

이달의 초점

사회보장영역에서 인공지능 기술 적용 현황과 과제

국내 사회보장 영역에서 인공지능 기술 적용 현황과 함의

|김은혜|

국외 사회보장 영역의 인공지능 기술 적용 현황과 규제

|김기태·신영규|

한국 인공지능기본법의 내용과 사회보장제도 내 한계

|양승엽|

사회보장 영역 인공지능 기술 적용의 순기능 및 위험성에 대한 검토

|김기태|



한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



국외 사회보장 영역의 인공지능 기술 적용 현황과 규제

Current Status and Regulation of Artificial Intelligence Applications in the Field of Social Security Abroad

김기태 한국보건사회연구원 사회보장정책연구실 연구위원

신영규 한국보건사회연구원 사회보장정책연구실 부연구위원

전 세계적으로 인공지능 기술이 사회보장 영역에서 폭넓게 활용되고 있다. 미국, 영국, 오스트리아, 뉴질랜드 등 복지국가들은 챗봇, 머신러닝 기반 수급자 식별, 자동 청구 시스템 등 다양한 방식으로 인공지능 기술을 보건복지 정책 현장에서 활용하고 있다. 반면 네덜란드(SyRI), 호주(Robodebt), 덴마크(Gladsaxe)의 사례를 보면 인공지능 기술은 인권 침해와 차별, 정보 유출 등 새로운 유형의 문제를 초래하고 있다. 이러한 부작용에 대응하는 규제의 대표 사례로 미국 바이든 정부의 행정명령 14110을 소개했다. 해당 규제는 트럼프 정부의 등장에 따라 폐기됐지만, 여전히 인공지능 규제로서 국제적인 표준 성격이 있는 점을 감안했다. 행정명령은 기술의 안전성과 형평성을 확보하기 위한 8대 원칙과 보건복지부 등 주요 부처의 구체적 실행 지침을 포함하고 있다. 한국 사회보장 영역에는 설명 가능성, 이의 제기권, 알고리즘 점검 체계 개선 등을 고려한 인공지능 규제안을 마련할 필요가 있다. 사회보장 관련 데이터 품질 검증 및 개선, 그리고 위험 관리를 담당할 조직 강화와 인력 확보를 검토해야 한다.

1 들어가며

공공영역 가운데 사회보장은 인공지능의 발전과 적용이 가장 활발하게 관찰되는 분야 가운데 하나다(Zaber et al., 2024). 인공지능 기술은 다수의

복지국가에서 이미 급여 자격 심사, 급여액 산정, 급여 지급 등의 과정에서 적용되고 있다. 참고로, 인공지능은 다음과 같이 정의된다. “인공지능 시스템이란 명시적 또는 암묵적 목표를 달성하기 위해 입력받은 정보를 바탕으로 예측, 콘텐츠, 추천 또는

의사결정과 같은 출력을 생성하는 방법을 추론하고, 이를 통해 물리적 또는 가상 환경에 영향을 미칠 수 있는 기계 기반 시스템.”(OECD, 2024a, p. 4.) 미국의 경우 미국 연방정부에서 인공지능을 활용하는 사례 가운데 3분의 1 이상이 보건복지부 업무 영역에 속한다(Burt, 2024. 10. 17.). 인공지능 기술은 사회정책 영역에서 효율성, 과학성, 중립성의 증진에 기여할 것이란 기대가 크기 때문이다(Zaber et al., 2024). 아동수당 및 실업수당 등 복지급여를 중심으로 오스트리아, 뉴질랜드, 미국, 영국 등 다수의 국가에서 인공지능 기술을 적용하고 있다(정세정 외, 2023).

그럼에도 사회보장 영역에 인공지능이 초래할 영향 및 정책적 대응에 대한 국내 연구는 희소하다. 디지털 및 정보화의 사회정책적 함의에 관한 연구는 일부 있었고(김수영, 2016; 김수영, 김수완, 2022; 정세정 외, 2023), 보건복지 현장에서의 인공지능 기술 적용에 관한 연구(조남경, 송기호, 2023; 정유채, 2022)는 일부 수행된 바 있다. 성윤희 (2022)는 사회보장 영역에서 인공지능이 초래할 변화를 검토한 드문 연구이지만, 인공지능을 “4차 산업혁명의 요체이며 지식 기반 사회의 핵심 화두”로 파악하는 기능적 부분에 초점을 뒀다.

한국이 공공행정에서의 인공지능 활용이 뒤쳐진 것도 아니다. 김기태 외(2024)에서는 한국의 사회보장 영역에서 정보통신기술(ICT)이 활용된

36개 사례를 확인했다. 한국은 공공 분야에서 디지털 인프라도 잘 갖춘 나라다. 한국은 경제협력 개발기구(OECD)(2024b)가 발표하는 디지털 정부 지수 순위에서 압도적 1위를 차지했다. 한국은 여섯 개 평가 영역 가운데 데이터 기반 정부, 플랫폼 정부, 개방형 정부, 선제적 정부 등에서 1위를 차지했고, 디지털 우선 정부, 국민 주도형 정부 등 2개 부문에서 2위를 기록했다. 영국 미디어 업체인 토터스미디어가 제시한 ‘Global AI Index’에서는 미국, 중국, 싱가포르, 영국, 프랑스에 이어 6위를 차지했다(Cesareo & White, 2023).

한국은 사회보장 영역에서 인공지능 기술을 디지털 인프라 위에서 빠른 속도로 적용시키고 있지만, 인공지능 기술 적용이 초래할 법적·윤리적 부작용 문제에 대해서는 상대적으로 방임적인 태도를 보였다. 관련 연구와 정책이 한국에서 제시된 바가 희소한 것은 이러한 상황을 증명한다. 이 글은 이러한 문제의식에서 출발했다. 다른 복지국가들의 사회보장 영역에 대한 ① 인공지능 기술 적용 현황 및 문제점을 살펴보고, ② 관련 규제 동향을 미국의 과거 바이든 정부의 행정명령 14110을 중심으로 분석한 뒤 이를 통해 ③ 인공지능 기술의 사회보장 적용 과정에서 짚어야 할 정책적, 제도적 함의를 간략히 검토한다.

2 국외 사회보장 영역의 인공지능 기술 적용 현황¹⁾

가. 국외 기술 적용 사례 및 영역

사회보장 영역에서 인공지능이 사용되기 시작한 시점은 2008년 이후지만 본격적인 확산기는 2017년 이후였다(Zaber et al., 2024). 인공지능이 활용되는 사회보장 영역은 가족급여, 보건, 산업 재해, 연금, 실업, 보편적 급여 등 다양하다. 2020년 이전에 사회보장 영역에서 주로 활용된 인공지능 기술은 챗봇이었다. 챗봇은 2020년 이후에도 코로나 범유행 상황에서 폭증한 급여 신청을 기관들이 대응하는 과정에서 활용도가 더욱 높았다. 일부 기관은 챗봇의 기능을 강화하는 과정에서 생성형 인공지능(generative artificial intelligence)을 활용하기 시작했다.

Zaber et al.(2024)는 사회보장 관련 기관들의 인공지능 활용 영역을 다음과 같이 제시했다(pp. 22-23).

첫째, 서비스 제공(Service delivery). 기관이 다양한 유형의 고객에게 더 쉽게 접근할 수 있는 채널을 활용하여 보다 적절하고 나은 서비스를 제공할 수 있다.

