도로정책 Brief+

04
April 2024
No. 165

이슈

- 도로·교통 분야 AI의 활용과 도전 과제

해외포커스

- 거대언어모델(LLM)과 도시/교통 연구
- 자율주행 기술 고도화를 위한 국내외 자율주행 데이터셋 현황

숫자로 보는 도로정책

- 국민들은 정책 의사결정과 수행 과정에 얼마나, 어떻게 참여하고 싶어할까

국토교통부 KRIHS 국토연구원

이슈

도로·교통 분야 AI의 활용과 도전 과제

심지섭
국토연구원 부연구위원
gis.up@krihs.re.kr

인공지능 기반 도로·교통의 미래

인공지능(AI)의 활용은 도로·교통 분야에서도 혁신적 변화를 가져오고 있다. 도로와 교통수단을 이용하는 다양한 이용자, 공급자, 관련 분야 연구자들까지도 미래의 도로·교통 시스템을 인식하고 활용하는 방식이 변화하고 있으며, 그 기저에는 빅데이터와 AI 기술 개발이 자리하고 있다. AI 기술은 도로·교통 분야의 두 가지 주요 목표인 교통 효율성 향상과 사고 예방에 큰 역할을 하며, 지속 가능한 미래로의 전환을 돕는 핵심적인 기술이다. 나아가 AI는 스마트 도로 및 인프라 구현에도 필수적인 요소이며, 차량-도로 간 통신을 통한 다양한 서비스를 개발하는 데 필수적인 요소이다. 이에 본고에서는 지속 가능한 도로·교통 운영 관리를 위해 AI를 도입한 다양한 정책 및 기술·서비스 사례를 살펴보고, AI 기술의 활용에 있어 주의해야 할 부분과 도전 과제들을 고찰하고자 한다.

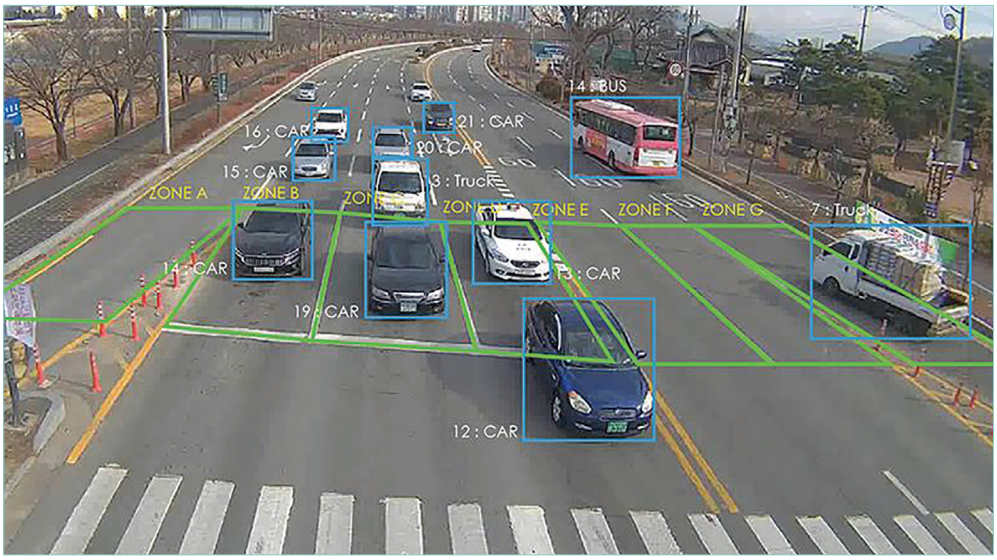
교통 운영 효율성 향상을 위한 AI 활용

교통 운영 효율화의 첫 단계는 도로를 주행하는 자동차의 현황, 즉 교통량을 파악하는 것이다. 교통량 조사는 도로교통량 조사지침에 따라 상시조사를 실시하거나 조사원을 통한 수시조사를 통해 수행된다. 교통량 조사의 조사 항목은 방향별·시간대별·일별 교통량을 12종의 차종으로 구분하여 조사하는 것이 원칙이며, 기존의 조사 방식은 조사 시기에 따라 결과의 차이가 발생할 수 있고 시간과 비용이 많이 소요된다는 단점이 존재한다.

이에 행정안전부에서는 2022년 10월부터 '인공지능 기반 CCTV 영상인식 교통량 분석모델' 프로젝트를 추진했다. 이 프로젝트에서는 전국 도로에 설치된 CCTV 영상을 바탕으로 교통량과 차종을 자동으로 파악할 수 있는 데이터 분석 모델을 개발하였으며, 도로교통량 조사 기준인 12종의 차종과 차량이 통과하는 차선의 위치를 포함한 차선·차종별 교통량을 자동 집계한다. 모델 개발을 위해 51개 CCTV에서 수집된 약 20TB의 영상 데이터를 활용하였고, 차종 분류 모델의 정확도는 98.7%에 이르는 것으로 나타났다. 개발된 모델은 행정안전부의 '범정부 데이터 분석 시스템'에 탑재되어 지자체를 비롯한 여러 수요 기관에서 활용할 수 있도록 표준화되었으며, 분석모델을 개별

기관에서 직접 활용이 가능한 형태로 배포하여 사용성을 높였다. 기존 교통량 조사의 경우 전국 3,900여 지점에서 육안 조사 등을 통해 수행되었으며, 지점에 따라 일일 조사 비용이 수십만 원에 이르렀다. AI 기반의 교통량 조사를 통해 연간 수십억 원 이상의 비용을 절감할 수 있을 뿐만 아니라, 조사 시점·기상 상황 등 여러 변수를 고려할 수 있는 신뢰도 높은 교통량 데이터의 확보를 기대할 수 있게 되었다.

