

AI와 데이터 시대의 사회복지: 행정데이터 활용의 가능성과 과제¹⁾

이종성^{a)}

a) Jongseong Lee
서울대학교 사회복지학과 교수
jslee1@snu.ac.kr

국문 초록과
영문 제목 및 영문 초록은 생략합니다.

I. 들어가며

2022년 OpenAI에서 개발한 ChatGPT의 등장 이후 빅데이터와 인공지능(AI)이 사회 전반에 걸쳐 큰 변화를 가져오고 있다. 마치 2007년 미국의 아이폰 출시가 스마트폰 중심의 생태계를 형성하며 인간 행동 양식 변화에 지대한 영향을 미쳤듯이, 대규모 언어모델(LLM)의 대중화는 AI 중심의 새로운 생태계를 형성하면서 사회 구성원의 행동 패턴 변화를 예고하고 있다. 이러한 인간 행동 양식의 변화는 필연적으로 다양한 사회적 서비스 수요의 변화와 밀접히 연결되며, 특히 인간의 존엄성과 복지라는 가치 중심적 접근을 특징으로 하는 사회복지 영역 역시 이러한 변화에서 예외가 될 수 없다. 따라서 기술 혁신으로 인한 사회적 변화가 인간 중심적 가치를 저해하지 않으면서도, 오히려 사회적 가치를 높이는 방향으로 AI 기술과 조화롭게 통합될 수 있도록 사회복지 연구자와 실무자들의 신중한 접근과 논의가 요구되는 시점이다.

특히, 최근 변화하는 정책 환경 속에서 사회복지 서비스 및 정책에 대한 데이터 기반의 실증적 접근(evidence-based)의 중요성이 점차 부각되고 있다. 특히 지역 간 인구구조의 불균형, 사회서비스 수요의 변화, 재정 건전성에 대한 우려 등 급격한 환경 변화는 기존 전달체계와 정책 설계 방식에 구조적인 조정을 요구하고 있다. 이러한 변화는 서비스 공급주체 및 방식에 대한 재검토를 넘어, 사회보험을 포함한 제도 전반의 지속 가능성을 점검하고 대응 방안을 모색하는 급격한 변화의 시작이라 보인다. 이러한 맥락에서 데이터와 AI라는 최근의 조류가 오히려 사회복지 영역에 주는 시사점이 타 영역에 비해 더욱 크다고 볼 수 있다.

지금까지 사회복지 연구 분야에서 AI·데이터와 관련된 논의는 돌봄 로봇과 같은 하드웨어 중심 사례를 중심으로

활발히 제기되어 왔다. 따라서 본 원고에서는 그간 상대적으로 덜 언급되어왔던 빅데이터 중심의 최근 변화와 향후 과제에 대한 내용을 공유함으로써, 사회복지 분야의 실증적 접근 확대 가능성을 논의하고자 한다. 이는 데이터라는 새로운 기술적 현상에 대한 일차적 이해 증진을 목적으로 할 뿐 아니라, 향후 활발히 이루어질 사회복지 서비스 및 정책 개편 논의에서 사회복지 연구자들이 데이터 중심 접근을 활용할 수 있는 토대를 제공하고 하는데 그 목적이 있다.

II. 사회복지 행정데이터 현황

1. 행정데이터 개요

최근 이슈가 되는 인공지능(AI) 기술의 확장적 적용에 있어 핵심적인 자료는 바로 데이터(data)이다. 데이터는 일반적으로 뉴스, SNS 댓글, 로그기록 등 비정형 데이터(unstructured data)와 서버에 응답 등 정형 데이터(structured data)로 구분된다. 최근에는 인공지능 기술과 결합할 수 있는 대용량의 빅데이터에 대한 관심이 높아지고 있으며, 많은 수의 사회복지 연구에서도 수십 년 치 신문기사 등 비정형 빅데이터를 활용한 분석 시도가 주로 이루어졌다.²⁾ 그러나 2020년 이른바 '데이터 3법'의 개정과 함께 정형 데이터인 행정데이터에 대한 활용 가능성이 본격적으로 논의되기 시작하였다.³⁾⁴⁾

특히 사회복지 영역은 상대적으로 공공부문 중심의 서비스 제공과 재정 집행이 이루어져 온 만큼, 다른 영역보다 행정데이터 활용 잠재력이 비교적 높다. 한국의 사회복지제도는 정부 주도로 공적 재원이 투입되어 왔으며, 공적 재원이 '지나간' 자리는 곧 데이터라는 기록으로 누적되어왔기 때문이다. 특히 사회복지 분야는 전국 단위

행정데이터의 구축이 조기에 이루어졌으며, 이는 국세청, 신용카드 등 민간 기반 데이터와 비교해도 상당한 규모를 보인다. 이에 따라 사회복지 연구자들은 정책 설계 및 평가 등에 있어 다양한 유형의 행정데이터를 적극적으로 활용할 수 있는 가능성을 보유하고 볼 수 있다.

