

SSIS

이슈&트렌드

발행인 김현준 | 발행처 한국사회보장정보원 | 발행일 2026년 6월 30일

2026 Vol. 05

| 제도·전달체계 |

AI 전환 시대의 사회보장 행정 혁신

서울대학교 사회복지학과 | 이종성 조교수

요 약

01. 데이터 기반 패러다임과 AI 전환의 도래

- 저출생·고령화, 재정 제약, 복지 수요 확대 속에서 현재의 서베이 데이터 중심, 신청주의 기반, 사후대응 중심의 전달체계로 인해 측정 오차, 사각 지대, 처리 지연이라는 구조적 한계가 지속될 가능성이 있음.
- 행정빅데이터와 AI라는 도구의 결합은 증거 기반 정책을 넘어 위기를 사전에 감지·대응하는 예측적·적응형 사회보장을 가능하게 할 잠재력이 있음.

02. AI·행정빅데이터 기반 사회보장 행정 혁신: 동향과 기회

- OECD 국가들은 AI를 ① 예측·조기경보, ② 자동화된 행정처리, ③ 맞춤형 서비스 등의 세 영역에 적용 중이며, 정책 형성-대상 선정-집행-평가의 전 주기에서 AI를 도입하는 실제 사례들이 관찰됨.
- 상대적으로 인프라가 취약한 일부 개발도상 국가들에서도 적응형 사회보호의 실현 사례가 실증적으로 관찰됨.

03. AI·행정빅데이터 기반 사회보장 행정의 위험

- 다수 선행연구들이 지적하듯 처리 과정의 불투명성, 데이터 품질, 알고리즘 편향 등 AI의 기술적 위험성과, 책임의 모호성, 프라이버시 침해 등 AI의 거버넌스 위험성이 동시에 상존함.
- 최근 이슈가 된 일부 실패사례(네덜란드, 호주 등)는 위험의 본질이 기술 자체의 한계가 아니라 적절한 거버넌스·통제 시스템의 부재에 있을 수 있다는 가능성을 제시함.

04. 책임 있는 AI 사회보장을 위한 제안

- 기존의 ‘속도·비용 절감’ 중심 성과체계에서 ‘권리 보호·서비스 품질’ 중심으로의 전환이 필요하며, ① 기본권 기반 서비스 품질 관리, ② 이용자 권리 기반 성과지표로의 전환, ③ ‘AI 복지 생태계’ 거버넌스 확립을 검토할 필요가 있음.
- 특히, 한국사회보장정보원은 행정빅데이터 분석 허브이자 데이터 거버넌스 중추로서, 데이터 기반 정책 선도와 권리 보호 중심 알고리즘 규제 제도화를 주도할 필요가 있음.

I 들어가며: 데이터 기반 패러다임과 AI 전환의 도래¹⁾

• 사회보장 환경의 구조적 변화

- 초고령화의 가속, 1인 가구 증가, 노동시장 이중구조 심화로 복지 수요가 양적·질적으로 팽창하는 반면, 가용할 수 있는 재정 및 행정 자원은 상대적으로 정체되어 사회보장 분야의 수요-공급의 구조적 불일치가 심화됨. 더불어, 기존의 신청주의, 사후적 대응(reactive) 중심의 전달체계는 복지 사각지대, 처리 지연, 행정 부담 누적이라는 한계를 지속적으로 노정함.

• 증거 기반 정책에서 AI·행정데이터 기반 정책으로

- 현대 복지국가는 한정된 자원의 효율적 배분을 위해 정확한 수치·통계에 기초한 증거 기반 정책(Evidence-Based Policy)을 지향해 왔음. 그러나, 전통적 서베이 데이터는 개인의 주관적 인식 및 행동 파악, 다차원적 정보 결합, 민주적 기능 등의 강점이 있으나, 2000년대 이후 지속적인 응답률 하락, 측정오차, 소수 집단 누락, 정책시차 등의 한계에 대한 지적이 있어옴.
- 반면, 행정데이터는 전수(全數) 자료로서 표본 오차가 없고, 플랫폼 노동자, 희귀질환자, 최빈곤 계층 등 소수 집단까지 분석이 가능하며, 나아가 AI와 결합하여 예측적 사회정책의 가능성을 현실화하고 있음. OECD(2025) 역시 AI가 위기를 사전에 감지·대응하는 적응형 사회보호(Adaptive Social Protection)를 가능하게 한다고 분석함.

