

# End-to-End 자율주행 연구 동향 및 발전 방향: 실험 환경 및 평가 방법을 중심으로

이용재<sup>1)</sup> · 장한빈<sup>2)</sup> · 임준희<sup>3)</sup> · 김동석<sup>\*1)</sup>

한국과학기술원 초천식모빌리티대학원<sup>1)</sup> · 한국과학기술원 기계공학과<sup>2)</sup> · 한국과학기술원 로봇공학학제전공<sup>3)</sup>

## End-to-End Autonomous Driving Research Trends and Future Directions: Focused on Experimental Environments and Evaluation Methods

Yongjae Lee<sup>1)</sup> · Hanbin Jang<sup>2)</sup> · Joonhee Lim<sup>3)</sup> · Dongsuk Kum<sup>\*1)</sup>

<sup>1)</sup>Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon 34051, Korea

<sup>2)</sup>Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

<sup>3)</sup>Robotics Program, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

(Received 23 October 2025 / Revised 4 November 2025 / Accepted 6 November 2025)

**Abstract :** End-to-End(E2E) autonomous driving systems have recently emerged as a significant research trend to overcome the limitations of conventional modular architectures. The research on E2E autonomous driving systems involves experimental environments and evaluation methods, all of which play a crucial role in determining the overall performance of an algorithm. Therefore, it is important to establish appropriate environments and evaluation methods that can accurately reflect real-world performance. This study analyzed recent trends, and proposed the future directions of E2E autonomous driving systems. To achieve this, datasets, simulators, and worldmodels are examined in experimental environments, while evaluation metrics are discussed in developing suitable evaluation methods.

**Key words :** Autonomous vehicle(자율주행 자동차), End-to-End system(종단간 시스템), Dataset(데이터셋), World model(월드모델), Simulator(시뮬레이터), Evaluation(평가)

### 1. 서론

기존 자율주행 기술은 인지, 예측, 계획, 제어로 구성된 모듈형 시스템을 기반으로 발전해왔다.<sup>1)</sup> 이러한 시스템은 각 모듈마다 개별적인 학습 목표들을 가지기 때문에 전체 시스템이 일관된 목표를 향해 학습하지 못한다. 또한 모듈 간 정보 전달 과정에서 데이터의 손실이 발생할 수 있으며, 이는 최종 제어 모듈에까지 누적되어 오류를 유발할 수 있다. 이러한 문제점은 복잡한 주행 환경에서 시스템의 안정성을 저해하는 요인으로 작용한다. 이를 극복하기 위해 제안된 접근 방식이 End-to-End(E2E) 자율주행 시스템이다. E2E 시스템은 센서 입력으로부터 경로 또는 제어 명령을 직접 산출하는 통합 학습 구조를 가지고 있다. 이러한 구조는 중간 모듈 간 정보 변환 과정 없이 통

일된 목표로 학습할 수 있기 때문에 누적 오류가 적고 학습과 추론의 효율성이 높다는 장점이 있다. 따라서 최근

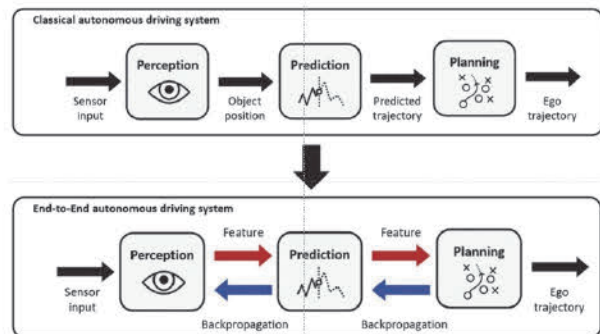


Fig. 1 Trend shift from classical system to end-to-end system

\*Corresponding author, E-mail: [dskum@kaist.ac.kr](mailto:dskum@kaist.ac.kr)

<sup>1)</sup>This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

자율주행 시스템은 Fig. 1과 같이 모듈형 시스템에서 E2E 시스템으로 변화되는 추세이며, E2E 시스템은 최근 학계와 산업계에서 중요한 연구 흐름으로 자리매김하고 있다.<sup>2)</sup>

E2E 자율주행 연구는 알고리즘의 학습 및 검증이 이루어지는 실험 환경과, 성능을 정량적으로 분석하기 위한 평가 방법을 기반으로 수행된다. 이 두 요소는 알고리즘의 성능을 결정짓는 핵심 요인으로, 동일한 알고리즘이더라도 적용된 환경이나 평가 방법에 따라 결과가 크게 달라질 수 있다. 따라서 객관성, 현실성, 실효성을 갖춘 실험 환경과 평가 방법을 설정하는 것이 연구의 신뢰성을 확보하기 위한 핵심 과제라 할 수 있다.

실험 환경은 데이터셋, 시뮬레이터, 그리고 최근 주목 받고 있는 월드모델로 나눌 수 있다. 데이터셋은 실제 주행 과정에서 취득한 다양한 센서 정보를 기반으로 구성된 것으로 모델의 학습과 평가에 활용되는 현실 기반의 실험 환경이다. 시뮬레이터는 가상의 도로, 차량, 보행자, 기상 조건 등을 인공적으로 재현하여 자율주행 알고리즘을 안전하고 반복적으로 실험할 수 있는 환경이다. 월드모델은 실제 환경의 동역학과 시각적 변화를 학습하여 새로운 상태나 장면을 예측하고 생성할 수 있는 학습 기반의 실험 환경이다. 따라서 데이터셋은 현실성을, 시뮬레이터는 다양성과 안전성을, 월드모델은 확장성에 장점을 가지며, 이 세가지 요소가 E2E 자율주행 연구의 핵심적인 실험 환경을 구성한다.

평가 방법은 Open-loop와 Closed-loop 방법으로 나눌 수 있다. Open-loop 평가 방법은 모델이 출력한 값을 실제 측정 한 값(정답 데이터)과 비교하는 방법이다. 이 과정에서 모델의 출력은 차량의 움직임에 반영되지 않기 때문에

환경과 상호작용하지 않는다는 특징을 가진다. 반면, Closed-loop 평가 방법은 모델이 출력한 값을 차량 제어에 반영하여 주행시킨 뒤, 주행 결과를 기반으로 모델의 성능을 측정하는 방법이다. 모델의 출력이 차량의 움직임에 반영되기 때문에 환경과 상호작용한다는 특징이 있다. 이러한 특징 때문에 Closed-loop 평가 방법은 Open-loop 평가 방법보다 실제 주행상황을 더 정확하게 반영할 수 있다. 최근에는 실제 데이터로도 Closed-loop와 유사하게 평가할 수 있는 비반응형(Non-reactive) Closed-loop 평가 방법이 제시되며 새로운 평가 패러다임이 확산되고 있다. 이러한 여러 평가 방법에 적용되는 평가지표는 모델의 성능을 정량적으로 측정하는 핵심 기준으로, 어떤 지표를 적용하느냐에 따라 모델의 우수성 평가 결과가 달라질 수 있다. 따라서 평가 방법의 특징을 고려하여 공정하고 표준화된 평가 지표를 마련하는 것이 중요하다.