행정안전부 CCTV 영상 분석 기반의 차선·차종별 교통량 집계 예시



자료: 대한민국 정책브리핑(2023)

교통량 조사 자동화 이외에도, AI를 활용해 교통 운영의 효율성을 향상시킬 수 있는 정책과 기술이 지속적으로 등장하고 있다. 도시 내 혼잡을 줄이기 위해 대규모의 교통량 자료를 기반으로 신호·경로를 최적화하거나, 긴급차량 우선순위·대중교통 우선신호 등 데이터 기반의 의사결정을 수행하기도 한다. 이를 위해 정부 부처, 연구기관, 민간 기업들에서는 기존 도로 인프라와 인공지능 기술을 활용해 비용을 절감하고 효율성을 높이기 위한 다양한 정책과 기술을 개발하고 있다.

국토교통부에서는 2021년부터 '스마트시티 챌린지' 사업을 통해 도시 전역의 스마트화를 위한 종합 솔루션 구현 사업을 추진했다. 그 중 대구광역시는 'AI 기반 도심교통 서비스'를 통해 AI 기반의 교통량 조사, 신호 제어, 차량 분산 유도 등 다양한 교통 운영 효율화 서비스를 개발해 왔다. 2021년 예비사업 수행 이후, 2022년부터 2024년까지 본사업을 추진 중이며, 사업 수행 주체인 대구광역시와 함께 한국과학기술원(KAIST), 카카오모빌리티, EMG 등 다양한 기관이 산·학·연 거버넌스를 구성하여 참여하고 있다. 해당 사업에서는 1) AI 기반의 최적 경로 안내 서비스, 2) 차량 신호등 잔여시간 표출 서비스, 3) 보행자 신호등 잔여시간 표출 서비스, 4) 교통약자 보호구역 우회경로 안내 서비스, 5) 긴급차량 이동경로 확보를 위한 우회경로 안내서비스 등 스마트 교차로 영상 자료와 민간 내비게이션 데이터를 이용한 다양한 서비스를 개발한다. 이와 같은 AI 기반 스마트 교통체계 구축을 통해 연간 500억 원 이상의 사회적 비용을 절감할 수 있을 것으로 기대하고 있으며, 향후 해당 서비스를 해외로 확산하는 계획 역시 수립되어 있다.

국토교통부-대구광역시 'AI 기반 스마트 교통체계' 서비스 예시

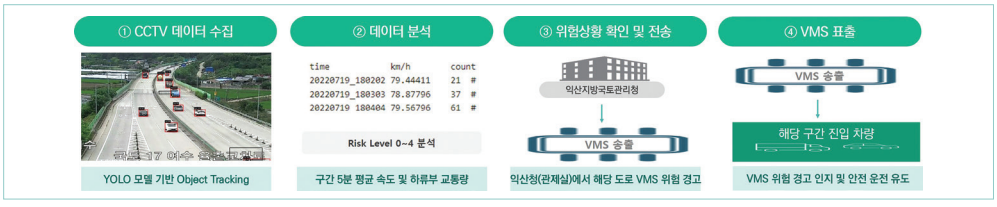
자료: 국토교통부 스마트시티 종합 포털

AI로 지키는 교통 안전

도로·교통 분야 AI의 활용은 교통 운영에만 국한되지 않는다. AI의 장점은 학습된 모델과 데이터를 기반으로 미래 상황에 대한 예측이 가능하다는 점이며, 데이터의 양이 많아질수록, 모델의 성능이 높아질수록 더 신뢰도 높고 객관적인 결과를 보여준다. 특히 교통 사고의 경우 인적 요인으로 인한 사고의 발생 비율이 매우 높으며, 이에 따라 위급 상황 시 AI의 도움은 사고 예방과 안전에 큰 도움을 줄 수 있다.

국토교통부와 한국교통안전공단에서는 2022년부터 'T-Safer'라는 이름의 AI 기반 교통사고 위험도 예측 시스템을 개발하여 시범 적용한 바 있다. 해당 시스템에서는 다양한 공간정보 데이터와 교통 정보를 이용해 도로 구간별 위험도를 예측한다. 또한 CCTV, 노변 검지기, GPS 등 다양한 출처로부터 수집된 데이터를 기반으로 교통상황을 분석하고, VMS를 통해 운전자에게 위험 정보를 제공한다.

국토교통부-한국교통안전공단 'T-Safer'



자료: 국토교통부 보도자료

아울러 교통 법규 위반 단속을 위해 AI 기술을 활용하는 사례도 있다. 경찰청 및 도로교통공단에서는 AI 영상 분석을 통해 기존의 단속 장비로 단속이 어려웠던 오토바이의 교통 법규 위반 행위를 단속하는 시스템을 구축하였으며, 여러 지자체를 필두로 스마트 인프라 구축을 통한 사고 예방 대책을 적극적으로 수립하고 있다. 일부 지역에서는 보행자 안전을 위해 CCTV 영상에서 보행자의 움직임을 실시간으로 감지하고 분석하여 돌발 상황을 인식하고, 운전자·보행자에 위험 정보를 제공하는 시스템을 구축하여 교통사고 발생률을 크게 줄인 바 있다.

도로·교통 분야 AI 활용의 도전 과제

AI 기술은 교통 효율성 향상과 사고 예방이라는 두 마리 토끼를 놓치지 않을 수 있도록 도와주는 훌륭한 도구이다. 그러나 AI 기술의 잠재력을 최대한 효과적으로 활용하기 위해서는, 몇 가지 주의해야할 점과 도전 과제가 남아있다.

먼저, AI의 근간은 대용량의 데이터라는 점을 상기해야 한다. 대규모 데이터에는 민감정보가 포함될 수 있으므로, 데이터·개인정보 보호에 대한 철저한 조치가 필수적이다. 특히 교통 데이터의 경우 개인의 위치 데이터나 이동 패턴 등 민감한 정보가 포함되므로, 데이터 처리와 AI 기술 활용에 있어 각별한 주의를 요한다. 최근에는 이러한 문제를 극복하기 위해 연합학습(Federated Learning)과 같이 개인정보 보호에 장점을 갖는 AI 모델의 개발과 활용도 논의된 바 있다.

