

AI 시대, 보행안전 데이터의 사각지대와 개선 방안

남궁지희
건축공간연구원 부연구위원

AI 전환과 정보 격차

불과 얼마 전까지만 해도 낯설고 먼 존재로 여겨지던 인공지능(AI)이 어느덧 우리의 일상에 깊숙이 파고들어 왔다. AI는 방대한 양의 정보를 단기간에 처리할 수 있고, 심화된 학습과 추론을 통해 적합한 결과물을 생성해냄으로써 인간 역량을 보완 또는 대체하는 도구로 막강한 잠재력을 지닌다. 이에 따라 단순히 기존과 동일한 업무를 더 빠르고 정확하고 효과적으로 처리하는 수준을 넘어, 업무를 처리하는 기본적인 절차나 체계 자체가 AI 활용에 적합한 형식으로 재구조화될 것이라는 예상이 제기된다.

AI의 상용화와 고도화를 위해서는 아직 풀어야 할 과제들이 많다. 특히 AI의 인지와 학습·추론의 기반이 되는 데이터 자체가 부족하거나 편향될 경우 불완전한 학습으로 인한 거짓 정보 생성이 왜곡된 의사결정으로 이어질 우려가 있다. AI는 복잡한 현실 세계의 상호작용을 있는 그대로 이해하는 것이 아니라 정형화된 데이터의 형식을 거쳐 받아들인다. 데이터의 영역에서 제대로 측정되고 관리되지 않는 요소들은 이후 학습과 의사결정 과정에서 원천적으로 배제되고, 양질의 데이터가 구축된 영역에서는 관련 지식과 기술이 계속 축적된다. 정보의 공백과 불균형을 직시하지 않고서는 ‘기울어진 공론장’을 바로잡기가 어렵다.

정책 현장에서 활용할 수 있는 데이터가 양적·질적으로 확장되고 있지만, 여전히 많은 사각지대와 제약 요인들이 남아 있다(남궁지희, 2023). AI 활용의 중요성과 기술적 의존도가 높아질수록 정보의 격차와 그에 따른 부정적 영향도 함께 커질 수밖에 없다. 이 글에서는 ‘보행안전’

보행안전 데이터의 사각지대와 한계

을 중심으로 관련 데이터의 사각지대와 문제점을 살펴보고, AI 활용 가능성과 개선과제를 제시하고자 한다.

2026년 2월 매사추세츠공과대학교(MIT) 연구진은 도시 단위에서 보행자의 통행을 체계적으로 분석할 수 있는 모형을 ‘최초’로 개발했다고 발표하였다(Sevtsuk et al, 2026). 기존의 보행 행태 분석 모형에서는 국지적이고 불완전한 보행 네트워크 자료와 제한적인 보행량 실측 데이터로 인해 모형의 검증과 보정이 어려웠으며, 실제 보행자의 선택과 경험에 영향을 미치는 다양하고 복합적인 요인들을 충분히 반영하지 못한다는 한계가 있었다(Zafri & Sevtsuk, 2026).

연구진은 뉴욕시 전역을 대상으로 보도와 횡단보도, 보행자 통로가 연결된 완전한 보행 네트워크를 구축한 후 구간별로 실측된 보행량 데이터를 활용하여 보행 유발 및 경로 선택의 요인을 반영한 분석기법을 정교화하였다. 이를 토대로 실제 측정 지점뿐 아니라, 도시 전역의 보행량을 추정할 수 있는 모형을 제시하였다(Dizikes, 2026). ‘최초’라는 표현이 다소 도발적으로 들릴 수 있지만, 이는 지금까지 도시 및 교통 분야의 논의가 차량 중심의 접근에 치우쳐 있었고, 보행자의 움직임을 이해하기 위한 데이터와 분석 도구는 상대적으로 부족하였다는 문제 인식을 압축적으로 보여준다.