본 연구에서는 E2E 자율주행 연구의 실험 환경과 평가 방법을 종합적으로 분석하고, 현재 기술의 한계와 향후 보완 방향을 제시한다. 또한, 연구자들이 새로운 알고리즘을 설계하고 평가할 시 참고할 수 있는 기초 자료를 제시한다.

## 2. End-to-End 자율주행 연구의 실험 환경

E2E 자율주행 시스템의 학습 및 평가는 데이터셋, 시뮬레이터, 그리고 월드모델로 구성된 환경에서 이루어진다. 알고리즘은 주어진 환경 내에서만 학습되고 평가되기 때문에, 이는 학습 효율성과 성능 평가에 중대한 영향을 미친다. 따라서 빠른 학습과 높은 현실성을 제공하기 위해 고도화된 실험 환경들이 제시되고 있다.

Table 1 Comparison of representative datasets, simulators

Type	Name	Multi-agent	GPU accel	Pre-constructed 3D environment	Real image	State prediction method	Image rendering method
Data set	nuScenes, <sup>4)</sup> nuPlan, <sup>5)</sup> Waymo open dataset <sup>6)</sup>	X	X	O	O	Log replay	Log replay
	CARLA Leaderboard <sup>13)</sup>	X	X	O	X	IDM	Unreal engine
	Bench2Drive <sup>14)</sup>	X	X	O	X	RL	Unreal engine
Simulator	CARLA <sup>15)</sup>	X	X	O	X	IDM	Unreal engine
	MetaDrive <sup>19)</sup>	O	O	O	X	RL	Pandar3D
	Waymax <sup>20)</sup>	O	O	X	X	IDM	X
	Gigaflow <sup>21)</sup>	O	O	X	X	RL	X
	GPUDrive <sup>22)</sup>	O	O	X	X	RL	X
	Scenario dreamer <sup>23)</sup>	O	O	X	X	Diffusion	X
	DRIVEARENA <sup>25)</sup>	X	X	X	O	MCTS	Diffusion
	NAVSIM <sup>78)</sup>	O	X	X	O	Log-replay / IDM	X / 3DGS
NVIDIA DRIVE Sim <sup>27)</sup>	X	X	O	X	IL	Omniverse RTX	

## 2.1 데이터셋

데이터셋은 실제 차량을 주행하여 취득한 데이터들의 집합으로 자율주행 알고리즘의 학습과 평가를 위해 초기부터 활용되어 온 현실 기반의 실험 환경이다. 특히 E2E 자율주행 모델은 이미지나 라이다 등 다양한 센서 입력을 기반으로 동작하기 때문에, 이와 같은 정보들이 데이터셋의 핵심 구성 요소로 작용한다. 이러한 데이터셋은 모델의 일반화 성능을 향상시키기 위해 다양한 시나리오를 반영하는 방향으로 발전하고 있다.

일반적으로 데이터셋은 현실 기반과 시뮬레이션 기반으로 구분된다. 현실 기반 데이터셋은 실제 차량을 운행하여 수집된 데이터셋으로 높은 현실성을 제공하지만, 다양한 환경에서의 데이터 확보가 어렵고 수집 비용이 크다는 한계가 있다. 반면 시뮬레이션 기반 데이터셋은 가상 환경에서 자율주행 알고리즘을 운용해 데이터를 생성하므로, 안전하게 다양한 시나리오 확보가 용이하나 현실과의 센서 값 또는 주행 환경에 대한 간극이 존재한다.<sup>3)</sup>

대표적인 E2E 자율주행 시스템을 위한 현실 기반 데이터셋으로는 nuScenes, nuPlan, Waymo open dataset<sup>46)</sup>이 있다. 이들 모두 센서가 장착된 차량이 실제 도심 환경을 주행하면서 취득한 데이터로 구성되어 있으며, 사용된 센서의 종류, 개수, 배치 방식 등에 따라 차이를 보인다. nuScenes는 초기엔 인지 및 예측 알고리즘 개발을 목적으로 공개되었지만, 이후 경로 계획 및 판단에 필요한 정보가 추가되면서 현재는 E2E 연구에서 가장 널리 활용되는 데이터셋이 되었다. 단, 직진 위주의 주행 시나리오가 다수를 차지하여 도출된 결과의 신뢰성에 한계가 있을 수 있다. nuPlan은 E2E 벤치마크 중 하나인 NAVSIM<sup>78)</sup>의 기반이 되는 데이터셋으로, 이미지 데이터를 제외한 모든 센서 정보가 공개되어 있으며 이미지 정보는 순차적으로 업로드 되고 있다. 일부 연구에서는 사용되는 데이터 및 시나리오에 따라 val14,<sup>9)</sup> test14,<sup>10)</sup> test14-inter<sup>11)</sup> 등으로 구분하여 알고리즘을 평가한다. 마지막으로 Waymo open dataset은 2025 Vision-based End-to-End Driving Challenge<sup>12)</sup>에 활용되는 데이터셋으로, 현실에서 발생하기 어려운 비정상적 상황의 데이터를 다수 포함하여 알고리즘의 일반화 성능과 강건성을 향상시키는 데 유용하다.

한편, 시뮬레이션 기반 데이터셋으로는 CARLA Leaderboard<sup>13)</sup>와 Bench2Drive<sup>14)</sup>가 대표적이다. 이들 모두 CARLA<sup>15)</sup> 시뮬레이터 환경에서 특정 자율주행 알고리즘을 실행해 수집된 주행 데이터로 구성되며, 현실에서 재현이 어려운 극단적 주행 상황(예: 돌발 보행자 출현, 악천후 등)을 포함할 수 있다는 장점이 있다. CARLA Leaderboard는 CARLA 시뮬레이터의 공식 벤치마크 환경에서 수집된 데이터셋으로, 다양한 기상 조건과 도시 환경

을 포함하고 있으며 객체들의 움직임은 TrafficManager와 Python API을 통해 구현된다. 이후 Town05,<sup>16)</sup> Longest6<sup>17)</sup> 등 기존 데이터셋의 구성 요소들을 변형한 다양한 버전이 다수의 연구에서 학습 및 평가용으로 사용되고 있다. 그러나 구성 시나리오가 다소 단순하여 실제 환경을 완전히 반영하기 어렵다는 한계가 있다. 이를 개선하기 위해 제안된 Bench2Drive는 보다 복잡하고 현실적인 주행 상황을 반영한 시뮬레이터 기반 데이터셋으로, 강화학습 기반 알고리즘인 Think2Drive<sup>18)</sup>를 활용해 데이터를 수집하였다. 공개 이후 다양한 연구에서 채택되며 시뮬레이션 기반 데이터셋의 대표적인 벤치마크로 활용되고 있다.

