국제사회보장리뷰 2025 봄호 Vol. 32 pp.75~89

DOI: 10.23063/2025.03.5

유럽연합 인공지능법이 한국 사회보장에 주는 함의[1)]

Implications of the EU AI Act for South Korea's Social Security System

김기태 (한국보건사회연구원 사회보장정책연구실 연구위원)
신영규 (한국보건사회연구원 사회보장정책연구실 부연구위원)
Kim, Ki-tae (Korea Institute for Health and Social Affairs)
Shin, Young-Kyu (Korea Institute for Health and Social Affairs)

인공지능의 급격한 발전에 따른 충격이 크다. 인공지능의 활용도를 높이면서, 파괴력을 제어하기 위한 규제도 전 세계적으로 모색되고 있다. 유럽연합이 2024년에 제정한 인공지능법은 인공지능에 관해 국제적으로 가장 포괄적인 법적 프레임워크다. 인공지능법은 위험성에 따라 인공지능을 네 가지 유형으로 나눠 규제하고 있다. 사회보장 영역에 한정해서 보면, 사회적 평점(social scoring)은 가장 위험한 '수용할 수 없는' 위험으로 원칙적으로 금지된다. 개인 정보에 근거해서 급여 자격을 판정하는 사회보장제도는 일종의 사회적 평점 부여 과정을 거칠 수밖에 없다. 법 집행 과정에서 논란이 예상된다. 사회보장 영역을 포괄하는 '필수 민간 및 공공 서비스 분야'의 인공지능 사용도 두 번째로 위험한 '고위험' 영역으로 분류된다. 엄격한 규제의 대상이 된다. 한국은 인공지능 기본법이 2024년 12월에 제정됐으나, 사회보장 영역에 대한 구체적인 언급이 없다. 한국의 사회보장 영역에서 인공지능 기술이 이미 적용되는 점을 고려하면, 정책적 대응이 시급히 필요하다.

1. 들어가며

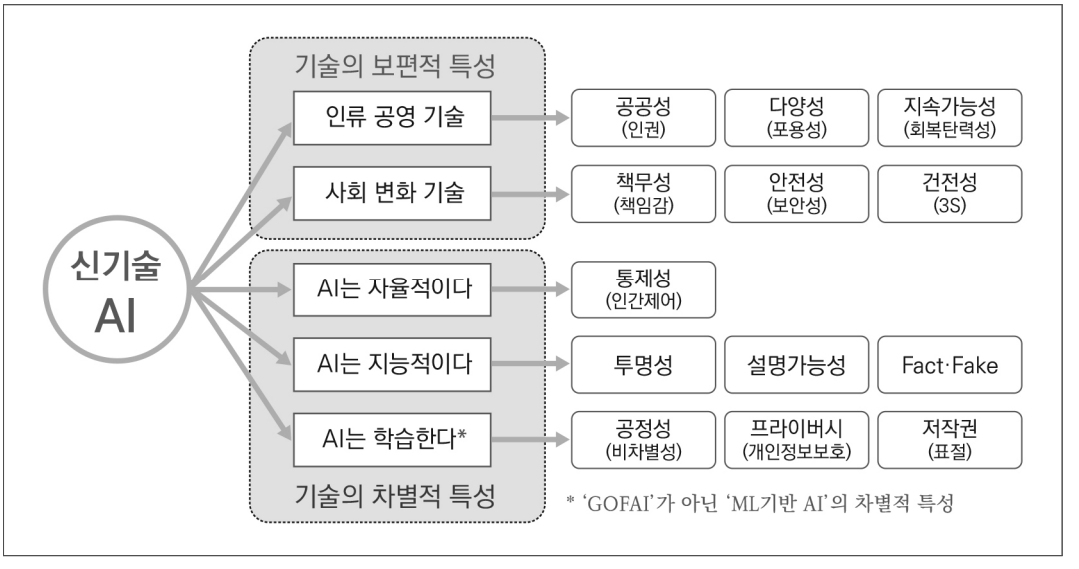
인공지능(Artificial Intelligence, 이하 AI)의 급격한 발전으로 인한 충격이 크다. 인공지능은 규칙이나 기준을 명시적으로 따르지 않고, 빅데이터의 경향과 관계를 학습한다. 그 결과 생성

1) 이 글은 김기태, 신영규, 김은하, 김명주, 변소연. (2024). 사회보장행정에서 인공지능 적용 동향과 함의(한국보건사회연구원)의 4장 2절 내용을 일부 수정·보완해서 작성하였다.

된 결정, 예측, 추천, 콘텐츠의 생성 과정을 인간이 이해하기는 어렵다. 여기서 인공지능의 '블랙박스'(Zaber et al., 2024, p.19) 문제가 대두된다. 인간이 인공지능의 입력값과 출력값은 알지만, 과정은 알 수 없다. 인공지능이 산출한 출력값을 신뢰할 근거도 제시하기 어렵다. 인공지능의 압도적인 능력을 염두에 둔 윤리 문제가 대두될 수밖에 없다. 여기서 윤리는 공공성, 다양성, 지속가능성, 책무성, 안정성, 건전성, 통제성, 투명성, 설명 가능성, 공정성 등을 아우른다(김명주, 2024, [그림 1] 참고). 인공지능이 가지는 특성에 따른 필연적인 결과다. 인공지능은 인류 공영을 위한 기술이면서 동시에 자율성, 지능성 및 학습능력을 갖기 때문이다.

이러한 배경에서 전 세계 국가 및 지역 정부 및 국제기구에서 인공지능에 대한 규제 혹은 가이드라인을 앞다투어 내놓고 있다. 인공지능이 초래할 수 있는 사회적 위험이 거대할 수 있고, 동시에 비가시적, 불확정적이기 때문이다. 대표적인 예가 World Economic Forum(2023)의 '책임 있는 생성형 AI에 대한 Presidio 권고사항', European Parliament(2024)의 인공지능법, OECD(2023)의 인공지능 원칙, UNESCO(2021)의 인공지능 윤리에 관한 권고, 미국 백악관의 AI 권리 장전 설계(White House, 2022), 미국 캘리포니아주에서 시도했던 AI 규제법안(SB 1047) 등이다. 이 글에서는 지난 2024년에 발효된 유럽연합 인공지능법의 내용을 간단히 살펴보고, 인공지능법이 사회보장 영역에 미칠 영향을 살펴보고자 한다. 무수한 규제 가운데 유럽

| 그림 1. 인공지능의 특징과 관련된 원칙 |



출처: "AI 윤리와 규제"[발표 자료], 김명주, 2024. 8. 7. p.39. 사회보장행정에서의 인공지능 적용 동향과 함의 세미나.