우리나라의 상황도 이와 크게 다르지 않다. 유관 분야에서 데이터에 대한 투자 규모나 관련 기술의 발전 속도와 비교할 때 보행 관련 데이터의 기반은 아직 작고 미흡한 수준에 머물러 있다(남궁지희 외, 2024). 자동차의 경우 장기간 축적된 주행기록 데이터를 바탕으로 최적 경로와 도착 예상시간, 혼잡도와 사고위험 구간 등 운행에 필요한 각종 정보가 운전자에게 실시간으로 제공되지만, 보행자가 어디를, 얼마나, 어떻게 걷는지를 제대로 측정하려는 시도는 드물다. 교통량이나 통행 실태에 관한 정기적 조사에서도, 통행수단별 분담률을 산출할 때도 보행자의 존재와 역할은 생략된다.

공간정보나 지능형 교통체계의 고도화 과정에서도 보도나 이면도로의 보행환경은 우선순위에 벗어났다. 자율주행차량을 위해 정밀도로 지도를 제작하고 실시간 소통이 가능한 스마트 도로 인프라를 구축하는

시대이지만, 보행공간이 어디에 어떻게 조성되어 있는지, 기초적인 현황 조차 제대로 파악되지 않고 있다. 여기에 더해 보행자의 통행과 활동을 위한 공간이 충분한지, 서로 원활하게 연결되는지, 경로의 단절이나 우회 등으로 불편이 발생하는 곳은 어디인지, 안전이나 편의시설이 부족하거나 부적합한 곳은 없는지 등 보행환경과 경험의 질을 좌우하는 현장의 구체적인 상황에 대해서는 쓸 만한 데이터가 거의 전무하다고 볼 수 있다. 이런 상황에서 물리적 환경 및 시설 차원의 개선이 적재적소에 효과적으로 이루어지기는 어렵다.

보행자의 안전에 가장 기본적인 교통사고 데이터조차도 완벽하지 않다. 교통사고분석시스템(TAAS)에서 공개되는 ‘경찰DB’는 경찰에 공식 접수·처리된 교통사고에 한해 사고 발생 위치와 사고 유형 등 속성정보를 제공하며, 교통사고 통계의 기초자료에 해당한다. 심각한 인적 피해나 법규 위반사항 없이 보험사나 공제조합을 통해 처리된 사고는 통합DB로 관리하고 있지만 경찰DB와 공식 통계에는 포함되지 않는다.

2022년 자료를 기준으로, 경찰DB와 통합DB의 사상자 수를 비교해 보면 사망자의 100%, 중상자의 75.2%, 경상자의 24.7%가 집계되었는데, 가장 경미한 부상신고자의 경우 단 1.7%만이 경찰DB에 접수되었다. 또한 전체 사상자 중에서 공식 통계에 집계된 비율은 14.7%에 불과하였고, 집계되지 않은 사상자가 무려 165만 6,000명에 이르렀다.

이 밖에도 보험사를 거치지 않고 합의된 사고들, 실제 사고로 이어진 않았지만 아슬아슬하게 위험을 피한 상황들, 잠재적이고 상시적인 위험 요인들까지 고려한다면, 공식적인 교통사고 통계 데이터에는 평소 우

경찰DB와 통합DB 사상자 수 집계 결과 비교(2022년)

구분	경찰DB(A) ¹⁾	통합DB(B) ²⁾	집계율(A/B)	집계 외(A-B)
사망	2,735	2,735	100.0%	0
중상	51,715	68,794	75.2%	17,079
경상	212,430	859,239	24.7%	646,809
부상신고	17,658	1,009,752	1.7%	992,094
합계	284,538	1,940,520	14.7%	1,655,982

출처: 아래 원자료에 제시된 심각도 등급별 사상자 수를 비교하여 작성.
 1) 한국도로교통공단 교통사고분석시스템, 교통사고(경찰DB) - 사고일반 통계 - 2022년도 자료 조회 결과.
 2) 한국도로교통공단 내부자료(2024); 박경욱 외(2025, p.112 표 3-55)에서 재인용.

