 설정하여 수행한 것으로, 실제 존재하는 특정 기업·서비스를 지칭하지 않습니다.

1. 제품 · 서비스 개요	<ul style="list-style-type: none"> · 교육기관에서 학생 평가에 활용할 수 있도록 설계 · 학생이 제출한 객관식 · 단답형 · 서술형 답안 및 에세이 과제를 분석하여 점수를 산출하고, 필요 시 학습 역량 향상을 위한 제한적 피드백을 제공
2. 목적 및 필요성	<ul style="list-style-type: none"> · (목적) 교사 · 교수의 교육 활동을 보조하여, 학습자에게 보다 질 높은 교육을 제공할 수 있도록 지원 · (필요성) AI 도입으로 교사·교수의 직무 변화가 불가피한 상황에서, 평가의 일정 영역을 시스템이 담당하게 함으로써 교사는 종합적 평가와 피드백 제공에 집중할 수 있고, 학생들이 역량을 성장시키는 데 집중할 수 있도록 도움을 주는 서비스 필요
3. 적용 분야	<ul style="list-style-type: none"> · 초·중·고·대학 등 교육기관의 정규·보충 수업 과정 · 온라인 학습 플랫폼 · 개별 학습자의 자율 학습 진단과 피드백 제공
4. 사용 절차	<ol style="list-style-type: none"> 1. 교육기관이 서비스에 가입하고 평가 기준을 등록 2. 학생이 과제·답안을 제출하면 시스템이 자동으로 데이터를 분석 3. AI 모델이 점수와 피드백을 생성 4. 교사·교수와 학생이 결과를 확인하고 학습 계획에 반영 5. 개인 학습자는 직접 과제를 제출하고 결과·피드백을 확인 <p>▶ 단계 2~3에서 자동 분석 및 점수 산출에 AI 작동</p>
5. 운영기간 및 빈도	<ul style="list-style-type: none"> · 운영 기간: 학기·학년도 단위로 상시 운영 · 빈도: 평가 시점이나 과제 제출 시마다 반복 사용 가능
6. 이용자 · 영향받는 자의 범주 · 특성	<ul style="list-style-type: none"> · (직접 이용자) 교육기관 담당자, 교사 · 교수, 학생 · (간접 영향자) 학부모·교육청 등은 결과를 간접적으로 참고할 수 있는 이해관계자 · 교사·교수는 기본적인 시스템 사용 역량을 갖추고 있으나, AI 모델의 세부 작동 원리에 대한 전문 이해는 제한적일 수 있음
7. 이용자 선택권	<ul style="list-style-type: none"> · 교육기관은 서비스 도입 여부를 결정 · 기관 도입 이후에도 개별 교사·교수는 서비스 활용 여부를 선택 가능 · 개인 학습자는 직접 구독을 통해 서비스 이용 여부를 결정할 수 있음

8. 위험 시나리오(복수 작성 가능)

※ 필요 시 복수의 시나리오를 추가 기재하고, 시나리오별로 피영향자, 관련 기본권, 영향수준, 이행 중인 관리 방안 등 작성

시나리오(1)	한 대학의 미디어커뮤니케이션학과 학생 B가 제출한 에세이를 AI가 분석한 결과, AI 작성 확률이 40~70%로 산출되고 평가 점수는 C+로 부여됨 → 담당 교수 C는 해당 점수를 그대로 반영해 성적을 매겼으나, 이에 대해 학생 B는 스스로 작성한 과제라며 성적 정정을 요구함 → 교수는 AI 판별 결과의 한계를 고려해 직접 재평가하여 성적을 B로 수정했지만, B 학생은 이 수정 결과 역시 초기 AI 평가 점수의 영향(뎃내림 효과)을 받았다고 주장하며 학교 당국에 이의를 제기
피영향자	<div>직접 피영향자(영향받는 자)</div> <ul style="list-style-type: none"> ▪ AI 평가 서비스를 통해 평가받는 대상자(학습자) (이유) 교육기관 소속 학습자는 AI 평가 서비스의 판단이 자신의 성적에 직 · 간접적으로 영향을 미치므로 1차적 이해당사자에 해당