2. 사회보장 행정데이터)

많은 행정데이터들 중에서도 최근 주목받는 것은 사회보장 행정데이터이다. 사회보장 행정데이터는 보건복지부 산하 사회보장위원회 주도로 구축되었으며, 개인 단위의 가명처리를 통해 부처별·제도별로 흩어져 있던 자료를 통합한 국가 단위 빅데이터이다. 제도적 기반은 2021년 개정된 「사회보장기본법」 제42조로, 이 조항은 모집단 대표성을 갖춘 자료 제공 요청 권한을 위원회에 부여하였다(보건복지부, 2024). 2019~2020년 시범사업을 거쳐 2022년부터 본격적인 1차 데이터 구축이 시작되었고, 2023년 2기 사업이 완료되었다. 2024년 4월 기준, 전 국민의 약 20%에 해당하는 1천만 명 규모의 표본 데이터를 바탕으로 33개 기관의 정보를 연계한 자료가 확보되어 있다.

해당 데이터에는 연금, 복지, 고용, 교육, 주거, 의료, 건강 등의 분야 정보가 포함되어 있으며, 변수 수는 826개에 달한다. 변수 예시로는 성별, 연령, 소득, 장애 유무 등의 인구사회학적 정보부터, 복지수급 내역, 건강검진 결과, 고용 및 주거 이력까지 폭넓게 포괄된다. 특히 2기 데이터에서는 국민취업지원제도, 노인일자리 세부유형, 금융소득 정보 등 새로운 항목이 추가되는 등 변수가 577개에서 826개로 확대되었다. 이로 인해 개인 또는 가구 단위 생애 경로, 소득 및 복지수급 패턴, 건강 위험요인 분석 등 다양한 사회복지 분야의 실증 연구가 가능해졌다.⁶⁾

세부 이용 절차는 사회보장위원회 사무국에 연구 목적과 계획을 제출하고, 내부 심의 과정을 거친 후, 지정된 분석 환경 내에서 제한적으로 데이터를 활용하는 방식으로 구성되어 있다. 그러나, 아직까지는 신청·승인이 제한적으로 이루어져 활용의 제약요인으로 작용하고 있다. 데이터 보안은 가명 연결키 방식으로 보장되며, 분석 결과는 위원회 사무국의 검토 후 반출 가능하다.

한편, 사회보장 행정데이터는 몇 가지 한계를 지닌다. 우선 데이터 접근성이 낮고, 연구자에게 제공되는 정보가 제한적이다. 또한 기관 간 연계 과정에서 코드 불일치, 누락, 일관성 결여 등의 품질 문제도 발생한다.⁷⁾ 또한 복잡한 신청 절차와 장시간 소요되는 승인 과정은 연구자의 활용을 제약하는 요인이 되고 있다.

3. 건강보험 행정데이터

건강보험 행정데이터는 국민건강보험공단(NHIS)이 수집·관리하는 의료 및 건강 관련 데이터이다. 전 국민 단일 보험 체계가 운영된 2000년 이후 자격정보, 보험료, 진료 내역, 건강검진 결과 등이 장기적으로 축적되었으며, 2014년부터는 빅데이터 플랫폼을 통해 연구와 정책 개발에 활용되고 있다.

해당 데이터는 자격정보, 진료내역, 처방정보, 검진결과, 장기요양자료 등으로 구성되어 있다.⁸⁾ 자격자료에는 가입자 유형, 소득 분위, 장애 여부, 사망 여부 등이 포함되며, 진료자료에는 상병코드, 진료일수, 진료비 등이 포함된다. 검진자료에는 키, 체중, 혈압, 혈당 등의 생체지표가 포함되며, 장기요양 자료는 등급과 서비스 이용 내역을 포함한다. 해당 데이터에는 사회과학 연구에서 흔히 이용하는 변수가 비교적 제한적으로 포함되었다고 보는 의견도 존재하나, 다양한 추정 기법(예: 납입 보험료를 활용한 소득수준 추정)을 통하여 한계를 보완할 수 있다. 실제 활용 측면에서 주로 만성질환 예측, 의료비 분석, 지역별 건강통계 작성 등의 선행연구가 관찰되나, 의료급여, 장기요양 등 사회복지 제도 관련 정보도 포함하고 있으므로 복지 분야의 연구 가능성이 높은 데이터로 분류할 수 있다.