AI 기반의 의사결정에 있어서도 많은 고민이 필요하다. 자율주행 차량의 위험 상황 시 판단 기준에 대한 윤리적 고려나, 기술 오류 시의 책임 문제 등 사회적 합의를 필요로 하는 영역이 AI의 교통 분야 적용에 있어 필연적으로 존재한다. 이는 도로·교통이 우리의 안전과 생활에 매우 깊숙이 닿아있기 때문이다. 따라서 앞으로는 지속적인 법·제도 보완을 통해 도로·교통 분야 AI 활용의 사회적 합의점을 찾을 필요성이 있다. 이를 위해서는 정책 입안자, 기술 개발자, 이용자 간의 긴밀한 협력과 대화가 꼭 필요하다.

AI의 등장은 인류가 불을 처음 발견했을 때의 순간과 흡사하다. 새로운 도구의 발견이 인류에게 새로운 가능성을 열어준 것처럼, AI도 인류가 지금껏 도달하지 못했던 영역까지의 혁신을 약속하고 있다. 그러나 불이 그러하듯, AI도 그 강력함에 따른 위험을 내포하고 있을 수 있다. AI를 통한 막대한 혜택을 누리기 위해서는, 지속적인 연구개발 노력을 수행함과 동시에 그 힘을 책임감 있게 관리할 수 있는 방법을 찾아야 할 것이다.

참고문헌

1. 국토교통부 보도자료. 2022. 인공지능(AI) 활용, 교통사고 위험 예측한다. https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmspage=1&id=95087076
2. 국토교통부 스마트시티 종합포털. https://smartcity.go.kr/프로젝트/스마트-챌린지/스마트시티-챌린지/2021-예비사업 (2024년 3월 20일 검색)
3. 대한민국 정책브리핑. 2023. CCTV 영상 AI로 분석해 차종·교통량 파악. 4월 18일. https://www.korea.kr/news/policyNewsView.do?newsId=148913997
4. 동아일보. 2023. 보행자 나타나면 AI가 조명-경고등: 어르신 밤길 안전 지킨다. 5월 1일. https://www.donga.com/news/Society/article/all/20230430/119081747/1
5. 심지섭. 2023. 교통 빅데이터 활용 시 개인 정보 보호를 위한 연합학습 기반의 경로 선택 모델링. 한국ITS학회논문지. 제22권 제6호.
6. 전자신문. 2021. 오토바이도 무인단속추진: 후면번호판, '카메라+AI'로 단속. 1월 10일. https://www.etnews.com/20210108000102
7. 행정안전부 보도자료. 2023. 전국 도로교통량 조사에 인공지능(AI) 본격 활용. https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=104016#none

해외포커스

거대언어모델(LLM)과 도시/교통 연구

여 지 호
가천대학교 조교수
jihoyeo@gachon.ac.kr

생성형 AI의 등장과 발전

최근 몇 년간 인공지능(AI) 기술은 비약적인 발전을 거듭해 왔다. 특히 2023년 이후로 등장한 거대언어모델(Large Language Model, 이하 LLM)은 자연어 처리 분야에서 혁신을 불러일으키고 있다. LLM은 방대한 양의 텍스트 데이터를 학습하여 인간과 유사한 수준의 언어 이해와 생성 능력을 보여준다. 이러한 LLM의 발전은 도시 및 교통 분야의 연구에도 큰 영향을 미치고 있다. 본고에서는 LLM이 도시/교통 연구에 어떻게 활용되고 있는지, 그리고 앞으로의 발전 방향에 대해 살펴보고자 한다.

데이터 분석 및 연구 패러다임의 전환

초기 생성형 AI는 주로 가상의 데이터를 생성하고, 데이터의 결측치를 보정하는 데 사용되었다. 예를 들어, GAN(Generative Adversarial Network)은 두 개의 신경망(생성자와 판별자)을 경쟁적으로 학습시켜 실제 데이터와 유사한 가상의 데이터를 생성한다. 이는 데이터가 부족한 상황에서 모델 학습을 위한 데이터를 확장하는 데 유용하게 활용될 수 있다. 또한 오토인코더(Autoencoder)나 변분오토인코더(Variational Autoencoder) 등의 기법은 데이터의 저차원 표현을 학습함으로써 결측치를 복원하거나 노이즈를 제거하는 데 사용될 수 있다.

하지만 최근 다양한 LLM 모델이 등장하면서 생성형 AI를 활용하는 방식이 크게 변화하고 있다. 본질적으로 Generative AI 모델의 근간이 되는 Transformer를 사용하는 것은 같으나, GAN을 활용한 연구는 이미지, 합성 데이터, 그래프 등을 생성하는데 초점을 맞추었다면, LLM을 활용한 연구는 기존에 서비스되고 있는 LLM 모델의 API를 직접 사용하거나 커스터마이즈하여 활용하는 경향이 있다.

LLM은 방대한 양의 텍스트 데이터를 사전학습한 언어모델로서, 문장 생성, 질의응답, 텍스트 분류 등 다양한 자연어처리 태스크에 범용적으로 활용될 수 있다. 최근에는 LLM을 활용하여 도시/교통 분야의 비정형 데이터(뉴스 기사, 소셜미디어 포스트, 정책 문서 등)를 분석하거나, 전문가 지식을 모사하여 의사결정을 지원하는 연구들이 활발히 진행되고 있다. 그 중 이러한 LLM 모델을 도시/교통 분야에 활용하기 위한 다양한 사례를 소개하고자 한다.

LLM을 활용한 다양한 시도

LLM은 도시/교통 분야의 다양한 연구 영역에서 활용되고 있다. Urban Design 분야에서는 LLM을 활용하여 도시 계획안에 대한 시민들의 의견을 수집하고 분석하는 연구가 진행되고 있다. UrbanistAI는 생성형 AI를 활용하여 도시계획 과정에서 시민참여를 혁신하고자 한다. 전통적인 설문조사 방식의 한계를 극복하고, 시민들이 직접 도시 공간을 설계하는 과정에 참여할 수 있도록 실시간 이미지 생성 및 수정 기능을 제공한다. 시민들의 아이디어를 바탕으로 도시 개선안을 빠르게 시각화하고 반복적으로 수정해 나가는 협력적 설계가 가능해진 것이다. 아래 그림은 UrbanisAI(https://urbanistai.com/)가 제공하는 Generative AI를 사용하여 생성한 헬싱키 여름 거리 이미지 예시이다. 초기 이미지는 왼쪽에 그늘막과 녹지가 있는 일반적인 거리 풍경이다. 하지만 생성형 AI를 통해 점진적으로 변화를 주면서 다양한 시도를 하고 있는 것을 볼 수 있다. 이처럼 UrbanistAI의 생성형 AI는 시민들의 아이디어를 즉각적으로 시각화하고 공유함으로써, 도시 설계에 대한 직관적 이해와 활발한 소통을 가능하게 한다.