	간접 피영향자							
	<ul style="list-style-type: none">교수자 및 교육기관 (이유) 서비스가 주관식 답안·에세이 평가 등에 있어 한계가 존재할 가능성이 여전히 있으며, 교수자의 독자적 판단에 영향을 줄 가능성이 있음. 이에 따라 학생의 이의 신청 및 분쟁 발생 가능성 존재							
	의도치 않은/예상치 못한 피영향자							
	<ul style="list-style-type: none">교육 당국 및 AI 교육 서비스 제공자 (이유) AI 평가 서비스가 교사 업무 부담을 줄이고 학습자에게 피드백을 제공할 잠재력을 가지고 있으나, 서비스 결과에 대한 사회적 불신이 확산될 경우 교육 당국 및 산업 전반에 대한 신뢰에 부정적 영향을 줄 수 있음							
관련 기본권	영향 내용	영향수준						이행 중인 관리방안
		영향 규모		영향 기간		발생 가능성		
평등권	<ul style="list-style-type: none">AI 평가 서비스가 특정 문제·언어 패턴에 편향될 경우 일부 학생이 불리하게 평가받을 수 있음서비스의 공정성과 객관성에 대한 논란 발생 시 평가 결과 신뢰도 저하 가능	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	<ul style="list-style-type: none">중요 과제는 교사 표본 검토로 교차 확인객관식 문항, 단답형 채점 등에만 제한적으로 AI 활용
알권리	<ul style="list-style-type: none">설명가능성의 한계로 평가 대상자가 납득할만한 근거 제시 어려운 상황 발생 가능설명가능성 부족은 학습자에게 결과를 개선할 기회를 제한	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/>	<ul style="list-style-type: none">평가 근거·결과 리포트 제공성적 이의제기 및 재검토 절차 운영
사생활의 비밀	<ul style="list-style-type: none">학생이 제출한 과제·답안·성적 등 개인 학습 정보가 AI 시스템에 저장·분석되며, 보관·활용 과정에서 의도치 않게 외부로 유출될 경우 사생활 침해 우려평가 종료 후에도 데이터가 적절히 삭제되지 않으면 개인 학습 이력이 과도하게 축적·노출될 가능성 있음	보통/경미 중간 높음 매우 높음	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	단기 중기 장기 세대 간	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	낮음 중간 높음 매우 높음	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<ul style="list-style-type: none">학습 데이터 최소 수집 및 평가 종료 후 자동 삭제권한 있는 사용자만 접근 가능, 저장·전송 시 암호화
⋮	⋮	⋮		⋮		⋮		⋮

부록2

인공지능시스템 기능별 침해 발생 가능 작동원리 예시

기능	침해 발생 가능 작동원리	설명
데이터 수집	개인정보 수집	사용자 식별 및 서비스 개인화를 위한 개인정보 수집 예) 이름, 연락처, 주소, 생년월일 등 기본 인적 사항
	행동 패턴 분석	서비스 이용 과정에서 사용자 행동 로그 수집 예) 웹/앱 로그, 쿠키, 세션 데이터 자동 수집, 컴퓨터 비전 기술
분석 및 판단	패턴 인식 및 분류	입력 데이터를 알고리즘으로 분석하여 패턴 도출 및 분류 예) 통계적 모델링, 클러스터링, 분류 알고리즘 적용
	예측 및 추론	과거 데이터를 바탕으로 미래 상황 예측 예) 딥러닝, 신경망 등 예측 알고리즘 활용
자동화 결정	프로파일링	수집된 데이터를 종합하여 사용자 분류 및 프로파일 생성 예) 데이터 마이닝, 클러스터링을 통한 세분화
	자동화된 의사결정	인공지능 분석 결과를 바탕으로 인간의 개입 없이 의사결정 예) 임계값 설정, 규칙 기반 자동 처리
	콘텐츠 필터링	텍스트, 이미지, 영상을 분석하여 부적절 콘텐츠 검열 및 차단 예) 자연어 처리, 컴퓨터 비전 기술 활용
상호작용	개인화 추천	개인 프로파일과 유사 사용자 분석을 통한 추천 예) 협업 필터링, 콘텐츠 기반 필터링 알고리즘
	피드백 수집	서비스 개선을 위한 사용자 반응 수집

