

인공지능기본법상 고영향 인공지능 영향평가의 문제점 및 개선 방향에 관한 연구*

A Study on the Issues and Improvements of High-Impact AI Impact Assessments under the Framework Act on AI

정 성 응**

Jung, SeongWoong

이 혜 승***

Lee, Haesung

박 도 현****

Park, Do Hyun

<https://doi.org/10.35148/ilsilr.2026..63.173>

투고일: 2026. 3. 15. / 심사완료일: 2026. 4. 8. / 게재확정일: 2026. 4. 12.

* 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2019-0-01842, 인공지능대학원지원(광주과학기술원))과 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. RS-2025-02217259, 뇌인지 다학제 국제 거버넌스 기반 인공지능 편향성 검출-교정-설명가능 지능적 자율진화 플랫폼 개발)을 받아 수행된 연구임. 영향평가에 대한 문제의식을 일깨워주신 한국지능정보사회진흥원(NIA)과 논문 심사 과정에서 유익한 코멘트를 주신 익명의 심사위원 분들께 깊은 감사를 드린다.

** 제1저자: 광주과학기술원(GIST) 전기전자컴퓨터공학과 학사과정

First author: B.S. Student in the Department of Electrical Engineering and Computer Science at Gwangju Institute of Science and technology

*** 참여저자: 광주과학기술원(GIST) AI정책전략대학원 박사후연구원, 법학박사

Participating author: Post-doctoral Researcher in the Graduate School of AI Policy and Strategy at Gwangju Institute of Science and technology

**** 교신저자: 광주과학기술원(GIST) AI정책전략대학원 조교수, 법학박사, 변호사

Corresponding author: Assistant Professor in the Graduate School of AI Policy and Strategy at Gwangju Institute of Science and technology

목 차

- | | |
|---|------------------------------------|
| I. 들어가며
II. 고영향 인공지능 영향평가 개관
III. 타 영향평가 제도와의 비교 분석 | IV. 고영향 인공지능 영향평가의 개선 방향
V. 맺음말 |
|---|------------------------------------|

인공지능의 건전한 발전과 신뢰 기반 조성을 목표로 제정된 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 ‘인공지능기본법’)」이 2026년 1월 22일부터 시행되었다. 이 법의 목적인 인공지능 산업 진흥과 기본권 보호라는 두 마리 토끼를 모두 잡을 수 있으려면 위험이 큰 인공지능에 집중하여 규제를 시행할 필요가 있다. 이를 위해 유럽연합 인공지능법은 인공지능의 위험도를 네 단계로 분류하여 그중에서 핵심이 되는 고위험 인공지능에 대해 집중적 규제를 가하였고, 그 일환으로 기본권 영향평가를 도입하였다. 국내 인공지능기본법도 이와 유사하게 고영향 인공지능 위주의 규제 체계를 도입하였고, 그에 따른 기본권 침해 우려를 최소화하고자 고영향 인공지능 영향평가를 신설하였다. 그러나 현행 고영향 인공지능 영향평가는 인공지능기본법의 균형 잡힌 입법 취지를 담보하지 못하고 있다. 자율규제 형태로 제정되어 제도의 실효성을 확보하기 어려울뿐더러, 평가 시 인공지능이 가져다주는 편익을 고려하고 있지 않기 때문이다. 이에 본고는 국내의 여타 영향평가와 타 관할권 인공지능 영향평가 사례를 비교 분석하여 고영향 인공지능 영향평가 제도의 한계를 도출하고 보완책을 제시한다.

[주제어] 인공지능, 인공지능기본법, 고영향 인공지능, 영향평가, 유럽연합 인공지능법

I. 들어가며

인공지능 산업의 발전과 국민의 기본권 보호라는 두 가지 목적을 달성하고자 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 ‘인공지능기본법’)」이 2026년 1월 22일부터 시행되고 있다. 정부는 인공지능기본법의 성격을 인공지능 생태계를 조성하기 위한 법으로 규정하면서,¹⁾ 이 법의 핵심적 규제 대상인 고영향 인공지능의 범위를 최대한 좁게 해석하겠다는 입장을 밝혔다.²⁾ 또한 최근 발표된 인공지능 행동 계획에는 규제 조항을 개선하기 위한 인공지능기본법 개정 협의체를 구성하고 중복

1) 관계부처 합동, AI 분야 규제합리화 로드맵, 2025, 6쪽.

2) 조선비즈, 심민관, “과기정통부 “AI 기본법 규제 받는 ‘고영향 AI’ 대상 엄격히 해석·최소한 규제””, 2025. 12. 24, <<https://biz.chosun.com/it-science/ict/2025/12/24/TKMPT6PCMRFGJMZSKAMDMDM5CSUE/>>, 검색일: 2026. 4. 10.

규제를 해소한다는 내용이 포함되기도 하였다.³⁾ 이러한 정부의 기초를 종합할 때 인공지능기본법은 두 가지 목적 중 전자에 조금 더 무게를 두고 있는 것으로 보인다.

그러나 인공지능기본법이 오로지 산업 진흥에만 매몰되지 않으려면, 동전의 이면에 있는 또 다른 중요한 목표인 기본권 보호에도 소홀하지 않을 필요가 있다. 기본권 보호는 헌법에 명시된 국가의 의무로서 그 자체로도 중요한 가치이지만 이를 통해 인공지능에 대한 국민적 신뢰를 높여 인공지능 발전에 일조할 수 있기도 하다. 이에 따라 인공지능기본법은 기본권 보호를 달성하기 위해 제4장에 ‘인공지능윤리 및 신뢰성 확보’라는 표제하에 여러 조항(제27조 내지 제36조)을 규정하였다.

그중에서도 인공지능사업자에게 적용되는 고영향 인공지능 영향평가에 관한 제35조는 특히 주목할 필요가 있다. 이 제도는 인공지능기본법의 핵심적 규율 대상인 고영향 인공지능이 사람의 기본권에 미치는 영향을 평가하여 잠재적 위험을 사전에 방지하기 위해 도입된 것이다. 이미 여타 과학기술 관련 법규에서 시행 중인 영향평가 제도를 인공지능 맥락으로 옮겨와 기본권 보호를 충실히 하려는 것이 이 조항을 도입한 취지이다.⁴⁾

그럼에도 불구하고 현행 고영향 인공지능 영향평가 제도에는 여러 가지 한계가 있다. 첫째로 영향평가의 내용 및 방법에 관한 전반적 과정이 인공지능사업자의 자율에 맡겨져 있어 국민의 기본권 보호라는 제도의 목적 달성이 불투명하다는 것이다. 공공기관이 고영향 인공지능을 활용한 인공지능 제품 또는 서비스를 이용하려는 경우 영향평가를 수행한 것을 우선적으로 고려하도록 하여 인센티브를 일부 제공하고 있지만 이것만으로는 제도의 실효성을 담보하기가 쉽지 않다.

둘째로 영향평가 시행 과정에서 고영향 인공지능으로 얻을 수 있는 편익에 대한 분석이 누락되어 있다는 점이다. 채용 심사 과정에서 인간의 편향을 절감함으로써 평등권 실현에 기여하는 것처럼, 인공지능의 활용은 위험뿐만 아니라 편익도 가져다 줄 수 있다. 그러나 현 인공지능 영향평가 제도에는 인공지능의 위험만이 반영되어 있고 편익과의 비교형량 절차가 부재하다. 이는 정부가 공언한 인공지능 산업 진흥을 오히려 저해할 수 있는 역설적 상황을 초래하고 있다.

고영향 인공지능 영향평가의 이러한 한계를 극복하기 위해 본고는 고영향 인공지능 영향평가를 자율적 규율에서 강제적 의무로 전환할 것을 제안한다. 다만 이 과정에서 발생할 수 있는 부작용을 해결하기 위해 영향평가 시 고영향 인공지능의 위험에 더해

3) 국가인공지능전략위원회, 대한민국 인공지능 행동계획, 2026, 58-59쪽.

4) 과학기술정보통신부/정보통신정책연구원, 인공지능 영향평가 가이드라인, 2026, 3쪽.

편익을 평가하는 과정을 추가할 필요가 있다. 이를 통해 실효성을 담보하면서도 산업 진흥을 과도하게 저해하지 않는 균형 잡힌 제도를 구축하는 데 기여하는 것이 본고의 최종적 목표이다.

II. 고영향 인공지능 영향평가 개관

고영향 인공지능 영향평가 제도의 의의와 한계를 구체적으로 진단하기 위해서는 먼저 그 대상이 되는 고영향 인공지능의 개념을 명확히 규정하고 이를 규율하는 전반적 규제 체계를 분석할 필요가 있다. 이에 따라 인공지능기본법과 시행령의 관련 내용을 분석하고 그 세부적 사항을 구체화한 인공지능 영향평가 가이드라인을 고찰하여 고영향 인공지능 영향평가의 의의와 한계를 도출하도록 한다.

1. 고영향 인공지능의 개념 및 규제 체계

1.1 고영향 인공지능의 개념

인공지능기본법 제2조 제4호는 ‘고영향 인공지능’을 “사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위협을 초래할 우려가 있는 인공지능시스템으로서 다음 각 목의 어느 하나의 영역에서 활용되는 것”으로 정의한다. 여기서 인공지능시스템이란 동조 제2호가 정의한 내용을 따른다. 정부가 인공지능기본법 시행과 동시에 배포한 고영향 인공지능 판단 가이드라인은 고영향 인공지능의 요건을 다음과 같은 두 가지로 세분화하여 해석한다.⁵⁾

첫째는 해당 인공지능시스템이 동호 각목에서 규정한 11개 영역에서 활용되어야 한다는 것이다. 여기서 가목부터 차목까지는 교통, 교육 등 구체적 영역 10개를 예시한 것이고, 마지막 카목은 “그 밖에 사람의 생명·신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역으로서 대통령령으로 정하는 영역”으로 하위법령의 위임을 통한 영역의 추가를 규정한 것이다. 다만 법 시행과 함께 도입된 시행령은 새로운 영역을 규정하고 있지는 않다.