• 원고의 목적과 구성

- 본 고에서는 AI·빅데이터가 사회보장 정책 및 서비스의 형성-집행-평가를 어떻게 재편할 수 있는지를 기회와 한계 측면에서 검토하고자 함. 또한 이를 위해 OECD 국가 사례 등에서 관찰되는 동향과 기회, 내재적인 위험, 정착을 위한 3대 제언, 정보원의 역할과 방향 등을 순차적으로 논의함.

1) 본 원고는 저자의 “AI 전환 시대의 사회보장 서비스 혁신(2026 사회보장 AX 미래전략 심포지엄 발표자료)”, “전환인가, 유행인가? 사회정책에서의 AI 적용가능성과 한계 (미출간 원고)” 등 현재 작성 중인 원고의 일부 내용을 요약한 것이므로 인용을 가급적 자제해주시기를 바랍니다. 상기 문헌에서 인용한 부분에 대해서는 별도의 인용표시를 하지 않았습니다.

II AI·행정빅데이터 기반 사회보장 행정 혁신: 동향과 기회

OECD 국가의 사례: 기능별 접근 (OECD, 2024; OECD, 2025)

- 개요
 - 다양한 행정데이터와 AI의 결합이 제시하는 패러다임은 이미 OECD 주요국에서 기능별 혁신 사례로 구체화되어 사회정책 전반의 효율성을 높이고 있음. OECD(2024)에서는 사회보험 AI·빅데이터가 사회정책에 적용되는 영역을 ① 예측 및 조기경보, ② 자동화된 행정 처리, ③ 맞춤형 서비스 제공의 세 유형으로 분류하고 있으며, 이를 기능 중심으로 재편하여 살펴보면 아래와 같음.
- 프런트오피스(Front-office) 및 민원 지원 개선
 - 핀란드 사회보험청(Kela)의 다국어 챗봇은 신청자 상황에 맞추어 적절한 제도와 절차를 안내하여 접수 이전의 신청 오류를 완화함. 노르웨이는 코로나19 시기 챗봇으로 하루 8,000건 이상의 문의를 처리하며 서비스 중단 없이 민원 폭증에 대응함.
 - 한국 일부 지자체는 AI 업체와 제휴하여 취약계층 대상 자동 안부 전화(주1~2회)를 시행하고, 응답 패턴 분석을 통해 고독사 위험군·돌봄 필요군을 자동 분류하여 담당 공무원에게 연계함.
- 백오피스(Back-office) 행정 자동화 및 부정 탐지
 - 에스토니아는 전체 실업 보험 급여 결정의 일부를 자동 처리하여 처리 기간을 단축하였음. 캐나다, 오스트리아, 핀란드 등은 이메일·음성·문서 자동 분류 시스템 도입으로 민원 배분의 정확도를 향상시킴.
 - 영국 HMRC 등은 분석 시스템을 통해 복지급여, 세제 영역의 부정수급을 탐지하고 위험도 높은 사례에 조사를 집중하고 있으며, 포르투갈에서는 얼굴·음성 인식 기술로 해외 거주 연금 수급자의 생존 여부를 자동 확인하여 부정수급을 방지하는 것으로 알려져 있음.
- 예측 기반 선제적 개입 및 적응형 사회보호
 - 네덜란드에서는 부채 위기 조기 개입 프로그램(Vroeg eropaf)에서 알고리즘으로 채무 위험 가구를 조기 발견하여 예방적 개입을 제공함. 미국 LA 및 영국 Maidstone 등에서는 노숙 위험 예측 모델로 노숙 위험군을 조기 식별하고 선제적으로 개입함.
 - 개발도상국가로 분류되는 토고 등에서는 데이터가 부족한 상황에서도 위성·기후 데이터와 휴대전화 메타데이터를 활용하여 농촌 빈곤층을 식별하고 긴급 현금 지원을 집행하여, 통계 인프라 부족 국가에서도 적응형 사회보호가 가능함을 보여줌.
 - OECD 국가는 아니지만, 파키스탄에서는 수 백만명의 빈곤층이 발생한 2022년 대홍수를 계기로 AI 기반 재해 예측 모델과 연계한 조기 현금지원 시스템을 시범적으로 운영하여 선제적 사회보호의 가능성을 실증하고 있으며, 위성영상과 통신 데이터를 결합한 머신러닝(ML) 기반 빈곤 지도 시범운동을 통하여 전통적