리가 길에서 체감하는 안전의 극히 일부만이 담겨 있다. 교통사고 데이터만으로 보행안전 수준을 평가한다는 것은 어찌 보면 지극히 상식적이고 당연한 선택 같지만, 실상은 겉으로 드러난 빙산의 일각을 보고 전체 모습을 상상해온 것과 크게 다르지 않다.

보행안전 정책을 위한 AI 활용 가능성과 개선 과제

보행안전을 위한 분석 모형에 AI를 활용한다면, 몇 가지 이점을 기대할 수 있다. 우선 AI는 대규모·대용량 데이터를 빠르고 효율적으로 처리할 수 있기 때문에 보행환경과 보행자의 행태를 더욱 정밀하게 관찰하여 고해상도의 시뮬레이션을 구축할 수 있고, 분석의 공간적·시간적 범위를 확장하여 도시 및 전국 단위에서 유효성을 검증하거나, 장기간에 걸친 시계열적 변화를 비교 추적하기에 용이하다. 또한 AI를 활용하면 고차원의 변수와 다양한 형식의 데이터를 융합 분석할 수 있다. 전통적인 통계분석에서는 변수 수와 형식에 기술적인 제약이 있는 데다 데이터 전처리와 사전 검토, 변수 선택과 중요도 평가, 변수 간 상호작용과 비선형적 관계를 탐색하는 과정을 연구자가 직접 수행해야 하는 반면, AI는 별도의 설정 없이도 자동 학습을 통해 적절한 구조와 패턴을 발견해 발견해낸다. 더 많은 데이터를 융합할수록, 현장 실증을 통해 오류를 줄여나가는 심화학습 과정을 반복할수록, AI 모형의 힘은 더욱 강력해진다. 이는 우리가 보행에 대해 이미 알고 있는 것들을 그대로 재현하는 수준에서 그치지 않고, 지금까지 드러나지 않았던 법칙이나 문제들을 미리 발견하고 새로운 관점이나 방향으로 나아갈 수 있도록 해준다(박훈태, 2025).

다차원 데이터 간의 관계를 동시에 유연하게 탐색하는 접근 방식은 복잡한 도시환경과 행태적 특성, 각종 위험요인이 얽혀 있는 보행안전의 구조를 파악하는 데 유효하다. 예를 들어 환경적인 조건이 동일하다면, 보행량과 교통량이 많고 서로 자주 노출될수록 사고 발생 확률이 높아진다. 따라서 만약 특정 가로에서 보행자 교통사고가 줄어들었다면 더 안전한 환경으로 개선되었기 때문인지, 아니면 유동인구가 줄어들었기 때문인지를 구분할 필요가 있다. 앞서 언급한 MIT의 연구에서 뉴욕시 전역에서 가장 많은 사고가 발생하는 곳은 맨해튼의 중심부이지만, 보행량에 따른 노출 수준을 보정한 결과 맨해튼 주변과 외곽 지역에서 상대적으로 위험성

이 높은 구간들을 추가로 발견할 수 있었다(Sevtsuk et al, 2026). 고정적인 환경요인 외에 가변적인 보행량과 그에 따른 노출 변수를 고려함으로써 중심지에 집중되어 있던 관심과 자원을 실제 개선이 필요한 지역에 배분할 수 있게 된 것이다.

물론 종속변수에 해당하는 교통사고 데이터와 잠재적 영향요인에 해당하는 환경적·행태적 속성 데이터가 모두 불완전한 상황에서 AI를 활용한 분석과 추론 및 예측의 결과를 어디까지 신뢰할 수 있는지의 문제는 남아 있다. 경험적으로 사고가 많이 발생하는 구간들에서 공통의 속성과 경향을 도출하는 것까지는 유용하지만, 그 이상을 기대하기에는 아직 무리가 있다. 구간별 속성이나 공통점에 대한 판단은 모형에 투입된 데이터와 변수들의 한계를 넘어설 수 없기 때문이다. 결국 AI 기반 보행안전 분석 모형의 타당성과 활용성을 개선하기 위해서는 다음과 같은 과제가 선결되어야 한다.