5) 과학기술정보통신부/한국지능정보사회진흥원, 고영향 인공지능 판단 가이드라인, 2026, 20쪽.

둘째는 첫째의 요건을 갖춘 인공지능시스템이 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위협을 초래할 우려가 있어야 한다는 것이다. 여기서 영향이나 위협을 초래하는 대상은 직접적 대상인 ‘이용자(제2조 제8호)’ 또는 간접적 대상인 ‘영향받는 자(제2조 제9호)’를 포괄한다. 문언상 여기서의 영향에는 긍정적, 부정적인 것이 모두 포함될 수 있지만 가이드라인은 입법취지에 따라 주로 부정적 영향에 초점을 둔다고 명시하고 있다.⁶⁾ 이는 긍정적 영향이 큰 인공지능이 도리어 규제의 대상이 되는 역설적 상황을 회피하기 위한 해석으로 보인다.

1.2 고영향 인공지능 규제 체계

고영향 인공지능이 사람의 생명, 신체의 안전 및 기본권에 부정적 영향을 미치거나 위협을 초래할 우려를 방지하기 위해 인공지능기본법 제4장은 인공지능윤리 및 신뢰성 확보에 관한 내용을 다룬다. 여기서 인공지능윤리란 “안전하고 신뢰할 수 있는 인공지능사회를 구현하기 위하여 인공지능의 개발, 제공 및 이용 등 모든 영역에서 사회구성원이 지켜야 할 윤리적 기준(제2조 제11호)”을 말하고, 신뢰성은 인공지능윤리가 달성해야 할 핵심 가치를 의미한다.⁷⁾

〈표 II-1〉 인공지능기본법상 주체별 책무

구분		근거 조항	내용
정부		제27조	인공지능 윤리원칙 등
		제29조	인공지능 신뢰 기반 조성을 위한 시책의 마련
		제30조	인공지능 안전성·신뢰성 검·인증등 지원
인공지능사업자	모든 인공지능사업자	제32조	인공지능 안전성 확보 의무
		제36조	국내대리인 지정
	고영향 인공지능사업자	제31조	인공지능의 투명성 확보 의무
		제33조	고영향 인공지능의 확인
		제34조	고영향 인공지능과 관련된 사업자의 책무
제35조	고영향 인공지능 영향평가		
기타		제28조	민간자율인공지능윤리위원회 설치

6) 과학기술정보통신부/한국지능정보사회진흥원, 위의 책, 21쪽.

7) 다만 인공지능기본법의 표제, 제1조, 제4장의 표제 등에서는 ‘신뢰성’만을 핵심 가치로 제시하면서 제2조 제11호, 제3조 제1항 등에서는 ‘신뢰성’과 ‘안전성’을 병치하여 핵심 가치로 제시하는 것은 정합적이지 않다. 이러한 지적은 송다솜/김다영/노재경/박도현, “인공지능 신뢰성과 안전성의 경계 획정”, 저스티스 제206호, 한국법학원, 2025, 221-263쪽 참조.

인공지능기본법 제4장은 크게 정부, 인공지능사업자, 기타 주체의 책무로 나누어 조항을 구성하고 있다. 제27조, 제29조, 제30조는 정부 또는 정부의 일원인 과학기술정보통신부장관의, 제31조 내지 제36조는 인공지능사업자의 책무를 규정하였고, 제28조는 인공지능사업자에 일부 기관을 더한 포괄적 주체의 책무를 규정하였다. 이중 고영향 인공지능 규제는 제31조, 제33조 내지 제35조에 마련되었다. 위의 <표 II-1>은 이를 도식화하여 정리한 것이다.

2. 고영향 인공지능 영향평가의 이해

2.1 고영향 인공지능 영향평가의 의의

고영향 인공지능 영향평가의 명확한 정의는 없지만, 일반적으로 인공지능 영향평가란 인공지능시스템의 설계, 개발, 배포 과정에서 조직이나 사회에 발생할 수 있는 다양한 위험 수준을 식별하고 이를 완화하거나 제거하기 위한 관리 도구 또는 평가 기준을 가리키고, 그 궁극적 목표는 인공지능윤리를 실현하는 것이다.⁸⁾ 즉, 고영향 인공지능 영향평가는 특히 높은 수준의 위험을 초래할 수 있는 종류의 인공지능에 특화하여 위와 같은 작업을 진행하는 제도를 가리킨다. 이 제도는 인공지능기본법 제35조와 동법 시행령 제28조에 법적 근거를 둔다.

고영향 인공지능 영향평가는 다음과 같은 두 가지 특징을 가지고 있다. 첫째는 수범 주체가 정부에 국한되지 않고 인공지능사업자로 포괄적으로 규정되어 있다는 것이다. 여기서 인공지능사업자란 인공지능산업과 관련된 사업을 하는 법인, 단체, 개인 및 국가기관등을 의미하므로(인공지능기본법 제2조 제7호),⁹⁾ 정부에 비해 보다 포괄적 의미를 가지게 된다. 이는 다른 많은 영향평가 제도가 정부를 수범자로 국한하는 것과 비교할 때 이례적이지만, 정부가 인공지능사업자인 경우를 배제하고 있지는 않아 문제라고 보기는 어렵다.

둘째는 영향평가의 수행 여부를 인공지능사업자의 자율에 맡긴다는 것이다. 인공지능기본법 제35조 제1항은 인공지능사업자가 고영향 인공지능 영향평가를 위해 노

8) 이유봉, AI 영향평가를 위한 입법연구, 한국법제연구원, 2024, 35쪽.

9) 인공지능사업자는 재차 “인공지능을 개발하여 제공하는 자”인 인공지능개발사업자와 “인공지능개발사업자가 제공한 인공지능을 이용하여 인공지능제품 또는 인공지능서비스를 제공하는 자”인 인공지능이용사업자로 이원화된다(동호 각목).

력하여야 한다고 규정하여 자율규제 방식을 따르고 있다. 이를 보완하기 위해 동조 제2항은 국가기관등이 고영향 인공지능을 이용한 제품 또는 서비스를 이용하려는 경우에는 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하도록 하고 있지만, 이와 같은 인센티브는 자율규제를 뒷받침하기 위한 수단일 뿐이다. 이것은 다른 많은 영향평가와 비교할 때 이 제도의 한계를 초래하는 주된 요인이 된다.

2.2 고영향 인공지능 영향평가 가이드라인

인공지능기본법과 동법 시행령에 포함된 고영향 인공지능 영향평가 관련 조항은 원론적 내용을 제시한 것이다. 이에 따라 정부는 법 시행과 동시에 인공지능 영향평가 가이드라인을 배포하여 고영향 인공지능 영향평가의 개요 및 수행 단계별 주요 고려사항을 제시하였다. 이에 따르면 영향평가의 목적은 국민의 기본권을 보호함과 동시에 글로벌 상호운용성을 확보하는 데 있다고 한다.¹⁰⁾

국내 고영향 인공지능 영향평가 체계는 해외의 프레임워크를 참고하여 설계된 제도이다. 이러한 비교법적 접근은 국제사회에서의 신뢰성을 강화하고, 역외 규제에 대응하며, 글로벌 시장 진출 시 제도적 장벽을 최소화하는 데 도움이 된다. 인공지능 영향평가 가이드라인은 구체적으로 유럽연합(European Union, 이하 ‘EU’)이 인공지능법(이하 ‘AI Act’)에서 규정한 기본권 영향평가를 벤치마크 하였음을 명시하고 있다.¹¹⁾

인공지능 영향평가 가이드라인은 평가 대상, 시기, 수행 주체를 다음과 같이 규정한다.¹²⁾ 먼저 평가 대상은 앞서 살펴본 고영향 인공지능이지만, 가급적 그에 국한되지 않고 모든 인공지능을 대상으로 영향평가를 수행·실천할 것을 장려한다. 평가 시기는 신규 인공지능 제품 또는 서비스를 제공하기에 앞서 선제적으로 진행된다. 다만 제품이나 서비스의 핵심 기능, 영향받는 대상, 기본권에 중대한 변화가 발생할 경우 재시행을 권장한다. 끝으로 평가 수행 주체는 인공지능사업자이고, 인공지능개발사업자와 인공지능이용사업자가 상이할 경우 후자가 주체가 된다. 동법 시행령 제28조 제2항에 따라 인공지능사업자가 직접 영향평가를 수행하기가 여의치 않은 경우 전문성을 보유한 제3자에게 의뢰하여 실시할 수도 있다.

10) 과학기술정보통신부/정보통신정책연구원, 앞의 책, 3쪽.

11) 과학기술정보통신부/정보통신정책연구원, 위의 책, 3쪽.

12) 이하의 내용은 과학기술정보통신부/정보통신정책연구원, 위의 책, 46쪽을 요약한 것이다.

고영향 인공지능 영향평가의 구체적 수행 과정은 크게 다음과 같은 3단계로 구분된다. 사전 준비 단계(1단계), 본 평가 수행 단계(2단계), 사후 단계(3단계)가 그것이고, 각 단계별 주요 고려사항은 다음과 같다.

2.2.1 사전 준비 단계 (1단계)

사전 준비 단계는 영향평가의 필요성을 판단하고 평가대상을 구체적으로 규정하는 것을 목표로 한다. 먼저 해당 인공지능이 법에서 규정한 고영향 인공지능에 해당하고 영향평가의 요건이 적용되는지를 검토한다. 고영향 인공지능에 해당하는지 여부는 고영향 인공지능 판단 가이드라인을 바탕으로 인공지능사업자가 검토하되, 판별이 어려운 경우에는 인공지능기본법 제33조에 의거하여 과학기술정보통신부 장관에게 확인을 요청할 수 있다.