서베이 조사 대비 낮은 비용으로 고해상도의 빈곤 분포 추정 가능성이 높음. 몽골에서도 UNDP와 공동으로 AI 준비도 평가를 실시하였으며, 위성영상을 AI로 분석하여 농촌 빈곤층 밀집 지역인 게르(Ger)의 위치와 밀집도를 지도화하고 드론 촬영 영상을 딥러닝 모델(YOLO)로 분석하여 중요한 자산인 가축들의 개체수를 자동 식별하는 시범사업을 실시함.

■ AI 기반 사회정책의 가능성

- 정책형성 (Policy Formulation): 제한된 경험 중심에서 보완된 설계로
 - 공중의제(Public agenda)를 비정형데이터 등을 기반으로 수집하고 이를 통해 실시간으로 정책 수요를 포착하는 역할을 할 수 있음. 정책 의제 발굴에는 통상 1년 단위의 한정된 예산으로 소규모 설문조사를 진행하거나, 전문가 집단, 정부와 소통하는 이해관계자 집단, 또는 기사로 대표되는 언론이 중요한 역할을 해왔음. 그러나, 발굴된 정책 의제의 대표성이 특히 대상(국민)과 시간(현재)의 측면에서 추가 보완될 필요가 있다는 지적이 있어 왔으며, AI를 활용한 비정형 데이터 기반의 정책 수요 발굴이 이를 보완(할 가능성)을 제시함.
 - 예를 들어, Bae et al.(2023)의 민원 분석은 분노, 슬픔 등 정서 유형이 정부의 민원 대응 속도에 영향을 미침을 보여줌. Lee(2022)의 LDA 기반 고령화 담론 분석은 국가부채·노동시장 혁신 담론이 증가하는 반면 소득보장 담론은 감소하는 경향을 보였으며, 이러한 작업은 현재 정책 논의 공백과 담론 편향을 식별하는 정책 설계 도구로 활용될 가능성이 있음.
- 정책 대상자 선정(Targeting): 포괄적 지원에서 정밀 타겟팅으로
 - 대상자 선정의 측면에서는 기존 방식은 집행의 투명성 등을 고려하여 연령, 소득 등 투명하지만 단선적인 기준(eligibility rule)에 의존하는 경우가 많았으나, 이로 인해 신청주의의 한계 및 복지 사각지대 등의 이슈도 발생하고 있음. AI 도입과 이로 인해 촉진되는 다양한 종류의 데이터 융합은 사회적 위기의 선제적 감지와 신청주의 한계에 도움을 줄 수 있음. 예를 들어, 덴마크에서는 가족, 주거, 조세 데이터 교차 검증으로 대상자 선정 확인이 필요한 집단을 데이터로 사전 추출함.
 - 대상자 발굴의 측면에서는 이전에 비해 능동적인 접근(Out-reach)이 가능해지며, 실제로 뉴질랜드에서는 청년층의 장기적 복지 의존 확률을 데이터를 기반으로 사전에 계산하여 자립 지원 프로그램을 연계하는 방식이 가능해짐. 앞서 살펴본 파키스탄의 사례도 이와 유사함.
 - 마지막 급여 지급 및 사후관리 측면은 사회보장 정책의 건전성 및 신뢰성과 직결되는 영역이나, 동시에 행정 오류가 발생할 경우 시민의 기본권과 재산권을 직접 침해할 수 있는 고위험(high-risk) 영역이기도 함. 호주와 네덜란드 등의 해외 선례에서는 특히 이 영역의 실패 사례가 이슈화되는 경향을 보임.
 - 이는 최근 다양한 영역의 전문가들이 제시하고 있는 AI와 데이터에 기반한 자동적인 대상 선정 및 급여 지급 결정의 가능성과 연결되며, 최근 AI 기반 신청주의 폐지 주장과도 연관됨. 다만, 자동화된 결정(automated decision)의 허용범위에 관한 문제는 대상 집단, 행정행위의 성격(예: 침익 또는 수익), 잠재

적 집행부담(예:신청주의 폐지시의 미지급 결정에 대한 입증책임) 등이 종합적으로 고려되어야 하는 사안으로, 기술적 가능성과 효율성 이외에 반드시 고려되어야 할 요인들이 경시될 경우 사회적 이슈가 예상되므로 추후 심도있는 논의가 필요함.

