

Novelty Assessment Report

Paper: DrivingGen: A Comprehensive Benchmark for Generative Video World Models in Autonomous Driving

PDF URL: <https://openreview.net/pdf?id=OrgL5DsU0f>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

Video generation models, as one form of world models, has emerged as one of the most exciting frontiers in AI, promising agents the ability to imagine the future by modeling the temporal evolution of complex scenes. In autonomous driving, this vision gives rise to driving world models—generative simulators that imagine ego and agent futures, enabling scalable simulation, safe testing of corner cases, and rich synthetic data generation. Yet, despite fast-growing research activity, the field lacks a rigorous benchmark to measure progress and guide priorities. Existing evaluations remain limited: generic video metrics overlook safety-critical imaging factors; trajectory plausibility is rarely quantified; temporal and agent-level consistency is neglected; and controllability with respect to ego conditioning is ignored. Moreover, current datasets fail to cover the diversity of conditions required for real-world deployment. To address these gaps, we present DrivingGen, the first comprehensive benchmark for generative driving world models. DrivingGen combines a diverse evaluation dataset—curated from both driving datasets and internet-scale video sources, spanning varied weather, time of day, geographic regions, and complex maneuvers—with a suite of new metrics that jointly assess visual realism, trajectory plausibility, temporal coherence, and controllability. Benchmarking 14 state-of-the-art models reveals clear trade-offs: general models look better but break physics, while driving-specific ones capture motion realistically but lag in visual quality. DrivingGen offers a unified evaluation framework to foster reliable, controllable, and deployable driving world models, enabling scalable simulation, planning, and data-driven decision-making.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **benchmarking generative video world models for autonomous driving**

A total of **50 papers** were analyzed and organized into a taxonomy with **24 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **World Model Architecture and Generation Mechanisms**
- **Controllability and Conditioning Mechanisms**
- **Closed-Loop Simulation and Interactive Environments**
- **Data Generation and Augmentation for Downstream Tasks**
- **Evaluation Frameworks and Benchmarking**
- **Specialized Applications and Domain-Specific Adaptations**

Complete Taxonomy Tree

- benchmarking generative video world models for autonomous driving Survey Taxonomy
- World Model Architecture and Generation Mechanisms
 - Diffusion-Based World Models
 - Latent Diffusion with Controllable Generation (4 papers)
 - [7] Gaia-2: A controllable multi-view generative world model for autonomous driving (Russell Lloyd Ph.D., 2025) [View paper](#)
 - [12] MagicDrive-V2: High-resolution long video generation for autonomous driving with adaptive control (Gao Ruiyuan, 2025) [View paper](#)
 - [22] DrivingDiffusion: layout-guided multi-view driving scenarios video generation with latent diffusion model (Xiaofan Li, 2024) [View paper](#)
 - [28] Epona: Autoregressive Diffusion World Model for Autonomous Driving (Zhang Kai-Wen, 2025) [View paper](#)
 - Diffusion Models with Geometric and Spatial Constraints (3 papers)
 - [15] Cogen: 3d consistent video generation via adaptive conditioning for autonomous driving (Zhu, 2025) [View paper](#)
 - [20] PosePilot: Steering Camera Pose for Generative World Models with Self-supervised Depth (Jin Bu, 2025) [View paper](#)
 - [29] Data-driven Diffusion Models for Enhancing Safety in Autonomous Vehicle Traffic Simulations (LU Jinxiong, 2024) [View paper](#)
 - Reinforcement Learning Enhanced Diffusion Models (1 papers)
 - [27] Rlgf: Reinforcement learning with geometric feedback for autonomous driving video generation (Yan, 2025) [View paper](#)
 - Autoregressive and Transformer-Based World Models
 - GPT-Style Autoregressive Video Generation (3 papers)
 - [11] DrivingWorld: Constructing world model for autonomous driving via video GPT (Hu Xiaotao, 2024) [View paper](#)
 - [19] Worlddreamer: Towards general world models for video generation via predicting masked tokens (Wang Xiaofeng, 2024) [View paper](#)
 - [31] GAIA-1: A Generative World Model for Autonomous Driving (Hu, 2023) [View paper](#)
 - Masked Token Prediction World Models (1 papers)
 - [50] MaskGWM: A Generalizable Driving World Model with Video Mask Reconstruction (Jingcheng Ni, 2025) [View paper](#)
 - Multimodal and Geometric Representation Models