영향평가의 필요성이 확인된 뒤에는 평가대상을 구체적으로 규정한다. 이러한 작업은 본 평가에 앞서 대상 인공지능 제품 또는 서비스가 의도된 목적에 맞게 사용될 내용에 관하여 간략히 기술하는 것이다. 여기에는 (1) 제품·서비스 개요, (2) 목적 및 필요성, (3) 적용 분야, (4) 사용 절차, (5) 운영 기간 및 빈도, (6) 이용자·영향받는 자의 범주·특성, (7) 이용자 선택권 등에 관한 정보가 포함된다.¹³⁾

2.2.2 본 평가 수행 단계 (2단계)

본 평가 수행 단계는 영향평가의 대상이 되는 구체적 인공지능 제품 또는 서비스가 사람의 기본권에 미칠 수 있는 영향을 체계적으로 식별하고 분석하는 것을 목표로 한다. 크게 피영향자 분석, 평가대상 작동원리 분석, 기본권 영향 식별 및 분석의 세 가지 세부 단계로 나누어 평가를 진행한다.¹⁴⁾

2.2.2.1 피영향자 분석

여기서는 인공지능 제품 또는 서비스의 제공 결과로 영향을 받을 수 있는 사람, 집단, 환경을 ‘피영향자’로 규정한 뒤, 이들을 직접 피영향자, 간접 피영향자, 의도치 않은/예상치 못한 피영향자로 구분하여 식별한다. 이는 유네스코의 주요(Primary), 간접(Secondary), 예상치 못한/의도하지 않은(Unexpected/Unintended) 피영향자 구분 모델을 차용한 것이다. 다만 가이드라인은 이중 직접 피영향자란 인공지능 제품이나

13) 이상의 내용은 과학기술정보통신부/정보통신정책연구원, 위의 책, 12-14쪽을 요약한 것이다.

14) 이하의 내용은 과학기술정보통신부/정보통신정책연구원, 위의 책, 15-23쪽을 요약한 것이다.

서비스의 영향받는 자(인공지능기본법 제2조 제9호)만이 해당한다고 해석하는데,¹⁵⁾ 구체적 맥락에 따라 이용자(인공지능기본법 제2조 제8호)도 포함될 여지가 있다는 것이 보다 합리적 해석으로 보인다.

2.2.2.2 평가대상 작동 원리 분석

다음으로는 후술할 기본권 영향 식별 및 분석의 기초 자료로 활용하기 위해 인공지능 기반 제품 또는 서비스의 작동원리를 자세하게 조사한다. 전체적 인공지능 활용 과정을 데이터 수집, 분석·판단, 결정, 피드백 등의 단계로 구분하여 체계적 조사를 꾀한다. 이 단계에서는 가치 판단이 필요한 요소는 배제하고 오로지 기술적, 운영적 메커니즘을 객관적으로 규명하는 데만 집중한다.

2.2.2.3 기본권 영향 식별 및 분석

그 이후 기본권에 미치는 영향을 식별하고 분석하는 단계가 이어지는데, 이는 시나리오 기반 위협 분석, 관련 기본권의 식별, 기본권 영향 판단의 순서로 진행된다. 먼저 앞서 파악된 작동원리를 바탕으로 서비스 운영 중 발생할 수 있는 잠재적 위협을 최대한 많은 시나리오로 폭넓게 작성한다. 다음으로는 이 잠재적 위협이 헌법상 어떤 기본권과 연관되는지를 식별한다. 이때 체계적 평가를 위해 피영향 기본권을 ‘제품 또는 서비스 차원’과 ‘인공지능기술 일반 차원’으로 구분하여 구체적, 추상적 차원의 분석을 통합적으로 진행한다. 끝으로 앞선 단계에서 파악한 요소들을 종합하여 인공지능이 기본권에 미치는 영향을 분석한다.

2.2.3 사후 단계 (3단계)

사후 단계는 영향평가 완료 이후의 후속 조치와 문서 관리에 관한 것이다. 이 단계에서는 식별된 부정적 영향을 완화·제거하기 위한 개선·보안 계획을 수립하고, 피드백을 반영한 운영 정책 및 절차의 업데이트, 부정적 영향이 현실화될 경우 취할 조치의 사전적 마련 등을 인공지능사업자가 자율적으로 시행한다. 영향평가의 결과와 영향평가 수행 과정 중 수집·작성된 근거자료는 인공지능사업자가 자율적으로 보관 및 관리할 것이 권장되고, 결과의 공개 여부는 사업자가 자율적으로 판단한다.¹⁶⁾

15) 과학기술정보통신부/정보통신정책연구원, 위의 책, 15쪽.

16) 이상의 내용은 과학기술정보통신부/정보통신정책연구원, 위의 책, 24쪽을 요약한 것이다.

2.3 고영향 인공지능 영향평가의 문제점

먼저 인공지능기본법 제35조 제1항은 고영향 인공지능 영향평가의 수행 및 결과 공개 여부를 인공지능사업자의 자율적 판단에 맡기고 있어 제도의 강제성이 없다. 이를 보완하기 위해 동조 제2항은 국가기관등이 고영향 인공지능을 이용한 제품 및 서비스를 이용하려는 경우에는 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하도록 하는 인센티브를 두고 있기는 하다. 그러나 인공지능사업자 입장에서 과연 이것이 충분한 유인이 될지에 대해서는 다음과 같은 두 가지 의문이 있다.

첫째, 해당 인센티브는 제품 또는 서비스에 관한 것일 뿐, 영향평가 수행 그 자체의 비용부담에 관한 것이 아니다. 즉, 인공지능사업자는 추후 정부조달 과정에서 얻을 수 있는 불확실한 대가를 기대하고 현재의 확실한 비용을 지출해야 하는 셈인데, 이러한 논리가 실효성이 있을지는 의문이다. 둘째, 영향평가의 구조가 인공지능사업자 스스로 자신이 제공하는 제품·서비스에 대한 부정적 정보를 공개하게끔 되어 있는 근본적 한계가 있다. 따라서 부정적 정보는 축소하거나 은폐하고 긍정적 정보에만 초점을 두는 방식으로 영향평가가 왜곡되지 않을까 우려된다.

또한 이론상 ‘고영향’에 긍정적 영향이 포함될 수 있음에도 실무상 부정적 영향에만 집중하고 있다는 문제도 있다. 인공지능기본법과 동법 시행령은 영향평가 과정에서 의 영향을 부정적인 것으로만 국한하지 않았다. 따라서 고영향 인공지능의 일부 활용 맥락에서는 그것의 부정적 영향을 긍정적 영향이 상쇄할 여지가 얼마든지 있다. 그러나 인공지능 영향평가 가이드라인에 따른 위협 시나리오는 명시적으로 부정적 영향에 초점을 두어 이러한 가능성을 염두에 두고 있지 않다. 이에 따라 고영향 인공지능 영향평가는 기본권 보호를 위해 위협을 사전적으로 예방하는 데만 중점을 두어 기술과 산업의 진흥을 억제할 소지가 있어 보인다.

Ⅲ. 타 영향평가 제도와의 비교 분석

현행 국내법상 ‘영향평가’라는 명칭을 명시적으로 사용한 제도로는 고영향 인공지능 영향평가를 포함하여 총 37개가 운용되고 있다.¹⁷⁾ 개별 영향평가 제도의 시행

17) 이는 형식적 의미의 법률상 별개의 조문을 단위로 한 것이고, 산업전환에 따른 고용안정 지원 등에

과정에서 얻은 교훈이 현행 영향평가 체계 전반에 녹아있음을 고려한다면, 고영향 인공지능 영향평가를 독립적 제도가 아닌 전체 영향평가 체계 속에서 분석하는 것은 유의미한 일이다. 따라서 이하에서는 기존의 영향평가 체계와 고영향 인공지능 영향평가 간의 대내적 비교 분석을 통해 이 제도의 규범적 의의와 한계를 확인하도록 한다.

또한 인공지능 영향평가 가이드라인에 따르면 국내 고영향 인공지능 영향평가 제도는 타 관할권 규제 체계를 참고한 것이므로 그것이 어떤 내용으로 이루어졌는지 살펴볼 필요가 있다. 이러한 작업은 단일 관할권 내에서 특정 제도의 전수조사를 통한 대내적 비교 분석과는 구분되는 대외적 비교 분석이라는 점에서 학술적 가치가 있다.¹⁸⁾ 이하에서는 구체적으로 EU AI Act의 기본권 영향평가를 분석의 대상으로 삼아 그 의의와 한계를 확인하고, 이것이 국내 고영향 인공지능 영향평가 체계에 어떻게 녹아들었는지를 비교법적으로 검토한다.

1. 국내법상 영향평가 제도

1.1 영향평가 제도의 의의

영향평가는 어떤 행위 또는 작용이 결과에 미치는 인과적 효과를 분석하는 것으로, 수행 시기에 따라 정책이나 사업 시행 전에 이루어지는 사전적 영향평가와 시행 후에 이루어지는 사후적 영향평가로 나뉜다.¹⁹⁾ 사전적 영향평가의 대표적 사례로는 1977년 제정된 구 환경보전법 제5조에 근거를 두고 도입된 국내 최초의 영향평가인 환경영향평가제도가 있다. 사후적 영향평가의 대표적인 예로는 성별영향평가법 제10조에 근거하여 법령, 정책, 공공기관이 수행하는 사업이 성평등에 미칠 수 있는 영향을 평가하는 제도인 특정성별영향평가가 있다. 고영향 인공지능 영향평가는 인공지능기본법 제35조 제1항에서 명시적으로 사전적 영향평가를 표방하고 있다.

관한 법률 제8조 ‘고영향 영향 사전평가’의 경우도 그 취지를 고려하여 분석 대상에 포함하였다.

18) 타 관할권의 인공지능 영향평가 사례에 관한 선행연구로는 김법연, “공공분야 인공지능서비스의 영향평가제도 도입에 관한 연구”, 정보법학 제27권 제2호, 한국정보법학회, 2023, 177-186쪽; 김근혜/박규동, “AI 시스템의 위험 완화를 위한 정책적 접근방안 연구: AI 영향평가를 중심으로”, 규제연구 제32권 제2호, 한국규제학회, 2023, 105-134쪽 등이 있다. 다만 이 선행연구들은 EU AI Act가 최종 제정되기 이전의 것이므로 본 논문에서 기본권 영향평가를 검토하는 것은 별도의 의의를 가진다.

19) 왕승혜, 법령상의 영향평가제도 정비방안 연구, 한국법제연구원, 2024, 45쪽.