이러한 접근 방식은 헬싱키, 베를린, 클리블랜드, 프리스티나 등 전 세계 여러 도시의 공공 워크숍에서 활용되고 있으며, 생성된 이미지들은 전문 설계자들에게 영감을 주는 무드보드로 활용되고 있다. UrbanistAI는 최근 두바이 정부와 협력하여 도시의 3D 모델을 계획 법규에 따라 수정할 수 있는 생성형 AI 모델을 개발하고 있다. 이를 통해 두바이는 생성형 AI 모델을 학습시키는 최초의 도시가 될 전망이다. 편향 제어, 악용 방지, 지적재산권 보호, 데이터 프라이버시 확보 등의 과제는 남아있지만, 생성형 AI는 창의적이고 혁신적인 방식으로 우리의 삶의 공간을 변화시킬 수 있는 잠재력을 보여주고 있다.

Generative AI를 통해 생성한 다양한 가로 이미지

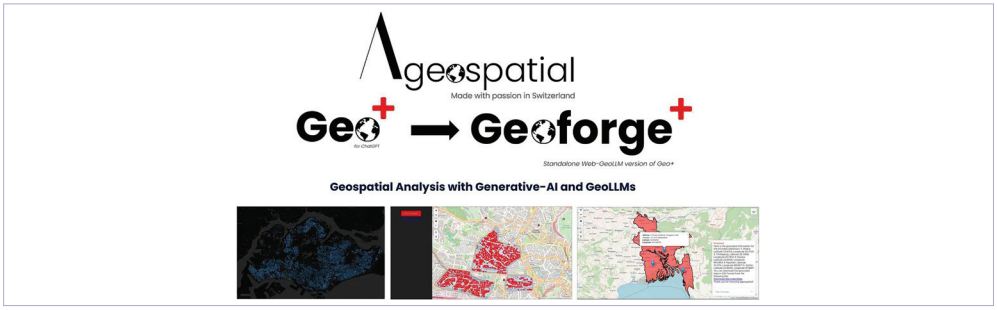
자료: https://urbanistai.com/

Spatial Data Analysis 분야에서는 위치정보가 포함된 데이터를 효과적으로 가공, 처리, 분석할 수 있는 플랫폼에 대한 관심이 높아지고 있다. 그 중에서도 GeoForge는 최근 가장 주목받는 플랫폼 중 하나이다. GeoForge는 OpenAI의 GPT-4를 기반으로 개발된 GeoLLM(Geospatial Large Language Model)을 활용하여, 일반 사용자도 쉽게 지도를 제작하고 공간 분석을 수행할 수 있도록 지원한다.

GeoLLM은 기존의 LLM과 달리, 지리공간 도메인에 특화된 어휘와 지식을 사전학습함으로써 공간 질의에 보다 정확하게 응답할 수 있다. 예를 들어, "서울에서 인구밀도가 가장 높은 지역을 찾아줘"와 같은 질문에 대해, 관련 데이터를 바탕으로 적절한 답변을 제시할 수 있다.

이러한 접근 방식은 기존의 공간 데이터 분석 과정을 크게 간소화하고 자동화할 수 있다는 장점이 있다. 사용자는 복잡한 GIS(Geographic Information System) 도구를 직접 다루지 않고도, 자연어로 문제를 기술하고 원하는 결과를 얻을 수 있게 된다. 또한 GeoLLM은 다양한 출력 형태(지도, 차트, 리포트 등)를 지원하기 때문에, 분석 결과를 효과적으로 커뮤니케이션하는 데에도 도움이 된다.

ChatGPT 기반의 공간데이터 분석 플러그인



자료: https://www.ageospatial.com/

WANG et al.(2023)은 생성형 AI를 활용한 자동화된 토지이용계획 생성 프레임워크를 제안하였다. 핵심은 토지이용계획을 어떻게 데이터로 관리하고 표현할 것인지 인데 위도, 경도, POI 카테고리로 구성된 3차원 텐서를 제시하였다. 또한, 본 연구에서는 대화형 AI 기반 Human-in-the-loop 방식의 도시계획 프레임워크를 제안하였다. 전문가의 요구사항을 자연어로 입력받아 도시계획안을 생성하고 전문가 피드백을 통한 반복적 개선이 가능한 청사진을 그렸다. 하지만, 이 논문은 방법론만 제시하였고, 구체적인 코드나 소프트웨어를 공개하지는 않았다. 아직은 아이디어나 구상 단계에 머물러 있는 것을 알 수 있다.

한계와 발전방향

LLM은 Text-to-Text, Text-to-Code, Text-to-Image, Text-to-Video 등 다양한 영역으로 발전하고 있다. 하지만 도시/교통 계획에 직접 활용되기 위해서는 보다 체계화된 구조를 가질 필요가 있다. 현재까지는 주로 데이터 가공, 처리, 분석, 시각화 등에 LLM이 활용되고 있으며, 고차원적인 해석이나 정책적 의사결정에는 아직 제한적으로 사용되고 있다.

LLM을 도시/교통 연구에 활용함에 있어 오픈소스 기반의 분석 방법론과 소프트웨어가 중요해질 것으로 보인다. 실제로 현재 제공되는 GeoAI 서비스들도 Python의 geopandas, QGIS 등을 중심으로 구현되고 있다. 웹에 존재하는 데이터로 학습하는 LLM의 특성상, 오픈소스 기반의 도구들이 상용 소프트웨어에 비해 LLM과의 호환성이 더 높을 것으로 예상된다.