## 2.2 시뮬레이터

시뮬레이터는 도로, 차량, 보행자, 기상 조건 등 다양한 주변 환경 요소를 인공적으로 재현한 가상의 실험 환경이다. 실제 도로 주행에서 발생할 수 있는 위험을 피하면서도 안전하게 자율주행 알고리즘을 학습하고 평가할 수 있다는 특징이 있다. 초기의 시뮬레이터는 단순한 물리 모델에 기반하여 제한된 환경을 재현하는 수준에 머물렀으나, 최근에는 다중 에이전트 학습(Multi-agent training), GPU 가속(GPU acceleration), 인공지능 기반 이미지 렌더링(Image rendering) 등을 통합하여 다양한 시나리오를 빠르게 학습하고 현실과 유사한 환경을 제공하는 방향으로 발전하고 있다.

먼저 CARLA<sup>15)</sup>는 Fig. 2와 같이 Unreal engine을 기반으로 실제 도시 구조를 모사한 오픈소스 시뮬레이터이며, 도로 환경, 보행자, 차량, 기상 조건 등을 설정하여 다양한 주행 환경을 재현할 수 있다. 이후 등장한 MetaDrive<sup>19)</sup>는 Pandar3D를 기반으로 환경을 구성하였으며, 다중 에이전트 학습을 지원해 학습 효율성을 향상시켰다. 한편 Waymax, Gigaflow, GPU Drive, Scenario dreamer<sup>20,23)</sup> 등은 기존 현실 데이터셋을 기반으로 이미지 정보를 제외한 상태 정보(State information)를 시뮬레이션하는 방식을 사용한다. 이러한 접근은 실제 데이터를 사용하기 때문에 주



Fig. 2 CARLA: simulation for E2E training and evaluation

Table 2 Comparison of representative world models

Type	Method	Prediction target	Core architecture
State predict network	MILE <sup>28)</sup>	BEV, Image	Probabilistic generative model
	GenAD, <sup>29)</sup> WoTE <sup>30)</sup>	BEV	Transformer
State predict network + Generative model	VISTA, <sup>33)</sup> DrivingDiffusion <sup>39)</sup>	Image	Diffusion
	InfiniCube <sup>40)</sup>	Image	Diffusion + 3DGS
	MUVO, <sup>41)</sup> GAIA-1 <sup>31)</sup>	Image	Transformer
	OccWorld <sup>42)</sup>	Image, Occupancy, Pointcloud	Transformer
	NVIDIA cosmos <sup>43)</sup>	Occupancy	Diffusion + Transformer
	DriveEnv-NeRF <sup>44)</sup>	Image	NeRF
	Dreamer series <sup>34,37)</sup>	Image	Diffusion + LLM, 4DGS

변 객체의 움직임에 대하여 현실성이 높다는 장점을 가진다. 이러한 시뮬레이터들은 다중 에이전트 병렬 학습 및 GPU 가속을 지원함으로써 대규모 학습 효율성을 크게 향상시켰으며, 특히 Scenario dreamer의 경우 Diffusion<sup>24)</sup> 모델을 활용해 새로운 상태 정보를 생성함으로써 기존 데이터셋에 없는 다양한 환경 데이터를 학습 및 평가에 활용할 수 있다. 한편 DriveArena<sup>25)</sup>와 NAVSIM은 각각 Diffusion 모델과 3DGS(3D Gaussian Splatting)<sup>26)</sup> 기반의 인공지능 렌더링 기술을 적용하여, 현실 기반 데이터셋에서 자차량의 위치 변화에 따라 바뀌는 새로운 환경에 대한 이미지를 생성할 수 있다. 이를 통해 시뮬레이터와 현실 환경 간의 시각적 도메인 차이(Domain gap)를 효과적으로 완화할 수 있다. 특히 NAVSIM은 병렬적인 Closed-loop 평가를 지원하는 시뮬레이터로서 기존 Closed-loop 방법 보다 빠르게 평가를 수행할 수 있다. 다만 NAVSIM과 DriveArena 모두 생성된 이미지의 현실성 측면에서 여전히 한계점이 존재한다. 마지막으로 NVIDIA DRIVE Sim<sup>27)</sup>은 NVIDIA에서 개발한 자율주행 시뮬레이션 플랫폼으로 Omniverse 기반의 RTX renderer 기술을 이용해 현실과 유사한 환경을 구성한다. 실제와 유사한 센서 데이터를 생성하고 다양한 시나리오들을 시뮬레이션 할 수 있지만 플랫폼 전체가 완전히 개방되어 있지 않아 수정에 제약이 있으며, 학습용 파이프라인이 제공되지 않는다는 한계가 있다.

### 2.3 월드모델

월드모델은 현재 상태 또는 장면을 액션(Action)과 함께 입력 받아 다음 상태 또는 장면을 예측하는 학습 기반의 실험환경이다. Fig. 3과 같이 시뮬레이터는 미리 정의된 물리 규칙에 따라 동작하는 물리 엔진이 다음 상태를 출력하고, 이를 렌더링 엔진이 입력 받아 환경을 재현한다. 반면, 월드모델은 학습된 상태 예측 네트워크가 다음

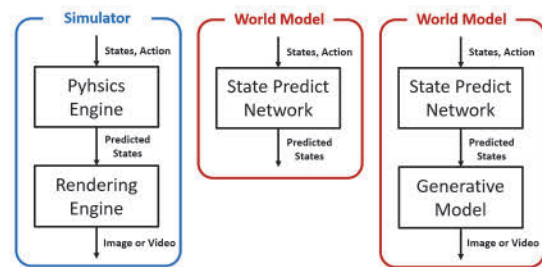


Fig. 3 Comparison between simulator and world model

상태를 추론하며, 예측된 상태 값이 생성모델의 입력으로 사용되어 환경을 재현한다는 점에서 차이를 보인다. 이때 상태 예측 네트워크뿐만 아니라 상태 예측 네트워크와 생성모델을 모두 포함하는 구조 역시 월드 모델로 정의된다. 월드모델은 이러한 학습 기반의 환경 재현을 통해 기존 현실 기반 데이터셋으로부터 현실성이 높은 새로운 환경을 생성함으로써 E2E 학습의 일반화 성능을 크게 향상시킬 수 있다. 이러한 가능성에 힘입어 최근에는 다양한 접근법을 기반으로 한 월드모델 연구가 활발히 진행되고 있다.