- 정책 집행(Implementation): 지연에서 자동화로

- 정책 집행 과정은 전형적인 Back-end 과정의 효율화에 해당하여, 행정 서류 처리 부담과 인력 부족으로 인한 만성적 행정부담을 효율적으로 개선할 가능성이 있음. 예를 들어, 절차(Process) 측면에서는 사회보장급여 청구의 부분적 자동화가 가능해짐. 미국 사회보장국(SSA)에서는 장애인 연금 청구에 의료 문서 등의 자동 분석 AI를 도입하여 처리 기간을 단축하였고, 싱가포르에서는 AI 알고리즘이 이미지, 텍스트 분석을 통해 민원을 정확한 담당 부서로 자동 배정하여 민원을 신속히 해결한 사례가 관찰됨. 전달(Delivery)의 측면에서는 24/7로 확장된 서비스가 가능해지며 핀란드 사회보장청(Kela) 등 다양한 국가에서 시행하는 AI 챗봇 서비스가 이에 해당함.
- 특히, 집행 과정에서 AI 기반 행정 자동화와 현장 실무자의 행정 부담 경감 등을 통해 절약되는 자원은 해당 서비스에 대한 재투자로 연결되어야 하며, 이를 전제로 AI는 인간적 유대와 대면 돌봄 시간을 회복하는 보조 도구(Augmentation tool)로 기능할 수 있음. 이는 OECD 등에서 강조하는 다채널 전략(Omni-channel)의 핵심으로, 영국과 일본의 사례 등에서 사회서비스 실무자들의 초과근무 시간 감소 및 생산성 향상과도 이어짐.

- 정책 평가(Evaluation): 사후약방문에서 실시간 피드백으로

- 정책 평가 측면에서는 데이터 축적까지의 수년의 행정 시차가 발생하는 사후 평가 중심의 체계의 한계를 벗어나, AI 도입 이후에는 실시간 피드백 루프, 빅데이터·소셜 네트워크 분석 등을 수행할 수 있는 가능성이 높아지는 것으로 보임.
- 예를 들어 정책반응에 대한 실시간 모니터링이 가능해지며 싱가포르에서는 공공 민원·소셜 미디어를 대상으로 실시간 감성 분석(Sentiment analysis)을 진행하여 정책 집행 상의 문제를 식별함. 더불어 Causal ML 등을 통해 '이 정책이 평균적으로 효과적인가?'를 넘어 '하위 그룹별(30대 저소득 여성 가장 vs 50대 남성 해고자) 이익의 크기가 어떻게 다른가?'에 대한 질문과 평가가 이전에 비해 보다 정교한 수준으로 가능해짐.

III AI·행정빅데이터 기반 사회보장 행정의 위험성

■ AI 알고리즘의 내재적 한계 및 기술적 위험성

- 개요
 - 앞서 살펴본 AI 기반의 혁신적 자동화와 행정 효율성 제고 사례의 이면에는 알고리즘 오작동, 기존 불평등 재생산 등 구조적 위험이 상존하며, 이는 AI 알고리즘 등 기술적 위험성과 AI 거버넌스 상의 위험성으로 나누어 살펴볼 수 있음. 이 중에서 전자는 아래와 같이 구분할 수 있음.
- 투명성 및 설명가능성의 부족
 - 알고리즘의 의사결정 과정을 이해하기 어려운 블랙박스(Black Box) 문제가 다양한 전문가 등을 통해 지적되고 있으며 급여의 한수 및 탈락 과정을 시민과 공무원이 역추적하기 어려워짐에 따라서 행정절차법상 적법 절차 원리(Due process)를 저해한다는 지적이 존재함 (Kuzniacki, 2023).
- 데이터 품질의 문제
 - 데이터 결측, 정합성 부족, 시계열 단절 등의 데이터 품질 문제가 데이터 기반 분석결과의 신뢰성을 저해하는 핵심 위험 요인으로 작용함. AI 모델은 대규모의 정확한 데이터를 필요로 하나, 현재 주로 사용하는 서버이 데이터 등은 표본 크기가 작고 주관적 편향에 대한 교정 매커니즘이 부족하여, 이러한 데이터에 기반하여 AI 모델이 산출한 결과의 신뢰성에 대한 지적이 존재함.
- 알고리즘의 편향성
 - 과거 데이터에 내재된 사회적 불평등이 AI 학습을 통해 강화되어 특정 인종, 계층, 지역에 불리한 결과를 초래함. 미국 사법부의 재범 위험 예측 도구는 흑인에게 거짓양성 판단을 내릴 확률이 백인보다 약 2배 높았음 (Angwin et al., 2016). 이외에도 출입국 행정 알고리즘에서도 유사한 문제가 존재함