둘째, 자동화 및 사례 관리(Automation and case management). 사회보장 기관이 인공지능

을 사용하여 사례 처리 과정을 자동화하고, 개인의 서비스 후속 조치를 위한 고충 대응을 지원한다.

셋째, 전향적이고 능동적인 사회보장(Prospective and proactive social security). 사회보장 기관에 인사이트, 비전 및 잠재적 결과를 파악할 수 있는 전망 분석을 위한 도구를 제공한다. 기관은 이러한 도구를 활용하여 코호트와 개인의 삶을 선제적이고 능동적으로 개선하기 위한 접근 방식을 수립한다.

넷째, 위험 관리 및 예방(Risk management and prevention). 기관이 위험을 식별하고 위험을 완화하거나, 위험에 대응할 수 있는 역량을 강화한다. 더 나은 서비스를 제공할 수 있도록 회원 정보를 분석할 수 있도록 지원한다.

다섯째, 평등과 공정성(Equality and fairness). 공정성과 형평성의 원칙을 지킬 수 있는 기관의 의무 측면에서 다양한 인공지능 솔루션, 프로그램, 대응에 대해 평가한다.

Zaber et al.(2024)이 제시하는 범주들은 명료하거나 배타적이지 않다. 영역별로 사례가 제시되지 않아서 직관적인 이해가 쉽지도 않다. 그러나 [그림 1]을 보면 다섯 개 영역에서 파생되는 인공지능 기술의 활용 내용을 보다 상세하게 파악할 수 있다. 다섯 영역 가운데 하나인 서비스 제공 분야에서 인공지능은 가장 활발하게 사용된다.

서비스 제공(Service delivery) 가운데서도 챗

1) 이 부분 전반부는 Zaber et al. (2024). Artificial Intelligence in Social Security Organizations(International Social Security Association)를 정리한 것이다.

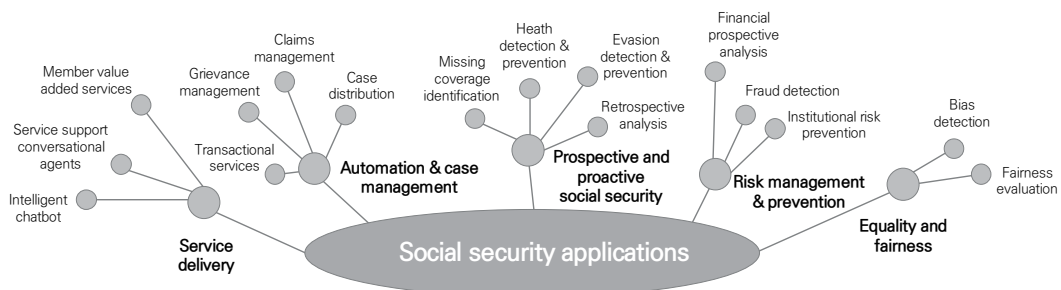
봇은 인공지능이 가장 활발하게 사용되는 영역이다. 흥미롭게도 남미 국가에서 챗봇의 활용도가 높았다(ISSA, 2021). 브라질, 아르헨티나, 파나마, 우루과이 등에서 국민들의 급여 관련 문의나 민원을 응대하는 데 챗봇이 활용됐다. 이를테면 코로나19 범유행 위기 상황에서 브라질 국립사회보장원(INSS: Instituto Nacional de Seguro Social)은 인공지능 기반의 가상 비서 챗봇 Helô를 도입했다. INSS가 관리하는 애플리케이션인 Meu INSS에 대한 문의에 응답하도록 설계된 Helô는 사용자와의 상호작용을 개선하고 더 복잡한 응답도 가능하도록 기능이 점진적으로 확장됐다. 초기 평가 결과 Helô는 첫 달에 100만 건 이상의 상담을 처리했다(ISSA, 2021). 이후 3년 동안 Helô는 90개 주제에 대해 660만 건의 상담을 처리했고, 사용자 만족도에서 0~5점 척도 기준으로 긍정적인 3.8점을 받았다(INSS, 2023).

챗봇의 활용은 소위 ‘선진국’에 국한되지 않는

다. 말레이시아의 직원연금기금(EPF)은 ELYA 스마트 챗봇을 운영하고 있다(Zaber et al., 2024; ISSA, 2020). ELYA는 자연어 처리(NLP)를 사용하면서 인공지능으로 구동되고 라이브 채팅으로 지원되는 사회보장 기관 최초의 이중언어가상비서(VA)다(ISSA, 2020). 고객이 직접 EPF 상품 및 서비스에 대한 정보에 액세스할 수 있도록 하는 방식으로 상담 센터의 부담을 덜어 주는 것이 목표였다. ELYA는 기본 챗봇에서 3단계의 업그레이드를 거쳐 복잡한 문의를 위한 자문 챗봇까지 가능하게 진화했다. ELYA는 실시간 상호작용, 다국어 지원, 연중무휴 24시간 상담 서비스를 제공했다.

복지급여 수급자 포착에도 인공지능은 이미 활용되고 있다. 캐나다 고용사회개발부(Employment and Social Development Canada)는 저소득 노인을 위한 급여인 보장소득보조금(Guaranteed Income Support) 영역에 인공지능 기술을 적용했다(Zaber et al., 2024). 급여 수급 자격이 있는

[그림 1] 사회보장 영역에서의 인공지능 기술 적용 분야



출처: "Artificial Intelligence in social security organizations", Zaber et al., 2024, International Social Security Association, p. 23.

노인을 신속하게 식별하기 위한 목적이었다. 머신러닝 모델을 통해 2000명 이상의 수급 대상자를 식별했다. 정확도는 92~98%였다(Zaber et al., 2024).

오스트리아의 사회보험연합은 청구 자동 처리를 지원하고 의사와 매칭하는 인공지능 기반 시스템을 구현했다(Zaber et al., 2024). 인공지능을 활용한 의료비 환급 관리 사례가 예시될 수 있다. 인공지능 솔루션을 도입하기 전에는 수작업 처리 시간이 길어 환급받기까지 몇 달씩 기다려야 했다. 플랫폼을 도입한 후에는 추가 인력 없이도 처리 시간이 며칠로 크게 단축됐다. 인공지능 기반 접근 방식은 광학 문자(Optical Character) 및 개체 인식과 같은 기술과 오픈소스 도구 및 언어를 사용하여 문서를 수집하고 처리함으로써 효율성과 투명성을 개선했다.

나. 기술 적용이 노출한 문제들

인공지능이 그리는 사회보장의 미래가 장밋빛으로만 그려지지는 않는다. 네덜란드, 덴마크, 호주의 사례는 인공지능 적용에 따른 인권 침해, 정보 유출, 공공성 훼손에 관한 도전을 보여 준다.