월드모델 연구는 크게 두 가지 방향으로 구분할 수 있다. 첫째는 E2E 구조 내부에서 상태 예측 네트워크를 직접 활용하는 방식이며, 둘째는 생성 모델을 추가로 사용해 새로운 장면을 생성하는 방식이다. 먼저 E2E 내부에서 월드모델을 활용하는 방식은 대표적으로 MILE과

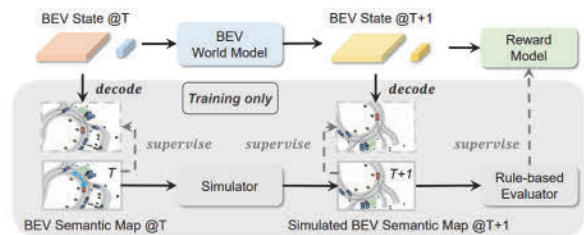


Fig. 4 World model for predicting next states in E2E planning<sup>27)</sup>

GenAD, WoTE 등이 있다.<sup>28-30)</sup> MILE은 확률론적 생성 모델(Probabilistic generative model)을 활용하여 잠재공간(Latent space)에서 BEV(Bird Eye View) 표현을 예측하고, 이를 통해 다음 상태를 확률적으로 샘플링하여 경로를 계획한다. GenAD와 WoTE는 Transformer 기반의 시계열 모델을 사용하여 과거 상태 및 제어 입력을 인코딩하고, 미래 BEV 장면을 예측하여 주행 경로를 결정한다. 이들 방식은 모두 Fig. 4와 같이 BEV 기반의 잠재표현을 사용하므로 계산 효율성이 높고, 복잡한 환경에서도 빠른 예측이 가능하다는 장점이 있다. 반면, 새로운 장면을 생성하기 위해 월드모델을 활용하는 방식은 대표적으로 GAIA, VISTA, DriveDreamer 시리즈 등이 있다.<sup>31-37)</sup> 이들은 예측타겟으로 이미지, BEV, Occupancy map 등 다양한 표현을 사용하며, Diffusion, NeRF,<sup>38)</sup> 3DGS과 같은 기술을 적용해 현실적인 이미지를 생성한다. VISTA와 DrivingDiffusion<sup>39)</sup> 등은 Diffusion 모델을 기반으로 영상 시퀀스로부터 다음 시점의 이미지를 예측하며, 시각적 연속성을 보장하도록 학습된다. InfiniCube<sup>40)</sup>는 Diffusion에 3DGS을 결합하여 지도정보, 텍스트 등으로부터 조건을 주어 주행영상을 생성하고 이를 3D 공간 데이터로 변환하여 대규모 주행 환경을 다양하게 재구성할 수 있다. 반면, MUVO,<sup>41)</sup> GAIA-1,<sup>31)</sup> OccWorld<sup>42)</sup> 등은 Transformer를 사용하여 다양한 입력 정보를 통합하여 하나의 통합적 표현으로 만든 후 이를 이용해 이미지, Occupancy map, Pointcloud 등을 예측한다. NVIDIA Cosmos<sup>43)</sup>는 Diffusion과 Transformer를 결합한 방법으로, 연속적인 시간대의 흐름을 학습하는 Transformer 기반 모델과 세밀한 시각적 복원을 담당하는 Diffusion 디코더를 통합하여, 시간적 연속성과 시각적 정밀도를 동시에 확보하였다. DriveEnv-NeRF<sup>44)</sup>는 NeRF를 이용해 다중 시점 장면을 3차원으로 복원하며, 시점 이동에 따라 일관된 영상을 생성한다. 이는 시각적 현실성을 제공하면서 가상 자율주행 환경 구축에도 활용될 수 있다. 마지막으로 DriveDreamer 시리즈<sup>34-37)</sup>는 Diffusion 기반 구조에 LLM 및 4DGS를 결합하여, 모델이 장면 내 객체의 의미적 관계를 인식하고 조건에 따라 특정 요소의 구성과 배치를 조절할 수 있도록 한다. 이를 통해 다중 시점 장면 생성과 실시간 환경 복원이 가능하다.

#### 2.4 발전 방향

현재 E2E 자율주행 실험 환경은 다양성과 비용, 현실성 간의 균형 문제를 안고 있다. 현실 기반 데이터셋은 실제 센서 및 주행 데이터를 제공할 수 있지만, 수집 범위의 제한과 높은 비용으로 인해 데이터 수집 효율이 낮고 다양한 상황을 포괄하지 못하는 한계를 지닌다. 반면 시물

레이터는 가상환경에서 저비용으로 다양한 상황의 데이터를 생성할 수 있는 장점이 있으나, 현실의 시각적 요소와 물리적 상호작용을 완전히 반영하지 못해 도메인 간 격차가 발생한다. 이러한 문제들을 해결하기 위한 접근으로 최근 등장한 월드모델은, 실제 주행 데이터를 학습하여 새로운 상태나 장면을 생성할 수 있다. 하지만 완벽하게 현실 환경의 복잡한 물리적, 시각적 특성을 재현하지는 못한다. 이러한 한계점으로 인해, 실험 환경에서 학습과 평가를 수행한 자율주행 알고리즘이 실제 환경에 적용될 때 성능 저하가 발생할 수 있다. 따라서 향후 연구에서 도로 노면 표시와 표지판 등 주행 환경의 3차원 일관성을 보장하는 현실에 가까운 월드모델의 개발이 우선되어야 한다고 본다. 이러한 월드모델을 활용해 현실성을 갖춘 가상 데이터를 생성하면, 데이터 수집 비용을 낮추는 동시에 실제 환경에서 확보가 어려운 고위험, 희귀 시나리오(예: 근접 충돌, 차선 이탈 추월 등)를 포함한 실험 환경을 구성할 수 있다. 궁극적으로는 현실 기반 데이터셋으로 학습된 월드모델을 시뮬레이터에 통합하는 방향으로 발전하여 단일 프레임워크 안에서 현실성과 시나리오 다양성을 동시에 확보하는 통합된 자율주행 평가 환경을 확립해야 한다.

### 3. E2E 자율주행 시스템의 평가방법

E2E 자율주행 시스템의 평가 방법은 크게 Open-loop와 Closed-loop으로 구분된다. 두 방법은 각각 예측 정확도와 주행 안정성 측면을 중점적으로 평가하며, 사용되는 데이터셋 및 벤치마크에 따라 적용되는 평가지표의 구성과 특징에 차이가 있다. 평가 지표는 모델의 성능을 정량적으로 측정하는 기준으로, 어떤 지표를 채택하는가에 따라 성능 평가 결과가 달라질 수 있다. 따라서 다양한 측면의 성능을 공정하고 일관되게 평가할 수 있는 표준화된 평가지표의 개발이 중요하다.