향후 LLM을 도시/교통 연구에 적용할 때는 두 가지 접근 방식이 가능할 것이다. 하나는 Lamma 등의 공개된 코드를 활용하여 도메인 지식에 맞는 데이터로 직접 LLM을 학습시키는 것이고, 다른 하나는 이미 상용화된 LLM 모델을 활용하는 것이다. 전자의 경우 도메인에 특화된 모델을 구축할 수 있다는 장점이 있는 반면, 대규모 컴퓨팅 자원과 전문 지식이 필요하다는 단점이 있다. 후자의 경우 별도의 학습 과정 없이 손쉽게 LLM을 활용할 수 있지만, 도메인 특화 성능은 상대적으로 떨어질 수 있다. 저자 역시 어떤 방식이 도시/교통 분야에 더 적합할지 확신하기 어려우며, 모든 가능성을 열어두고 접근할 필요가 있다고 생각한다.

글을 마치며

지금까지 LLM이 도시/교통 연구에 미치는 영향과 활용 사례, 그리고 발전 방향에 대해 간략히 살펴보았다. LLM은 도시/교통 분야의 데이터 분석 및 연구 패러다임에 큰 변화를 가져오고 있으며, 다양한 영역에서 혁신적인 활용 사례들이 등장하고 있다. 하지만 LLM을 맹신하거나 무분별하게 사용하는 것은 경계해야 한다. LLM이 제시하는 결과를 비판적으로 검토하고, 도메인 지식에 기반하여 해석하는 연구자의 역량이 여전히 중요하다.

또한 최근에는 LLM으로 생성된 텍스트를 판별하는 기술도 등장하고 있다. 학술 연구의 진실성과 책임성을 담보하기 위해, LLM을 활용한 연구 결과 제시 시 투명하고 엄격한 기준이 마련되어야 할 것이다.

LLM은 도시/교통 연구에 많은 기회와 가능성을 제공하고 있지만, 동시에 신중하게 접근해야 할 과제들도 안고 있다. 연구자들은 LLM의 장점을 최대한 활용하되, 그 한계를 인지하고 보완해 나가야 한다. LLM과 연구자의 지혜로운 협업을 통해 도시/교통 분야의 혁신과 발전이 이루어지기를 기대해 본다.

참고문헌

1. Nelson, RJ. Naing, KM. Beroche, H. (2023). Generative AI for Urban Governance. Urban AIhttps://urbanai.fr/wp-content/uploads/2023/10/Generative-AI-Report.pdf
2. Xu, M., Niyato, D., Chen, J., Zhang, H., Kang, J., Xiong, Z., ... & Han, Z. (2023). Generative AI-empowered simulation for autonomous driving in vehicular mixed reality metaverses. IEEE Journal of Selected Topics in Signal Processing.
3. Yan, H., & Li, Y. (2023). A Survey of Generative AI for Intelligent Transportation Systems. arXiv preprint arXiv:2312.08248.
4. Wang, D., Lu, C. T., & Fu, Y. (2023). Towards automated urban planning: When generative and chatgpt-like ai meets urban planning. arXiv preprint arXiv:2304.03892.
5. “SmartCityGPT”: How Generative AI Creates Smart and Sustainable Cities”https://nonsmartcity.medium.com/smartcitygpt-how-generative-ai-creates-smart-and-sustainable-cities-4d00ce73da10
6. “GeoForge: Geospatial Analysis with Large Language Models (GeoLLMs)”
7. https://medium.com/@ageospatial/geoforge-geospatial-analysis-with-large-language-models-geollms-2d3a0eaff8aa
8. “The Power of Generative AI in Urban Planning: Text2Map Revolution”https://medium.com/urban-ai/the-power-of-generative-ai-in-urban-planning-text2map-revolution-38544f02fb29

해외포커스

자율주행 기술 고도화를 위한 국내외 자율주행 데이터셋 현황

김형주
차세대융합기술연구원 경기도자율주행센터 선임연구원
hyungjoo@snu.ac.kr

서론

정부는 '2030 미래차 산업 발전 전략'을 통해 27년 전국 주요 도로에서 레벨 4 자율주행차 상용화를 목표로 하고 있다. 이를 성공적으로 달성하기 위해서는 자율주행 차량의 학습과 성능 향상을 위한 학습 데이터셋이 중요하다. 자율주행은 라이다, 카메라, 레이더 등 다양한 센서 기반으로 주변 상황을 인지·판단하고 제어함으로써 차량 스스로 목적지까지 운행하는 기술로 각 단계별로 레벨 0(비자동화), 레벨 1(운전자 보조), 레벨 2(부분 자동화), 레벨 3(조건부 자율주행), 레벨 4(고도 자율주행), 레벨 5(완전 자율주행)까지 구분된다(Society of Automotive Engineers). 각 레벨의 차이점은 주로 제어의 주체와 주행 책임에서 구분되며 레벨 0~2는 운전자가 주행의 주요 책임을 지고 시스템은 주행을 보조하는 역할을 한다. 반면, 레벨 3부터는 차량이 주행의 상당 부분 또는 전부를 자동으로 처리할 수 있으며, 레벨이 올라갈수록 인간 운전자의 개입이 점점 더 줄어들게 된다.