세부 이용절차로는 건강보험 빅데이터를 이용하고자 하는 공공·민간 연구자는 국민건강정보 자료제공 서비스(NHIS)를 통해 자료 이용을 우선 신청해야 한다. 일반적으로 연구계획서와 IRB 승인서, 연구자 신원·기관 증명서류 등을 제출하여 공단·평가원의 심의위원회 심의를 거쳐야 한다. 승인 이후에는 원격접속(표본DB)을 이용하거나, 전국에 분포된 오프라인 센터 내의 연구분석 공간(맞춤형 DB)을 활용할 수 있다.

건강보험 행정데이터는 방대한 양의 데이터를 제공하나 일부 한계도 존재한다. 첫째, 데이터 접근이 여전히 엄격히 통제되며 사용료가 부과될 수 있어 연구자에게 부담으로 작용할 수 있다. 둘째, 원자료가 행정 목적(진료비 심사)을 위해 수집된 것이므로 임상 징후나 중증도 정보 등이 상대적으로 부족하여 연구 활용에 한계가 있다. 마지막으로, 국정감사 등을 통해서 민간보험사예의 데이터 제공에 대한 이슈가 지속적으로 제기됨에 따라 향후 활용 범위와 규제가 변화할 수 있다는 불확실성도 남아 있다.

4. 가명정보 결합제도

가명정보 결합제도는 2020년 '데이터 3법' 개정을 통해 도입되었으며, 가명처리된 개인정보를 서로 다른 기관 간에 결합하여 통계작성, 과학적 연구, 공익적 기록보존 등의 목적에 활용할 수 있도록 허용하고 있다. 이 제도는 개인정보보호법 제28조의3을 기반으로 하며, 개인정보처

리자는 정보주체의 동의 없이도 결합전문기관을 통해 가명정보를 결합하고 분석할 수 있다. 상기 제도는 개인 식별이 불가능하도록 정보를 가명화한 상태에서 데이터의 연계 분석을 가능케 함으로써 개인정보 보호와 데이터 활용 간의 균형을 모색하였다.

해당 제도에서 중요한 점은 자료제공기관 외에 별도로 존재하는 결합전문기관이다. 현재까지 건강보험공단, 사회보장정보원 등 약 20개 이상의 기관이 결합전문기관으로 지정되어 있다. 이들 기관은 연구자 또는 기관이 신청한 가명정보 결합 절차를 수행하며, 결합 결과를 분석에 활용할 수 있도록 가명처리된 형태로 제공한다. 결합 대상은 각 자료제공기관이 보유한 가명정보이며, 공통 결합키(암호화된 개인식별자)를 생성하여 데이터 간 연계를 수행한다. 예컨대 보험사 고객정보와 사회보장정보원의 복지수급 데이터를 연계하면 소득·복지·건강 간의 종단적 분석이 가능하다. 앞서 언급한 사회보장 및 건강보험 행정데이터와 달리, 자료의 원천이 다양하여 통일된 코드북 등 표준화된 정보가 존재하지 않으나, 한편으로는 다양한 자료제공기관의 다양한 정보를 활용할 수 있는 가능성을 가장 획기적으로 열어둔 제도라 볼 수 있다.⁹⁾

세부 이용 절차는 상대적으로 복잡하다. 데이터 결합을 원하는 연구자나 기관은 우선 개별 자료제공기관으로부터 해당(행정)데이터 이용에 관한 승인을 얻어야 한다. 그 이후 개인정보보호위원회의 고시에 따라 결합신청서, 결합 목적 설명서, 개인정보 처리방안 등을 포함한 문서를 결합전문기관에 제출해야 한다. 결합전문기관은 제출된 문서를 바탕으로 결합 가능성을 사전 검토한 후 모의 결합을 실시하고, 실제 결합 작업을 수행한다. 이후 ‘반출심사위원회’의 승인을 거쳐 분석 가능한 형태의 자료를 제공하게 된다. 자료는 지정된 분석 공간 내에서만 접근이 가능하며, 외부 반출 시에는 재가명처리 또는 익명처리를 필수적으로 거치는 등 다소 엄격한 절차를 거쳐야 한다.