#### 3.1 Open-loop 평가지표

Open-loop 평가 방법은 Fig. 5와 같이 주어진 주행 데이터를 기반으로 모델이 출력한 경로와 정답 데이터를 비교하여 예측 정확도를 평가하는 평가지표를 통해 성능을 측정한다. 이 방법은 실제 주행 환경에서 수집된 데이터를 사용하므로, 동일한 조건 하에서 모델의 경로 편차나 충돌률을 정량적으로 측정하고 비교할 수 있어 객관성이 높다. 하지만 평가 시 피드백 제어가 반영되지 않기 때문에, 실제 주행 안정성보다는 예측 정확도 중심의 평가로 한정된다.

Table 3 Comparison of open loop evaluation metrics across datasets

Name	Displacement error	Collision rate	Off-road rate	Precision	Heading error	Diversity
nuScenes <sup>4)</sup>	O	O	O	X	X	X
Waymo open dataset <sup>6)</sup>	O	O	O	O	X	O
Argoverse <sup>45,46)</sup>	O	X	O	O	X	X
nuPlan <sup>5)</sup>	O	X	O	X	O	X

대표적인 Open-loop 평가 방법에 사용되는 데이터셋으로는 nuScenes,<sup>4)</sup> Waymo Open Dataset,<sup>6)</sup> Argoverse,<sup>45,46)</sup> nuPlan<sup>5)</sup>이 있다. nuScenes는 nuTonomy에서 공개한 다중 센서 기반 데이터셋으로, 약 20초 길이의 1,000개 시나리오를 포함하며, 카메라, 라이다, IMU, GPS 등의 센서 정보를 제공한다. 주요 평가지표로는 ADE(Average Displacement Error), FDE(Final Displacement Error), CR(Collision Rate), CCR(Curb Collision Rate) 등이 있으며, 많은 E2E 자율주행 연구에서 이 지표들을 사용하여 예측 정확도와 안전성을 비교 평가한다. Waymo open dataset은 약 10만 개의 20초 길이 주행 장면을 포함한 대규모 데이터셋으로, 카메라와 라이다 기반의 다양한 주행 데이터들을 제공한다. 이 데이터셋에서는 Rater feedback score가 주요 평가지표로 활용되며, 이는 인간 평가자의 피드백 점수를 반영하여 다양한 가능한 경로에 대한 예측 결과의 품질을 정량적으로 평가한다. Argoverse 1과 2는 Argo AI에서 공개한 데이터셋으로, 약 25만 개의 주행 시나리오를 포함하고 있다. 평가에는 MinADE, MinFDE, mAP\_F(mean Average Prediction\_Forecasting) 등이 사용된다. 특히 mAP\_F는 예측 시점 및 종료 시점에서 실제 경로와 예측 경로의 일치를 바탕으로 True positive 및 False negative를 정의하여 계산되며, 예측된 경로 집합의 품질을 평가할 수 있는 핵심 지표로 활용된다. nuPlan은 Motional에서 개발한 학습 및 평가용 벤치마크로, 약 1,312시간의 실제 주행 데이터를 포함한다. 주요 평가지표는 ADE, FDE, AHE(Average

Heading Error), FHE(Final Heading Error), MR(Miss Rate) 등이 있으며, 각 지표는 중요도에 따라 가중합되어 하나의 종합 점수로서 계산된다.

### 3.2 Closed-loop 평가지표

Closed-loop 평가 방법은 Fig. 6과 같이 모델이 출력한 값을 환경에 반영하고, 그 상호작용 결과를 평가지표에 따라 평가하는 방식이다. 이 방법은 환경과의 실시간 상호작용을 포함하므로, 제어되는 차량의 움직임에 따른 객체들의 반응을 반영할 수 있다. 따라서 단순 예측 성능을 넘어 주행 안정성, 효율성, 승차감 등 종합적인 정책 품질을 평가할 수 있다는 장점이 있다.

대표적인 Closed-loop 벤치마크로는 CARLA Leaderboard,<sup>13)</sup> Bench2Drive,<sup>14)</sup> nuPlan Closed-loop Mode,<sup>5)</sup> NAVSIM v1,<sup>7)</sup> NAVSIM v2<sup>8)</sup> 등이 있다. CARLA Leaderboard는 CARLA 시뮬레이터 기반의 대표적인 벤치마크로, 다양한 도시 환경과 주행 시나리오를 포함하고 있다. 평가지표로는 IS(Infraction Score), RC(Route Completion), DS(Driving Score)가 사용된다. IS는 충돌, 신호 위반, 차선 이탈 등 주행 중 발생하는 위험 요소를 평가하고 RC는 주행 경로의 완주율을 평가하며 DS는 IS와 RC의 곱으로 계산된다. Bench2Drive도 CARLA 시뮬레이터를 기반으로 한 벤치마크로 CARLA Leaderboard 보다 더 복잡한 상호작용을 하는 주행 시나리오를 제공한다. 주요 평가지표로는 DS, Success rate, Efficiency, Comfort 등

Open Loop Evaluation

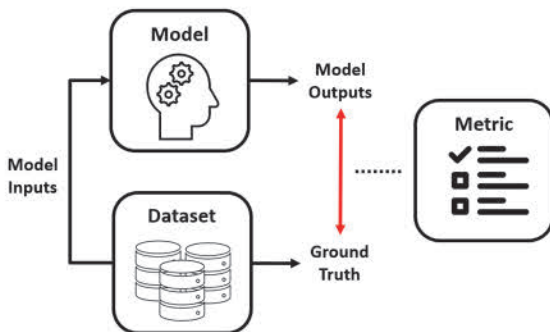


Fig. 5 Open-loop evaluation process

Closed Loop Evaluation

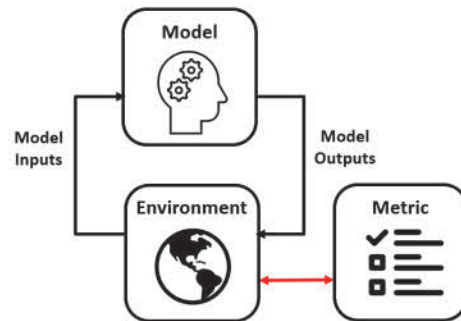


Fig. 6 Closed-loop evaluation process

Table 4 Comparison of closed-loop evaluation metrics across benchmark

Name	Collision	Comfort	TTC	Off-road	Traffic sign	Speed limit	Success	Driving direction	Blocking
CARLA Leaderboard <sup>13)</sup>	O	X	X	O	O	O	O	X	O
nuPlan <sup>5)</sup>	O	O	O	O	X	O	O	O	X
Bench2Drive <sup>14)</sup>	O	O	X	O	O	O	O	X	O
NAVSIM v1 <sup>7)</sup>	O	O	O	O	X	X	O	X	X
NAVSIM v2 <sup>8)</sup>	O	O	O	O	O	X	O	X	X