■ AI 거버넌스가 내포하는 위험성

- 책임의 모호성
 - 알고리즘 오류 피해 발생 시 IT 기업·정부·실무자 간 책임 전가가 가능하며, 알고리즘의 독점적 비공개(영업 비밀)로 정부조차 시스템을 완전히 통제할 수 있는 가능성이 존재함. (OECD, 2024).
- 프라이버시와 시민권에 대한 위험
 - 취약계층의 소비, 통신, 지리 정보를 전방위적으로 수집하는 과정에서 입법적 통제와 사회적 숙의를 우회한 정책 집행의 우려가 존재함. 즉, “누구에게 민원을 제기하고, 설득해야 할 것인가?”의 문제로, 권리를 지닌 시민이 수동적인 신청자로 전락한 가능성이 존재함.

Ⅳ 책임 있는 AI 사회보장을 위한 제안

■ 개요

- AI·빅데이터는 사회보장 서비스의 효율성과 포괄성을 제고할 가능성이 있으나(Ⅱ 장), 제도적으로 통제되지 않은 도입은 다양한 이슈를 확산할 가능성이 일부 존재함(Ⅲ 장). 특히, AI 논의의 급격한 확산 과정에서 기존 행정 체계와의 조응성이 상대적으로 경시되고 있는 것으로 보이며, 이는 결과적으로 사회정책·서비스의 직접 대상자인 취약계층 보호와도 밀접한 관련이 있음.
- 따라서 현 시점에서의 핵심적인 정책 과제 중 하나는 이미 거스를 수 없는 대세인 ‘도입 여부’가 아니라, ‘어떻게 신뢰할 수 있는(trustworthy) 방식으로 정착시킬 것인가’로 보이며, 이를 위해서는 서비스 품질관리, 성과 측정 개념의 전환, 행정 거버넌스의 고도화 등을 같이 고려할 수 있음.

■ 서비스 품질 관리

- 정책과 서비스 등 행정 처리의 속도를 더이상 서비스 성과의 유일한 요소로 보기 어려움. 정확성, 일관성, 권리구제 가능성, 알고리즘에 의한 소수자 차별 방지 등 품질관리 측면이 행정 서비스의 새로운 평가 척도로 기능할 필요가 있음. EU(2024) 부속서 등에서도 공공급여(예:복지 수당, 주거 지원 등) 결정에 관여하는 AI를 고위험 시스템으로 언급하여 품질에 대한 규제와 영향 평가의 필요성을 제시함.

■ 성과 측정 개념의 전환

- 예산·처리 시간 단축에 집중한 기존 성과 지표를 넘어서 OECD 가이드라인 및 WHO 윤리 원칙 등을 준용한 권리에 기반한 성과 지표로 전환을 검토할 필요. 예를 들면, 접근성, 품질, 공정성, 책임성 등 실제 이용자 및 수급자의 권리에 기반한 지표체계 등 다양한 성과지표를 고민할 수 있음.