먼저 네덜란드의 사례를 살펴보겠다. SyRI (Systeem Risico Indicatie)는 중앙정부와 지방정부가 사회보장 및 소득 관련 제도, 세금 및 사회보험료, 노동법 분야에서 부정행위를 방지하고 대처하기 위해 도입한 인공지능 시스템이다. SyRI는 다양한 공공기관의 데이터를 연결하고 분석하여 위험 보고서를 생성함으로써 재정의 오용을 방지하고, 부당한 이익을 탐지하는 데 사용된다(Rechtbank Den Haag, 2020). 이 시스템에 참여하는 기관은 지방정부, 네덜란드 조세 및 관세청, 사회보험은행

[표 1] 사회보장 영역에서의 인공지능 기술 적용 사례

사회보장 영역	인공지능의 해법	상세 내용
연금	연금 행정 자동화	연금 관리 및 연금 지급을 자동화하여 퇴직자에게 적시에 급여 지급, 관리 비용 절감, 효율성 개선
아동 돌봄 및 교육	인공지능 아동 돌봄 및 교구 활용	스마트 모니터링 시스템으로 어린이의 안전 보장. 인공지능 기반 교육 플랫폼으로 어린이 개인화한 학습 경험 제공 및 발달 과정 모니터링
급여 수급자 확인	자동화한 수급자 확인	인공지능 기반 시스템으로 본인 확인 절차 간소화. 신원, 문서, 데이터를 신속하고 정확하게 확인하여 부정 수급 가능성 차단
고령 돌봄	돌봄을 위한 로봇 활용	인공지능 기반 로봇이 노약자에게 신체적, 인지적 지원 제공 및 일상 업무 지원, 복약 알림, 말벗 기능 수행
정서적 돌봄	정신건강을 위한 챗봇	인공지능 기반 챗봇이 정서적 지원 및 상담 서비스 제공

출처: "Artificial Intelligence in social security organizations", Zaber et al., 2024, International Social Security Association, p. 29의 <표 3> 내용을 필자가 번역하고 일부 편집함.

(Social Insurance Bank), 이민 및 귀화 서비스, 고용보험청, 사회 및 고용감찰청 등과 같은 감독 기관이다. 이 기관들로부터 수집되는 데이터는 보건, 재정, 교육, 재정 지급, 고용 등 방대한 영역에 걸쳐 있다. 알고리즘을 통한 데이터 분석 결과를 바탕으로 특정 사례에 대한 '위험 보고서'가 제출되면 관련자는 부정행위의 가능성이 있는 것으로 판단되어 정부의 조사 대상으로 간주된다(Appelman et al., 2021, p. 263).

이러한 알고리즘을 활용한 분석은 부정 수급을 적발하는 데 효과적이라는 것이 네덜란드 정부의 입장이다. 이 시스템은 주로 부정 수급 가능성이 클 것으로 예상되는 가난하거나 취약한 사람들에 초점을 맞추어 조사를 수행한다는 점에서 논란이 많았다. 이와 관련하여 Platform Bescherming Burgerrechten, Nederlands Juristen Comité voor de Mensenrechten 등 네덜란드 6개 시민 단체는 SyRI의 문제점을 지적하며 공동 행동을 시작했다. 단체들은 2018년 3월 네덜란드 정부를 상대로 이 시스템이 "모든 개인은 사생활과 가족생활, 주거와 통신을 존중받을 권리가 있다"고 규정한 유럽인권협약(European Convention of Human Rights) 제8조를 위반했다는 취지로 헤이그 지방법원에 소송을 제기했다(van Bekkum & Borgesius, 2021). 2020년 헤이그 법원은 정부 기관들이 개인 정보를 공유하는 과정에서 투명성이 부족하고 개인 정보에 대한 적절한 안전장치가 부족하다는 이유로 SyRI가 유럽인권협약을 위반한다고 판결했다.

호주의 로버데트 스캔들(Robodebt Scandal)도 복기할 필요가 있다. 호주는 2016년부터 개개인의 복지수당 초과 지급분의 계산 및 초과분 환수 통보를 자동화한 로버데트 시스템을 도입했다(홍승헌, 황하, 2024). 이 과정에서 데이터 매칭 알고리즘이 핵심적 역할을 담당했지만, 47만 건에서 오류가 발생한 것으로 추정됐다. 액수로는 총액 10억 호주달러였다. 데이터에 기반한 알고리즘을 과신함으로써 수많은 사회적 약자들에게 해악을 끼친 대표적 사례다. 2020년 시스템은 공식적으로 폐지됐다(홍승헌, 황하, 2024). 정확성이 결여된 인공지능 기반 사회보장 행정이 초래할 수 있는 끔찍한 결과다.

2018년 덴마크의 지방정부인 글라삭세(Gladsaxe)는 사회경제적으로 취약한 아동을 추적하기 위해 알고리즘을 활용한 정책 실험을 추진하였다. 덴마크의 수도 코펜하겐 근교에 위치한 글라삭세 정부는 실업 및 의료를 비롯한 여러 사회경제적 영역의 데이터를 결합하여 200개 이상의 위험 지표를 분석하는 머신러닝 모델을 구축하였고, 이를 활용하여 가정폭력이나 학대의 위험이 큰 아동을 찾고 시도했다. 이 모델은 가정이나 부모의 상황들을 점수화했다. 예를 들면 정신질환의 경우 3000점, 실업에는 500점, 예약된 의사의 진료 불참에는 1000점, 예약된 치과 진료 불참에는 300점 등이 부여되었다. 글라삭세 알고리즘 모델은 이러한 점수를 바탕으로 위험한 상황에 처한 아동들은 판별하고자 했다(Thapa, 2019).

이러한 시도는 시민단체, 학계의 신랄한 비판을 받

있고, 대중의 심각한 반발을 낳았다. Mchangama와 Liu(2018)는 인공지능을 활용한 복지행정이 효율성을 높일 수는 있지만, 그 과정에서 취약계층을 대상으로 하는 무분별한 자료 수집 및 그에 근거한 개입 과정에서 개인의 자유와 민주주의의 가치가 희생되고 있다며 “덴마크 복지국가가 인공지능 때문에 사살하고 있다”라고 주장했다. 이러한 비판에도 불구하고 덴마크 정부는 위험에 처한 아동을 조기에 발견할 수 있다는 장점을 강조하며 글라삭세 모델을 전국으로 확대한다는 계획을 세웠다(Bendixen, 2018; Jørgensen, 2021). 그러나 글라삭세 모델에 대한 공개적인 검증이 이루어지면서 상황은 바뀌었다. 이 모델을 통해 작성되는 개별 평가서에 포함된 다양한 정보가 부모들 모르게 활용되고 저장된 사실이 밝혀졌다. 알고리즘에 근거한 위기 아동 판별의 부정확성 문제도 제기됐다(Jørgensen, 2021). 이로 인해 2018년 12월 글라삭세 모델은 중단되었다(Algorithm Watch & Bertelsmann Stiftung, 2020).

3 인공지능 기술의 사회보장 분야 적용에 대한 규제: 미국 바이든 정부의 행정명령 14110

가. 행정명령 14110의 배경 및 개요

전 세계 국가 및 지역 정부, 국제기구가 인공지능에 대한 규제를 내놓고 있다. 인공지능이 초래할 수

있는 사회적 위험이 거대할 수 있고, 동시에 비가시적 불확정적이기 때문이다. 대표적인 예가 World Economic Forum(2023)의 The Presidio Recommendations on Responsible Generative AI, European Union(2024)의 AI Act, OECD(2023)의 AI Principles overview, UNESCO(2021)의 Recommendation on the Ethics of AI, 미국 백악관의 Blueprint for an AI Bill of Rights(White House, 2022), 미국 캘리포니아주가 시도했던 인공지능 규제 법안(SB 1047) 등이다. 여기에서는 미국의 이전 행정부인 바이든 정부의 행정명령 14110을 살펴보고자 한다. 유럽연합의 인공지능법과 바이든 정부의 행정명령 14110을 제외한 다른 형태들은 대부분 법적 구속력이 없는 권고나 가이드라인 형태를 띠기 때문이다. 이 중에서도 굳이 임기가 지난 바이든 정부의 행정명령 14110을 소환하는 이유는 해당 규제가 “인공지능 법제를 유럽형·미국형으로 분류할 때 미국형의 표준”(양승엽, 2025, p. 1)으로 간주되기 때문이다. 유럽형 표준에 해당하는 유럽연합의 인공지능법이 지나치게 규제에 편중됐다는 지적도 고려했다(Petrosyan & Ataliotou, 2024).