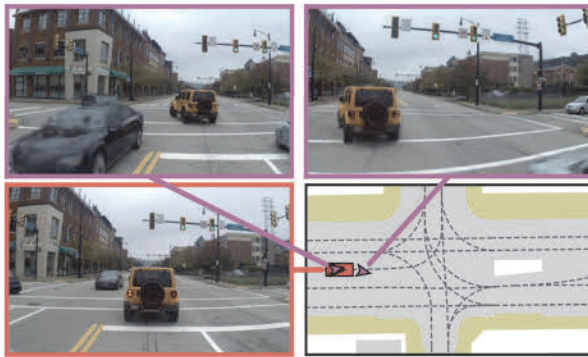


Fig. 7 Pre-generative images made by NAVSIM v2<sup>8)</sup>

이 사용되며, 주행 성공률뿐 아니라 주行的 부드러움까지 정량화 할 수 있다. 특히 Bench2Drive는 현실 주행 환경에서 발생하기 어려운 극단적인 상황들을 포함하고 있어 현재 많은 최신 연구들에서 평가에 활용되고 있다. nuPlan Closed-loop mode는 자체 구축된 데이터셋 기반 시뮬레이션 환경에서 이미지 없이 상태 정보만을 기반으로 평가를 수행한다. 평가 모드는 비반응형과 반응형으로 나뉘며, 전자는 기존 데이터셋의 객체 움직임을 그대로 재현하고, 후자는 주변 객체가 차량의 움직임에 반응하여 더 현실적인 시나리오를 구현한다. 주요 평가지표로는 Scenario score가 사용되며, 이는 주행 성공률, 안정성, 효율성 등의 요소를 중요도에 따라 가중합하여 계산된다. NAVSIM v1은 nuPlan 데이터셋을 기반으로 구축된 시뮬레이터로, 현실 이미지를 입력으로 사용하여 모델이 생성한 경로를 차량이 비반응형 환경에서 추종하며, 그 주행 결과를 기반으로 성능을 평가한다. 주요 평가지표는 PDMS(Predictive Driver Model Score)로, nuPlan의 Closed-loop mode와 유사하게 다양한 평가항목을 중요도에 따라 가중합하여 점수를 계산한다. NAVSIM v2는 기존 버전에 AI기반 생성 기술을 결합한 확장 버전으로, 기존 데이터셋에 존재하지 않는 새로운 이미지 장면을 생성하여 보다 다양한 시나리오 평가가 가능하다. 평가지표는 EPDMS(Enhanced PDMS)가 사용되며, 이는 기존 PDMS에 생성 이미지로부터 생성된 경로의 실효성을 추

가로 반영한 지표다. 평가과정은 Fig. 7과 같이 기존 이미지(주황)와 AI 생성 이미지(보라)를 바탕으로 PDMS를 계산한 후, 생성된 이미지의 품질을 반영해 EPDMS를 산출하는 방식으로 진행된다. 이와 같은 접근은 현실 기반 데이터셋의 한계를 보완하고, 시뮬레이션 환경의 현실성과 다양성을 동시에 확보할 수 있다는 점에서 중요한 의미를 가진다.

### 3.3 발전 방향

현재 E2E 자율주행 시스템의 평가는 Open-loop 평가 방법과 Closed-loop 평가 방법으로 이루어지고 있으나, 각각의 한계로 인해 완전한 주행 성능을 검증하기에는 미흡한 실정이다. Open-loop 평가 방법은 실제 주행 데이터를 기반으로 예측 정확도를 정량적으로 비교할 수 있으나, 주행 중의 상호작용이나 비정상 상황을 충분히 반영하지 못한다는 한계가 있다. 반면 Closed-loop 평가 방법은 시뮬레이터 환경에서 실시간 상호작용을 포함하여 주행 안정성을 검증할 수 있으나, 주변 객체의 반응과 시간적 현실성이 부족하여 실제 도로 상황과의 차이가 발생한다. 이로 인해 예측 성능과 주행 안정성 간의 불균형이 발생하며, 현실 주행에서의 일반화 능력을 정확히 평가하기 어렵다. 따라서 향후 연구에서는 두 평가 방식을 고도화하고 상호 보완적으로 결합한 통합 평가 체계 구축이 요구된다. Open-loop의 현실 데이터 기반 정확도 검증과 Closed-loop의 실시간 주행 안정성 평가를 통합함으로써, 주변 차량과 상호작용하는 차량에 대한 주관적인 평가와 전문가 데이터와의 정합성에 대한 객관적인 평가를 종합적으로 평가할 수 있을 것이라 기대된다. 또한 각각의 알고리즘들이 하나의 환경에서 일부 평가지표만을 사용해 평가되는 것이 아닌 Fig. 8과 같이 여러 환경에서 다양한 평가지표들을 사용해 평가하고 통합하여 공정하고 일관된 비교 체계를 마련하는 것이 중요하다. 나아가, 시뮬레이터와 월드모델을 결합하여 모델이 자율적으로 시나리오를 생성하고 해당 환경에서 알고리즘이 평가될 때 생성된 시나리오의 신뢰도 및 재현가능성 등을 평가할 수 있는 평가지표의 개발도 필수적이다. 즉, 시나리오



Fig. 8 Distributed evaluation framework for fair comparison in robotics

의 현실성과 주행 알고리즘의 일반화 성능을 동시에 평가할 수 있는 접근이 요구된다.

#### 4. 결론

현재 E2E 자율주행 연구는 현실 기반 데이터셋의 효율성과 다양성 부족, 시뮬레이션 및 월드모델의 현실 재현 한계, 그리고 평가지표의 대표성 부족 및 표준화 부재 등의 이유로, 알고리즘의 실제 성능을 완전히 파악하기 어려운 상황이다. 그럼에도 불구하고, 연구자들은 이러한 한계를 극복하기 위해 다양한 노력을 기울이고 있다. 현실 기반 데이터셋의 확장, 월드모델과 시뮬레이터의 결합을 통한 현실적인 시나리오 생성, 그리고 주행 안정성, 효율성, 승차감 등 복합 요소를 통합적으로 평가할 수 있는 새로운 지표 개발 등이 활발히 이루어지고 있다. 그러나 이러한 노력들은 아직 초기 단계에 머무르고 있다. 드물게 발생하는 주행 상황에 대한 데이터 확보, 시뮬레이터 및 월드모델의 정밀도 향상, 통합된 평가기준의 표준화와 같은 핵심 과제들은 여전히 해결이 필요한 상태이다. 궁극적으로 이와 같은 과제들이 해결되고 유기적으로 결합될 때, 자율주행 학습 및 평가 체계는 현실성과 일반화 성능을 동시에 확보하며, 신뢰성 높은 E2E 자율주행 시스템 개발의 기반이 될 것으로 기대된다.