이는 데이터의 보안성과 개인정보 보호 측면에서는 발전된 조치이나, 동시에 연구자 입장에서는 데이터 활용을 저해하는 구조적 장벽으로 작용할 가능성도 있다. 연구자 입장에서는 우선 절차가 복잡하고 시간·비용 부담이 크다. 결합신청 시 제출서류가 많고, 개인정보보호 심사를 거쳐야 하므로 적지 않은 준비가 필요하다. 또한 결합 과정에서 가명식별키 정확성 확보가 어려워 결합률이 낮아질 경우에는 연구 상에서 특정 인구 집단이 배제되는 문제가 발생할 수 있다. 또한 개인정보 보호차원에서 결합 후에도 엄격한 재가명·반출심사가 요구되기 때문에, 연구자가 즉시 결과를 이용하기 어려워 연구의 진행속도가 상대적으로 느릴 수 있다는 단점이 존재한다.

III. 사회복지 행정데이터의 실천적 활용 과제

사회복지 분야에서 행정데이터의 체계적인 활용은 정책적 문제 해결과 서비스 개선에 실질적인 기여를 할 수 있다. 특히 행정데이터는 사회적 위험에 대한 선제적 대응, 지역 격차 해소, 공공-민간 협업 확대 등 다양한 수요에 대응할 수 있는 정책적 대안 마련의 강력한 수단으로 기능할 수 있다.

첫째, 행정데이터는 복지 위기가구 조기 발굴과 같은 선제적 대응 체계 구축에 활용될 수 있다. 이미 보건복지부에서 사회보장정보원의 행정데이터를 중심으로 연계하여 사업을 수행하고 있으나, 현재는 공변량이 높은 데이터들만 활용하여 예측력이 높지 않다는 단점이 존재한다. 따라서 중장기적으로 통신사·카드사 등 다양한 민간 데이터를 결합하여 소득 감소, 의료비 급증, 고용 불안정 등의 징후를 분석하여 복합 위험 가구를 조기에 탐지하고 적절한 개입 시행에 도움이 되는 방향으로 개편을 검토할 수 있다. 이와 관련된 사례로 영국 Greater Manchester의 Early Help System 등을 참고할 수 있다. 또한, 복지사업 수급 이력과 성과를 연계한 통합 데이터베이스 구축을 통해 정책효과 분석이 가능하며, 미국 Social Security Administration의 연구 사례는 이를 뒷받침하는 대표적 예다.

둘째, 지역 및 지자체 단위 분석을 통해 맞춤형 복지정책 설계가 가능하다. 행정데이터는 큰 표본 수로 인해 각 읍면동에 거주하는 개인들이 충분히 데이터에 포함되어 있다. 예를 들어, 복지 수급률, 빈곤율, 장애 등록률 등 주요 지표를 시각화한 지역 데이터 지도를 활용하면, 복지 사각지대와 중복지원을 직관적으로 확인할 수 있을 것이다. Harvard Opportunity Atlas는 이러한 지역 격차를 시각화된 도구로 표현한 대표적 사례이다. 또한 지역의 고령화율, 실업률, 출산율 등 다양한 변수와 행정데이터를 통합 분석하여 수요예측 모델을 고도화할 수 있으며, OECD의 Regional Well-Being Indicators는 성과지표 개발에 유용한 참고 자료가 된다.

셋째, 중장기적으로 공공-민간 연계를 통해 데이터 활용 범위를 확장할 수 있다. 민간 데이터(예: 통신, 금융)와의 결합을 통해 복합 취약계층을 보다 정밀하게 탐지하고, 맞춤형 복지서비스 설계가 가능해진다. 예를 들어, 협약을 맺은 사회서비스 스타트업에 공공·행정데이터 API를 제공하는 방식으로, 고령자 돌봄, 청년 금융, 저소득 건강관리 등 다양한 영역에서 민간의 혁신을 유도할 수 있다. 에스토니아의 X-Road 기반 민간 협업 생태계는 이러한 접근의 선도적 사례다. 궁극적으로는 데이터 기반의 맞춤

형 사회서비스 전달체계를 구축하여 행정의 효율성과 포용성을 동시에 제고할 수 있을 것이다.

IV. 사회복지 행정데이터의 활용 촉진을 위한 과제

앞서 살펴보았듯 사회복지 분야 행정데이터의 실천적 잠재력은 상당히 높은 수준이나, 일부 기관은 데이터의 완결성과 보안성에 대한 우려로 공개에 소극적인 태도를 보이고 있다. 이러한 우려는 일정 부분 타당한 측면이 있으나, 동시에 이미 안정적으로 구축되어있는 행정데이터 활용 가능성을 위축시켜 사회복지 영역의 실증 기반 확대를 저해할 수 있다. 따라서 정책적 활용과 개인정보 보호 간 균형을 유지하는 다양한 방안을 검토해 볼 필요가 있다.