- Occupancy and Voxel-Based World Models (2 papers)
 - [2] Muvo: A multimodal generative world model for autonomous driving with geometric representations (Daniel Bogdoll, 2025) [View paper](#)
 - [9] Occworld: Learning a 3d occupancy world model for autonomous driving (Wenzhao Zheng, 2024) [View paper](#)
- BEV and Multi-View Representation Models (2 papers)
 - [18] Driveworld: 4d pre-trained scene understanding via world models for autonomous driving (Chen Min, 2024) [View paper](#)
 - [46] End-to-End Driving with Online Trajectory Evaluation via BEV World Model (Li, 2025) [View paper](#)
- Language-Conditioned and Foundation Model Integration (3 papers)
- [3] Drivedreamer-2: Llm-enhanced world models for diverse driving video generation (Guosheng Zhao, 2025) [View paper](#)
- [17] Prospective role of foundation models in advancing autonomous vehicles (WU JianHua, 2024) [View paper](#)
- [42] MindDrive: An All-in-One Framework Bridging World Models and Vision-Language Model for End-to-End Autonomous Driving (Bin Sun, 2025) [View paper](#)
- Controllability and Conditioning Mechanisms
 - Trajectory and Ego-Motion Control (3 papers)
 - [4] Drivedreamer: Towards real-world-drive world models for autonomous driving (Xiaofeng Wang, 2024) [View paper](#)
 - [16] Dreamforge: Motion-aware autoregressive video generation for multi-view driving scenes (Mei, 2024) [View paper](#)
 - [36] Unleashing generalization of end-to-end autonomous driving with controllable long video generation (Ma, 2024) [View paper](#)
 - Layout and Scene Structure Control (1 papers)
 - [38] Unimlvg: Unified framework for multi-view long video generation with comprehensive control capabilities for autonomous driving (Chen Rui, 2024) [View paper](#)
 - Adaptive and Context-Aware Control (1 papers)
 - [43] GEM: A Generalizable Ego-Vision Multimodal World Model for Fine-Grained Ego-Motion, Object Dynamics, and Scene Composition Control (Mariam Hassan, 2024) [View paper](#)
- Closed-Loop Simulation and Interactive Environments
 - Closed-Loop Generative Simulation Platforms (2 papers)
 - [1] Drivearena: A closed-loop generative simulation platform for autonomous driving (Yang Xuemeng, 2025) [View paper](#)
 - [6] Glad: A streaming scene generator for autonomous driving (Xie Bin, 2025) [View paper](#)
 - World Model-Based Planning and Reinforcement Learning (5 papers)
 - [13] Think2drive: Efficient reinforcement learning by thinking with latent world model for autonomous driving (in carla-v2) (Qi-feng Li, 2024) [View paper](#)
 - [21] Adawm: Adaptive world model based planning for autonomous driving (Wang Hang, 2025) [View paper](#)
 - [23] Mitigating covariate shift in imitation learning for autonomous vehicles using latent space generative world models (Popov Alexander, 2024) [View paper](#)
 - [35] Enhance sample efficiency and robustness of end-to-end urban autonomous driving via semantic masked world model (Zeyu Gao, 2024) [View paper](#)
 - [49] Planning with Adaptive World Models for Autonomous Driving (Arun Balajee Vasudevan, 2025) [View paper](#)
- Data Generation and Augmentation for Downstream Tasks
 - Safety-Critical and Corner Case Generation (3 papers)
 - [26] ReSim: Reliable World Simulation for Autonomous Driving (Yang Jiazhi, 2025) [View paper](#)
 - [34] Drivinggen: Efficient safety-critical driving video generation with latent diffusion models (Zipeng Guo, 2024) [View paper](#)
 - [47] Gendds: Generating diverse driving video scenarios with prompt-to-video generative model (Yongjie Fu, 2024) [View paper](#)
 - Diverse and Realistic Scenario Synthesis (2 papers)
 - [14] Generalized predictive model for autonomous driving (Jiazhi Yang, 2024) [View paper](#)
 - [33] DiffRoad: Realistic and Diverse Road Scenario Generation for Autonomous Vehicle Testing (Zhou Junjie, 2024) [View paper](#)
 - Geometry-Aware and 3D-Consistent Data Synthesis (2 papers)
 - [25] Geosim: Realistic video simulation via geometry-aware composition for self-driving (Chen Yun, 2021) [View paper](#)
 - [39] Vehiclesim: realistic and 3D-aware video editing with one image for autonomous driving (Beike Yu, 2025) [View paper](#)
 - Simulator-Conditioned and Hybrid Data Generation (2 papers)
 - [32] Assessing Quality Metrics for Neural Reality Gap Input Mitigation in Autonomous Driving Testing (Stefano Carlo Lambertenghi, 2024) [View paper](#)
 - [41] Simworld: A unified benchmark for simulator-conditioned scene generation via world model (Li Xinqing, 2025) [View paper](#)
- Evaluation Frameworks and Benchmarking
 - Comprehensive Benchmarking Frameworks ★ (2 papers)
 - [0] DrivingGen: A Comprehensive Benchmark for Generative Video World Models in Autonomous Driving (Anon et al., 2026) [View paper](#)
 - [37] Worldsimbench: Towards video generation models as world simulators (Qin Yi-ran, 2024) [View paper](#)
 - Survey and Taxonomic Reviews (4 papers)
 - [5] A survey of world models for autonomous driving (Feng Tuo, 2025) [View paper](#)
 - [8] Exploring the interplay between video generation and world models in autonomous driving: A survey (Fu Ao, 2024) [View paper](#)
 - [30] Understanding world or predicting future? a comprehensive survey of world models (Jingtao Ding, 2025) [View paper](#)
 - [40] Foundation Models in Autonomous Driving: A Survey on Scenario Generation and Scenario Analysis (Gao Yuan, 2025) [View paper](#)
- Specialized Applications and Domain-Specific Adaptations
 - Interpretability and Explainability (1 papers)
 - [10] Textual explanations for self-driving vehicles (Kimi¹/₄]Jinkyu, 2018) [View paper](#)
 - Domain Adaptation and Generalization (1 papers)
 - [48] Low-Rank Sparse Generative Adversarial Unsupervised Domain Adaptation for Multitarget Traffic Scene Semantic Segmentation (Mohsen Saffari, 2024) [View paper](#)
 - Trajectory Prediction via Video Generation (1 papers)
 - [45] Trajectory prediction using video generation in autonomous driving (DT Iancu, 2022) [View paper](#)
 - Embodied AI and Vehicular Networks (1 papers)
 - [44] Generative Diffusion-Based Contract Design for Efficient AI Twin Migration in Vehicular Embodied AI Networks (Yue Zhong, 2025) [View paper](#)