대부분의 영향평가 제도는 전자의 방식을 채택함으로써 잠재적 위협의 선제적 관리라는 ‘사전 예방’의 가치를 지향한다. 그러나 사후적 영향평가가 역시 실제로 발현된 위협을 진단하고 이를 피드백한다는 점에서 위험 관리의 본질적 목적과 밀접하게 맞닿아 있다. 즉, 사전적 영향평가와 사후적 영향평가는 시점의 차이만을 가질 뿐, 위협의 적정 관리라는 단일한 목표를 지향한다. 나아가 사후적 영향평가가 그로 인해 발생할 신규 위협원을 차단하는 데 활용될 수 있으므로 양자의 구분이 모호해지는 경우가 많기도 하다. 이러한 이유로 이하에서는 이들을 통합적 관점에서 분석하도록 한다.²⁰⁾

1.2 영향평가 제도의 현황

현재 국내에서 시행 중인 영향평가 제도는 총 37개이고, 이중 명칭에 ‘기본법’이 명시된 법률에 규정된 것은 총 11개이다. 기본법은 개별법과 상하위의 관계에 있지 않지만, 국가 정책의 방향을 제시하고 추진한다는 그 기능을 고려할 때 추상적 성격이 강한 편이다.²¹⁾ 즉, 기본법에 근거한 영향평가 제도는 개별법에 근거한 영향평가 제도보다 적용 범위가 넓고, 보편적 성격을 갖는다.

〈표 III-1〉 기본법에 근거한 영향평가 제도

분야	영향평가 명칭	근거 기본법 조항
환경	환경영향평가	환경정책기본법 제41조
	기후변화영향평가	기후위기 대응을 위한 탄소중립·녹색성장 기본법 제23조
	기후보건영향평가	보건의료기본법 제37조의3
	해양이용영향평가	해양환경 보전 및 활용에 관한 법률 제20조
사회	성별영향평가	양성평등기본법 제15조
	고용영향평가	고용정책기본법 제13조
	어문규범의 영향평가	국어기본법 제12조
	문화영향평가	문화기본법 제5조 제5항
과학	고영향 인공지능 영향평가	인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 제35조
	지능정보서비스 등의 사회적 영향평가	지능정보화 기본법 제56조
	기술영향평가	과학기술기본법 제14조

20) 같은 맥락에서 여러 법제에 산재한 영향평가와 리스크평가 제도 역시 통합적으로 고찰할 필요성을 제기한 선행연구로 윤혜선, “AI 거버넌스의 새로운 축(軸)-영향평가제도에 관한 연구-”, 공법연구 제53집 제2호, 한국공법학회, 2024, 377-424쪽 참조.

21) 박영도, 기본법의 입법모델연구, 한국법제연구원, 2006, 24쪽.

기본법에 근거한 영향평가 제도는 영향평가의 대상이 활용되는 분야를 기준으로 환경, 사회, 과학으로 분류할 수 있다. 환경 분야에는 4개, 사회 분야에는 4개, 과학 분야에는 3개의 영향평가가 마련되어 있다. 요컨대 고영향 인공지능 영향평가는 인공지능기본법이라는 기본법에 근거한, 과학 분야의 영향평가 제도에 속한다. 기본법에 근거한 영향평가 제도 11개의 자세한 내용은 위 <표 III-1>과 같다.

본고는 기본법에 근거한 11개 영향평가 제도 중 개별 법률에서의 활용 빈도가 특히 높은 환경영향평가와 기술영향평가를 자세히 검토한다. 환경영향평가는 국내에 가장 먼저 도입된 영향평가 제도로 거의 모든 개발 사업에서 필수적으로 적용되고 있다. 기술영향평가 또한 신기술을 대상으로 매년 수행되어 이행 빈도가 높은 편이고 고영향 인공지능 영향평가와 같은 과학 분야에 속해있다. 따라서 이들 두 영향평가 제도를 상세히 살펴보는 것은 고영향 인공지능 영향평가 제도의 개선 방향을 제언하는 데 있어 중요한 가치를 지닌다.

1.2.1 환경영향평가

환경영향평가는 환경에 영향을 미치는 사업의 계획이 제출되었을 때 정부가 그것이 환경에 미치는 영향을 미리 평가하여 문제를 제거하거나 감소시키는 방안을 마련하는 제도로(환경영향평가법 제2조 제2호), 그 법적 근거는 환경정책기본법 제41조이다. 여기에는 소규모 환경영향평가도 포함되어 환경에 영향을 미치는 사실상 모든 사업에서 필수적으로 진행된다. 이와 같이 환경영향평가의 대상이 매우 광범위하기에, 환경영향평가는 해당 제도의 법적 근거를 직접 규정한 환경정책기본법 이외에 총 16개의 개별 법률에서 규정한다.

환경영향평가는 사전예방 원칙(precautionary principle)에 입각한 대표적 제도 중 하나이다. 사전예방 원칙이란 환경오염 또는 파괴처럼 미리 개입하지 않을 경우 회복할 수 없는 중대한 악결과를 가져올 가능성이 있는 때에는, 비록 그 가능성이 불확실하고 과학적 입증이 부족한 경우에도 위험을 예방하는 조치를 취할 수 있다는 원칙이다.²²⁾ 즉, 환경영향평가는 그 영향력이 광범위하고 중대한 평가대상의 위험을 사전에 식별하고 제거하기 위해 도입된 제도인 셈이다.

22) 권은정, “인공지능 서비스 영향평가의 체계와 방법론 - 「지능정보화 기본법」의 사회적 영향평가 제도를 중심으로-”, 경제규제와 법 제16권 제1호, 서울대학교 법학연구소, 2023, 34쪽.

1.2.2 기술영향평가

기술영향평가는 신기술의 발전이 사회에 미치는 영향에 대한 정부 주도의 사전적 영향평가 제도로 과학기술기본법 제14조 제1항에 법적 근거를 두고 있다. 기술영향평가는 과학기술정보통신부 장관이 매년 시행하도록 되어 있다(동법 시행령 제23조 제3항). 동법 시행령 제23조 제2항 제3호는 영향평가 결과 해당 기술이 부작용을 초래할 가능성이 있는 경우 그 방지 방안 등을 필수적으로 포함하도록 하고 있어, 이것이 환경영향평가와 마찬가지로 잠재적 위험에 대한 사전 예방 조치임을 확인할 수 있다.

기술영향평가의 평가 대상은 신기술과 기술적·경제적·사회적 영향과 파급효과 등이 큰 기술이다(동법 시행령 제23조 제1항). 그리하여 입법자는 기술영향평가 제도를 과학기술기본법이라는 기본법에 도입하여 그것의 보편적 성격을 강조하였다. 이러한 특성으로 인해 기술영향평가는 환경영향평가와 같이 여타의 개별 법률에서 인용되는 모습을 확인할 수 있고, 그 범위는 총 4개 법률에 달한다.

1.3 비교 분석

이상의 분석을 종합하자면 국내 영향평가 체계는 평가 대상의 영향력이 보편적이고 광범위한 경우 기본법에 근거를 두고, 적용 분야가 구체적으로 특정되는 경우 개별법에 그 근거를 두는 이원적 구조를 취한다. 기본법에 근거한 영향평가 제도는 해당 평가 대상이 사회구조에 초래할 수 있는 광범위한 위험을 대상으로 삼는다. 환경영향평가, 기술영향평가 등 파급효과가 모든 국민에게 미치고 중대성이 인정되는 제도가 그 대표적 사례이다.

고영향 인공지능은 사회에 보편화될 경우 그 파급효과가 광범위하고 국민의 기본권에 중대한 영향을 미칠 가능성이 높으므로 이를 기본법에 둔 것은 입법기술상 타당한 선택으로 보인다. 특히 고영향 인공지능 영향평가의 실시 시점을 신규 인공지능 제품 또는 서비스 제공 전으로 명시하여 그 잠재적 위험을 선제적으로 식별·방지하고자 한 데는 사전예방 원칙이 일정 부분 투영되어 있음을 시사한다. 그럼에도 불구하고 고영향 인공지능 영향평가는 자율규제 방식에 의해 운영되어 형식과 실질이 괴리되어 있는 상태이다.

2. EU AI Act상 기본권 영향평가

국내 인공지능기본법의 전반적 규율에는 세계 최초로 공식 제정된 인공지능 일반 법인 EU AI Act가 많은 영향을 준 것으로 알려져 있다. EU AI Act는 인공지능의 위험도에 따른 차등화된 규제를 적용하고 있어, 고영향 인공지능에 해당하는지에 따라 규제 수준을 달리하는 인공지능기본법과 구조적 유사성을 가진다. 이에 더해 고영향 인공지능 영향평가는 EU AI Act의 기본권 영향평가를 참고한 제도이기도 하다.²³⁾ 따라서 고영향 인공지능 영향평가의 한계를 정밀하게 진단하고 실질적 보완책을 제시하기 위해서는 그 근간이 된 EU AI Act의 전반적 규제 체계 및 기본권 영향평가의 의의와 한계를 분석하는 과정이 필수적으로 요구된다.

2.1 EU AI Act 규제 체계

2024년 8월 1일 발효된 EU AI Act의 핵심적 특징은 인공지능 시스템을 그 잠재적 위험에 따라 ① 수용 불가 위험, ② 고위험, ③ 제한적 위험, ④ 최소 위험의 네 가지 종류로 분류하고 차등적으로 의무를 부과한다는 점이다. 이중 실무에서 가장 큰 비중을 차지하고 있는 고위험 인공지능에 대한 규제 조치는 매우 상세하고 구체적이다.

EU AI Act의 제3장은 고위험 인공지능에 대한 규제 조치를 자세히 규정하고 있다. 제1절은 무엇이 ‘고위험 인공지능’으로 분류되는지에 관한 규칙을 크게 두 가지로 제시한다. 구체적으로 부속서 I에 열거된 조화 입법(harmonisation legislation)에 따라 제3자의 적합성 평가를 받아야 하거나 부속서 III에 열거된 인공지능 시스템은 고위험으로 분류된다(제6조).