연합의 인공지능법에 주목하는 이유는, 앞서 제시한 규제 가운데 거의 유일하게 법적인 강제력을 가지고 초국가적으로 집행되는 법이기 때문이다.

2. 유럽연합 인공지능법 제정 의미 및 절차, 구성

유럽연합의 인공지능법(Artificial Intelligence Act)은 세계 최초로 제정된 AI에 관한 포괄적인 법적 프레임워크다(European Artificial Intelligence Act, 2024). 인공지능법은 유럽연합에서 자주 사용되는 간접적인 방식인 지침(directive)이 아닌 법(act)으로 제정됐다. 따라서 모든 회원국에 직접적이고 즉각적인 규제(regulations)로 적용되고, 국가별로 별도의 법률 제정 없이 효력을 발휘한다.

인공지능법은 2021년 4월 유럽연합 집행위원회(European Commission)가 유럽연합 지역 내에서 AI 규제를 위한 제안서를 발표하면서 시작됐다(라기원, 2024). 이 제안서를 바탕으로 유럽연합 이사회(European Union Council)는 다양한 논의와 협의를 진행하였고, 유럽연합 의회(European Union Parliament)는 윤리적 쟁점, 생체 인식 기술, 고위험 애플리케이션에 중점을 두고 인공지능법안 작업을 시작했다. 2022년에는 유럽연합 이사회 의장단이 회원국들 사이에 쟁점이 되는 여러 조항에 대한 타협안을 도출했다. 이후에도 인공지능법안에 대한 보완 조치는 지속적으로 이루어졌고, 2023년 6월에 유럽연합 의회는 법안을 최종적으로 확정했다. 2024년 2월에는 법률의 이행을 감독하기 위한 유럽 인공지능사무국(European Artificial Intelligence Office)도 출범했다. 5월 들어 유럽연합 이사회는 이 법률의 채택을 공식화했고, 7월에 유럽연합 관보에 공식적으로 인공지능법의 채택 선언이 게재되었다(European Artificial Intelligence Act, 2024).

새롭게 제정된 유럽연합 인공지능법의 목표는 유럽을 비롯한 세계 어느 지역에서도 신뢰할 수 있는 AI가 개발될 수 있도록 모든 AI 시스템이 인간의 기본권, 안전, 윤리 원칙 등을 존중하게 하고, AI 모델의 위험성을 관리하는 데 있다(European Artificial Intelligence Act, 2024). 따라서 이 법은 구체적인 상황별로 AI 개발자와 배포자가 지켜야 할 명확한 요건과 의무를 명시하고 있다. 물론 법이 AI 개발을 지원하려는 의도도 담고 있다. AI 관련 혁신 정책 패키지 및 AI 관련 협력 계획을 명시하고 있고, AI 개발 및 활용과 관련하여 기업, 특히 중소기업의 행정적 부담과 비용을 줄이는 방안도 포함한다(European Artificial Intelligence Act, 2024).

인공지능법은 모두 13개 장의 113개 규정과 13개의 부속서로 구성되어 있다. 각 장의 제목과 내용은 〈표 1〉과 같다.

| 표 1. 유럽연합 인공지능법의 구성 |

장	내용
제1장	**총칙:** 법 적용 대상 및 범위, 용어 정의, AI 리터러시
제2장	**AI 활용 관련 금지행위:** AI 시스템의 활용이 금지되는 경우와 예외
제3장	**고위험 AI 시스템:** 제1절: AI 시스템의 고위험 분류 기준 제2절: 고위험 AI 시스템의 준수사항 제3절: 고위험 AI 시스템 제공자·배포자·기타 이해관계자의 준수사항 제4절: 승인기관(통보기관), 인증기관 제5절: 표준·적합성 평가·인증서·등록 기준
제4장	**특정 AI 시스템의 제공자 및 배포자에 대한 투명성 의무:** 사람과 직접 상호작용할 목적으로 사용되는 AI 시스템의 제공자 및 배포자의 투명성 의무
제5장	**범용 AI 모델:** 제1절: 범용 AI 모델의 분류 규칙 제2절: 범용 AI 모델 제공자의 준수사항 제3절: 시스템적 위험이 있는 범용 AI 모델 제공자의 준수사항
제6장	**혁신 지원 방안:** 규제 샌드박스 및 실제 조건에서의 고위험 AI 시스템 시험이 가능한 예외와 스타트업 등 중소기업에 대한 산업 진흥 제도
제7장	**거버넌스:** 제1절: 유럽연합 수준의 거버넌스 – AI사무국, 유럽AI위원회, 자문포럼, 독립전문가 과학패널 제2절: 국가 관할기관 – 각 회원국의 관할 기관 지정
제8장	**고위험 AI 시스템을 위한 EU 데이터베이스:** 유럽연합 집행위원회가 관리하는 고위험 AI 시스템에 대한 EU 데이터베이스
제9장	**사후 모니터링, 정보공유, 시장감독:** 제1절: 사후 모니터링 제2절: 중대한 사고에 대한 정보 공유 제3절: 시장감독기관 및 집행위원회의 규범 집행을 위한 수단 제4절: 구제수단 제5절: 범용 AI 모델 제공자에 대한 감독, 조사, 집행 및 모니터링
제10장	**행동규범 및 지침:** 고위험 AI 시스템을 제외한 시스템에 적용할 수 있는 행동규범 및 지침 마련
제11장	**권한의 위임과 위원회 절차:** 집행위원회 권한 위임의 근거, 집행위원회 보조위원회 근거 규정
제12장	**처벌:** AI 시스템에 대한 규범 위반, 범용 AI 모델에 대한 규범 위반에 관한 처벌 규정
제13장	**종칙**

주: EU Artificial Intelligence Act 및 "유럽연합 인공지능법(EU AI ACT)의 특성과 쟁점: 우리나라 인공지능 입법에 대한 시사점", 라기원, 2024, AI Issue Paper를 바탕으로 표를 재구성함.