바이든 정부를 뒤이은 트럼프 정부는 2025년 1월 20일 행정명령 14110을 폐지했다. 트럼프 정부의 2025년 1월 23일 발행된 대통령령 행정명령 14179 ‘미국의 인공지능 리더십을 위한 장벽 제거(Removing Barriers to American Leadership in Artificial Intelligence)’는 안전 고려 사항보다

‘미국의 글로벌 인공지능 지배력 유지 및 강화 (sustain and enhance America’s global AI dominance)’를 명시적으로 우선시켰다(White House, 2025. 1. 23.). 트럼프 정부의 인공지능 정책은 인공지능에 대한 지원과 규제에서 지원으로 지나치게 편중됐다는 평가를 받고 있다(Masood, 2025). 반면 바이든 정부의 행정명령은 혁신뿐 아니라 사회적 안전, 윤리, 국제 협력, 국가안보를 동시에 고려한 균형 잡힌 규제, 비교적 긍정적인 평가를 받는다(Masood, 2025; 양승엽, 2025). 행정명령 14110을 보면, 미국 및 글로벌 차원에서의 AI 규제의 규범적 미래 모습에 대한 추정이 가능하다.

바이든 정부는 미국 연방정부 최초로 관리예산처(OMB)의 각서 초안과 함께 안전하고 신뢰할 수 있는 인공지능의 개발 및 사용에 관한 행정명령 (Executive Order of the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 행정명령 14110)을 2023년 10월에 발표했다. 행정명령 14110은 전체 13조로 구성됐다(US Executive Order No. 14110, 2023, 표 2). 1조에서는 행정명령의 목적을 제시하고, 3조에서는 주요 용어에 대한 개념을 정의했다. 2조가 전체 구조에서 핵심이라고 볼 수 있는데, 인공지능 기술 활용에서의 8대 원칙을 제시했다. 이 명령에서는 8대 원칙이 뒤에 이어지는 4~11조의 내용을 구성하도록 했다. 즉 1원칙 ‘인공지능 기술의 안전과 보안 확보’가 4조의 내용이 되고, 2원칙 혁신과 경쟁 촉진이 5조의 내용이 된

다. 이러한 순서로 제시된 3~8의 원칙은 다음과 같다. 노동자 지원(6조), 형평성과 시민권 증진(7조), 소비자·환자·승객·학생 보호(8조), 개인정보 보호(9조), 연방정부의 인공지능 활용 증진(10조), 해외에서의 미국 리더십 강화(11조) 등이다. 그 밖에 12, 13조는 시행 및 일반 규정에 관한 내용을 담았다.

내용을 조금 더 살펴보자. 행정명령 1조(목적)에서는 인공지능이 불러올 혜택과 위험을 모두 제시한 뒤 무수한 이점을 활용하기 위해 정부, 민간, 학계, 시민사회의 협력을 강조했다. 2조에서는 8개 원칙 가운데 ‘혁신, 경쟁, 협력의 촉진(2원칙)’ 혹은 ‘글로벌 사회, 경제, 기술 발전의 선도(8원칙)’보다 ‘안전과 보안(1원칙)’, ‘노동자 지원(3원칙)’ 같은 가치들이 선순위에 있는 점이 눈에 띈다.

3조에서는 주요한 개념들을 정의했다. 행정명령의 대상이 되는 ‘기관(agency)’은 미국 법전 44편 35장에 제시된 기관의 정의에 따랐다. 여기서 기관은 ‘정부 행정부(대통령 행정실 포함), 군부, 정부 법인, 정부 통제 법인 또는 기타 기관 또는 독립 규제 기관’(US Exec. Order No. 14110, 2023, p. 75193)이다. 인공지능은 다음과 같이 정의됐다. “인간이 정의한 일련의 목표에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 권장 사항 또는 결정을 내릴 수 있는 기계 기반 시스템. 인공지능 시스템은 기계 및 인간 기반 입력을 사용하여 실제 및 가상 환경을 인식하고, 자동화된 방식의 분석을 통해 이러한 인식을 모델로 추상화하며, 모델 추론을 사용하여 정보 또는 행동에 대한 선택을 형성”(White

[표 2] 미국 행정명령 14110의 사회보장 관련 주요 내용

구분	내용
제6조 노동자 지원	- 인공지능이 노동시장에 끼치는 영향력 및 실업자 지원 방안을 담은 보고서 제출 - 고용 및 직무평가, 모니터링 등에 활용되는 인공지능이 노동자의 복지에 미칠 수 있는 잠재적 피해를 완화하고 이익을 극대화하는 데 활용될 수 있는 사용자 원칙과 모범 사례 개발 및 공표
제7조 형평성과 시민권 증진	- 형사 사법 시스템에서의 인공지능 및 시민권 강화를 위해 인권 침해와 차별 없이 형사 사법 시스템에 인공지능을 활용하는 방법을 설명하는 보고서 제출 - 정부 복지 프로그램에서 시민권 보호를 위해 인공지능을 활용할 때 공평하고 정당한 결과가 도출되는지에 대한 분석 및 그러한 인공지능 활용법에 대한 지침 개발 - 광범위한 경제적 영역에서의 인공지능 및 시민권 강화를 위해 고용, 부동산, 금융, 장애 등과 관련된 맥락에서 차별 금지를 달성하기 위한 지침 개발
제8조 소비자, 환자, 승객, 학생 보호	- 독립 규제 기관을 통한 모델의 투명성 강화와 설명 능력 제고를 포함한 여러 조치를 통해 미국 소비자 보호 - 특히 헬스케어, 공중보건, 인적 서비스, 교통, 교육, 통신 등의 영역에서 인공지능의 책임 있는 개발, 배치, 이용 보장 강조

주: 볼드 및 밑줄은 필자가 강조를 위해 사용.

출처: “인공지능 시대의 경쟁력 강화를 위한 AI 규제 연구”, 김정욱 외, 2024, 경제인사회연구원, pp. 145-146를 일부 수정 및 보완.

House, 2023. 10. 30., p. 75193)하는 것이다.

나. 사회보장 관련 행정명령의 규제들

행정명령 가운데 사회보장과 관련된 부분은 7조(형평성과 시민권 증진)와 8조(소비자, 환자, 승객, 학생 보호) 부분이다. 6조도 노동자의 복지에 관한 내용을 일부 담고 있다. 또 10조에서는 미국 보건복지부(Department of Health and Human Services)를 포함한 기관들의 관리 지침을 제시하고 있다.

먼저 7조에서 2항은 ‘정부의 급여 및 프로그램과 관련한 시민권의 보호’에 대한 내용이다. 기관들은

연방정부의 프로그램과 급여를 집행하는 과정에서 인공지능 기술의 활용으로 발생할 수 있는 ‘불법적인 차별 및 기타 피해를 예방 및 해결’(prevent and address unlawful discrimination and other harms)”(US Exec. Order No. 14110, 2023, p. 75212)하기 위해 각 기관에 있는 민권 및 시민 자유 사무소(civil rights and civil liberties office)를 활용해야 한다. 여기서 민권 및 시민 자유 사무소는 미국에서 시민의 기본권 및 자유에 영향을 미칠 수 있는 주요 기관에 마련된 부서를 지칭한다.²⁾ 이 원칙에 따라 행정명령에서는 두 개 부처의 장관직을 직접 호명하면서 관련 지침을 제시하고 있다. 두 장관 가운데 하나가 다른 아닌 보건복지부 장관

2) 미국의 보건복지부라 할 수 있는 Department of Health and Human Services(2023)에서는 Office for Civil Right(OCR)가 행정 집행 과정에서의 차별과 인권 보호 문제를 다룬다. 이를테면 OCR의 한국어 누리집(<https://www.hhs.gov/ocr/get-help-in->

(The Secretary of Health and Human Services)이다.