#### 후 기

본 연구는 국토교통부의 재원으로 국토교통과학기술원의 지원을 받아 수행된 연구임(RS-2021-KA162184).

#### References

- 1) S. Kim, Y. Kim, H. Jeon, D. Kum and K. Lee, "Autonomous Driving Technology Trend and Future Outlook: Powered by Artificial Intelligence," Transactions of KSAE, Vol.30, No.10, pp.819-830, 2022.
- 2) L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger and H. Li,

"End-to-End Autonomous Driving: Challenges and Frontiers," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.46, No.12, pp.10164-10183, 2024.

- 3) H. Jo, J. Kim and S. W. Yoon, "Proposal of CARLA-Based Test Scenarios for Autonomous Vehicle Data Acquisition in a Virtual Environment," Transactions of KSAE, Vol.32, No.3, pp.267-272, 2024.
- 4) H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, "NuScenes: A Multimodal Dataset for Autonomous Driving," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.11621-11631, 2020.
- 5) N. Kamchanachari, D. Geromichalos, K. S. Tan, N. Li, C. Eriksen, S. Yaghoubi, N. Mehdipour, G. Bernasconi, W. K. Fong, Y. Guo and H. Caesar, "Towards Learning-Based Planning: The NuPlan Benchmark for Real-World Autonomous Driving," IEEE International Conference on Robotics and Automation, pp.629-636, 2024.
- 6) S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou, Z. Yang, A. Chouard, P. Sun, J. Ngiam, V. Vasudevan, A. McCauley, J. Shlens and D. Anguelov, "Large-Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset," IEEE/CVF International Conference on Computer Vision, pp.9710-9719, 2021.
- 7) D. Dauner, M. Hallgarten, T. Li, X. Weng, Z. Huang, Z. Yang, H. Li, I. Gilitschenski, B. Ivanovic, M. Pavone, A. Geiger and K. Chitta, "NAVSIM: Data-Driven Non-Reactive Autonomous Vehicle Simulation and Benchmarking," Advances in Neural Information Processing Systems, 2024.
- 8) W. Cao, M. Hallgarten, T. Li, D. Dauner, X. Gu, C. Wang, Y. Miron, M. Aiello, H. Li, I. Gilitschenski, B. Ivanovic, M. Pavone, A. Geiger and K. Chitta, "Pseudo-Simulation for Autonomous Driving," Conference on Robot Learning, 2025.
- 9) D. Dauner, M. Hallgarten, A. Geiger and K. Chitta, "Parting with Misconceptions about Learning-Based Vehicle Motion Planning," Conference on Robot Learning, pp.1268-1281, 2023.
- 10) J. Cheng, Y. Chen, X. Mei, B. Yang, B. Li and M. Liu, "Rethinking Imitation-Based Planners for Autonomous Driving," IEEE International Conference on Robotics and Automation, pp.14123-14130, 2024.
- 11) H. Liu, L. Chen, Y. Qiao, C. Lv and H. Li, "Reasoning Multi-Agent Behavioral Topology for Interactive Autonomous Driving," Advances in Neural Information Processing Systems, 2024.
- 12) Waymo LLC, Waymo Open Dataset Challenge: End-to-

- End Driving, Waymo, <https://waymo.com/open/challenges/2025/e2e-driving/>, 2025. (Accessed on 2025).
- 13) CARLA Simulator, CARLA Leaderboard, CARLA, <https://leaderboard.carla.org/>, 2025. (Accessed on 2025).
  - 14) D. Dauner, M. Hallgarten, T. Li, X. Weng, Z. Huang, Z. Yang, H. Li, I. Gilitschenski, B. Ivanovic, M. Pavone, A. Geiger and K. Chitta, "Bench2Drive: Towards Multi-Ability Benchmarking of Closed-Loop End-to-End Autonomous Driving," *Advances in Neural Information Processing Systems*, pp.819-844, 2024.
  - 15) A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez and V. Koltun, "CARLA: An Open Urban Driving Simulator," *Conference on Robot Learning*, 2017.
  - 16) A. Prakash, K. Chitta and A. Geiger, "Multi-Modal Fusion Transformer for End-to-End Autonomous Driving," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.7073-7083, 2021.
  - 17) K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz and A. Geiger, "TransFuser: Imitation with Transformer-Based Sensor Fusion for Autonomous Driving," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.45, No.11, pp.12878-12895, 2023.
  - 18) Q. Li, X. Jia, S. Wang and J. Yan, "Think2Drive: Efficient Reinforcement Learning by Thinking with Latent World Model for Autonomous Driving (in CARLA-v2)," *European Conference on Computer Vision*, pp.142-158, 2024.
  - 19) Q. Li, Z. Peng, L. Feng, Q. Zhang, Z. Xue and B. Zhou, "MetaDrive: Composing Diverse Driving Scenarios for Generalizable Reinforcement Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.45, No.3, pp.3461-3475, 2023.
  - 20) C. Gulino, J. Fu, W. Luo, G. Tucker, E. Bronstein, Y. Lu, J. Harb, X. Pan, Y. Wang, X. Chen, J. D. Co-Reyes, R. Agarwal, R. Roelofs, Y. Lu, N. Montali, P. Mouglin, Z. Yang, B. White, A. Faust, R. McAllister, D. Anguelov and B. Sapp, "Waymax: An Accelerated, Data-Driven Simulator for Large-Scale Autonomous Driving Research," *Advances in Neural Information Processing Systems*, pp.7730-7742, 2023.
  - 21) M. Cusumano-Towner, D. Hafner, A. Hertzberg, B. Huval, A. Petrenko, E. Vinitsky, E. Wijmans, T. Killian, S. Bowers, O. Sener, P. Krähennühl and V. Koltun, "Robust Autonomy Emerges from Self-Play," *arXiv preprint arXiv:2502.03349*, 2025.
  - 22) S. Kazemkhani, A. Pandya, D. Cornelisse, B. Shacklett and E. Vinitsky, "GPU DRIVE: Data-Driven, Multi-Agent Driving Simulation at 1 Million FPS," *International Conference on Learning Representations*, pp.33069-33085, 2025.
  - 23) L. Rowe, R. Girgis, A. Gosselin, L. Paull, C. Pal and F. Heide, "Scenario Dreamer: Vectorized Latent Diffusion for Generating Driving Simulation Environments," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.17207-17218, 2025.
  - 24) J. Ho, A. Jain and P. Abbeel, "Denoising Diffusion Probabilistic Models," *Advances in Neural Information Processing Systems*, Vol.33, pp.6840-6851, 2020.
  - 25) X. Yang, L. Wen, Y. Ma, J. Mei, X. Li, T. Wei, D. Fu, P. Cai, M. Dou, B. Shi, L. He, Y. Liu and Y. Qiao, "DriveArena: A Closed-Loop Generative Simulation Platform for Autonomous Driving," *arXiv preprint arXiv:2408.00415*, 2024.
  - 26) B. Kerbl, G. Kopanas, T. Leimkühler and M. Steinberger, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, Vol.42, No.4, pp.1-14, 2023.
  - 27) NVIDIA Corporation, NVIDIA DRIVE Sim, NVIDIA, <https://developer.nvidia.com/drive/simulation>, 2025. (Accessed on 2025).
  - 28) A. Hu, G. Corrado, N. Griffiths, Z. Murez, C. Gurau, H. Yeo, A. Kendall, R. Cipolla and J. Shotton, "Model-Based Imitation Learning for Urban Driving," *Advances in Neural Information Processing Systems*, pp.20703-20716, 2022.
  - 29) W. Zheng, R. Song, X. Guo, C. Zhang and L. Chen, "GenAD: Generative End-to-End Autonomous Driving," *European Conference on Computer Vision*, pp.87-104, 2025.
  - 30) Y. Li, Y. Wang, Y. Liu, J. He, L. Fan and Z. Zhang, "End-to-End Driving with Online Trajectory Evaluation via BEV World Model," *arXiv preprint arXiv:2504.01941*, 2025.
  - 31) A. Hu, L. Russell, H. Yeo, Z. Murez, G. Fedoseev, A. Kendall, J. Shotton and G. Corrado, "GAIA-1: A Generative World Model for Autonomous Driving," *arXiv preprint arXiv:2309.17080*, 2023.
  - 32) L. Russell, A. Hu, L. Bertoni, G. Fedoseev, J. Shotton, E. Arani and G. Corrado, "GAIA-2: A Controllable Multi-View Generative World Model for Autonomous Driving," *arXiv preprint arXiv:2503.20523*, 2025.
  - 33) S. Gao, J. Yang, L. Chen, K. Chitta, Y. Qiu, A. Geiger, J. Zhang and H. Li, "Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability," *Advances in Neural Information Processing Systems*, pp.91560-91596, 2024.
  - 34) X. Wang, Z. Zhu, G. Huang, X. Chen, J. Zhu and J. Lu, "DriveDreamer: Towards Real-World-Driven World Models for Autonomous Driving," *European Conference on Computer Vision*, pp.55-72, 2024.