3. 유럽연합 인공지능법의 특징 및 적용 범위

유럽연합의 인공지능법 제3조에서는 인공지능을 다음과 같이 정의하고 있다. "다양한 수준의 자율성을 가지고 작동하도록 설계되었으며, 배포 후 적응성(adaptiveness)을 보일 수 있는 기계

기반 시스템을 의미한다. 이 시스템은 명시적 또는 암묵적 목적을 위해 입력받은 데이터를 바탕으로 예측, 콘텐츠, 추천 또는 물리적 혹은 가상 환경에 영향을 미칠 수 있는 결정을 포함한 출력을 생성하는 방법을 추론한다"(European Artificial Intelligence Act, Article 3, (1)).

인공지능법은 유럽연합 역내 시장에서 AI를 활용한 상품과 서비스의 기준을 통일함으로써 상품과 서비스의 국경 간 자유로운 이동을 보장하고자 한다(라기원, 2024). 즉, 회원국들이 개별적으로 AI에 대한 규제를 강화하는 것을 막아 AI 시스템의 개발, 수입 또는 사용에 있어 법적 확실성을 보장하는 것을 목적으로 한다.

가. 인공지능법의 특징

인공지능법의 특징은 유럽평의회 인공지능 고위급 전문가그룹(European Commission AI HLEG, 2019)이 제시한 "신뢰할 수 있는 AI를 위한 윤리 지침(Ethics guidelines for trustworthy AI)"을 바탕으로 규범이 형성됐다는 점이다. 인공지능법 전문은 윤리 지침이 제시하는 다음의 일곱 가지 원칙에 근거한다. 인간 주도·감독(Human agency and oversight), 기술적 견고함과 안전성(Technical Robustness and safety), 프라이버시와 데이터 거버넌스(Privacy and data governance), 투명성(Transparency), 다양성과 차별 금지 및 공정성(Diversity, non-discrimination and fairness), 사회·환경적 복지(Societal and environmental well-being), 책임성(Accountability) 등이다(EU AI Act, (27)). 위와 같은 원칙들은 모두 직간접적으로 사회보장 영역에도 연결됨을 추정할 수 있다.

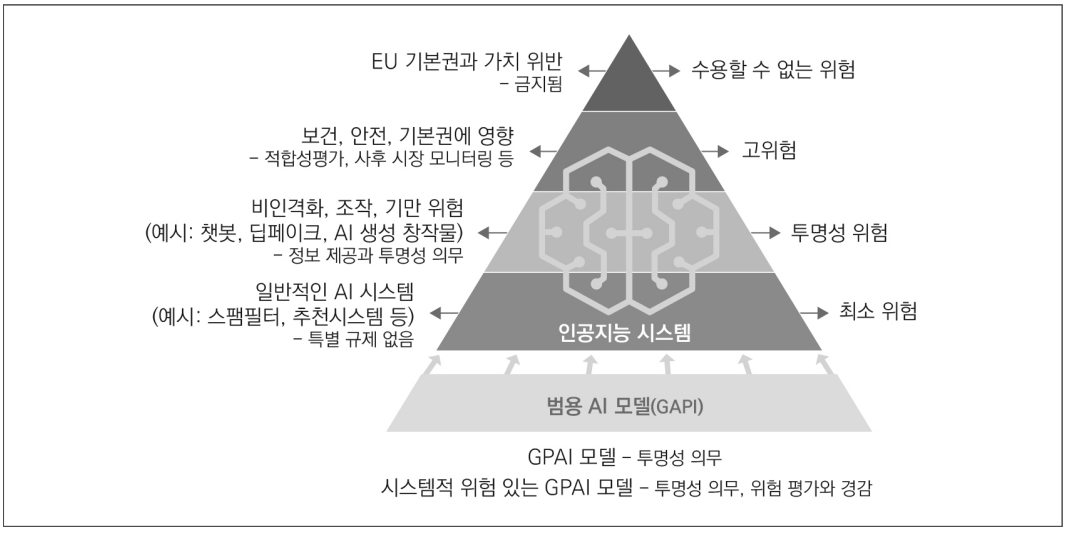
인공지능법의 가장 큰 특징은 위험성 차원에서 AI 시스템을 분류한 다음, 그에 따라 규제의 내용을 차등화한다는 점이다. 인공지능 시스템이 내포하는 위험성의 강도와 범위에 비례하여 규제 유형과 정도가 규정되는 것이다.

인공지능법에서 핵심 개념인 위험성은, 인공지능의 개념 정의에 바로 뒤이어 다음과 같이 제시된다. "위험의 발생 가능성과 그로 인한 피해의 심각성의 결합"(European Artificial Intelligence Act, Article 3, (2)). 유럽연합은 인공지능 모델의 성격과 용도 등을 고려해서 위험성의 경중을 진단하고, 그에 따른 규제의 수준도 연동한다. 이 법에서는 AI 시스템의 위험성 수준을 네 가지로 구분한다([그림 2] 참고). 네 가지 위험성 수준은 '수용할 수 없는 위험성(unacceptable risk)[2]', '고위험성(high risk)', '제한된 위험성(limited risk)', '최소한의 위험성(minimal risk)' 등이

2) 유럽연합의 인공지능법 본문에서는 '수용할 수 없는(unacceptable)'에 더해 '금지된(prohibited)'이라는 표현도 자주 등장한다. 또한 '제한된 위험성'은 투명성(transparency) 위험으로도 사용된다.

다. 가장 높은 수준의 '수용할 수 없는 위험성'을 가진 인공지능은 활용 자체가 금지된다. 다음으로 '고위험성'을 가진 인공지능 활용에는 엄격한 준수사항이 요구된다. 세 번째인 '제한된 위험성'을 내포하고 있는 AI의 활용은 투명성 의무가 부과되고, 최소한의 위험성이 있는 경우에는 자율적 규제가 따라붙는다(윤혜선, 2024.9.23.). 위의 네 가지 위험성을 순서대로 하나씩 살펴보겠다.

| 그림 2. 유럽연합 인공지능법에서 규정하는 위험의 위계 |



출처: "Artificial intelligence act", Madiega, 2024, European Parliament, p.1의 그림을 라기원(2024)이 번역해서 제시함. 저작권 2024, European Parliament, 한국법제연구원.