행정명령 7조 2항 (b)호에 따르면 정부 지급 급여의 공평성 제고를 위해 보건복지부는 급여 및 서비스를 시행할 때 자동화 또는 알고리즘 시스템의 사용을 촉진하는 계획을 행정명령 시점 기준으로 180일 이내에 발표해야 한다. 그러한 지침의 목적은 다음과 같은 절차를 개선하기 위함이다. 첫째, 자격을 갖춘 수급자의 접근성에 대한 평가, 둘째, 인공지능 시스템의 존재에 대한 수급자 대상 통지, 셋째, 급여의 부당한 거부를 감지하기 위한 정기적인 평가, 넷째, 전문기관 직원(expert agency staff)이 적절한 수준의 재량권을 유지하도록 하기 위한 절차, 다섯째, 급여 거부에 대해 급여 신청자가 사람인 심사자에게 이의를 제기할 수 있도록 하는 절차, 여섯째, 급여 프로그램에서 사용 중인 알고리즘 시스템이 공평하고 공정한 결과를 달성하는지 분석하는 것이다(US Exec. Order No. 14110, 2023, pp. 75212-75213).

행정명령의 8조(소비자, 환자, 승객, 학생 보호)에서는 직접적으로 관련 정부 부처들을 호명하면서 일일이 지침을 제시했다. (a)항은 독립 규제 기관(independent regulatory agencies), (b)항은

보건복지부, (c)항은 교통부, (d)항은 교육부, (e)항은 연방통신위원회에 주는 지침이다. 이 가운데 (b)항인 보건복지부 대상 지침이 가장 길다. 대략 3쪽 분량의 8조 내용 가운데 (b)항만 2쪽가량을 차지한다. 보건복지부가 수행하는 업무의 민감성이 반영된 결과로 추정된다.

보건복지부에 대한 지침 사항을 담은 (b)호는 ‘의료, 공중보건 및 복지³⁾ 부문에서 안전하고 책임감 있는 인공지능의 배포와 사용을 보장하기 위해’라는 목적을 제시하면서 전체 네 개의 세부 ‘목’을 명시하고 있다. (i)목은 보건복지부가 설치해야 할 태스크포스에 대한 지침이고(90일 시한), (ii)목은 보건복지부의 인공지능 기반 기술의 품질 평가 판단 기준 개발에 대한 지침이다(180일 시한). (iii)목은 정부의 재정 지원을 받는 보건복지 서비스 제공 업체에 대한 규제 관련 고려 사항을 담고 있다(180일 이내). 그리고 (iv)목은 의료 영역의 인공지능 관련 안전 프로그램 수립을(365일 이내), (v)목은 신약 개발에서 인공지능 사용에 대한 규제 전략을 제시하고 있다(365일 이내).

해당 내용을 하나씩 살펴보면 다음과 같다. 먼저 (i)목에서는 보건복지부 안에 인공지능 태스크포스의 구성이 명령일로부터 90일 이내에 이뤄져야

other-languages/korean.html)에서는 아래와 같이 안내하고 있다. “미국 Department of Health and Human Services (HHS, 보건복지부) 산하 Office for Civil Rights(OCR, 민권담당국)는 … 비차별, 양심 및 종교적 자유, 대상 주체에서 의료 정보 프라이버시에 대한 시민들의 기본권을 보호하는 연방 민권 법률, 양심 및 종교적 자유 법률, Health Insurance Portability and Accountability Act(HIPAA, 건강보험 이전 및 책임에 관한 법률), Privacy, Security, and Breach Notification Rules(개인정보 보호, 보안 및 침해 고지 규칙), Patient Safety Act and Rule(환자안전법 및 규칙)을 집행합니다.”

3) 원문에는 이 대목에서 ‘healthcare, public health and human services’라고 제시했다. 미국 보건복지부에서 ‘human services’ 부분은 전통적인 아동복지, 노인복지, 빈곤정책 등을 담당하는 점을 고려해 ‘복지’라고 번역했다.

함을 규정한다. “보건복지부 장관은 국방부 및 보건부 장관과 협의하여 보건복지 부문, 즉 연구 및 발전, 의약품 및 기기 안전, 의료 전달 및 재정, 공중보건 영역에서 인공지능 및 인공지능이 탑재된 기술의 책임 있는 배치 및 사용에 대한 정책과 프레임워크를 포함한 전략계획을 수립해야 한다.” (US Exec. Order No. 14110, 2023, p. 75214) 전략계획에는 필요시 적절한 규제 조치를 포함할 수 있다. 보건복지부 내의 인공지능 태스크포스는 구성 시점 이후 365일 이내에 이러한 전략계획을 짜야 한다. 전략계획에는 다음과 같은 영역에서 적절한 지침과 자원이 마련되어야 한다(US Exec. Order No. 14110, 2023, pp. 75214-75215).

첫째, 품질 측정, 성과 개선, 급여 집행 등 의료서비스 제공 및 재정 분야에서 예측 및 생성 인공지능(predictive and generative AI) 기반 기술의 개발, 유지, 사용. 여기에서 인공지능이 생성한 결과를 현장에 적용하는 과정에 대한 인간의 적절한 감독도 함께 고려해야 한다.

둘째, 보건복지 부문에서 인공지능 기반 기술의 장기적인 안전성 및 실제 성과 모니터링. 여기서 모니터링의 대상은 임상적으로 관련이 있거나 중요한 수정 사항 및 성과가 된다. 규제 기관, 개발자 및 사용자에게 제품 업데이트를 알릴 수 있는 수단도 제공되어야 한다.

셋째, 보건복지 부문에서 사용되는 인공지능 기반 기술에 형평성 원칙 반영. 이를 위해 새로운 모델을 개발할 때 영향을 받는 인구에 대한 분리된 데

이터와 대표 인구 데이터 세트를 사용한다. 기존 모델의 알고리즘이 가질 수 있는 차별과 편견 관련 부작용을 모니터링하고, 현재 시스템이 가질 수 있는 차별과 편견을 식별하고 완화하기 위함이다.

넷째, 개인 식별 정보를 보호하기 위해 소프트웨어 개발 라이프사이클에 안전, 개인정보 보호 및 보안 표준 반영. 보건복지 부문에서 인공지능으로 심화한 사이버 보안 위협을 해결하기 위한 조치도 포함해야 한다.

다섯째, 보건복지 현장에서 사용자가 적절하고 안전한 인공지능 사용을 결정하는 데 도움이 되는 문서의 개발, 유지 관리 및 가용성 확보.

여섯째, 지역 단위 보건복지 관련 기관과 협력하여 인공지능 기술 사용에 대한 긍정적인 현장 사용 사례와 모범 사례를 촉진하기 위한 작업 수행.

일곱째, 행정 부담 감소를 포함하여 보건 및 인적 서비스 부문에서 업무 효율성과 만족도를 증진하기 위한 인공지능 사용 사례 파악.