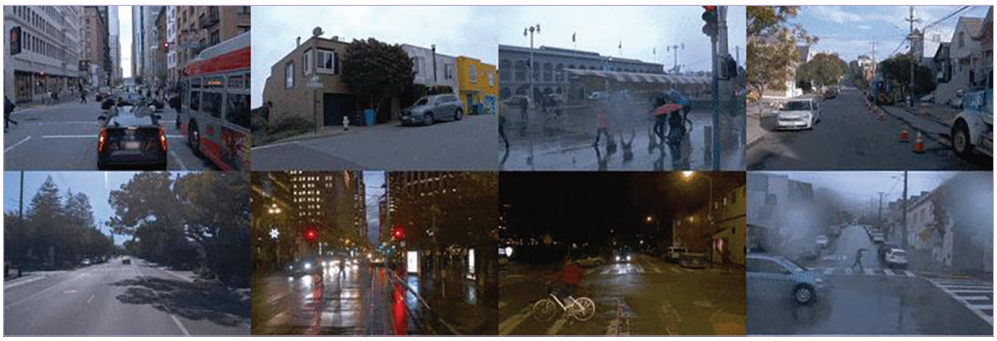
이렇듯 자율주행 기술 레벨이 고도화됨에 따라 차량이 복잡한 도로 환경에서 더욱 안전하고 효과적으로 주행할 수 있도록 다양한 센서 기반의 인공지능 기술이 요구된다. 라이다, 카메라, 레이더 등의 센서는 차량 주변 환경의 정밀한 정보를 수집하여, 차량 주변 환경을 인지하고 적절한 주행 결정을 내린다. 이러한 고도화된 사물 인식 기능을 개발하고 최적화하기 위해서는 방대한 양의 학습 데이터셋이 필요하다. 학습 데이터셋은 자율주행 차량의 인공지능 시스템이 인지, 판단, 제어 등의 과정을 통해 학습하고 개선할 수 있도록 다양한 환경과 주행 조건에서 수집된 데이터의 집합이다. 이러한 데이터셋은 주로 차량의 센서에 의해 수집되며, 차량과 주변 환경에 대한 정보를 포함한다. 대표적으로 차량위치, 주변차량 및 보행자위치, 도로상황, 교통신호상태, 기상조건 등이 포함된다. 따라서, 본고에서는 국내외 자율주행 데이터셋에 대한 사례조사를 토대로 자율주행 차량의 학습과 성능 향상을 위한 학습 데이터셋의 중요성을 제시하고자 한다.

해외 자율주행 데이터셋 사례

대표적인 해외 자율주행 학습 데이터셋으로는 Waymo Open Dataset이 존재한다. 알파벳 자회사인 웨이모가 2019년에 공개한 자율주행 데이터셋은 25개 도시에서 약 1,000만 마일에 달하는 자율주행 테스트를 진행하며 데이터를 수집하였다.

데이터셋에는 인지 데이터와 모션 데이터로 구성되어 있으며, 인지 데이터는 10Hz로 수집된 각 20개의 세그먼트 2,030개로 구성되어 있다. 모션 데이터의 경우 10Hz에서 20초 동안의 객체 트랙과 세그먼트가 포함하는 영역에 대한 지도 데이터를 포함하는 103,354개의 세그먼트(2천만 프레임 이상), 574시간의 데이터로 구성되어 있다. 이러한 연속주행 영상은 다양한 도로 환경에서 객체의 행동을 추적하고 예측하는 모델을 개발하는 데 활용될 수 있다. 특히 Waymo 데이터셋을 활용하여 20~23년까지 Waymo Open Dataset Challenge를 개최하여 자율주행 인지·판단·제어 분야별 기술 고도화에 기여하고 있다.

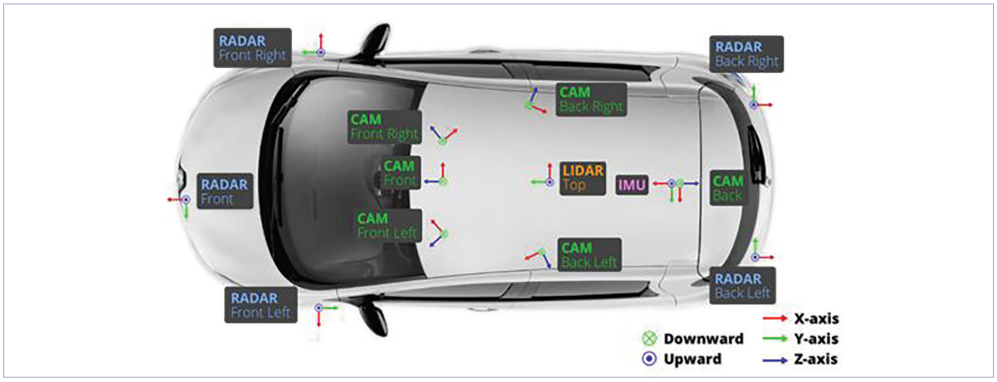
다양한 시공간적 범위를 포함하는 Waymo Open Dataset



자료: https://waymo.com/open/

두 번째로 Motional에서 개발한 nuScenes 자율주행 데이터셋은 3D 객체 주석을 갖춘 대규모 공개 자율주행 데이터셋이다. 데이터셋에는 Motional 테스트 차량으로 교통이 혼잡한 미국 보스턴과 싱가포르 도로구간에서 수집한 140만 개의 카메라 이미지, 39만 개의 라이다 정보, 140만 개의 레이더 정보, 140만 개의 object bounding box가 포함되어 있다. 주행 장면은 센서 당 2Hz로 수집된 20초 영상으로 물체 단위로 23개의 클래스와 8개의 속성으로 주석이 달려 있으며 다양한 교통상황과 위험상황을 포함한다. 수집된 데이터셋은 교통량이 많고 위험상황이 많이 포함되어 복잡한 시나리오 상황을 연구하기 위해 사용되고 있다. 특히 nuPlan Planning Challenge를 개최하여 nuScenes 데이터셋을 활용한 자율주행 차량별 인지·판단·제어 기술을 평가하여 자율주행 시스템의 안전성 검증에 기여하고 있다.

Motional 데이터 수집 차량



자료: https://www.nuscenes.org/nuscenes

경기도 판교 자율주행 데이터셋 사례

자율주행 인공지능 학습 데이터셋의 국내 대표사례는 경기도 판교제로시티 자율주행차 시범운행지구를 들 수 있다. 판교 자율주행 데이터셋은 국내 기업 및 연구기관의 자율주행 기술 개발에 공공데이터로 활용되고 있으며 국내 자율주행 산업생태계 조성 및 기술 상용화에 기여하고 있다. 국내 최초 자율주행 실증단지인 판교제로시티는 제1, 2판교 테크노밸리 지역으로 일반차와 자율주행차가 혼재되어 운영되고 있다. 특히 통합관제센터를 중심으로 자율주행 차량과 도로인프라 정보 공유를 위한 통신환경이 구축되어 있으며, 대표적으로 신호현시 서비스, 보행자케어 서비스, 자율주행 도로감시 서비스, 도로환경 감시 서비스 등이 포함된다.