첫째, 가장 높은 수준의 '수용할 수 없는 위험성(unacceptable risk)'의 경우를 살펴보자. 수용할 수 없는 위험성은 AI 시스템이 사람들의 안전, 건강, 기본권에 명백한 위협이 되는 경우다. 인공지능법 제5조 1항에서는 수용할 수 없는 위험성이 있다고 간주하는 사례를 (a)~(h)에 걸쳐 여덟 가지를 제시했다. 짧게 요약하면 다음과 같다(Artificial Intelligence Act Article 5 (1)). 첫째, 사람의 의식을 넘어서는 잠재적인 기술을 활용하거나 의도적으로 조작적이거나 기만적인 기술을 사용해서, 개인 또는 집단의 행동을 실질적으로 왜곡해서 중대한 피해를 초래하거나 그럴 가능성이 있는 경우, 둘째, 연령, 장애 또는 사회경제적 환경에 기인한 취약점을 악용하여 행동을 왜곡함으로써 심각한 피해를 유발하거나 그럴 가능성이 있는 경우, 셋째, 자연인 또는 집단의 사회적 행동이나 알려진, 추론된 또는 예측된 개인적 또는 성격적 특성을 기반으로 일정 기

간 평가하거나 분류하기 위해 사회적 평점(social scoring)을 사용한 경우다. 넷째, 프로파일링 또는 성격 특성만을 기반으로 하여 개인의 범죄 행위를 예측하는 경우, 다섯째, 인터넷이나 폐쇄회로(CCTV) 영상에서 무차별적으로 얼굴 이미지를 스크랩하여 얼굴 인식 데이터베이스를 구축하는 경우, 여섯째, 의료 또는 안전을 제외한 목적으로 직장이나 교육 기관에서 개인의 감정을 추론한 경우, 일곱째, 인종, 정치적 견해, 노동조합 가입 여부, 종교적 혹은 철학적 신념, 성생활, 성적 지향 등과 같이 민감한 개인적 특성을 유추하기 위한 생체 인식 분류 시스템(biometric categorisation systems)인 경우, 마지막으로 법 집행 목적으로 공공장소에서 실시간 원격 생체 인식 정보('real-time' remote biometric identification)를 수집한 경우다. 여덟 가지 가운데 사회보장 영역과 가장 밀접한 내용은 세 번째인 사회적 평점 부분이다. 관련 내용은 아래 4절에서 살펴보겠다.

수용할 수 없는 위험성 관련 금지 행위를 위반한 경우에는 최대 3500만 유로(약 500억 원)와 전년 회계연도 기준 전 세계 연매출액의 최대 7% 가운데 더 높은 금액이 과징금으로 부과된다(윤혜선, 2024 .9. 23.). 단, 중소기업이 위반했을 때는 위 두 기준 가운데 더 낮은 금액이 과징금으로 부과된다. 유럽연합 소속 기관, 대리인, 조직의 경우에는 최대 150만 유로(약 20억 원)가 부과된다.

두 번째로 고위험성(high risk)을 살펴보겠다. 인공지능법 제6조 제3항에 따르면 고위험성을 가진 인공지능 시스템은 자연인의 건강, 안전 또는 기본권에 위해를 가할 중대한 위험을 내포하는 시스템이다. 인공지능법 부속서 Ⅲ에서 제시된 여덟 개 영역의 고위험성을 요약하면 다음과 같다.[3] ① 생체인식 및 분류, 감정인식 시스템[4], ② 주요 사회기반시설, ③ 교육, 직업훈련 영역, ④ 고용, 근로자 관리 및 자영업자에 대한 접근, ⑤ 필수 민간 및 공공 서비스, ⑥ 인간의 기본권과 관련된 법 집행, ⑦ 이주, 망명 및 국경 통제 관리, ⑧ 사업절차 및 민주적 절차 등이다. 여덟 가지 가운데 ⑤ 필수 민간 및 공공 서비스 분야(essential private and public services)는 사회보장 영역과 직접적으로 연관된다. 해당 내용도 아래 4절에서 살펴보겠다.

고위험 인공지능 시스템은 시장에 출시되기 전 준수사항으로 다음의 일곱 가지가 제시된다

3) 유럽연합의 인공지능법에서는 고위험성 인공지능을 부록 I과 부록 III에서 각각 다르게 제시하고 있다. 부록 I에서 제시하는 내용은 사회보장 영역과 무관해서 별도로 설명하지는 않는다.

4) 생체인식 및 감정인식 시스템은 앞서 '수용할 수 없는 위험'으로 원칙적으로 금지되지만, 유럽연합 혹은 개별 회원국의 법에 따라 예외적으로 허용되는 경우는 '고위험군'으로 분류된다(유럽연합 인공지능법 부록 III 1호). 공공장소에서 실시간 생체인식이 예외적으로 인정되는 경우로는 실종된 아동을 수색해야 하는 경우, 구체적이고 임박한 테러 위협을 방지해야 하는 경우, 중대한 범죄 행위의 가해자 또는 용의자를 탐지, 위치 파악, 식별 또는 기소해야 하는 경우 등이다. 유럽 일부 국가에서는 인공지능 사용에 따른 프라이버시 침해 등의 논란을 우회하는 하나의 방식으로 이와 같은 인공지능 사용 목적에 보안(security)이나 안전(safety)을 중심에 두고 접근하는 경향이 일고 있다(van Bekkum & Borgesius, 2021).

(European Artificial Intelligence Act, 2024; 윤혜선, 2024. 9. 23.). ① 적절한 위험 평가 및 완화 시스템 구축, ② 리스크 및 차별적 결과를 최소화하기 위한 고품질 데이터 구축 및 관리, ③ 결과의 추적 가능성을 보장하기 위한 활동 기록(logging), ④ 시스템과 그 목적에 대한 모든 정보를 제공하는 상세한 문서화(규제 준수 여부 확인 목적), ⑤ 사용자에게 명확하고 적절한 정보 제공, ⑥ 위험을 최소화하기 위한 적절한 인간의 감독 조치, ⑦ 높은 수준의 견고성, 보안 및 정확성이다. 유럽연합에서는 인공지능 시스템을 둘러싼 이해관계자를 제공자, 배포자, 공인 대리인, 수입업자 등으로 분류하고, 각자에 대한 의무를 구체적으로 명기하고 있다. 이를테면 제공자(provider)에게는 유럽연합 적합성 선언 작성, CE 마크 작성 등의 17개 의무가 부과된다. 사회보장 영역에서 작동하는 다수의 인공지능 시스템은 이러한 규제에 놓이게 될 가능성이 높다.