- 35) G. Zhao, X. Wang, Z. Zhu, X. Chen, G. Huang, X. Bao and X. Wang, "DriveDreamer-2: LLM-Enhanced World Models for Diverse Driving Video Generation," Proceedings of the AAAI Conference on Artificial Intelligence, Vol.39, No.10, pp.10412-10420, 2025.
- 36) G. Zhao, C. Ni, X. Wang, Z. Zhu, X. Zhang, Y. Wang, G. Huang, X. Chen, B. Wang, Y. Zhang, W. Mei and X. Wang, "DriveDreamer4D: World Models Are Effective Data Machines for 4D Driving Scene Representation," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.12015-12026, 2025.
- 37) C. Ni, G. Zhao, X. Wang, Z. Zhu, W. Qin, G. Huang, C. Liu, Y. Chen, Y. Wang, X. Zhang, Y. Zhan, K. Zhan, P. Jia, X. Lang, X. Wang and W. Mei, "ReconDreamer: Crafting World Models for Driving Scene Reconstruction via Online Restoration," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.1559-1569, 2025.
- 38) B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," European Conference on Computer Vision, pp.405-421, 2020.
- 39) X. Li, Y. Zhang and X. Ye, "DrivingDiffusion: Layout-Guided Multi-View Driving Scene Video Generation with Latent Diffusion Model," European Conference on Computer Vision, pp.469-485, 2024.
- 40) Y. Lu, X. Ren, J. Yang, T. Shen, Z. Wu, J. Gao, Y. Wang, S. Chen, M. Chen, S. Fidler and J. Huang, "InfiniCube: Unbounded and Controllable Dynamic 3D Driving Scene Generation with World-Guided Video Models," arXiv preprint arXiv:2412.03934, 2024.
- 41) D. Bogdoll, Y. Yang, T. Joseph, M. Yazgan and J. M. Zöllner, "MUVO: A Multimodal Generative World Model for Autonomous Driving with Geometric Representations," IEEE Intelligent Vehicles Symposium, pp.2243-2250, 2025.
- 42) W. Zheng, W. Chen, Y. Huang, B. Zhang, Y. Duan and J. Lu, "OccWorld: Learning a 3D Occupancy World Model for Autonomous Driving," European Conference on Computer Vision, pp.55-72, 2025.
- 43) N. Agarwal, A. Ali, M. Bala, Y. Balaji, E. Barker, T. Cai, P. Chattopadhyay, Y. Chen, Y. Cui, Y. Ding, D. Dworakowski, J. Fan, M. Fenzi, F. Ferroni, S. Fidler, D. Fox, S. Ge, Y. Ge, J. Gu, S. Gururani, E. He, J. Huang, J. Huffman, P. Jannaty, J. Jin, S. W. Kim, G. Klár, G. Lam, S. Lan, L. Leal-Taixé, A. Li, Z. Li, C. H. Lin, T. Y. Lin, H. Ling, M. Y. Liu, X. Liu, A. Luo, Q. Ma, H. Mao, K. Mo, A. Mousavian, S. Nah, S. Niverty, D. Page, D. Paschalidou, Z. Patel, L. Pavao, M. Ramezani, F. Reda, X. Ren, V. R. N. Sabavat, E. Schmerling, S. Shi, B. Stefaniak, S. Tang, L. Tchapmi, P. Tredak, W. C. Tseng, J. Varghese, H. Wang, H. Wang, H. Wang, T. C. Wang, F. Wei, X. Wei, J. Z. Wu, J. Xu, W. Yang, L. Yen-Chen, X. Zeng, Y. Zeng, J. Zhang, Q. Zhang, Y. Zhang, Q. Zhao and A. Zolkowski, "Cosmos World Foundation Model Platform for Physical AI," arXiv preprint arXiv: 2501.03575, 2025.
- 44) M. Y. Shen, C. C. Hsu, H. Y. Hou, Y. C. Huang, W. F. Sun, C. C. Chang, Y. L. Liu and C. Y. Lee, "DriveEnv-NeRF: Exploration of a NeRF-Based Autonomous Driving Environment for Real-World Performance Validation," arXiv preprint arXiv:2403.15791, 2024.
- 45) M. F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan and J. Hays, "Argoverse: 3D Tracking and Forecasting with Rich Maps," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.8740-8749, 2019.
- 46) B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes, D. Ramanan, P. Carr and J. Hays, "Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting," arXiv preprint arXiv:2301.00493, 2023.