첫째, 합성데이터(Synthetic Data)의 활용이 대안이 될 수 있다. 합성데이터는 실제 데이터를 바탕으로 통계적 특성을 유사하게 반영한 가상 데이터로, 개인정보 노출 위험 없이 데이터 분석을 가능하게 한다. AI 기술의 발전과 함께 대규모 학습데이터의 필요성이 증가하고 있으며, 합성데이터는 이러한 수요를 충족시키는 실효성 있는 개인정보 보호 강화 기술(Privacy Enhancing Technology)의 일종으로 부각되고 있다.¹⁰⁾ 국내에서도 개인정보보호위원회가 「합성데이터 생성·활용 안내서」를 발간하고 의료·금융 등 5대 분야에 대해 가이드라인을 제공하고 있다. 사회복지 분야 역시 주요 정보를 중심으로 합성데이터를 생성·개방함으로써 실증 연구의 기반을 마련하는 방안을 검토할 필요가 있다.

둘째, 사회보장 분야의 표준화된 표본 데이터 세트 구축이 필요하다. 예컨대 국민건강보험공단은 전수자료 일부를 추출한 표본 데이터를 연구자에게 제공하고 있으며, 이는 표준화된 구조로 설계되어 접근성과 활용성이 높다. 사회복지 분야에서도 전수 데이터를 기반으로 대표성을 갖춘 표본 데이터 구축이 가능하며, 이를 통해 정보보호와 연구 활용 간 절충점을 마련할 수 있다. 표본 데이터는 이용 절차가 간소화할 수 있고, 이용 이후 필요한 경우에 한하여 기관에서 맞춤형(user-driven) 데이터를 제공하게 할 수 있다는 점에서 데이터 수요자와 공급자 모두에게 유용하다.

셋째, 데이터 활용 초기단계에서는 전문가 워킹그룹 운영을 검토할 수 있다. 행정데이터는 설문조사에 비해 수집 목적이 다양하고 구조가 복잡하여, 해석과 활용에 어려움이 따르는 경우가 많다. 특히 코드북이나 설명 문서의 부재는 외부 연구자의 분석을 제약하며, 대부분의 데이터 오류가 데이터 기술방식(description)의 차이에서 비

롯됨에도 실제 행정의 오류로 오인될 가능성이 있어 현재 대다수 데이터 제공기관의 부담을 야기하고 있다. 다만 이러한 부담은 단순한 데이터 개방 연기가 아닌 공공기관·학계 연구소가 협업하여 데이터의 품질을 검토하고 해석 기준을 정립하는 워킹그룹 운영을 통해서 완화할 수 있다. 이를 통해 검증 효율성 증대, 데이터 문해도(literacy) 향상, 활용도 제고 등을 달성할 수 있으며, 실제로 미국 Census의 Administrative Data Pilots, OECD의 Social Data Innovation Program은 이러한 협업 모델의 긍정적 사례로 평가받는다.

V. 나가며

사회복지 영역에서 행정데이터의 활용은 단지 새로운 분석 수단의 도입을 넘어, 실증 기반 연구 혹은 접근이라는 관점에서 점차 중요성이 커지고 있다. 특히 AI 기술의 확산과 더불어, 데이터 중심의 접근은 공공성과 효율성을 동시에 제고할 수 있는 수단으로서 그 중요성이 더욱 부각되고 있다. 물론 AI, 데이터라는 도구가 가치 중심적 접근에서 태동된 사회복지 영역에 전가의 보도처럼 무비판적으로 활용될 수는 없는 한계는 분명 존재한다. 다만, 사회보장 행정데이터, 건강보험 빅데이터, 가명정보 결합 제도 등 최근 급속히 확장되고 있는 데이터 인프라는 사회복지 연구자들에게 보다 구조적인 시각에서 복지 현상을 분석할 수 있는 가능성과 설득력을 제공한다는 점 또한 부인하기는 어렵다.