나. 인공지능법의 적용 범위

유럽연합의 인공지능법에서 규정한 이해관계자들은 법의 적용 범위와 관련해서 중요한 의미가 있다. 인공지능법 2조는 법의 적용 범위를 규정하면서, 일곱 가지 유형의 이해당사자 가운데 하나로 "유럽연합에서 AI 시스템 출시·서비스화하거나 범용 AI 모델을 출시하는 제공자"라고 규정했다. 그러면서 동시에 '장소불문("irrespective of whether those providers are established or located within the Union or in a third country)"(Artificial Intelligence Act Article 2 (1))이라는 조건을 달았다. 즉, 유럽에서 서비스를 제공하는 경우라면, 인공지능 시스템을 출시하거나 적용하는 업체는 국적을 불문하고 규제의 대상이 된다.[5] 국제적으로 인공지능 산업을 대부분 선도하는 미국의 업체들도 유럽연합에 와서는 규제를 따라야 한다. 이는 한국에도 간접적으로 영향을 미칠 수밖에 없다.

이러한 고위험성 인공지능 시스템 준수사항을 위반하면 최대 1500만 유로(약 218억 원) 혹은 전년 회계연도 기준 전 세계 매출액의 최대 3% 중에 더 높은 금액이 부과된다. 중소기업이 고위험 관련 조항을 위반하면 두 기준 가운데 더 낮은 금액이 부과된다. 유럽연합의 기관, 에이전시, 기구의 경우, 과징금은 최대 75만 유로(약 10억 원)로 한정된다.

5) 물론 예외도 다음과 같이 있다. ① 군사, 국방, 국가 안보 목적의 AI 시스템, ② 유럽연합 또는 그 회원국과 사법 공조 협약을 체결한 제3국의 공공기관이 그 협약의 체계 안에서 사용하는 AI 시스템, ③ 과학적 연구 및 개발 목적으로만 특별히 개발되고, 서비스가 제공되는 AI 시스템과 AI 모델, ④ 출시 및 서비스 개시 전 AI 시스템을 연구, 테스트, 개발하는 활동, ⑤ 고위험성 AI 시스템이 아닌 무료 및 오픈소스 라이선스로 출시된 AI 시스템, ⑥ 사적 및 비전문적인 활동, ⑦ 일부 목적을 위한 법집행기관(경찰, 이민 당국 등)의 AI 시스템 사용 등에는 인공지능법 적용이 제외되거나 특례가 인정된다(윤혜선, 2024.9.23.).

유럽연합이 적시한 네 가지 유형의 위험성 가운데 세 번째는 제한적인 위험성이다. 제한적인 위험성이란 AI 사용에 있어 투명성이 부족한 것에 기인하는 위험성을 뜻한다. 인공지능법은 인간에게 정보를 제공하거나 신뢰를 높이기 위한 목적으로 투명성에 관한 구체적인 의무를 규정한다. 예를 들어 챗봇과 같은 AI 시스템을 사용할 때 인간은 기계와 상호작용하고 있다는 사실을 인지할 수 있어야 하고, 그래야 계속 챗봇을 사용할지를 결정할 수 있다는 것이다(European Artificial Intelligence Act, 2024). 전 세계적으로 사회보장 분야에서 가장 많이 사용되는 인공지능기술이 챗봇인 점을 참고할 필요가 있다(Zaber et al., 2024).

마지막으로 최소한의 위험성이 있다. 최소한의 위험성을 내포하는 AI 시스템에는 스팸 필터, 알고리즘을 통한 콘텐츠 추천시스템 등이 해당된다. 인공지능법은 이와 같이 최소한의 위험성에 해당하는 AI 시스템에 별도의 규제를 규정하지 않고, 자율적 규제를 허용하고 있다.

4. 유럽연합 인공지능법이 사회보장에 미칠 영향

가. 핵심 의제, 사회적 평점(social scoring)

유럽연합의 인공지능법에서 사회보장과 연관된 가장 민감한 대목[6]은 '수용할 수 없는 위험성(unacceptable risk)'에서 세 번째로 제시된 사회적 평점(social scoring)이다. 사회적 평점은 "개인의 데이터를 기반으로 개인을 평가하는 것을 의미하며, 여기에는 신용 행태, 교통 위반, 사회적 참여 등이 포함된다. 이러한 평가는 특정 서비스나 특권에 대한 접근을 규제하기 위해 사용된다."(Mosene, 2024). 유럽연합의 인공지능법을 본문 그대로 번역하면 다음과 같다(European Parliament, 2024, Chapter II, Article 5, 1. (c)).

> "자연인 또는 집단의 사회적 행동이나 알려진, 추론된 또는 예측된 개인적 또는 성격적 특성을 기반으로 일정 기간 대상을 평가하거나 분류하기 위한 목적으로 인공지능(AI) 시스템을 시장에 출시하거나, 서비스에 투입하거나 사용하는 행위로서, 이러한 사회적 평점이 다음 중 하나 이상의 결과로 이어지는 경우:
> (i) 데이터가 원래 생성되거나 수집된 맥락과 무관한 사회적 맥락에서 특정 자연인 또는 집단에 대해 불리하거나 부정적인 대우를 초래하는 경우.
> (ii) 특정 자연인 또는 집단의 사회적 행동 또는 그 행동의 심각성과 비교하여 부당하거나 과도하게 불리하거나 부정적인 대우를 초래하는 경우"

6) 한 가지 확인할 점은 있다. 해당 주제에 대한 분석을 담은 학술논문이나 보고서는 아직 찾기 어렵다. 현재로서는 법률 내용, 관련 시민단체 성명, 국내 법률 전문가 자문 등을 중심으로 관련된 영향을 추정하는 수밖에 없었다.