첫째, 공간정보 융합분석의 기반이 되는 표준화된 네트워크 모듈이 필요하다. 보행자가 통행 가능한 실제 경로를 노드와 링크 형식으로 구현한 ‘보행자길 네트워크’ 레이어는 보행 관련 공간정보 구축의 근간에 해당한다. 이를 토대로 보행자의 이동경로를 연속적으로 파악할 수 있고, 보행환경이나 행태 관련 다층적인 속성정보를 결합할 수 있다. 따라서 보도와 횡단시설, 이면도로와 단지 내 보행로 등 보행공간의 유형이나 관리주체가 다르더라도 하나의 통합된 네트워크로서 정제된 데이터 규격과 핵심 속성 정의, 좌표와 코드체계 표준화 등 일관된 기준을 정립할 필요가 있다. 이러한 표준 네트워크 데이터는 일종의 공공 인프라로서 새로운 모형 개발이나 학습 및 검증에 활용될 수 있도록 개방해야 한다. 이를 통해 보행 관련 연구와 정책의 투명성과 호환성을 확보하고 관련 지식과 경험이 효과적으로 축적·공유되는 선순환 구조를 기대할 수 있다.



보행자 교통사고 발생 현황(2007~2023)

시를 활용한 위험도 예측 결과

출처: 이면도로 보행환경 웹지도 시스템(n-streets.kr, 검색일: 2026.4.1.)에서 강남역 주변 조회 결과.

둘째, 모형의 신뢰도를 높이기 위해 AI가 학습할 수 있는 고품질 원천 데이터의 지속적인 확보와 안정적인 검증 체계 구축이 필수적이다. 특히 보행량, 교통량, 통과속도, 보차상충 빈도 등과 같은 핵심 변수들에 대해서 실제 관측된 값이 아닌 대리변수(proxy)나 추정값(imputed data)에 의존할 경우 입력값에 내재된 가정이 반복적으로 학습·재생산되면서 특정 패턴이 과도하게 강조되고 오차나 편향·왜곡이 증폭될 위험이 있다. 따라서 기존에 구축된 공간정보와 빅데이터를 최대한 활용하되, 현장 실측이 필요한 항목에 대해서는 표준화된 프로토콜을 마련해야 한다. 아울러 스마트센서·CCTV·모바일 데이터 등 쉽고 빠르고 저렴한, 또는 고성능·고효율의 조사기법과 장비 등이 더욱 다양하게 개발·보급될 필요가 있다.

나아가 유사한 조사항목 간의 유기적 연계, 기초 및 결과 데이터의 공유 활용을 통해 개별 조사 수행 비용을 줄이고 정책적 활용도를 높이는 방안이 마련되어야 한다. 다양한 출처와 형식의 데이터 처리 과정에서 발생하는 오차나 누락, 편향을 점검할 수 있도록 품질관리체계도 구축되어야 한다. 이는 학습 과정의 정확도를 높일 뿐 아니라, 결과물의 신뢰성과 검증 가능성을 확보하는 기반이 된다.

셋째, 모형의 설명력을 좌우하는 설명변수의 다변화와 내실화가 필요하다. 도면·이미지·영상 등 현장에서 수집한 원본 자료에 보행환경과 행태의 다차원적인 속성과 맥락 정보가 담겨 있더라도, 기계학습이 가능한(machine-readable) 형태로 식별 및 변환되지 못하는 정보는 존재하지 않는 것과 마찬가지이다. 거리 이미지나 항공사진에 컴퓨터 비전, 자연어 처리 등 AI 기법을 활용하여 보도 폭, 가로 활성화, 차로 구성, 가시성 등 물리적 환경 요소를 지표화하는 등 원 데이터에서 환경·행태 지표를 추출할 수 있는 기술적 기반은 고도화되고 있다. 이제는 AI에게 무엇을, 어떻게 학습시킬지 과제와 방향을 명확히 설정하고, 그 의미와 효용을 평가할 수 있도록 적절한 질문을 던지는 능력이 중요하다(박훈태, 2025). 예를 들어 동영상에서 보행자를 식별한다면 단순히 몇 명인지 세는 것에서 그치지 않고, 보행자 간의 간격이나 밀도, 보행속도와 궤적, 우회나 지체 여부, 회피나 상충의 빈도와 같은 특정한 의도를 담은 지표들을 다각적으로 정의함으로써, 보행자의 움직임의 원리나 경험의 실체에 더 가깝게 다가갈 수 있다.