고위험으로 분류된 인공지능 시스템은 제2절(제8조 내지 제15조)에 규정된 모든 요구사항을 충족해야 하고, 이를 달성하기 위한 의무는 제3절(제16조 내지 제27조)에서 여러 이해당사자에게 배분되어 있다.²⁴⁾ 기본권 영향평가는 이중 맨 마지막 조문인 제27조에 규정되어 있다. 이 조항은 국내 고영향 인공지능 영향평가의 원형으로 형식과 내용상 많은 부분 공통점을 가진다. 따라서 고영향 인공지능 영향평가의 의의와 한계, 그리고 보완책을 논하기 위해서는 EU AI Act의 기본권 영향평가의 내용에

23) 과학기술정보통신부/정보통신정책연구원, 앞의 책, 3쪽.

24) 보다 상세한 내용에 대한 해설은 최경진 외, EU 인공지능법, 박영사, 2024, 60쪽 이하 참조.

대해 분석하는 과정이 필수적으로 요구된다.

2.2 기본권 영향평가의 규율 내용

EU AI Act의 제27조는 기본권 영향평가를 규율하고 있다. 동조 제1항은 그 수범자를 고위험 인공지능 시스템의 배포자(deployer)로 명시한다. 배포자란 인공지능 시스템을 자신의 권한으로 사용하는 자연인 또는 법인, 공공기관, 대행사 또는 기타 단체로(제2조 제4항), 인공지능기본법상 인공지능이용사업자와 일부 차이가 있으나 밀접한 관련이 있다.²⁵⁾ 평가 시기는 고위험 인공지능 시스템의 배포 전으로 규정되어 있다.

제27조 제1항에는 기본권 영향평가에 포함되어야 할 주요 사항들을 명시하고 있다. 기본권 영향평가를 구성하는 항목은 아래 <표 III-2>와 같고,²⁶⁾ 고영향 인공지능 영향평가 가이드라인에서 제시된 내용과 상당 부분 유사하다.

<표 III-2> EU AI Act 기본권 영향평가 구성 항목

<ul style="list-style-type: none"> (a) 의도된 목적에 따라 고위험 인공지능 시스템이 사용될 배포자의 프로세스에 관한 설명 (b) 각 고위험 인공지능 시스템이 사용될 기간 및 횟수에 관한 설명 (c) 특정한 상황에서 그 사용으로 영향을 받을 수 있는 자연인 및 집단의 범주 (d) 제13조에 따라 제공자가 제공한 정보를 고려하여 이 항 제(c)호에 따라 확인된 자연인 또는 집단의 범주에 영향을 미칠 가능성이 있는 특정한 피해 위험 (e) 사용설명서에 따른 인적 감독 조치의 시행에 관한 설명 (f) 내부 지배구조 및 불만제기 메커니즘에 관한 약정을 포함하여 해당 위험이 실현되는 사례에서 취할 조치

EU AI Act 제27조 제3항은 기본권 영향평가 결과의 통지에 관한 사항을 규율하고 있다. 기본권 영향평가를 실시한 고위험 인공지능 시스템 배포자는 그 결과를 지정된 설문 양식에 맞춰 작성한 후, 관련 시장 감시 당국(market surveillance authority)에 통보하여야 한다. 이는 국내 고영향 인공지능 영향평가 사후 단계의 결과 공개 및 문서 보관에 대응되는 것이다. 다만 기본권 영향평가의 결과는 필수적으로 시장 감시 당국에 통지하도록 규정한 반면, 고영향 인공지능 영향평가는 결과 공개 여부를 인공

25) 이에 대한 보다 상세한 해석은 채은선, “우리나라 인공지능기본법의 주요 내용 및 관련 고려사항”, 경제규제와 법 제18권 제1호, 서울대학교 법학연구소, 2025, 97-100쪽 참조.

26) 세계법제정보센터, “유럽연합 인공지능법 번역본”, <https://world.moleg.go.kr/web/wli/lgslInfoReadPage.do?CTS_SEQ=51409&AST_SEQ=93>, 검색일: 2026. 4. 10.

지능사업자가 스스로 결정하도록 규정했다는 차이가 있다.

2.3 기본권 영향평가와 고영향 인공지능 영향평가의 차이점

국내 고영향 인공지능 영향평가 제도는 외견상 EU 기본권 영향평가와 유사한 형식을 가지고 있지만 그 실질에는 일부 차이가 있다. 첫째는 인공지능의 위험도를 네 단계로 분류한 뒤 2단계인 고위험 인공지능에 집중한 기본권 영향평가와 달리, 국내 인공지능기본법은 사실상 고영향 인공지능과 일반 인공지능시스템²⁷⁾만을 구분한 뒤 전자에 집중한 영향평가를 규정하였다는 것이다. 다만 특별히 위험도가 높은 것으로 평가된 인공지능 유형을 대상으로 규제의 초점을 맞춘 이들의 형식적 유사성과 달리 위험도 유형의 개수 차이는 그렇게까지 유의미한 것으로 보이지 않는다.

둘째는 영향평가 시행의 강제성 여부이다. EU의 기본권 영향평가는 고위험 인공지능 시스템 배포자에게 기본권 영향평가를 의무적으로 수행하도록 하고 있다. 그에 반해 국내 고영향 인공지능 영향평가는 인공지능사업자에게 수행에 대한 자율적 노력의무만을 부여하고 있다. 수행 여부뿐 아니라 수행 이후 절차에서도 기본권 영향평가의 결과는 시장 감시 당국에 필수적으로 통보되도록 규정되어 있으나, 고영향 인공지능 영향평가는 사업자의 자율에 맡겨져 있다.

이러한 점들을 종합해 볼 때, EU의 기본권 영향평가는 기본권 수호가 최우선 가치인 반면, 국내 고영향 인공지능 영향평가는 산업 진흥에 초점을 맞추고 있다고 볼 수 있다. 이에 대해 영향평가 대상이 4개 위험군 중 1개인 EU와 비교할 때 2개 위험군 중 1개인 한국의 규제 범위가 더 넓지 않은가 하는 의문이 제기될 수 있다. 그러나 위험군의 형식적 개수가 적더라도 실질적 범위를 좁게 해석할 수 있고, 그것이 현재 한국 정부의 입장이라는 점에서 이러한 지적은 타당하지 않다.

2.4 비교 분석

한 가지 유의할 점은 국내 고영향 인공지능 영향평가가 많은 면에서 EU AI Act 기본권 영향평가를 벤치마킹한 것은 사실이지만, EU의 제도에도 한계가 있다는 것이

27) 고영향 인공지능은 그 정의상 인공지능시스템의 일종이지만 법률에서 명시적으로 ‘고영향 인공지능 시스템’ 대신 ‘고영향 인공지능’이라는 표현을 활용하였기에 본문에서도 이에 따라 서술하였다.

다. 이와 관련하여 Claudio Novelli와 동료들은 EU AI Act가 가지고 있는 네 가지 한계를 제시하였다. 해당 연구에 의하면 EU AI Act는 위험(Risk)을 위험원(Hazard)이 실제 손해(Loss)로 이어질 가능성으로 정의하는 전통적 방식을 따르고 위험을 방지하기 위한 보호장치(Safeguards)가 많아질수록 실제 위험은 감소한다고 상정한다.²⁸⁾ 이에 따라 EU AI Act는 인공지능의 위험을 4단계 범주로 분류하여 규제 체계를 구조화하였는데, 이는 다음과 같은 4가지 한계를 가진다.²⁹⁾

첫째, EU AI Act에는 세부적 위험 평가 방법론이 부재하여, 특정 범주에 해당하는 개별 인공지능 시스템의 특수성을 고려하지 않고 위험도를 일률적으로 부여한다는 문제가 있다. 예컨대 같은 의료 인공지능이라도 환자의 치료 과정에 개입하는 수준이나 구체적 사용 상황에 따라 실제 위험도는 크게 다를 수 있다. 그러나 EU AI Act는 제6조 분류 규칙상 요건을 갖춘 인공지능 시스템을 일률적으로 고위험으로 분류한다.

둘째, EU AI Act는 인공지능의 위험을 지나치게 단순화하여 평가함으로써 법이 실질적으로 수호하고자 하는 핵심 가치들을 충분히 반영하지 못하고 있다. 현재의 평가 과정은 데이터셋의 적절성이나 기술 표준 준수 여부 등 공학적 지표에만 치중되어 있고, 민주적 가치나 인권 등 규범적 논의는 부족하다. 이는 사회적 가치를 위험이라는 단일 매개변수로 치환하는 과정에서 생겨난 일종의 오차에 해당한다.

셋째, EU AI Act의 방법론은 기존에 존재하는 보호장치를 상정하지 않고 위험만을 측정한다는 결점이 있다. 만일 어떤 인공지능이 위험을 산출하더라도 이를 적절히 억제할 제도적 대안이 실재한다면 굳이 이에 대한 규제를 도입할 이유가 없다. 이에 따라 정책 실무에서 보다 중요한 순위험을 측정하는 데 실패하여 인공지능의 위험을 실제보다 과대평가할 여지가 있다.

넷째, EU AI Act의 위험 평가 방법론은 인공지능 활용을 통해 기대되는 잠재적 편익과 위험 간의 비례적 평가 절차가 결여되어 있다. 만일 어떤 인공지능이 위험을 산출하더라도 이를 상쇄하거나 능가하는 편익이 존재한다면 이를 막연히 규제하는 대신 다른 접근을 사용할 수 있다. 그럼에도 편익에 대한 고려 없이 위험만을 기준으로 규제를 도입한다면 인공지능 산업 진흥을 크게 저해하는 요인이 될 수 있다.

이 중 첫째는 국내 인공지능 영향평가 가이드라인에 도입된 시나리오 기반 위험

28) Claudio Novelli/Federico Casolari/Antonino Rotolo/Mariarosaria Taddeo/Luciano Floridi, "AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act", *Digital Society Vol. 3 No. 1*, Springer, 2024, p. 4.

29) 이하의 논의는 Claudio Novelli et al., *Ibid.*, pp. 7-9를 요약한 것이다.