유럽연합의 인공지능법은 사회적 평점을 논의하면서 사회보장제도와 관련한 언급을 하지는 않았다. 유럽의회 산하의 연구기관인 유럽의회연구소(European Parliamentary Research Service)가 인공지능법을 설명하는 짤막한 해설서(Madiega, 2024)에도 사회보장제도에 대한 언급은 없다. 그럼에도 사회적 평점 부여는 사회보장과 일정한 연관을 가질 수밖에 없다. 개인 혹은 가구 단위의 소득, 재산, 가구원 등의 정보에 근거해서 빈곤, 실업, 은퇴, 상병 여부를 판단하고, 그에 근거해서 급여를 제공하는 사회보장제도는 급여 자격을 판정하는 과정에서 일종의 사회적 평점(social scoring)을 개인 혹은 가구에 부여할 수밖에 없다. 이를테면 국내의 국민기초생활보장제도에서도 가구의 소득, 재산, 부양의무자, 가구원 정보 등에 기반해서 수급 자격 및 급여액을 결정한다. 그러한 자료의 내용이 개인의 소득, 연령, 가구, 건강 등 개인적 정보를 담고 있다는 점에서 사회적 평점은 민감할 수밖에 없는 의제다.

유럽연합의 인공지능법 제정 과정에서도 사회적 평점과 관련한 우려가 제기됐다. 유럽의 인권 단체인 Human Rights Watch(2023. 10. 9.)는 다른 시민단체들과 함께 유럽이사회와 유럽의회에 사회적 평점 관련 규정 강화를 제안하면서 사회보장제도에서 나타날 문제를 다음과 같이 언급했다. “프랑스, 네덜란드, 오스트리아, 폴란드, 아일랜드에서의 조사 결과, AI 기반의 사회적 평점 부여 시스템이 사람들의 사회보장 지원 접근을 방해하고, 프라이버시를 침해하며, 빈곤에 대한 고정관념을 가지고 차별적인 방식으로 데이터를 프로파일링하고 있음이 드러났다.”(Human Rights Watch, 2023. 10. 9.) 이러한 문제들은 시민단체나 학계에서도 꾸준히 제기됐다. 프랑스(La Quadrature Du Net, 2023), 네덜란드(van Bekkum & Borgesius, 2021)와 덴마크(Jørgensen, 2021)에서도 논란이 일었다. 이를테면 정부에서 복지급여의 부정수급을 탐지하거나 위험가구를 발굴하는 과정에서 개인 정보를 무리하게 활용하거나(Appelman et al., 2021), 알고리즘이 취약계층에게 편향적으로 작동했다(van Bekkum & Borgesius, 2021)는 지적이다. 네덜란드에서는 이와 같은 문제들 때문에 정부에서 추진하던 사회보장 부정수급 추적 알고리즘인 시스템 위험도 표시(Systeem Risico Indicatie) 시스템이 법원에 의해 제동이 걸리면서 작동이 중단된 바 있다(Bekker, 2021).

나. 사회적 평점의 기준

유럽연합도 사회적 평점과 관련한 비판 및 부작용을 염두에 둔 것으로 보인다. 법률에 붙은 (i)과 (ii)의 내용은 인공지능의 사회적 평점이 ‘수용할 수 없는 위험’으로 간주되는 경우를 한정

했다. 즉, 특정 집단이나 개인에게 불리하거나 부정적이거나 부당한 대우를 초래할 때만 사회적 평점이 금지된다. 바꾸어 말하면, 사회보장제도에서 인공지능의 작동이 '순기능'을 하는 경우에는 규제의 대상으로 삼지 않는다는 뜻이다. 그렇지만 논란이 종료되기는 어렵다. 사회보장 영역에서 인공지능이 급여 대상자를 판단 혹은 추론하는 과정에서 데이터 편향성 등의 문제로 특정 집단에 대한 급여를 변경/중지/중단한다면, 이는 불리/부정/부당할 수 있다(강지원, 2024. 11. 7.). 실제로 덴마크나 네덜란드 등 다수의 국가에서 사회보장제도에 순기능할 것으로 기획된 인공지능 알고리즘들이 결과적으로 차별과 편향을 낳았다는 비판에 직면했다(Jørgensen, 2021; Bekker, 2021). 이 대목은 앞으로 인공지능 기술이 사회보장 영역에서 적용되는 과정에서 끊임없이 논점으로 부각될 것으로 예상된다.

다음으로 인공지능법에서 두 번째로 수위가 높은 '고위험(high risk)' 인공지능 영역을 보겠다. 여기에서는 다섯 번째 고위험 인공지능 영역으로 '필수 민간 및 공공 서비스 분야'가 제시됐다. 이 영역은 사회보장 영역을 직접적으로 언급하고 있다. 고위험 인공지능을 영역별로 상술한 유럽연합 인공지능법 부록(Annex) Ⅲ에서 사회보장을 다음과 같이 언급했다(Artificial Intelligence Act Annex Ⅲ, 5).

> "필수 민간 서비스 및 필수 공공 서비스와 혜택에 대한 접근 및 이용[7]:
> (a) 공공기관 또는 공공기관을 대신하여 사용될 목적으로, 자연인의 필수적인 공공 복지 급여[8] 및 서비스(의료 서비스 포함)에 대한 자격을 평가하거나, 해당 혜택 및 서비스를 부여, 축소, 취소 또는 회수하기 위해 사용되는 AI 시스템"

'고위험'으로 분류한 '필수 민간 및 공공서비스'는 앞서 살펴본 사회적 평점(social scoring)과 밀접하게 연관됨을 알 수 있다. 급여를 제공하는 사회보장 영역에서 자격 요건을 판정하기 위해서는 사회적 평점 산정이 대부분 필요하기 때문이다. 유럽연합법은 관련 고위험에 대해서 다음과 같이 상세하게 설명했다(Artificial Intelligence Act, (58)).