(b)항의 (ii)목에는 인공지능 기반 기술의 품질 평가 판단을 위한 전략 개발 내용을 담고 있다. 보건복지부는 행정명령 발표 180일 이내에 이러한 전략을 개발해야 한다. 전략의 내용은 앞의 (i)목에서 제시한 영역을 포괄해야 한다. 전략에는 인공지능 보증정책(AI assurance policy) 및 관련 인프라에 대한 요구 사항이 포함되어야 한다. 인공지능 보증 정책은 인공지능 기반 의료 도구의 성능을 평가하기 위함이다. 두 번째인 인프라에 대한 요구 사항은 인공지능 기반 의료 기술 알고리즘 시스템 성

능의 시판 전 평가 및 시판 후 감독을 실제 데이터와 대비해 수행할 수 있도록 하기 위해서다.

(b)항의 (iii)목은 정부의 재정 지원을 받는 보건 복지 서비스 제공 업체에 대한 규제 내용을 담고 있다. (iii)목의 내용에 따르면 연방의 재정 지원을 받는 보건복지 서비스 제공 업체는 인공지능 관련 기술을 적용하는 과정에서 연방차별금지법(Federal nondiscrimination laws)을 준수하도록 보건복지부가 적절한 조치(appropriate actions)를 고려해야 한다. 이러한 고려는 180일 이내에 이뤄져야 한다. 법에서는 여기서의 ‘적절한 조치’ 두 가지를 예시했다. 한 가지만 보면 인공지능과 관련된 연방 차별금지법 혹은 개인정보보호법 미준수에 대한 불만 또는 기타 신고에 대해 보건복지부가 지침을 발표하는 것이다.

(b)항의 (iv)목은 인공지능 안전 프로그램 수립에 관한 내용이다. 즉 보건복지부는 연방에 등록된 환자 안전기관(PSO: Patient Safety Organization)⁴⁾과 협력하여 인공지능 안전 프로그램을 수립해야 한다. 시한은 행정명령 발표일 기준으로 365일이다. 인공지능 안전 프로그램은 다음의 내용을 포함한다.

첫째, 의료 현장에서 인공지능 기술을 적용함으로써 발생하는 임상 오류를 식별하고 포착하기 위한

공통의 프레임워크 수립. 환자, 간병인 등에게 차별과 편견 등으로 피해를 입히는 등의 사고에 대한 중앙 추적 저장소(central tracking repository)⁵⁾의 사양(specifications) 수립.

둘째, 이렇게 수집된 데이터와 생성된 증거를 분석하여 적절한 경우 이러한 피해를 방지하기 위한 권고안, 모범 사례 또는 기타 비공식 지침 개발.

셋째, 이러한 권고안, 모범 사례 또는 기타 비공식 지침을 의료서비스 제공자를 포함한 적절한 이해관계자에게 배포.

(b)항의 (v)목에는 신약 개발에서 인공지능 사용에 관한 규제 전략을 담고 있다. 이 전략에 들어갈 내용도 다음과 같이 제시하고 있다. 첫째, 신약 개발의 각 단계에서 적절한 규제에 필요한 목적, 목표, 상위 원칙 정의, 둘째, 이러한 규제안을 구현하기 위한 향후 규칙 제정, 지침 또는 추가적인 법적 권한이 필요할 수 있는 영역 확인, 셋째, 이러한 규제안에 필요한 기존 예산, 자원, 인력 및 새로운 공공-민간 파트너십 파악 등이다. 지금까지 살펴본 행정명령 14110에 제시된 보건복지 분야 관련 내용을 종합하면 <표 3>과 같다.

참고로, 장애인에 관한 내용도 포함되어 있다. 행정명령은 “장애인이 시선 방향, 시선 추적, 보행

4) 환자안전기관(PSO)은 2005년 미국의 환자 안전 및 의료 질 향상법(Patient Safety and Quality Improvement Act)에 의해 만들어졌다(Agency for Healthcare Research and Quality, 2023). 환자의 안전을 개선하기 위해 의료 기관들이 자발적으로 참여한다. 미국 보건복지부 산하 의료연구·의료의 질 개선 담당 기관(Agency for Healthcare Research and Quality, AHRQ)가 등록 및 감독을 맡는다.

5) 여기서 중앙 추적 저장소(central tracking repository)에 대한 부가적인 설명이 없어서 어떠한 성격일지는 모호하다. 맥락으로 보아서는 의료 현장에 인공지능 기술이 적용되는 과정에서 발생한 문제나 피해 내용을 기록하고 추적 및 활용하는 데이터 저장소를 의미하는 것으로 추정된다.

분석, 손동작과 같은 생체 인식 데이터 사용으로 인한 불평등한 대우 등 위험으로부터 보호받는 동시에 인공지능의 잠재력을 활용할 수 있도록 건축 및 교통 장벽 준수 위원회(Transportation Barriers Compliance Board)는 적절하다고 판단되는 경우 대중의 참여를 요청하고, 지역사회 참여를 유도하며, 생체 인식 데이터를 입력하여 사용할 때 인공지능의 위험과 이점에 대한 기술 지원 및 권고를 하

고, 장애인이 정보통신기술 및 교통 서비스에 접근할 수 있도록” 권장한다(US Exec. Order No. 14110, 2023, p. 75214).

다. 행정명령의 위상 및 의미

남궁준(2024)은 미국 행정명령의 법적 위상에 대해 다음과 같이 설명했다. 다소 길지만 인용하면

[표 3] 행정명령 14110에서 제시한 보건의복지 분야 관련 내용

구분		미국 행정명령 14110
제7조 형평성과 권리 증진	2항 정부의 급여 및 프로그램과 관련한 시민권의 보호	- 기관들은 연방정부의 프로그램과 급여를 집행하는 과정에서 인공지능 기술의 활용으로 불법적인 차별 및 기타 피해를 예방 및 해결해야 함 - 정부 급여의 공정성 제고를 위해 자동화 또는 알고리즘 시스템의 사용을 촉진하는 계획 작성. 다음의 절차를 개선하기 위함임 ① 자격을 갖춘 수급자의 접근성에 대한 평가 ② 인공지능 시스템의 존재에 대한 수급자 대상 통지 ③ 급여의 부당한 거부를 감지하기 위한 정기적인 평가 ④ 전문 기관 직원이 적절한 수준의 재량권을 유지하는 절차 ⑤ 급여 거부에 대해 급여 신청자가 사람인 심사자에게 이의를 제기할 수 있도록 하는 절차 ⑥ 급여 프로그램에서 사용 중인 알고리즘 시스템이 공정하고 공정한 결과를 달성하는지 분석
제8조 소비자, 환자, 승객, 학생 보호	(b)호 “의료, 공중보건 및 복지 ⁶⁾ 부문에서 안전하고 신뢰성 있는 인공지능 배포와 사용을 보장하기 위해” (p. 75214)	- 미국 보건복지부 안에 인공지능 태스크포스 구성(90일 이내). 태스크포스 구성 이후 365일 이내 전략계획 개발. 다음 영역에서 지침과 자원 모색(i 목) ① 의료서비스 제공 및 재정 분야에서 예측 및 생성 인공지능(predictive and generative AI) 기반 기술의 개발, 유지, 사용 ② 인공지능 기반 기술의 장기적인 안전성 및 실제 성과 모니터링 ③ 인공지능 기반 기술에 형평성 원칙 반영 ④ 안전, 개인정보 보호 및 보안 표준 반영 ⑤ 안전한 인공지능 사용을 결정하는 데 도움이 되는 문서의 개발, 유지 관리 및 가용성 확보 ⑥ 긍정적인 현장 사용 사례와 모범 사례 촉진 ⑦ 업무 효율성과 만족도를 증진하기 위한 인공지능 사용 사례 파악 - 인공지능 기반 기술의 품질 평가 판단을 위한 전략 개발(180일 이내). 전략은 위의 i목 내용 포괄(ii 목) - 정부의 재정 지원을 받는 보건의복지 서비스 제공 업체에 대한 규제 고려(180일 이내) (iii 목) - 의료 영역에서 인공지능에 관한 환자 안전 프로그램 수립(365일 이내) (iv 목) - 신약 개발 시 인공지능 사용에 대한 규제 전략 제시(365일 이내) (v 목)

출처: “US Executive Order, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence”, 2023. White House에서 필자가 내용 정리함.