분석을 바탕으로 상당 부분 보완되었고, 둘째는 기본권 영향 식별 및 분석 과정에서 일정 정도 보완될 여지가 있으며, 셋째는 보호장치가 기본권 영향 자체를 제거했다거나 사후 단계에서 해소된 것으로 평가할 여지가 있어 보인다. 반면 넷째는 법률, 시행령, 가이드라인을 종합하더라도 여전히 이 문제가 잔존하고 있다. 바로 이 부분이 EU AI Act의 기본권 영향평가의 한계를 발전적으로 변용해야 하는 대목이기도 하다.

IV. 고영향 인공지능 영향평가의 개선 방향

1. 고영향 인공지능 영향평가 제도의 의무화

앞서 본 것처럼 EU AI Act는 고위험 인공지능 시스템의 배포자에게 기본권 영향평가 수행을 강제하고 그 결과를 시장 감시 당국에 의무적으로 통보하게 함으로써 규제의 실효성을 확보하였다. 이와 같은 접근을 취한 이유는 EU AI Act의 주된 입법 목적이 사람의 기본권 수호에 있기 때문이다. 이는 인공지능 산업의 역동성을 저해할 우려가 있다는 지적도 있기는 하지만, 기본권 보호를 통해 높아진 사회적 신뢰는 장기적으로 산업 발전에 긍정적 영향을 줄 수 있어 속단하기는 이르다.

국내 고영향 인공지능 영향평가 또한 인공지능기본법의 주된 입법 목적 중 하나인 신뢰 기반 조성을 달성하기 위해 사람의 기본권 보호를 위해 도입된 영향평가 제도의 실효성을 확보할 필요가 있다. 그런데 고영향 인공지능 영향평가는 수행 여부를 수범자인 인공지능사업자의 자율에 맡기고 있어 한계가 있다. 이와 같은 문제를 극복하고자 국가기관등에게 영향평가를 실시한 고영향 인공지능을 이용한 제품 또는 서비스를 우선적으로 고려하게끔 하는 인센티브를 마련하였다. 그러나 자율규제의 근본적 한계가 극복된 것은 아닐뿐더러, 위험원을 파악하는 영향평가를 통해 자신이 제공한 제품 또는 서비스의 한계가 드러날 수 있어 자율규제의 실효성이 의심되는 상황이다.

더구나 국내에서 시행 중인 37개 영향평가 제도를 전수조사한 결과 아래 <표 IV-1>과 같이 고영향 인공지능 영향평가를 비롯하여 단 5개의 영향평가 제도만이 자율규제 형태를 취하고 있다. 이는 영향평가와 자율규제가 구조적으로 그리 친숙하지 않음을 방증하는 것이다.

〈표 IV-1〉 자율규제 형태로 제정된 영향평가 제도

영향평가 명칭	근거 조항
디지털의료제품 영향평가	디지털의료제품법 제37조 제1항
지능정보서비스 등의 사회적 영향평가	지능정보화 기본법 제56조 제1항
장외발매소 영향평가	한국마사회법 제44조의3 제1항
화재안전 영향평가	화재의 예방 및 안전관리에 관한 법률 제21조 제1항
고영향 인공지능 영향평가	인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 제35조 제1항

위의 5개를 제외한 대다수 영향평가가 의무적 규정으로 제정된 이유는 영향평가 제도의 1차적 취지가 사람의 기본권 보호를 위한 위험관리에 있기 때문이다. 영향평가 제도의 대다수는 평가 시기를 사업이나 정책 시행 이전으로 정하여 잠재적 위협의 선제적 관리를 목표로 운용되고 있다. 즉, 이러한 목적이 성공적으로 달성되도록 하려면 영향평가 제도를 의무적 규정으로 두어 실효성을 확보할 필요가 있다는 것이다.

더구나 고영향 인공지능 영향평가를 제외한 4개의 영향평가에는 해당 제도가 자율규제 형태로 제정된 타당한 입법론적 사유가 존재한다. 이하에서는 자율규제 형태로 운용되는 나머지 4개 영향평가를 순차적으로 검토함으로써 각각의 제도가 자율규제의 형식을 취하게 된 입법론적 사유를 규명한다. 이를 고영향 인공지능 영향평가와 비교·분석하여 고영향 인공지능 영향평가가 자율규제가 아닌 의무적 규정의 형식을 취해야 함을 논증하고자 한다.

1.1 자율규제 형식의 영향평가 제도 분석

1.1.1 디지털의료제품 영향평가

디지털의료제품 영향평가는 디지털의료제품의 활용과 확산이 사회·경제·문화 및 국민보건에 미치는 영향을 조사 및 평가하는 제도를 가리킨다. 이 제도는 디지털의료제품의 안전성 관리를 위해 2024년 세계 최초로 제정된 디지털의료제품법 제37조에 법적 근거를 둔다. 영향평가의 주체는 국가기관인 식품의약품안전처장이며, 평가 시기는 동법 시행규칙 제48조 제5항에 의하여 식품의약품안전처장이 디지털의료제품과 관련된 사회·문화·경제적 환경의 변화 추세를 고려하여 정하도록 하고 있다.

디지털의료제품 영향평가가 자율규제의 형식을 취한 이유는 그것의 주된 목적이 디지털의료제품의 위험으로 인한 사람의 기본권 침해를 직접적으로 겨냥하고 있지

않기 때문이다. 동법 제37조 제2항에 따르면, 식품의약품안전처장은 영향평가의 결과를 디지털의료제품과 관련한 정책의 수립·시행에 반영하도록 노력해야 한다. 물론 여기에는 임상적 안전성 및 유효성, 국민 건강 및 보건의료 체계에 미치는 영향처럼 기본권 침해가 어느 정도 고려되기는 하지만, 정부의 정책을 통해 장기적이고 간접적으로 반영되는 것이지 사업자의 행위에 직접적으로 반영되는 것은 아니다.

1.1.2 지능정보서비스 등의 사회적 영향평가

지능정보서비스 등의 사회적 영향평가는 국민의 생활에 파급력이 큰 지능정보서비스 등의 활용과 확산이 사회·경제·문화 및 국민의 일상생활 등에 미치는 영향을 평가하는 제도로, 지능정보화 기본법 제56조에 법적 근거를 둔다. 지능정보화 기본법은 4차 산업혁명이 대두되던 2020년 이로 인한 부작용을 해결하기 위해 기존의 국가정보화 기본법을 전부개정한 것이다. 영향평가의 수행 주체는 국가 및 지방자치단체이고, 평가 시기에 대해서는 명확히 제시된 바가 없다.

지능정보서비스 등의 사회적 영향평가가 자율규제의 형식을 취한 이유는 이 제도가 지능정보사회의 기반을 조성하고 정부의 시책을 수립하는 데 있기 때문이다. 즉, 이는 사람의 기본권 침해를 직접 겨냥한 규제가 아닌 지능정보사회의 도래에 따른 사회적 변화에 발맞추어 국가 정책의 방향성을 제시하기 위한 기초적 작업에 해당한다. 제도의 명칭이 비교적 포괄적인 데다가, 동법 제56조 제1항 각호에서 규정된 영향평가의 항목이 대체로 추상적인 점도 이러한 결론을 뒷받침하는 논거가 될 수 있다.

1.1.3 장외발매소 영향평가

장외발매소 영향평가는 장외발매소가 주거환경이나 교육환경 등 지역사회에 미치는 영향을 평가하는 제도로, 한국마사회법 제44조의3에 법적 근거를 두고 있다. 그 수행 주체인 농림축산식품부 장관이 동법 시행규칙 제11조의2 제2항에 따라 7년 이내의 주기로 정기평가를 실시하거나 필요하다고 인정하는 경우 수시평가를 실시할 수 있다.

장외발매소 영향평가가 자율규제의 형식을 취한 이유는 장외발매소 설치에 대해 이미 승인제라는 강력한 규제가 마련되어 있기 때문이다. 동법 제6조 제2항은 장외발매소 설치 시 농림축산식품부 장관의 승인을 받도록 명시하고 있는데, 초등학교 인근 등 공익에 부합하지 않는 곳에 설치되는 경우를 대비하기 위한 것이다. 즉, 이와 유사

한 강력한 사전적 규제가 동반되어 있지 않은 상황에서도 막연히 장외발매소 영향평가의 자율규제적 방식을 원용하는 것은 곤란하다.

1.1.4 화재안전영향평가

화재안전영향평가는 화재발생 원인 및 연소과정을 조사·분석하는 등의 과정에서 법령이나 정책의 개선이 필요하다고 인정되는 경우 화재 위험성의 유발요인 및 완화방안을 평가하기 위한 제도로, 화재의 예방 및 안전관리에 관한 법률 제21조에 법적 근거를 두고 있다. 화재안전영향평가의 평가 주체는 소방청장이고, 평가 시기는 소방청장의 재량에 따르고 있다(화재안전영향평가 운영절차 등에 관한 규정 제3조).

화재안전영향평가는 그 이름 때문에 자칫 화재를 예방하기 위해 수행되는 영향평가라고 인식할 수 있으나, 실은 이미 발생한 화재의 원인 및 연소과정을 분석하는 과정에서 법령 및 정책의 개선 여지가 있을 때 수행되는 것이다. 해당 영향평가의 대상은 국가의 법령이나 정책으로 명시되어 있으므로, 이를 통해 화재안전영향평가가 화재와 관련한 잠재적 위협을 미리 식별하기 위한 절차가 아님을 알 수 있다. 실질적 화재의 예방은 건축법, 소방시설 설치 및 관리에 관한 법률 등에 마련된 강력한 규제 체계에 따라 엄격히 규율되고 있다. 즉, 이 경우에도 사실관계가 유사하지 않은 상황에 막연히 화재안전영향평가의 자율규제적 방식을 원용하는 것은 곤란하다.