7) 급여 수급에 관한 내용을 담은 (a) 외에도 보건의료 영역에서 다음과 같은 언급도 있다. 사회보장 영역과 일정한 연관성이 있지만, 이번 글에서는 해당 내용까지 다루지는 않는다. "(c) 생명 및 건강 보험의 경우, 자연인에 대한 위험 평가 및 가격 책정을 위해 사용되는 AI 시스템; (d) 자연인의 긴급 호출을 평가 및 분류하거나, 경찰, 소방관, 의료 지원을 포함한 긴급 대응 서비스의 출동 우선순위를 결정하거나 출동하는 데 사용되는 AI 시스템, 그리고 응급 의료 환자 분류 시스템"

8) 인공지능법에서는 'public assistance benefit'라는 용어를 사용했고, 사회보장 영역에서는 이를 공공부조 급여로 번역하는 것이 적정할 듯하지만, 법의 맥락에서는 공공부조에 한정되지 않는 공적인 복지급여 전체를 지칭하는 것으로 보았다.

> "필수적인 공적 복지급여와 서비스를 신청하거나 받는 자연인은… 공공기관 앞에서 취약한 위치에 있다. 이러한 급여와 서비스가 지급, 거부, 축소, 취소 또는 회수되어야 하는지 여부를 결정하기 위해 인공지능 시스템이 사용된다면, 해당 시스템은 사람들의 생계에 상당한 영향을 미칠 수 있고, 사회보호에 대한 권리, 차별 금지, 인간의 존엄성 또는 효과적인 법적 구제와 같은 기본권을 침해할 수 있으므로 고위험으로 분류되어야 한다."

여기까지만 보면 유럽연합은 사회보장제도에서의 인공지능 적용에서 엄격한 입장인 것으로 보인다. 앞서 살펴본, 다소 강한 규제 내용이 사회보장 영역에서 적용될 여지도 크다. 그렇지만 법의 다음 문장에서는 현행 사회보장제도에 인공지능 기술 적용의 가능성도 열어 둔다.

> "이 규정은 공공행정이 준수되고 안전한 AI 시스템의 광범위한 사용을 통해 이익을 얻을 수 있도록 혁신적인 접근 방식의 개발과 사용을 저해해서는 안 된다. 단, 해당 시스템이 법적 및 자연인에게 고위험을 초래하지 않아야 한다."(Artificial Intelligence Act, (58)).

사회보장 행정을 집행하는 정부 및 공공기관은 고위험 인공지능 시스템 과정에서 '배포자(deployer)로 구분될 가능성이 높다(강지원, 2024. 11. 7.). 배포자는 고위험성 인공지능 활용에 있어서 13가지의 의무를 이행해야 한다. 이를테면 데이터 보호 영향 평가, 기본권 영향 평가(fundamental rights impact assessment) 등이 예가 된다. 정부 및 공공기관 입장에서는 부담이 크다.

고위험 인공지능 관련 규정이 유럽 사회 및 사회보장에 미칠 영향에 대해서 유럽 사회는 다소 유보적인 입장을 보인다. 독일 사회보험 유럽대표부(Deutsche Sozialversicherung Europavertretung, 2024)는 "인공지능법 도입 이후 사회적 영향을 면밀히 모니터링해야 한다"고 논평했다. 이러한 반응은 아직 인공지능법의 내용이 전반적으로 모호하고, 구체적인 후속 조치도 아직 이루어지지 않았기 때문으로 보인다. 물론 유럽 노동조합 측에서는 인공지능법이 관련 대기업에 빠져나갈 구멍을 만들어 주었다는 비판도 가하고 있다(Vranken, 2023; Del Castillo, 2023). 따라서 인공지능법 규제의 실질적인 강도는 향후 추가로 마련될 가이드라인, 기준, 표준, 행동강령, 판례 등의 내용에 따라 결정될 것으로 보인다. 앞으로 인공지능법의 구체화를 위한 논의와 협상에서 기업, 노동자 단체, 인권 단체, 관료 집단 등 이해당사자 사이에 첨예한 대립과 갈등이 발생할 가능성이 크다. 그 결과에 따라 이 법이 사회보장에 미치는 의미와 영향력이 달라질 것으로 보인다.

5. 나가며

지금까지 유럽연합 인공지능법의 배경과 내용을 일람하고, 사회보장 영역에 미칠 영향을 살펴보았다. 유럽연합의 인공지능법이 한국에 주는 함의를 검토하기 위해서는 인공지능법이 가지는 두 가지 성격을 먼저 살펴볼 필요가 있다. 첫째, 이 법은 지금까지 나온 AI에 관한 가장 포괄적인 법적 프레임워크라는 점이다. 미국이나 중국 등이 인공지능 기술과 산업의 육성을 중심으로 접근한다면, 유럽연합은 인공지능이 가지는 파괴력을 제어하기 위해 상대적으로 인간중심적, 윤리적 접근을 취하고 있다. 인공지능이 가지는 잠재적인 위험성을 고려할 때, 전 세계가 인공지능에 대한 규제를 마련한다면 유럽연합의 인공지능법은 권위 있는 기준이 될 것이다. 둘째, 유럽연합의 인공지능법은 인공지능을 선도하는 미국 및 중국에 유럽이 던지는 견제구의 성격도 있다(Petrosyan & Ataliotou, 2024). 영국의 언론기관인 Tortoise Media(2024)에 따르면, 인공지능 기술 역량을 기준으로 한 국가별 순위는 미국이 압도적인 1위이고 다음으로 중국, 싱가포르, 영국, 프랑스, 한국, 독일, 캐나다 등의 순이었다. 유럽연합의 시도는 인공지능 규제에 대한 국제적인 표준을 주도하면서, 공적인 규제 과정에서 미국과 중국 등과의 기술적인 격차를 좁히려는 시도로도 해석된다(윤혜선, 2024. 9. 23.).