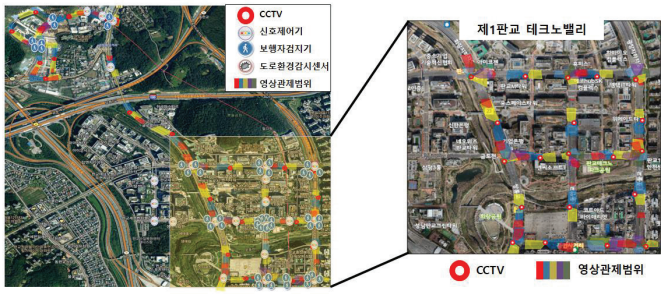
자율주행 데이터셋 구축을 위한 차량플랫폼은 제로셔틀, 스타렉스 개조차량, 판타G버스 등이 있으며 각 차량플랫폼은 라이다, 카메라, 레이더 등의 센서가 장착되어 다양한 도로환경, 교통상황, 날씨조건 등의 판교 데이터셋을 구축 중이다. 인프라 관점에서는 CCTV, 신호제어기, 보행자검지기, 도로환경감시센서 등이 구축되어 실시간으로 데이터를 수집하고 있다.

판교 자율주행 차량 데이터 수집

구분	자율주행 제로셔틀	스타렉스 개조차량	자율주행 판타G버스
차량			
실증내용	일반인 탑승 운행 및 자율주행 실증 연구	자율주행 데이터 수집 및 국산센서 개발지원	일반인 탑승 운행
센서구성	LiDAR(16ch 1개, 32ch 1개, 2D 4개), Radar 2개, Camera 4개 등	LiDAR(16ch 4개, 32ch 1개, 128ch 1개), Radar 2개, Camera 6개 등	LiDAR(16ch 1개, 32ch 4개), Radar 2개, Camera 8개 등

자료: 경기도자율주행센터

판교 자율주행 인프라 데이터 수집



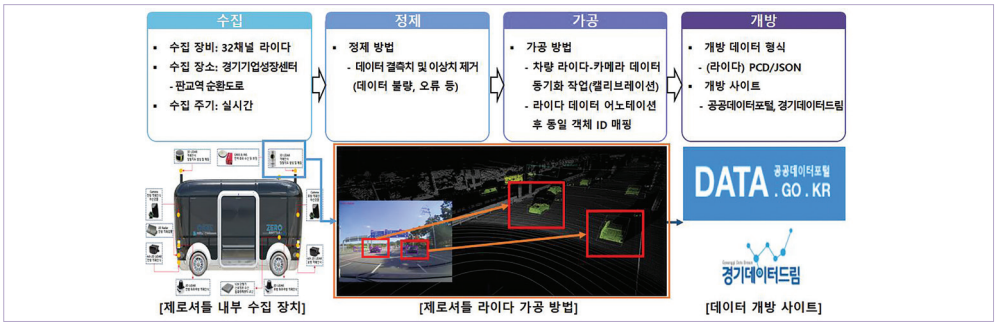
자료: 경기도자율주행센터

※ 경기도 판교 자율주행 데이터 개방 현황 : 약 549만 건 데이터 개방

- **2021년 개방데이터 약 4.5만 건**
  - 공공데이터 포털(https://www.data.go.kr/) 약 4.5만 건
- **2022년 개방데이터 약 215만 건**
  - 공공데이터 포털 약 190만 건
  - 공간융합 빅데이터 플랫폼(https://bigdata-geo.kr/) 약 25만 건
- **2023년 개방데이터 약 330만 건**
  - AI Hub(https://www.aihub.or.kr/) 약 39만 건
  - 공간융합 빅데이터 플랫폼 약 291만 건

판교 자율주행 데이터셋은 판교제로시티에서 수집되는 자율주행차 및 인프라 센서데이터를 가공 및 표준화하여 공공데이터로 개방되고 있다. 차량 관점에는 라이다, 카메라, GPS/INS, 객체위치인식정보, 차량제어정보, 차량상태정보, 기본안전메세지 등이 포함되며, 수집/정제/가공 과정을 통해서 공공데이터로 개방된다. 예를 들어 라이다 데이터 가공과정에서는 차량 라이다-카메라 데이터 캘리브레이션 및 어노테이션 작업 등이 포함된다. 인프라 관점에는 CCTV, 라이다, 레이더, 도로노면 감시정보, 자율주행 도로감시 서비스 결과, 보행자케어 검지내역 등이 포함된다. 인프라 데이터의 경우도 수집/정제/가공 과정을 거치며 CCTV 데이터 내 개인정보 비식별화와 객체 어노테이션 작업 등이 포함된다. 특히 판교 자율주행 데이터를 활용한 5개 도전분야(2D 객체검지, 3D 객체검지, CCTV 혼잡도 분류, 자율주행차량 경로추정, 자유주제)에 대한 경진대회('23.9-12월)를 개최하여 실제 현장데이터를 활용하는 경험과 자율주행 기술에 대한 이해를 심화시키는데 기여했다.

제로셔틀 라이다 데이터셋 구축 과정



자료: 경기도자율주행센터

결론

자율주행 차량이 실도로의 복잡하고 예측 불가능한 도로환경에서 안전하게 주행하는데 신뢰할 수 있고 다양한 시나리오를 포함하는 학습 데이터셋은 필수적이다. 자율주행 기술 고도화를 위한 다양한 도로 및 교통상황, 위험상황 등에 대한 데이터셋이 요구됨과 동시에 자율주행 시스템의 학습 및 테스트에 필요한 데이터의 양과 품질이 점차 중요해지는 시점이다. 특히 국내에서 부족한 실도로 환경의 도로종류별/교통상황별 위험상황 데이터셋 구축과 오픈소스 데이터셋의 접근성 개선, 그리고 경진대회를 통한 자율주행 데이터 활용도 제고 등이 필요하다. 이를 위해서 정부, 산업체, 연구기관 등이 협력하여 자율주행 데이터셋의 수집, 구축, 공유, 활용방안 수립이 요구된다.