- General Deep Learning Foundations (1 papers)
- [24] A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU (Thinagan Perumal, 2023) [View paper](#)

Narrative

Core task: benchmarking generative video world models for autonomous driving. The field has evolved around several interconnected branches that reflect both the technical challenges of building realistic simulators and the practical demands of downstream autonomy tasks. World Model Architecture and Generation Mechanisms explores the foundational neural architectures—ranging from diffusion-based approaches like DriveDreamer[4] and DriveDreamer Two[3] to transformer and occupancy-based representations such as OccWorld[9]—that enable photorealistic or semantically rich video synthesis. Controllability and Conditioning Mechanisms addresses how these models incorporate diverse inputs (e.g., trajectories, maps, textual commands) to steer generated scenarios, while Closed-Loop Simulation and Interactive Environments focuses on enabling agents to interact with the world model over multiple timesteps, as seen in DriveArena[1] and Muvo[2]. Data Generation and Augmentation for Downstream Tasks examines how synthetic rollouts can improve perception or planning modules, and Specialized Applications and Domain-Specific Adaptations covers tailored solutions for safety-critical or rare-event scenarios. Finally, Evaluation Frameworks and Benchmarking consolidates methods for systematically assessing realism, controllability, and utility across these diverse models.

Within this landscape, a particularly active line of work centers on comprehensive benchmarking frameworks that go beyond isolated metrics to evaluate multiple facets—visual fidelity, physical plausibility, and downstream task performance—in a unified manner. DrivingGen[0] exemplifies this direction by proposing a holistic suite of tests that measure not only perceptual quality but also how well generated videos support planning and control algorithms. It sits closely alongside WorldSimBench[37], which similarly emphasizes multi-dimensional evaluation, and contrasts with earlier efforts like DriveDreamer[4] that primarily targeted generation quality without extensive closed-loop or task-oriented benchmarks. By integrating diverse evaluation axes, DrivingGen[0] addresses a key open question: whether improvements in generative realism translate into tangible gains for end-to-end autonomy, thereby bridging the gap between pure synthesis research and practical deployment considerations.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Worldsimbench: Towards video generation models as world simulators

Authors: Qin Yi-ran, Shi Zhelun, Yiran Qin, Yu Jiwen, Zhelun Shi, et al. (30 authors total) | **Year/Venue:** 2024 | **URL:** [View paper](#)

Abstract

Recent advancements in predictive models have demonstrated exceptional capabilities in predicting the future state of objects and scenes. However, the lack of categorization based on inherent characteristics continues to hinder the progress of predictive model development. Additionally, existing benchmarks are unable to effectively evaluate higher-capability, highly embodied predictive models from an embodied perspective. In this work, we classify the functionalities of predictive models into a ...