사회보장 영역에서 인공지능 규제의 국제 표준으로서 유럽연합 인공지능법이 한국에 미칠 영향을 현재로선 단정하기 힘들다. 그러나 인공지능이 가질 파괴력을 고려할 때, 한국도 언제까지 규제에 손을 놓고 있을 수는 없다. 특히 한국에서 급여 대상자 및 급여 수준 사전에서 인공지능, 머신러닝 등 정보통신기술(ICT)이 다수 활용되고 있다. 김기태 외(2024)는 국내 사회보장 분야에서 ICT가 적용된 사례가 36건이라고 분석했다. 한국에서는 '인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 인공지능기본법)'이 2024년 12월에야 제정됐다. 인공지능 기술은 세계 6위 수준이지만(Tortoise Media, 2024), 규제의 측면에서는 이제야 첫걸음을 내디딘 셈이다. 인공지능기본법의 법안 구성 과정에서 사회보장 혹은 사회복지 분야에 대한 논의 또는 언급은 전무했다. 관련 연구도 희소하다. 정부에서도 복지정책 영역에서 인공지능에 대한 준비는 진행된 바가 없다. 인공지능법이 한국에 준용된다면 당장 정부와 공공기관은 배포자로서 데이터 보호 영향 평가, 기본권 영향 평가 등 이행할 의무가 산적해 있음을 고려할 필요가 있다. 관련 규제와 정책의 마련이 시급하다.

| Abstract |

AI's rapid advancement is making a far-reaching impact. Governments worldwide are seeking regulations to maximize AI's benefits while minimizing its risks. Enacted in 2024, the EU AI Act is the most encompassing transnational legal framework to date concerning AI use. It regulates AI use in four areas, classified by their respective risk levels. Within the field of social security, the Act entirely prohibits AI-based social scoring, labeling it as an "unacceptable" and highly harmful risk. However, as administering social security inherently involves some form of social scoring, disputes are likely as the law phases into effect. The use of AI in "essential private and public services," such as social welfare programs, is classified as "high risk," just below the "unacceptable" category, and is subject to strict regulations. Meanwhile, Korea's AI Framework Act, enacted in December 2024, does not address in any substantial way AI use in the area of social security. Given that AI technology is already in use in the field of social security, urgent policy action is necessary to address the associated challenges.

참고문헌

강지원. (2024.11.7.). **정부의 사회보장 서비스에서 AI의 활용과 EU/미국 일부 주 법의 시사점** [발표 자료]. 한국보건사회연구원 사회보장행정에서의 인공지능 적용 동향과 함의 세미나.

김기태, 신영규, 김은하, 김명주, 변소연. (2024). **사회보장행정에서의 인공지능 적용 동향과 함의**. 한국보건사회연구원.

김명주. (2024. 8. 7.). **AI 윤리와 규제** [발표 자료]. 한국보건사회연구원 사회보장행정에서의 인공지능 적용 동향과 함의 세미나.

라기원. (2024). 유럽연합 인공지능법(EU AI ACT)의 특성과 쟁점: 우리나라 인공지능 입법에 대한 시사점. **AI Issue Paper**, 24-20-3. 한국법제연구원.

성준호. (2016). 주민등록번호에 의존한 본인확인제도의 문제점: 각국의 개인식별번호제도 및 관련 법률의 검토를 통한 시사점. **공공사회연구, 6**(2), 208-246.

윤혜선. (2024.9.23.). **EU 인공지능법의 주요 내용과 함의** [발표 자료]. 한국보건사회연구원 사회보장행정에서의 인공지능 적용 동향과 함의 세미나.

인공지능 육성 및 신뢰 기반 조성에 관한 법률 (인공지능기본법)

Appelman, N., Fathaigh, R. O., & van Hoboken, J. (2021). Social welfare, risk profiling and fundamental rights: The case of SyRI in the Netherlands. *J. Intell. Prop. Info. Tech. & Elec. Com. L., 12*, 257-271.

Bekker, S. (2021). Fundamental rights in digital welfare states: The case of SyRI in the Netherlands. Netherlands Yearbook of International Law 2019: *Yearbooks in International Law: History, Function and Future*, 289-307.

Del Castillo, A. (2023). The AI Act: deregulation in disguise. *Social Europe*. https://www.socialeurope.eu/the-ai-act-deregulation-in-disguise

Deutsche Sozialversicherung Europavertretung. (2024). *Conference highlights the synergy between AI and the European Pillar of Social Rights*. https://dsv-europa.de/en/news/2024/03/ki-in-sozialer-sicherheit.html.

European Commission AI HLEG. (2019). *Ethics Guidelines for Trustworthy AI*. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

European Artificial Intelligence Act. (Regulation (EU) 2024/1689, 2024)

Human Rights Watch. (2023. 10. 9.). *EU: Artificial Intelligence Regulation Should Ban Social Scoring*. https://www.hrw.org/news/2023/10/09/eu-artificial-intelligence-regulation-should-ban-social-scoring

Jørgensen, R. F. (2021). Data and rights in the digital welfare state: the case of Denmark. Information, Communication & Society, 1-16.

La Quadrature du Net. (2022). *CAF : le numérique au service de l'exclusion et du harcèlement des plus précaires*. https://www.laquadrature.net/2022/10/19/caf-le-numerique-au-service-de-lexclusion-et-du-harcelement-des-plus-precaires/

Madiega, T. (2024). *Briefing: Artificial Intelligence Act*. European Parliamentary Research Service.

Mosene, K. (2024). One step forward, two steps back: Why artificial intelligence is currently mainly predicting the past. Digital Society Blog. Humboldt Institut für Internet und Gesellschaft. https://www.hiig.de/en/why-ai-is-currently-mainly-predicting-the-past/

OECD. (2023). OECD AI Principles overview. Paris: OECD.

Petrosyan, L., & Ataliotou, K. (2024). *A Tale of Two Policies: The EU AI Act and the U.S. AI Executive Order in Focus*. https://trilligent.com/a-tale-of-two-policies-the-eu-ai-act-and-the-us-ai-executive-order-in-focus/

Tortoise Media. (2024). *Global AI Ranking*. https://www.tortoisemedia.com/intelligence/global-ai#rankings

UNESCO. (2023). *Recommendation on the Ethics of Artificial Intelligence*.

Van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. *European Journal of Social Security, 23*(4), 323-340.

Vranken, B. (2023). *Big Tech lobbying is derailing the AI Act*. Social Europe. https://www.socialeurope.eu/big-tech-lobbying-is-derailing-the-ai-act에서 2024.12.14. 인출.

White House. (2022). *Blue Prints for an AI Bill of Rights*.

World Economic Forum. (2023). *The Presidio Recommendations on Responsible Generative AI*.

Zaber, M., Casu, O., & Brodersohn, E. (2024). *Artificial Intelligence in social security organizations*. International Social Security Association.