부록3

인공지능 영향평가 지원 안내

- 정보통신정책연구원(KISDI)은 「인공지능기본법」에 따른 영향평가 제도의 원활한 안착을 지원하기 위해 사업자의 영향평가 수행을 지원할 예정
- 인공지능 영향평가 가이드라인 제공, 사업자 문의사항에 대한 상담·컨설팅, 샘플 보고서 및 참고자료 제공 등 지원 범위를 단계적으로 확대해 나갈 계획

〈문의처〉

- 정보통신정책연구원(KISDI) 인공지능정책연구실
- 전화 : 043-531-4114
- 이메일 : aiaa@kisdi.re.kr

인공지능 영향평가 관련 FAQ

Q. 인공지능 영향평가란 무엇인가요?

- A. 인공지능 영향평가란 인공지능사업자가 고영향 인공지능을 이용한 제품이나 서비스를 제공하기 전, 해당 서비스가 사람의 기본권에 미치는 영향을 사전에 평가하는 절차를 말합니다(법 제35조제1항). 이는 인공지능을 이용한 제품 또는 서비스의 잠재적 리스크를 식별하고, 기본권을 존중·보호하는 방향으로 개선방안을 마련하는 데 목적이 있습니다.

Q. 어떤 사업자가 이 평가를 수행해야 하나요?

- A. 고영향 인공지능을 이용한 제품 또는 서비스를 제공하려는 인공지능사업자는 법 제35조 제1항에 따라 이 평가를 수행하기 위하여 노력하여야 합니다. 고영향 인공지능이 아닌 일반 인공지능을 이용했다면, 인공지능사업자는 이러한 노력 의무를 지지 않지만, 신뢰성 확보와 글로벌 규제 대응을 위해 자율적으로 수행할 것이 장려됩니다.

Q. 인공지능 영향평가는 언제 수행해야 하나요?

- A. 인공지능 영향평가는 신규 제품이나 서비스를 제공하기 ‘전’에 수행하는 것이 원칙입니다. 영향평가를 수행했다 하더라도, 제품이나 서비스의 핵심 기능이나 영향받는 대상에 중대한 변화가 생겼다면 재시행하는 것이 바람직합니다. 또한, 법 시행 이전부터 이미 제공 중인 제품 또는 서비스의 경우, 사업자가 자율적으로 판단하여 수행하는 것이 권장됩니다.

Q. 외부 기관에 의뢰하여 평가를 진행해도 되나요?

- A. 네, 가능합니다. 「인공지능기본법 시행령안」 제28조제2항은 인공지능사업자가 직접 영향 평가를 수행하거나 제3자에 의뢰하여 영향평가를 실시할 수 있다고 규정하고 있습니다. 이에 정보통신정책연구원(KISDI)은 제도의 원활한 안착을 위해 전문적인 지원을 준비하고 있으며, 본 가이드라인의 제공 외에 사업자 문의사항에 대한 상담·컨설팅, 샘플 보고서 및 참고자료 제공 등 지원 범위를 단계적으로 확대해 나갈 계획입니다.