1.2 의무적 규정 형식의 영향평가 제도 분석 - 환경영향평가

지금까지 살펴본 4개의 영향평가 및 고영향 인공지능 영향평가 제도를 제외한 나머지 32개의 영향평가는 모두 의무적 규정 형식을 취하고 있다. 이는 영향평가의 1차적 목적이 사람의 기본권에 초래될 수 있는 위협의 사전적 예방에 있기 때문이다. 32개 영향평가 중 특히 환경영향평가는 국내 실정법에 최초로 도입되었고 다수의 개별법에 원용되는 영향평가의 대표적 사례로, 고영향 인공지능 영향평가에도 많은 시사점을 줄 수 있다. 다만 환경영향평가 제도 자체에 대해서는 앞서 자세히 논의하였기에, 이하에서는 의무적 규정 형식에 대한 부분만을 간단히 검토해 본다.

앞서 살펴보았듯 환경영향평가는 개발사업 과정에서 진행되는 다양한 절차 중 하나가 아닌, 그 사업의 존폐를 결정할 만큼 비중이 큰 절차이다. 이처럼 환경영향평가를 중요한 규제 수단으로 활용하는 이유는 그것이 사전예방 원칙을 구현하는 위험관

리의 핵심적 수단이기 때문이다. 환경의 오염이나 파괴는 국민의 기본권에 큰 영향력을 끼치고 일단 발생한 경우 그 회복 또한 쉽지 않으므로 영향평가 제도를 통해 사전에 위협을 식별하여 예방할 필요가 있다. 이에 따라 환경영향평가를 자율규제가 아닌 의무적 규정 형식으로 규정한 것이다.

1.3 고영향 인공지능 영향평가에의 적용

국내 실정법상 시행 중인 37개 영향평가 제도를 비교 분석한 결과 고영향 인공지능 영향평가는 자율규제 형식으로 규정된 4개 영향평가 제도와 같은 평면에 놓기 어렵다는 결론이 도출되었다. 먼저 디지털의료제품 영향평가나 지능정보서비스 등의 사회적 영향평가는 국가 정책의 방향성이라는 추상적 내용의 기초자료로 활용될 목적을 가진 것이고, 수범자도 공적 주체로 명시되어 있다. 장외발매소 영향평가나 화재안전 영향평가의 경우, 이미 여타 조항을 통해 강력한 사전 규제가 확립되어 있어 단지 영향평가가 자율규제 형식으로 되어 있다는 점에만 주목한다면 본질을 놓치게 된다.

오히려 의무적 규정 형식을 따르는 나머지 32개 영향평가 중 대표적 사례인 환경영향평가 사례가 고영향 인공지능 영향평가에 많은 시사점을 준다. 이것이 의무적 규정으로 도입된 핵심적 이유는 사람의 기본권 보호를 위한 사전예방 원칙을 구현하는 수단이기 때문이다. “사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위협을 초래할 우려가 있는 인공지능시스템”이라는 고영향 인공지능의 정의를 고려하면 그것이 국민의 기본권에 방대한 영향력을 끼치고 일단 위협이 현실화될 경우 그 회복이 쉽지 않다는 사전예방 원칙의 취지가 공통적으로 적용될 여지가 있다. 물론 ‘고영향’의 범위를 획정하는 것이 쉽지 않고 이로부터 과잉규제의 위험성이 예상되기는 하지만, 그것은 고영향 인공지능이라는 범주화 자체에 대해 가능한 비판이지 고영향 인공지능 영향평가 제도의 의무적 규정화에 대한 적절한 반론이라고 보기는 어렵다.

다만 이는 일반론에 불과하므로, 이러한 결론을 적용할 때는 다음의 사항에 유의할 필요가 있다. 먼저 가급적 고영향 인공지능의 유형과 적용 맥락을 세분화하여 EU AI Act의 고위험 내지 금지 범주에 준하는 대상을 세밀하게 골라내는 것이 바람직할 것이다. Novelli 등이 지적한 것처럼 이는 EU AI Act의 한계이기도 한데, 이와 유사한 체계를 가진 한국 인공지능기본법의 구조상 동일한 문제가 발생할 수 있기 때문이다.

나아가 이것이 산업 진흥에 방점을 두고 의도적으로 ‘고위험’ 대신 ‘고영향’이라는 표현을 사용한 인공지능기본법의 입법 취지를 살리는 길이기도 하다.

2. 비례적 평가 절차의 도입

상술한 논의를 통해 고영향 인공지능 영향평가를 의무적 규정으로 전환해야 한다는 결론을 수용하더라도 이 과정에서 한 가지 유의할 지점이 있다. Novelli 등이 지적한 넷째 문제인 영향평가 과정에서 고위험 인공지능이 산출하는 편익을 위험과 형량하는 절차가 부재하다는 것이다. 이것은 국내 인공지능기본법에서 더욱 부각되는 문제인데, 애초부터 ‘고위험’ 인공지능에 집중한 EU와 달리 국내에서는 이 표현을 ‘고영향’으로 수정하였기 때문이다. 개념 정의상 여기서의 영향은 긍정적인 것과 부정적인 것을 모두 포괄함에도 인공지능 영향평가 가이드라인은 부정적 영향, 즉, 위험에 주로 주목하겠다는 입장을 취한다. 이는 산업 진흥에 방점을 둔 입법 취지와 모순되는 것이다.

EU AI Act가 위험과 편익을 비례적으로 형량하지 못한다는 한계에 대하여 Novelli 등은 2단계 해결 방안을 제시했다. 이는 위험 시나리오 구축과 비례성 검사의 단계로 구분된다. 이하에서는 Novelli 등이 제시한 2단계 해결 방안을 자세히 살펴보고 이를 국내 고영향 인공지능 영향평가의 한계를 극복하는 데 적절히 변용해 보도록 한다.

2.1 이론적 논의

Novelli 등의 2단계 해결 방안의 골자는 개별 인공지능 시스템이 운용되는 구체적 상황을 반영하여 위험 시나리오를 도출하고 기존에 인공지능에 부여된 위험도가 타당한지를 비례성 원칙에 따라 검증하는 것이다. 먼저 1단계는 개별 인공지능 시스템의 특수성을 반영한 구체적 위험 시나리오를 구축하여 위험도를 결정한다. 이후 제2단계는 결정된 위험도의 적정성을 평가하기 위해 비례성 검사를 수행한다.

2.1.1 위험 시나리오 구축 (1단계)

EU AI Act의 인공지능 위험도 분류가 문제가 되는 이유는 여러 맥락에서 활용되는 동종의 인공지능에 한 가지 위험도를 일률적으로 부여하기 때문이다. 예컨대 같은

의료 인공지능이라도 사람의 생명권에 직접적 영향을 미치는 것과 그렇지 않은 것은 구분할 필요가 있으나 EU AI Act는 이들을 일률적으로 고위험 인공지능으로 분류한다. Novelli 등은 개별 인공지능의 활용 맥락을 고려하여 위험도를 부여하기 위한 방법론으로 위험 시나리오를 정성적으로 도출한 뒤 다음 과정을 통해 정량화해야 한다고 주장하였다.³⁰⁾

이들은 먼저 인공지능의 위험을 위험 결정 요인(Determinants)이라는 상위 범주와 하위 범주인 위험 동인(Drivers)으로 구분하였다. 위험 결정 요인은 위험원(Hazard), 노출(Exposure), 취약성(Vulnerability), 대응(Response)의 네 가지 핵심 요소로 구성되고, 각각은 구체적 상황을 반영하는 하위 범주인 동인에 의하여 구체화 된다. 이는 ‘기후변화에 관한 정부 간 협의체(Intergovernmental Panel on Climate Change, IPCC)’의 위험 평가 프레임워크를 도입한 것인데, 인공지능 위험 또한 기후변화와 마찬가지로 구체적 변수에 따라 가변적이라는 점에 착안한 것이다.

위험 결정 요인과 위험 동인을 각기 다른 범주로 구분하여 얻을 수 있는 이점은 이들 간의 상호작용을 체계적으로 분석할 수 있다는 점이다. 구체적으로 이 프레임워크는 크게 3가지 종류의 상호작용을 분석하여 기존에 정성적으로 도출한 위험 시나리오를 정량화한다. 첫째는 이질적 위험 결정 요인 간의 상호작용 분석으로, 4가지 위험 결정 요인 중 2가지 이상이 서로에게 미치는 영향을 통해 어떤 위험이 도출되는지를 확인하는 것이다. 둘째는 단일 위험 결정 요인 내에서 위험 동인 간의 상호작용 분석으로, 이를 통해 위험의 발현 양상을 구체화한다. 셋째는 인공지능과 무관한 외부 위험과의 상호작용 분석으로, 시장 위험, 법적 책임 위험, 인프라 위험 등의 외부 요인에 따른 위험을 분석하는 단계이다.

이렇게 수행된 세 가지 상호작용 분석 결과를 바탕으로 발생 가능성(Likelihood)과 피해 심각도(Severity)를 도출하여 해당 시나리오에 맞는 적절한 위험도를 부여한다. 이때 기존에 정성적으로 도출한 위험 시나리오는 위의 두 가지 매개변수를 통해 정량화된다. 이러한 과정은 개별 인공지능 시스템의 위험 수준이 기존 EU AI Act가 설정한 4가지 위험 범주 중 어디에 배치되어야 하는지를 결정하는 근거가 된다. 다만, 이렇게 결정된 위험도는 오직 위험 시나리오에 의해서만 결정된 것이므로 인공지능의 편익을 비례적으로 형량하는 과정을 재차 거칠 필요가 있다.

30) 이하의 내용은 Claudio Novelli et al., *Ibid.*, pp. 9-14를 요약한 것이다.

2.1.2 비례성 검사 (2단계)

1단계에서 구축된 위험 시나리오를 바탕으로 2단계에서 비례성 검사가 수행된다.³¹⁾ 비례성 검사의 핵심은 법철학자 알렉시(Robert Alexy)가 제시한 비례성 원칙 중 협의의 비례성을 수식화한 가중치 공식을 실무적으로 적용하는 데 있다. 협의의 비례성이란 ‘특정한 가치(P_y)를 보호하여 얻는 이득이 대립하는 가치(P_x)를 제한하여 발생하는 손해보다 클 때 해당 수단이 정당화된다는 원칙이다. 인공지능 규제 맥락에 이를 적용하면 어떤 인공지능 시스템에 특정 위험 범주 R 규제를 집행하여 P_x 가 제한되는 정도를 I_x , P_y 를 수호하기 위한 해당 규제의 중요성을 C_y 로 정의한 뒤 여기에 각각의 추상적 가중치 W_x , W_y 를 곱하여 전자를 후자로 나눈다. 이로부터 도출된 $W_{x,y}$ 를 통해 R 규제의 정당성을 평가할 수 있다. <수식 1>은 이를 도식화한 것이다.