다음과 같다. “연방의회가 새로운 법안을 통해 AI 사용 규제를 위한 새로운 기관을 설립하거나 기존 규제 당국에 새로운 권한을 부여하지 않고 있다. 따라서 연방 행정기관들은 AI 관련 규제가 필요할 경우 그들이 이미 가지고 있는 기존 권한, 즉 관련 법령 3(의 재해석)에 근거해 AI 규제를 가능하게 하는 방향으로의 관할 사무 집행을 할 수 있을 뿐이다. 그런데 대통령이 발령한 동 행정명령은 (향후 그러한 방향으로의 사무 집행이 법원의 심사를 통해 취소되는 것은 별론으로 하더라도) 현재 발효 중인 법령을 보다 넓게 (재)해석하고 부여된 권한을 보다 적극적으로 행사할 것을 관련 행정기관 모두에 주문함으로써 포괄적·입법적 입법을 통한 AI 규제와 어느 정도 유사한 효과를 가져올 수 있을 것으로 기대”(p. 188)하고 있다.

미국의 보건복지부는 행정명령 발표 이전인 2021년에 최고인공지능책임자(CAIO)를 지명했다(Alder, 2024. 5. 29.). 지난 2024년 12월 기준 보건복지부의 CAIO는 미키 트리파티(Micky Tripathi)였다. 트리파티 차관보는 행정명령이 정하는 보건복지부의 인공지능 대응 태스크포스도 이끌고 있다.⁶⁾ 보건복지부는 행정명령 8조 (b) (ii)에서 규정하는 부처 차원의 인공지능 대응 전략계획을 2025년 1월에 제시하겠다고 밝혔다(Burt,

2024. 10. 17.). 트리파티가 지난 2023년 미국 하원의 에너지 및 상업위원회(House Committee on Energy and Commerce)에 출석해 발언한 내용을 보면 미국 보건복지부도 백악관이 내놓은 입장을 맞추기는 벅찼던 것으로 보인다(Mitchell, 2024. 10. 9.). 미국 연방정부에서 인공지능을 활용하는 사례 가운데 3분의 1 이상이 보건복지부 업무 영역에 속한 점을 고려하면(Burt, 2024. 10. 17.) 바이든 행정부의 행정명령 14110에 담긴 보건복지 관련 규제 내용은 앞으로 전 세계의 보건복지 분야 인공지능 적용에서 규제의 전범으로 활용될 것으로 보인다.

4 나가며

공공 영역 중에서도 특히 사회보장 분야에서는 인공지능 기술의 적용이 빠르게 확산되고 있다. 급여 지급, 자격 심사, 사례관리 등 다양한 분야에서 인공지능이 활용되고 있다. 이러한 기술 적용에는 명암이 교차한다. 실제로 네덜란드의 SyRI, 호주의 로버렛, 덴마크 글라삭세 모델 사례는 인공지능 기반 사회보장 행정이 법적·윤리적 문제를 초래한 사례로 평가된다. 이에 따라 규제의 정당성과 투명성, 설명 가능성, 사후 구제 절차의 중요성이 부각된다.

6) 원문에서는 이 대목에서 ‘healthcare, public health and human services’라고 제시됐다. 미국 보건복지부에서 ‘human services’ 부분은 전통적인 아동복지, 노인복지, 빈곤 정책 등을 담당하는 점을 고려해서 ‘복지’라고 번역했다.

7) 미국 이전 정부의 보건복지부 조직도(www.hhs.gov/about/agencies/orgchart/index.html)를 보면 미키 트리파티 차관보가 이끄는 ‘기술정책차관보실 및 국가 건강 정보 기술 조정관 사무소’는 제시됐지만, CAIO 직위나 인공지능 태스크포스 조직은 제시되지 않고 있다. 조직도만 보면 인공지능 태스크포스 조직이 신설됐다고 보기는 어려워 보이며, 기존 기술 정책 차관보실에서 해당 업무를 떠안았던 것으로 추정된다.

인공지능 기술 적용에 대한 규제 대표적 사례로 미국 바이든 행정부의 행정명령 14110을 점검했다. 당시 정부는 안전하고 신뢰할 수 있는 인공지능의 개발 및 사용을 목표로 8대 원칙을 제시하고, 보건복지부 등 주요 부처에 구체적인 규제 이행 방안을 지시했다. 행정명령은 유럽연합의 인공지능법과 비교될 만큼 포괄적인데, 미국형 규제 모델의 표준으로 간주된다.

지금까지의 논의에 근거해 정책적 함의를 두 가지 제시하고자 한다. 인공지능 기술은 사회보장 영역에서 사회적 약자에게 불리한 결정을 내릴 수 있는 편향 문제가 내재하고 있다. 예컨대 네덜란드의 SyRI 시스템은 저소득층을 주요 표적 대상으로 삼아 심각한 인권 침해 논란을 일으켰다. 이러한 사례는 인공지능 기술을 적용할 때 공정성과 투명성을 보장할 제도적 장치가 필요함을 시사한다. 미국 바이든 행정부의 행정명령 14110은 정부 프로그램에서 인공지능 기술을 활용할 때 차별을 방지하고 공정한 결과를 유도하기 위해 다양한 규범과 절차를 제시한다. 한국의 사회보장 행정에서 인공지능 기술이 폭넓게 사용되지만, 이에 관한 규제는 사실상 전무하다. 인공지능 기반 행정 결정의 설명 가능성과 시민의 이의 제기권 보장, 알고리즘 편향에 대한 정기적 점검 체계를 갖추는 것이 시급하다. 이는 기술에 대한 사회적 신뢰를 확보하고 복지행정의 공공성을 강화하기 위한 선결 조치다.