또한 기계가 읽을 수 없는 분석 대상에 대한 세밀한 이해와 해석, 통찰이 요구되는 정책적 판단이나 창의적 대안이 필요한 영역에서는 ‘사람’의 관점과 역할이 더욱 중요해질 것이다(남궁지희, 2023). 그만큼 지표의 의미가 원래의 공간적·사회적 맥락에서 동떨어지지 않도록 실제 이용자와 전문가의 의견, 현장 경험과 안전에 미치는 영향 등을 균형 있게 고려할 필요가 있다.

종합해 보면 표준화된 보행 네트워크 모듈, 학습용 실측 데이터의 체계적 축적과 검증, 설명변수의 다변화와 내실화를 통해 지속가능한 데이터 인프라를 구축하고 분석 및 검증 과정이 유기적으로 연계되는 기반을 마련하는 것이 핵심 과제라 할 수 있다. 이를 통해 보행안전 정책에서 AI의 도입과 활용이 단지 조사분석의 도구로써 기술적인 편의성이나 효율성, 정확도 향상의 수준을 넘어 ‘사고 중심, 사후 대응 위주’의 정책에서 종합적이고 정교한 분석을 통해 ‘사전 예측과 예방, 선제적 관리 중심’의 정책으로 전환하는 계기가 되기를 바란다.

참고문헌

- 1 남궁지희. (2023). 보행 안전 정책과 데이터의 진화. 안전충북포커스, 2023 15호, 30-39.
- 2 남궁지희, 최가윤, 채정은, 강수연. (2024). 데이터 기반 보행정책 활성화를 위한 공공 데이터 현황과 개선과제. 건축공간연구원.
- 3 박경욱, 우승국, 심재익, 조민경, 이동윤. (2025). 사람의 생명가치를 고려한 교통사고비용 추정방법론 개선 연구. 한국교통연구원.
- 4 박훈태. (2025). AI 기술 변화의 시대, 보행 행태와 공간을 이해하는 새로운 방법. 건축과 도시공간 2025년 가을호, 15-24.
- 5 이면도로 보행환경 웹지도 시스템. n-streets.kr(검색일: 2026.4.1.)
- 6 한국도로교통공단 교통사고분석시스템. 교통사고(경찰DB) - 사고일반 통계 - 2022년도 자료 조회 결과. https://taas.koroad.or.kr/sta/acs/exs/typical.do?menuId=WEB_KMP_OVT_UAS_ASA(검색일: 2025.9.20.)
- 7 한국도로교통공단 내부자료. (2024). 사고 심각도별 사고 건수 및 사상자 수(2022년).
- 8 Dizikes, P. (2026.2.6). I'm walking here! A new model maps foot traffic in New York City. MIT News. <https://news.mit.edu/2026/new-model-maps-foot-traffic-new-york-city-0206> (검색일: 2026.3.10.)
- 9 MIT City Form Lab. (n.d.). Projects - NYCWalks. <https://cityform.mit.edu/projects/nycwalks>(검색일: 2026.3.10.)
- 10 Sevtsuk, A., Basu, R., Liu, L., Alhassan, A. & Kollar, J. (2026). Spatial distribution of foot traffic in New York City and applications for urban planning. *Nature Cities*, 3, 136-145. <https://doi.org/10.1038/s44284-025-00383-y>
- 11 Zafri, N. M., & Sevtsuk, A. (2026). Advancing Pedestrian Models: A Comparative Review and Vision for the Future. *Journal of the American Planning Association*, 1-18. <https://doi.org/10.1080/01944363.2026.2618643>