<수식 1> 가중치 공식

$$W_{x,y} = \frac{I_x \times W_x}{C_y \times W_y}$$

공식의 각 변수에 실질적 수치를 대입하는 과정에서는 1단계 위험 시나리오 구축을 통해 도출한 위험도를 활용한다. 최소 위험, 제한적 위험, 고위험, 수용 불가 위험 인공지능마다 각각 I_x 에 1, 2, 4, 16의 기하급수적 수치를 대입한다. C_y , W_x , W_y 역시 동일한 기준에 따라 점수를 대입한다. Novelli 등은 최종 결과값인 $W_{x,y}$ 가 1보다 작거나 4보다 큰 경우 EU AI Act의 위험 통제 조치가 현저히 불균형하다고 평가할 수 있다고 주장한다.

2.2 고영향 인공지능 영향평가에의 적용

고영향 인공지능 영향평가가 인공지능이 사람의 기본권에 미치는 긍정적 영향에 대한 분석을 누락한 채로 진행된다면 대부분의 고영향 인공지능에 대해 강력한 규제가

31) 이하의 내용은 Claudio Novelli et al., *Ibid.*, pp. 15-18를 요약한 것이다.

불가피해질 것이다. 이에 따라 Novelli 등이 제시한 2단계 해결 방안, 특히 비례성 검사는 인공지능 산업 진흥을 도모하기 위한 유효적절한 대안이 될 수 있다. 그러나 국내 인공지능기본법과 EU AI Act는 그것의 형식적 유사성에도 불구하고 용어, 체계 등에서 상당 부분 차이가 있으므로 국내 실정에 맞게 변용하여 도입해야 한다.

예컨대 영향평가 과정에 고영향 인공지능이 사람의 기본권에 미치는 긍정적 영향을 식별하는 편익 시나리오 도출 단계를 추가할 수 있다. 여기에서 ‘편익’의 범위는 피영향자의 기본권에 직접적으로 미치는 이익부터 사회 일반에 간접적으로 미치는 긍정적 영향까지 정책적 필요성에 따라 달리 설정하는 것이 가능하다. 이로부터 고영향 인공지능 영향평가의 수행 주체인 인공지능사업자가 자신이 제공한 인공지능 제품이나 서비스가 비례성 검사를 통과함을 증명하고자 하는 인센티브 구조를 설계할 수 있다.

또한 이 방안은 고영향 인공지능 영향평가의 투명성을 강화할 수 있다는 부수적 효과도 기대할 수 있다. 인공지능사업자의 입장에서 위험과 달리 편익은 이를 투명하게 공개할 때 대중의 신뢰를 얻을 수 있으므로 영향평가의 결과를 널리 알리려는 동기가 있기 때문이다. 여기에 편익을 위험과 상쇄할 수 있도록 하는 절차를 둘 경우 위험을 숨기려는 동기를 일정 정도 견제할 수 있다. 나아가 단순히 영향평가를 시행하였다는 사실을 넘어, 편익을 최대한 확보한 것으로 평가된 인공지능 제품이나 서비스를 국가기관이 우선적으로 고려할 필요가 있기도 하다.

V. 맺음말

본 연구는 인공지능기본법상 고영향 인공지능에 대한 규제 체계를 분석하고 중요한 규제 수단인 고영향 인공지능 영향평가의 구체적 내용을 검토하였다. 그 결과, 고영향 인공지능 영향평가는 자율규제 형태로 제정되어 제도의 실효성을 확보하기 쉽지 않고 고영향 인공지능이 기본권에 미치는 긍정적 영향과 부정적 영향을 비례적으로 형량하는 절차가 부재한다는 한계가 있었다. 이는 기존 국내 영향평가 제도와 EU AI Act의 기본권 영향평가와의 비교 분석을 통해 구체적으로 확인되었다.

이러한 두 가지 한계를 극복하기 위해 우선 자율에 맡겨진 고영향 인공지능 영향평가를 의무적 규정으로 전환하여 규범의 실효성을 확보해야 함을 주장하였다. 다만,

강제된 영향평가가 산업의 위축으로 이어지지 않으려면 모든 고영향 인공지능이 아니라 고영향 인공지능의 유형과 적용 맥락을 세분화할 필요가 있다. 아울러 위험 시나리오에만 치중된 현재의 영향평가 실무 관행에 편익 시나리오 도출 단계를 추가하여 합리적 인센티브 구조를 설계해야 할 것이다.

다만 본 연구의 결론과 달리 자율규제와 의무적 규정화 사이의 어떤 중간지점이 우리 사회가 처한 여러 여건을 고려할 때 최적의 정답지일 가능성 역시 상존한다. 이는 어떤 제도가 시행되기에 앞서 불완전한 정보만을 바탕으로 정책적 의사결정을 내리는 과정에서 불가피한 현상이다. 따라서 본고가 제시한 결론은 어디까지나 가능한 한 가지 잠정적 대안일 뿐임을 유념할 필요가 있고, 후속 논의를 통해 보다 진일보한 정책적 대안이 마련될 필요가 있다.

참고문헌

1. 단행본

- 박영도, 기본법의 입법모델연구, 한국법제연구원, 2006.
- 왕승혜, 법령상의 영향평가제도 정비방안 연구, 한국법제연구원, 2024.
- 이유봉, AI 영향평가를 위한 입법연구, 한국법제연구원, 2024.
- 최경진/강지원/김법연/김형준/박신욱/오정익/채기현/채은선, EU 인공지능법, 박영사, 2024.

2. 학술지

- 권은정, “인공지능 서비스 영향평가의 체계와 방법론 - 「지능정보화 기본법」의 사회적 영향평가 제도를 중심으로-”, 경제규제와 법 제16권 제1호, 서울대학교 법학연구소, 2023, 29-50쪽.
- 김근혜/박규동, “AI 시스템의 위험 완화를 위한 정책적 접근방안 연구: AI 영향평가를 중심으로”, 규제연구 제32권 제2호, 한국규제학회, 2023, 105-134쪽.
- 김법연, “공공분야 인공지능서비스의 영향평가제도 도입에 관한 연구”, 정보법학 제27권 제2호, 한국정보법학회, 2023, 171-219쪽.
- 송다솜/김다영/노재경/박도현, “인공지능 신뢰성과 안전성의 경계 획정”, 저스티스 제206호, 한국법학원, 2025, 221-263쪽.
- 윤혜선, “AI 거버넌스의 새로운 축(軸) -영향평가제도에 관한 연구-”, 공법연구 제53집 제2호, 한국공법학회, 2024, 377-424쪽.
- 채은선, “우리나라 인공지능기본법의 주요 내용 및 관련 고려사항”, 경제규제와 법 제18권 제1호, 서울대학교 법학연구소, 2025, 93-113쪽.
- Claudio Novelli/Federico Casolari/Antonino Rotolo/Mariarosaria Taddeo/Luciano Floridi, “AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act”, *Digital Society Vol. 3 No. 1*, Springer, 2024, pp. 1-29.

3. 신문기사

- 조선비즈, 심민관, “과기정통부 “AI 기본법 규제 받는 ‘고영향 AI’ 대상 엄격히 해석·최소한 규제””, 2025. 12. 24, <<https://biz.chosun.com/it-science/ict/2025/12/24/>>

TKMPT6PCMRFGJMZSKAMDM5CSUE/>, 검색일: 2026. 4. 10.

4. 웹자료

세계법제정보센터, “유럽연합 인공지능법 번역본”, <https://world.moleg.go.kr/web/wli/lgsInfoReadPage.do?CTS_SEQ=51409&AST_SEQ=93>, 검색일: 2026. 4. 10.

5. 기타자료

과학기술정보통신부/정보통신정책연구원, 인공지능 영향평가 가이드라인, 2026.

과학기술정보통신부/한국지능정보사회진흥원, 고영향 인공지능 판단 가이드라인, 2026.

관계부처 합동, AI 분야 규제합리화 로드맵, 2025.

국가인공지능전략위원회, 대한민국 인공지능 행동계획, 2026.

[Abstract]

A Study on the Issues and Improvements of High-Impact AI Impact Assessments under the Framework Act on AI

Jung, SeongWoong^{*}

Lee, Haesung^{**}

Park, Do Hyun^{***}

The Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation for Trust (Framework Act on AI), which was enacted to establish a sound development and a foundation for trust in artificial intelligence (AI), came into force on January 22, 2026. For the purposes of this law, it is necessary to classify AI according to risk levels and implement selective regulation to ensure both the development of the AI industry and the protection of fundamental rights. To this end, the European Union's AI Act classifies the risk levels of AI and imposes substantive regulatory measures specifically on high-risk AI. Similarly, South Korea's Framework Act on AI imposes substantive regulatory measures after determining whether an AI system qualifies as high-impact AI. In addition, it introduces a high-impact AI impact assessment to mitigate concerns about potential infringements on fundamental rights by such systems. However, the current high-impact AI impact assessment does not guarantee the balanced legislative intent of the Framework Act on AI. This is because it is enacted as a form of self-regulation, which makes it difficult to ensure the effectiveness of the system, and the assessment does not take into account the benefits that AI brings. This paper aims to identify the

^{*} First author: B.S. Student in the Department of Electrical Engineering and Computer Science at Gwangju Institute of Science and technology

^{**} Participating author: Post-doctoral Researcher in the Graduate School of AI Policy and Strategy at Gwangju Institute of Science and technology

^{***} Corresponding author: Assistant Professor in the Graduate School of AI Policy and Strategy at Gwangju Institute of Science and technology

limitations of the current high-impact AI impact assessment by examining the domestic impact assessment framework and relevant cases from other jurisdictions, and to propose complementary measures.

[Key Words] Artificial Intelligence, Framework Act on AI, High-impact AI, Impact Assessment, EU AI Act