다만, 실제 분석 과정에서는 여러 현실적인 제약이 존재한다. 행정데이터는 구축 및 연계 단계에서의 복잡성, 문서화 부족, 기관 간 코드 불일치 등의 한계를 내포하고 있으며, 이러한 점은 설문조사 기반 2차 자료에 비해 분석 설계와 수행에 상당한 시간과 노력을 요하는 원인이 된다. 또한 최근 연구자들이 처한 논문 실적 중심의 평가 환경을 고려하면, 이처럼 많은 준비와 시간이 소요되는 행정데이터 활용은 기대만큼의 즉각적인 연구성과를 보장하지 않는다는 점에서 부담이 되는 측면도 분명히 존재한다.

그럼에도 불구하고, 사회복지 연구자에게는 단순한 연구 성과 창출을 넘어, 미래 복지의 기초를 설계하고 사회적 위험으로부터 인간의 존엄성을 보호하는 방안에 대해 고민하는 책무가 주어져 있다고 생각한다. 이러한 맥락에서 행정데이터의 활용 가능성을 탐색하는 것은 복지제도의 지속 가능성을 모색하고 사회적 약자를 실질적으로 지원할 수 있는 정책 기반을 구축하는 데 중요한 출발점이 될 수 있다. 이러한 점에서 더 많은 연구자들이 장기적인

관점에서 행정데이터의 활용에 적극적으로 참여하고, 데이터 기반 사회복지학의 체계화에 동참해 주기를 소망한다.

머스크(Musk)가 구축의 필요성을 언급하기도 하였으며, 2030년까지 약 3조832억 규모로 성장이 예측되는 분야이다.

Notes

1) 본 원고는 학술적이고 구조적인 논의보다는 노인·장애인·저소득층 행정데이터를 구축·활용·분석한 과정에서의 필자의 개인적 경험과 문제의식을 공유하기 위한 취지에서 작성되었다.

2) 빅데이터(Big data)는 전통적 방식으로는 수집, 저장, 처리, 분석이 어려운 방대한 데이터를 의미하며, 일반적으로 방대한 데이터의 양(Volume), 빠른 생성 및 처리 속도(Velocity), 종류의 다양성(Variety)의 측면에서 기존 데이터와 차별화된다(Laney, 2001; 과학기술정보통신부, 2017).

Laney, D., 2001, 3D data management: Controlling data volume, velocity and variety, META Group.

과학기술정보통신부, 한국정보화진흥원, 2017, 빅데이터 이해와 활용.

3) 행정데이터는 정부 및 공공기관이 보유한 자료로서, 전국민을 포함하는 대규모 자료를 구조화된 형태로 축적한 것이 특징이다. 모집단 전체에 대한 분석이 가능하다는 점, 무응답이나 기억 오류에서 자유롭다는 점, 시계열 구조를 가질 수 있다는 점에서 설문조사 기반의 자료와 구별된다. 반면, 단일 기관의 자료는 분석 범위가 제한적이며, 자료 간 결합 없이는 활용에 한계가 존재한다는 단점도 존재한다(유종성, 2023).

유종성, 2023, 사회보장 행정데이터 활용사례와 향후 과제. 보건복지포럼 325: 6-19.

4) 데이터 3법은 개인정보 보호법, 정보통신망법, 신용정보법을 개정한 것으로, 가명정보의 개념을 도입하고 통계작성, 과학적 연구, 공익적 기록보존 등의 목적에 한해 정보주체의 동의 없이도 활용할 수 있도록 규정하고 있다. 이러한 제도 변화는 사회복지 분야를 포함한 사회과학 연구의 자료 활용 범위를 크게 확장시켰다.

5) 세부 사항은 보건복지부(2024), 이현주(2022)에 자세히 기술되어 있다.

보건복지부, 2024, 부처별 흩어진 행정데이터 연계, 사회보장 행정데이터 구축 사업 보고회 개최 [보도자료].

이현주, 2022, 사회보장제도 근거 강화와 사회보장 행정데이터 구축: 사회보장 행정데이터 연계·활용을 위한 과제, 보건복지포럼.

6) 위 이현주의 2022년 글.

7) 위 이현주의 2022년 글.

8) Park, I. (2022). How to use health insurance data effectively for healthcare research. Journal of Health Informatics and Statistics, 47(Suppl 2), S31-S39.

9) 물론 개별 자료제공기관으로부터의 신청과 승인을 받아야 하는 절차는 동일하므로 이용절차가 획기적으로 개선되었다고 보기 어렵다는 의견도 존재한다.

10) AI 성능 향상을 위해서 많은 양의 훈련데이터가 필요하나, 실제 법적 제약 등으로 데이터가 제한되는 현상을 극복하기 위하여 최근 AI 발전과 관련하여 테슬라의 일론