■ 행정거버넌스의 고도화

- AI 기반 사회보장은 다차원 데이터에 기반해야 하므로, 단일한 정부 부처가 모든 것을 통제하기 어려움. 따라서, 국가 기관, 민간 IT 기업, 지역사회, 시민사회 단체 간의 일종의 다원화된 AI 복지 생태계(AI Welfare Ecosystem)에서 실제 정책과 서비스가 작동하는 방안을 고려할 수 있음. 예를 들어 위기가구 발굴 과정에서는 유사한 공공서비스 정보 외에도 핸드폰 소액결제 내역, 지역사회복지기관 이용 내역 등이 유용하게 활용될 수 있음. 다수 실증 연구 결과에서도 국가별 성과를 결정하는 것은 AI 기술 도입 수준이 아닌, 각 주체 간 조정(Coordination) 및 리더십 역량인 것으로 나타남.

V 나가며: 한국사회보장정보원의 역할과 방향

I 개요

- 한국사회보장정보원은 산재된 사회보장정보를 체계적으로 통합·관리·활용하여 포용적 복지를 실현하는 사회보장 정보 플랫폼 선도기관으로서, 공정하고 따뜻한 정보서비스를 제공하여 국민의 삶의 질과 행복에 기여하고 있음. 이를 위해 사회보장 체계의 인프라를 총괄적으로 운영하고, 공공 및 민간 기관 간 데이터 상호운용성을 확보하며, 행정데이터 기반 보건복지 정책 개발 및 고도화 지원을 수행해왔음. 최근 AI라는 기술의 재료가 되는 데이터(Data)를 총괄하는 기관으로써 정보원의 역할 확대가 중요하게 요구되는 시점이며, 이에 다음과 같은 역할을 제안함.

II 인공지능(AI)·빅데이터 중심의 생태계 구축

- 행정빅데이터 중심으로 사회정책 패러다임을 선도
 - 최근 전 세계적인 서베이 조사 응답률 하락 및 측정 오차 등 전통적 통계의 위기에 대응하여, 표본 오차가 없고 정밀한 포착이 가능한 전수 ‘행정빅데이터 분석 허브’ 역할을 수행할 필요가 있음.
 - 개인정보보호 우려와 데이터의 활용을 조화롭게 절충하기 위해 사회보장 합성데이터(Synthetic data)를 구축하고, 다수 부처의 정보가 수집되는 특성을 활용하여 태국의 사례처럼 사회보장 데이터 레이크(Data lake) 구축을 통한 행정 고도화를 추진할 수 있음.
- 사회서비스 가치사슬(Value Chain)별 AI 융합 생태계 주도
 - 사회서비스 및 정책의 각 단계별 가치사슬에 따라 AI 도입의 주도적인 역할을 수행할 수 있으며, 간단한 예시는 아래의 표로 제시할 수 있음.

서비스 단계	AI 도입에 따른 정보원의 주도적 역할 및 방향	적용 기대 효과 및 시사점
상담	다국어 지원 24시간 AI 챗봇 및 자가 진단 인터페이스 지원	물리적 접근성 확장, 단순 안내 채널 전환을 통한 초기 민원 응대 예산 절감
사례관리	사례관리의 실시간 전사, 서식 자동화, 위험 신호 탐지 알고리즘 구현 및 검증	사회복지 현장인력의 행정 부담 완화를 통한 대면 서비스 시간 확대
급여 지급 및 사후관리	조세·금융 등 이종 행정데이터 교차 검증을 통한 자격 자동 심사 모델링	복지 재정 누수 방지, 적극동적 직권 급여 체계 전환 대비

- 정책 생애주기별 실시간 데이터 피드백 제공
 - DW의 사례관리 면담 데이터, 혹은 SNS 민원 텍스트 등을 자연어 처리(NLP) 및 감성 분석을 활용하여 실시간으로 분석하여 중장기적으로 국민적 정책 수요와 아젠다를 선제적으로 발굴할 수 실시간 루프 시스템을 구축할 수 있음.