참고문헌
1. 미국 자동차공학회 SAE(Society of Automotive Engineers) J3016
2. Waymo Open Dataset, https://waymo.com/open/
3. nuScenes 자율주행 데이터셋, https://www.nuscenes.org/nuscenes
4. 경기도자율주행센터, https://ggzerocity.or.kr/

숫자로 보는
도로정책

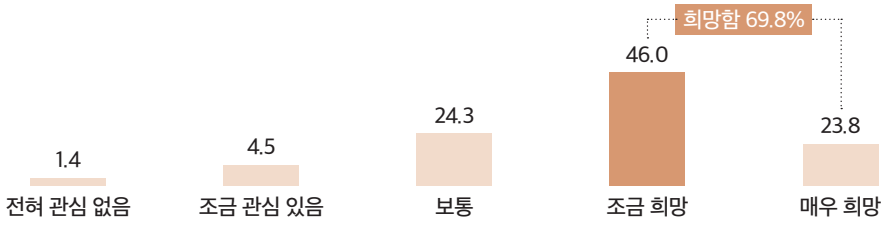
국민들은 정책 의사결정과 수행 과정에 얼마나, 어떻게 참여하고 싶어할까

김 민 영
국토연구원 전문연구원
mkim@krihs.re.kr

조사 개요

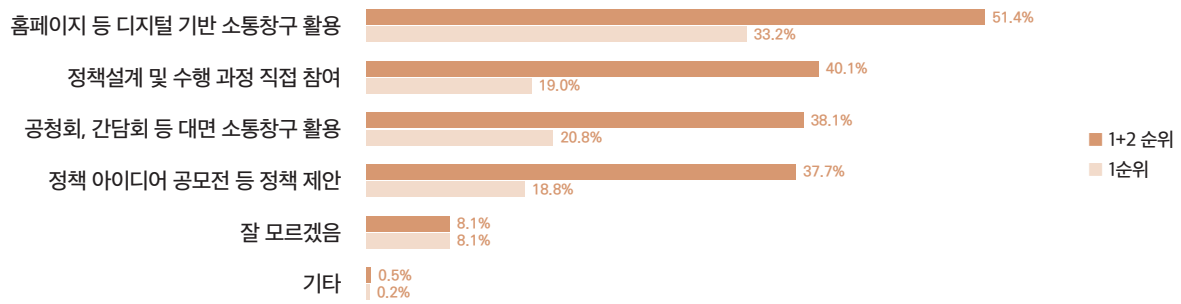
- **조사명**: 국토 정책분야 주요 이슈 일반국민 여론조사
- **대상(표본수)**: 전국 만19~69세 일반국민 1,000명
- **조사기간**: 2023.12.15.~2023.12.20
- **조사방법**: 구조화된 설문지를 이용한 온라인 조사
- **수행기관**: 국토연구원 / 조사기관: 엠브레인퍼블릭

공간정책 관련 참여 희망 정도



- 공간정책 관련 정책 참여 희망 정도를 질문한 결과, '매우 희망'(23.8%), '조금 희망'(46.0%) 등 10명 중 7명 가량이 '희망한다'(69.8%)고 응답한 반면, '전혀 관심없다'는 대답은 1.4%에 불과하여 정책 참여 열기는 대체로 높은 것으로 보임

희망하는 정책 참여 방법



- 정책 참여를 희망한다는 응답자(986명)를 대상으로 희망 참여 방법을 질문한 결과, '홈페이지 등 디지털 기반 소통창구 활용'(51.4%)이 가장 높게 나타났으며, '정책설계 및 수행 과정 직접 참여'(40.1%), '공청회, 간담회 등 대면 소통창구 활용'(38.1%), '정책 아이디어 공모전 등 정책 제안'(37.7%) 등의 순으로 나타남(1+2순위 응답 기준)

응답자 연령대별·직업별 희망 정책 참여 방법

(단위: %, 1+2순위 응답 기준)

- 60대를 제외한 모든 응답자 연령대별, 직업별 결과에서, '홈페이지 등 디지털 기반 소통창구 활용'이 가장 응답 비율이 높았음
  - PC, 스마트폰 등 디지털기기 이용률이 높은 만큼 디지털 기반 소통창구를 활용한 참여 희망 비율이 높았으나, 60대의 경우 여전히 '직접 참여' 또는 '대면 소통창구를 활용'한 참여를 더 선호하는 것으로 나타남
  - 향후 디지털 리터러시(digital literacy) 격차가 줄어들고 고령층도 디지털 환경에 익숙해지면 일반국민의 정책 참여를 유도하는 데 디지털 기반 소통창구의 활용은 더욱 중요해질 것으로 판단됨
- 디지털 기반 소통창구 이외에, 40대 이하 연령층과 직업별로 화이트칼라, 학생, 프리랜서 그룹에서 '정책 아이디어 공모전 등 정책 제안'을 통한 참여 희망 비율이 상대적으로 높게 나타남
  - 이들은 도입을 검토 중이거나 추진 중인 정책에 대해 의견을 제시하는 수준을 넘어, 자신의 아이디어가 정책화되기를 희망하는, 보다 적극적인 참여를 원하는 것으로 해석할 수 있음
  - 이러한 적극적인 참여 의사가 디지털 기반 소통창구와 결합하면 더 큰 시너지를 발휘할 수 있을 것으로 기대됨

국토연구원 홈페이지(www.krihs.re.kr)
홈페이지를 방문하시면 도로정책Brief의 모든 기사를 볼 수 있습니다.
홈페이지에서 회원가입을 하시면 메일링서비스를 통해 도로정책Brief를 받아 볼 수 있습니다.

• 발행처 | 국토연구원 • 발행인 | 심교언
• 주소 | 세종특별자치시 국책연구원로 5 • 전화 | 044-960-0269 • 홈페이지 | www.krihs.re.kr

※ 도로정책Brief에 수록된 내용은 필자 개인의 견해이며 국토교통부나 국토연구원의 공식적인 견해가 아님을 밝힙니다.

OPEN 출처표시 상업용금지 변경금지
공공누리 공공저작물 자유이용허락