한국의 사회보장 관련 당국에도 인공지능 품질 검증 및 개선을 맡으면서 위험 관리 체계를 갖춘 조

직과 인력이 필요하다. 보건복지 분야에서 인공지능이 다루는 데이터는 건강, 소득, 가족관계 등 가장 민감한 개인정보를 포괄한다. 인공지능에 따른 급여 판정 등은 개인과 가구의 생명과 생계를 좌우한다. 기술 적용 과정에서의 안전성과 책임성 확보가 필수적이다. 의료서비스나 사회복지급여의 자동화 결정이 오작동하거나 편향될 경우 회복하기 어려운 피해가 발생할 수 있다. 미국 행정명령 14110이 미국 보건복지부에 인공지능 대응 태스크포스 설치를 명령하고, 인공지능 기반 기술의 성능과 형평성, 개인정보 보호 수준을 지속적으로 모니터링하는 전략계획 수립을 요구하는 점을 유의할 필요가 있다. ■

참고문헌

- 김기태, 김명주, 김은하, 신영규, 변소연. (2024). **사회보장행정에서의 인공지능 적용 동향 및 함의**. 한국보건사회연구원.
- 김수영, 김수완. (2022). 데이터 복지국가의 도래와 쟁점: 국가, 시민, 시장의 관계 지형을 중심으로. **한국사회복지정책학회 춘계학술대회자료집** (pp. 91-118).
- 김수영. (2016). 사회복지정보화의 윤리적 쟁점-사회보장정보시스템을 통한 데이터감시를 중심으로-. **한국사회복지학**, 68(1), 193-224.
- 김정욱, 김종립, 최윤경, 김민정, 유성희, 최현이, 김승현, 진승화, 김윤경, 박정후, 추수진. (2024). **인공지능 시대의 경쟁력 강화를 위한 AI 규제 연구**. 경제·인문사회연구원.

- 남궁준. (2024). 인공지능과 알고리즘 관리에 대한 노동 법적 규율: 미국 바이든 정부의 AI 행정명령을 중심으로. **사회법연구**, 53, 183-218.
- 성윤희. (2022). 인공지능 (AI) 과 사회복지법제에 대한 소고. **사회복지법제연구**, 13(2), 119-148.
- 양승엽. (2025). 한국 인공지능기본법의 의의, 내용과 한계.
- 정세정, 김기태, 박윤경, 우선희, 최준영, 이영수. (2023). **한국 복지국가의 재구조화를 위한 연구 - I. 디지털 복지국가의 딜레마**. 한국보건사회연구원.
- 정유채. (2022). AI 와 블록체인 기반 의료 복지 서비스 사례 연구 1. **복지경영학연구**, 11, 77-86.
- 조남경, 송기호. (2023). 사회복지의 상담기록, 좀 더 활용할 수 있을까? '머신러닝'을 통한 사회복지 상담 텍스트 활용 가능성의 점검. **한국사회복지조사연구**, 79, 5-26.
- 홍승헌, 황하. (2024). 누구를 위한 디지털 전환인가? 자동화된 복지행정의 위험성. **정부학연구**, 30(2), 61-84.
- Agency for Healthcare Research and Quality. (2023). *Patient Safety Organization(PSO) Program*. <https://pso.ahrq.gov/>
- Alder, M. (2024, 5. 29). HHS names acting chief AI officer as it searches for permanent official. *FedScoop*. <https://fedscoop.com/hhs-names-acting-chief-ai-officer/>
- Algorithm Watch & Bertelsmann Stiftung. (2020) *Automating Society Report 2020*. <https://automatingsociety.algorithmwatch.org>
- Appelman, N., O'Fathaigh, R., & van Hoboken, J. (2021). Social welfare, risk profiling and fundamental rights: The case of SyRI in the Netherlands. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 12, 257-271.
- Bendixen, M. (2018) Denmark's 'anti-ghetto' laws are a betrayal of our tolerant values. *The Guardian*. <https://www.theguardian.com/commentisfree/2018/jul/10/denmark-ghetto-laws-niqab-circumcision-islamophobic>
- Burt, C. (2024. 10. 17.). HHS Strategic Plan for AI Coming Soon. *ExecutiveGov*. <https://executivegov.com/2024/10/department-of-health-and-human-services-strategic-ai-plan/>
- Cesareo, S., & White, J. (2023). *The Global AI Index*. Tortoise Media.
- Department of Health and Human Services. (2023). Office for Civil Right. <https://www.hhs.gov/ocr/get-help-in-other-languages/korean.html>
- INSS. (2023). *Helô, assistente virtual do INSS completa três anos*. <https://www.gov.br/inss/pt-br/noticias/helo-assistente-virtual-do-inss-completa-tres-anos>
- ISSA. (2020). ELYA: *The bilingual virtual assistant of the Employees Provident Fund - Empowering customers to self-serv, anytime, anywhere*. <https://www.issa.int/gp/208625>
- ISSA. (2021). *The application of chatbots in social security: Experiences from Latin America* (Analysis). International Social Security Association.
- Jørgensen, R. F. (2021). Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society*, 1-16.
- Masood, A. (2025, May). *Comparative analysis of executive approaches to artificial*

- intelligence governance: Lessons from competing orders. Medium. <https://medium.com/@adnanmasood/comparative-analysis-of-executive-approaches-to-artificial-intelligence-governance-lessons-from-f9b0202e2745>
- Mchangama, J., & Liu, H.-Y. (2018). The Welfare State Is Committing Suicide by *Artificial Intelligence*. *Foreign Policy*. <https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/>
- Mitchell, B. (2024. 10. 9.). *HHS developing new AI strategic plan*. <https://fedscoop.com/hhs-developing-new-ai-strategic-plan/>
- OECD. (2024a). *Explanatory Memorandum on the Updated OECD Definition of an AI System*. OECD.
- OECD. (2024b). 2023 OECD Digital Government Index: Results and key findings (OECD Public Governance Policy Papers, No. 44). OECD Publishing. <https://doi.org/10.1787/1a89ed5e-en>
- Petrosyan, L., & Ataliotou, K. (2024). A Tale of Two Policies: The EU AI Act and the U.S. AI Executive Order in Focus. *Trilligent*. <https://trilligent.com/a-tale-of-two-policies-the-eu-ai-act-and-the-us-ai-executive-order-in-focus/>
- Rechtbank Den Haag. (2020). *SyRI legislation in breach of European Convention on Human Rights*. *de Rechtspraak*. <https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx>
- Thapa, E. P. (2019). *Predictive Analytics and AI in Governance: Data-driven government in a free society – Artificial Intelligence, Big Data and Algorithmic Decision-Making in government from a liberal perspective*. European Liberal Forum.
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.
- 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.
- White House. (2022). *Blueprint for an AI Bill of Rights*. White House.
- White House. (2023. 10. 30). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence [Executive order 14110].
- White House. (2025. 1. 23.). *Removing Barriers to American leadership in artificial intelligence*. <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>
- World Economic Forum. (2023). *The Presidio Recommendations on Responsible Generative AI*. World Economic Forum.
- Zaber, M., Casu, O., & Brodersohn, E. (2024). *Artificial Intelligence in social security organizations*. International Social Security Association.

Current Status and Regulation of Artificial Intelligence Applications in the Field of Social Security Abroad

Kim, Ki-tae

Shin, Young-kyu

(Korea Institute for Health and Social Affairs)

AI technologies are widely used in social security systems worldwide. In many welfare states, including the US, the UK, Austria, and New Zealand, AI-powered tools—such as chatbots, machine learning-based models for identifying eligible individuals, and automated claims systems—assist in various aspects of health and social welfare policy administration. By contrast, initiatives like the Dutch SyRI, Australia’s Robodebt, and Denmark’s Gladsaxe have raised concerns regarding discrimination, human rights violations, and data leaks. Executive Order 14110, issued during the Biden Administration, exemplifies regulatory responses to such issues. Although later rescinded following the advent of the second Trump Administration, this order, featuring eight guiding principles for safe and equitable AI development, along with concrete action directives for the Department of Health and Human Services and other key federal agencies, still resonates as an international standard in AI regulation. As for social security in Korea, it is necessary to establish AI regulations taking into account the issue of explainability, the right to appeal, and the need to improve the algorithm audit system. It is also important to strengthen organizations and ensure sufficient staffing for verifying and improving the quality of social security data, as well as managing associated risks.