■ 신뢰할 수 있는 데이터 거버넌스 및 리스크 통제

- 잠재적 시스템 에러(System Error) 통제 및 사전 예방
 - 최근 네덜란드, 호주 사건 등 안전장치 없는 행정 자동화가 야기한 해외 사례를 참고하여, 이미 해외 빅테크 회사들이 자체적으로 개발하고 있는 알고리즘 차별(인종·계층 편향 등) 검증 및 식별 매커니즘 등의 개발·도입하여 사전 예방 체계를 구축할 수 있음.
 - 의사결정 과정이 불투명한 일종의 블랙박스 문제를 완화하기 위해 최근 논의되고 있는 ‘설명 가능한 AI(XAI)’를 도입하고, 무엇보다 데이터 품질 상시 모니터링 체계를 고도화할 필요가 있음.
- 권리 보호 중심의 알고리즘 규제 제도화
 - 사회보장 수급 자격 영구 박탈이나 아동 강제 분리 등 국민의 생존권과 기본권에 중요한 영향을 미치는 고위험 AI에 대해서는 도입 전 ‘알고리즘 영향평가(AIA)’를 제도화 하는 방안 등도 검토할 수 있을 것으로 보임.
 - 또한, 일선 사회복지 현장 공무원이 기계적 판단을 무비판적으로 맹신하는 자동화 편향(Automation Bias)을 방지하기 위해, 최종 판단과 책임은 인간이 지는 ‘Human-in-the-Loop’ 원칙 구현을 위한 가이드라인 등 제도화 기반을 마련할 필요가 있음. 최근 OECD 뿐 만 아니라 EU 등 각 국가에서는 사회보장 등 대면 서비스 행정 영역의 AI 도입 및 활용과 관련하여 공통적으로 이러한 원칙을 반영하고 있음.
 - 무엇보다 성과 측정 기준을 기존의 예산 감축 및 처리 시간 단축에서 ‘접근성(포용성), 품질(신뢰도), 공정성(차별방지), 책임성(통제력)’ 등의 수혜자 권리 기반의 지표로의 전환을 검토할 필요가 있음. AI 도입으로 업무 처리 속도의 편차가 완화될 가능성이 높아 보이므로, 새로운 성과 측정 지표의 개발 및 보급을 고민할 시점으로 보임.

참고문헌

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Bae, Y., Woo, B.-D., Jung, S., Lee, E., Lee, J., Lee, M., & Park, H. (2023). The relationship between government response speed and sentiments of public complaints: Empirical evidence from big data on public complaints in South Korea. SAGE Open, 13(2).
- Bedingfield, W. (2020, August 19). Everything that went wrong with the botched A-Levels algorithm.
- Booth, R. (2024, March 4). Scrap plans to scan accounts of benefit claimants or risk new scandal, MPs told. The Guardian.
- Card, D., Chetty, R., Feldstein, M., & Saez, E. (2011). Expanding access to administrative data for

- research in the United States. SSRN. <https://ssrn.com/abstract=1888586>
- Dominitz, J., & Manski, C. F. (2024). Using total margin of error to account for non-sampling error in election polls: The case of nonresponse. arXiv.
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689.
- Henley, J., & Booth, R. (2020). Welfare surveillance system violates human rights, Dutch court rules. The Guardian.
- Holmes, C., / Royal Commission into the Robodebt Scheme. (2023). Report of the Royal Commission into the Robodebt Scheme.
- Kuźniacki, B., Almada, M., Tyliński, K., & Górski, Ł. (2022), Towards eXplainable AI (XAI) in Tax Law, World Tax Journal, 14(4).
- Lee, S. C. (2022). Topic modeling of Korean newspaper articles on aging via latent Dirichlet allocation. Asian Journal for Public Opinion Research, 10(1), 4-22.
- Meyer, B. D., & Mittag, N. (2017). Misclassification in binary choice models. Journal of Econometrics. 200(2). 295-311.
- OECD. (2024). Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. OECD Publishing.
- OECD. (2025). AI and the Future of Social Protection in OECD Countries. OECD Artificial Intelligence Papers, No. 42.
- van Bakkum, M., & Zuiderveen Borgesius, F. J. (2025). Digital welfare fraud detection and the Dutch SyRI judgment. arXiv.
- Verdin, R., O'Reilly, J., & McDonnell, A. (2023). Digital Welfare Ecosystems in Europe: Social protection systems preventing social exclusion and enhancing opportunities to participate in the digital economy. EUROSHIP Working Paper No. 23.
- 이종성. (2026a). “AI 전환 시대의 사회보장 서비스 혁신”, 2026 사회보장 AX 미래전략 심포지엄 발표자료, 4월 29일 서울.
- 이종성. (2026b). “전환인가, 유행인가? 사회정책에서의 AI 적용가능성과 한계”. 미출간 원고.