Q. 제품 또는 서비스로 인해 영향받는 자를 어디까지 고려해야 하나요?

- A. 본 평가 단계의 첫 번째 핵심은 피영향자를 다각도로 식별하는 것입니다. 따라서 인공

지능에 의해 직접적으로 생명이나 기본권에 영향을 받는 자뿐만 아니라, 이들과의 관계나 환경 변화로 인해 근접하게 영향을 받는 간접 피영향자와 의도치 않은 또는 예상치 못한 피영향자까지 모두 고려하는 것이 권고됩니다. 특히 의도하지 않은 영향을 사전에 완벽히 파악하기는 어렵더라도, 다양한 가능성을 열어두고 진행하는 브레인스토밍은 잠재적 위험에 대한 선제적 대비 체계를 구축한다는 점에서 중요한 의미가 있습니다.

Q. 기술적 작동원리를 분석할 때 무엇을 중점적으로 검토하여야 하나요?

- A. 작동원리를 분석할 때는 데이터 수집, 분석 및 판단, 결정, 피드백 등 인공지능 활용 모든 단계에서 인공지능 기술이 작동하는 방식에 집중하여 조사해야 합니다. 이를 통해 향후 단계별로 적절한 보호 조치와 완화 방안을 설계할 수 있는 기초 자료를 확보할 수 있습니다.

Q. 위험 시나리오는 어떤 기준으로 작성하는 것이 효과적인가요?

- A. 동일한 기술이라도 활용 분야에 따라 위험의 종류가 다르므로, 실제 제품 또는 서비스가 목표로 하는 분야와 상황을 기준으로 구체적인 시나리오를 작성해야 합니다. 예를 들어, 동일하게 데이터를 분석하여 패턴을 찾아내고, 미래를 예측하는 기술이라도 ‘중환자실 위험 징후 감지’를 위해 활용되는 경우, 진단 오류로 인한 환자의 생명권 침해가 주된 위험이지만, ‘환자 답변용 챗봇’의 경우, 데이터 보안 실패에 따른 민감 정보의 유출이 주요 위험이 됩니다. 이 과정에서는 발생 가능성이 낮더라도 최대한 다양한 위험을 폭넓게 제시하는 것이 필요하며, 필요한 경우 외부 전문가의 조력을 받아 심층적인 시나리오를 확보하는 것이 권장됩니다.

Q. 인공지능이 침해할 수 있는 기본권은 어떻게 식별하나요?

- A. 평가의 편의를 위해 ‘인공지능 기술 일반 차원’과 ‘제품·서비스 차원’으로 구분하여 식별하는 것이 좋습니다. 인공지능 기술의 일반적인 작동 방식에서 기인하는 개인정보자기결정권이나 사생활의 비밀 같은 기본권은 물론이고, 서비스의 활용 맥락에 따라 생명권, 평등권, 표현의 자유, 직업의 자유 등이 침해될 가능성이 있는지 포괄적으로 식별해야 합니다.

Q. 식별된 위험이 기본권에 영향을 미치는지 어떻게 판단하나요?

- A. 식별된 위험 시나리오를 바탕으로 기본권에 미치는 영향의 현실성을 종합적으로 판단해야 합니다. 위험 시나리오가 실제 기본권의 보호 영역을 제약하는지 판단하고, 그 영향의 규모와 기간, 발생 가능성을 자체적으로 평가합니다. 이 과정에서 해당 제품 또는 서비스와

관련성이 높은 시나리오를 선별하고, 기본권에 대한 부정적 영향을 완화 또는 제거하는 데 현재 이행 중인 관리 방안이 충분한지를 검토하여 사후 단계의 후속 조치로 연결해야 합니다.

Q. 평가 결과를 공개해야 하나요?

- A. 평가 결과를 공개할 것인지는 **인공지능사업자가 자율적으로 결정**할 수 있습니다. 다만, **사회적 신뢰 확보와 이용자 이해 증진**을 위해 평가 결과의 **공개가 권장**됩니다. 또한, 법 제35조제2항은 국가기관 등이 제품 또는 서비스 도입 시 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하여야 한다고 규정하고 있으며, 이때 영향평가 결과의 제출이 필요할 수 있으므로 **관련 문서를 자율적으로 보관**하는 것이 좋습니다.