Relationship Analysis

Both papers belong to the Comprehensive Benchmarking Frameworks category, presenting multi-dimensional evaluation systems for generative video world models. They overlap in assessing visual quality, temporal consistency, and trajectory plausibility for driving scenarios, with both introducing novel metrics beyond standard FVD. The key difference is that DrivingGen focuses exclusively on autonomous driving with driving-specific metrics (safety, trajectory controllability, agent-level consistency) and a curated 400-sample driving dataset, while WorldSimBench evaluates world simulators across three broader embodied scenarios (open-ended environments, autonomous driving, and robot manipulation) with a dual framework combining explicit perceptual evaluation and implicit manipulative evaluation through closed-loop task performance.

Contributions Analysis

Overall novelty summary. The paper proposes DrivingGen, a comprehensive benchmark for evaluating generative driving world models across multiple dimensions including visual quality, trajectory plausibility, temporal consistency, and controllability. Within the taxonomy, it resides in the 'Comprehensive Benchmarking Frameworks' leaf under 'Evaluation Frameworks and Benchmarking'. This leaf contains only two papers total (including DrivingGen), indicating a relatively sparse research direction. The sibling paper is WorldSimBench, suggesting that holistic, multi-dimensional evaluation frameworks for driving world models represent an emerging but not yet crowded area of investigation.

The taxonomy reveals that most research activity concentrates on model architectures and generation mechanisms, with substantial work in diffusion-based models (11 papers across three sub-leaves) and data generation for downstream tasks (10 papers across four sub-leaves). The evaluation branch sits somewhat apart from these technical development efforts. Neighboring leaves include 'Survey and Taxonomic Reviews' (4 papers) which provide broader field overviews, and the various architecture categories which propose the models that benchmarks like DrivingGen aim to assess. The scope_note for this leaf explicitly excludes papers proposing models without comprehensive evaluation frameworks, clarifying that DrivingGen's focus on systematic assessment distinguishes it from generation-focused work.

Among the three contributions analyzed, the benchmark dataset contribution examined 10 candidates and found 1 potentially refutable prior work, suggesting some overlap with existing evaluation datasets. The novel metrics contribution examined 10 candidates with none clearly refuting it, indicating this aspect may be more distinctive. The comprehensive model evaluation contribution examined only 2 candidates with no refutations found. Given the limited search scope of 22 total candidates examined, these statistics suggest moderate novelty for the metrics and evaluation methodology, while the dataset contribution faces more substantial prior work within the examined literature.

Based on the limited top-22 semantic search results, DrivingGen appears to occupy a relatively underexplored niche focused on holistic benchmarking rather than model development. The sparse population of its taxonomy leaf and the moderate refutation rates suggest the work addresses a recognized gap, though the small candidate pool means potentially relevant evaluation frameworks outside the search scope remain unexamined. The analysis captures the paper's positioning within known benchmarking efforts but cannot assess novelty against the broader evaluation literature.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: DrivingGen benchmark with diverse evaluation dataset

Description: The authors introduce DrivingGen, a comprehensive benchmark that includes a carefully curated evaluation dataset covering diverse driving conditions such as varied weather (rain, snow, fog), times of day (dawn, day, night), global geographic regions, and complex driving maneuvers. This dataset addresses the limited diversity in existing benchmarks like nuScenes and OpenDV.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. One million scenes for autonomous driving: Once dataset

URL: [View paper](#)

Brief Assessment

Once Dataset[65] focuses on 3D object detection with 1 million lidar scenes and does not address video generation world models or the specific diversity dimensions (weather, time of day, geographic regions, driving maneuvers) that DrivingGen emphasizes for evaluating generative video models in autonomous driving.

2. Towards a Transitional Weather Scene Recognition Approach for Autonomous Vehicles

URL: [View paper](#)

Brief Assessment

Transitional Weather Recognition[62] focuses on classifying intermediate weather transition states (sunny to rainy, etc.) using a specialized dataset (AIWD6), not on creating a comprehensive benchmark for driving world models with diverse geographic regions, times of day, and complex driving maneuvers as in DrivingGen.

3. S2R-Bench: A Sim-to-Real Evaluation Benchmark for Autonomous Driving

URL: [View paper](#)

Brief Assessment

S2R Bench[60] focuses on perception robustness evaluation under sensor anomalies and corruption in real-world scenarios, not on generative world models or video generation benchmarks with diverse driving conditions.

4. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding

URL: [View paper](#)

Brief Assessment

ACDC[66] focuses on semantic segmentation tasks under adverse conditions (fog, nighttime, rain, snow) with pixel-level annotations, not on generative video world models for autonomous driving. The datasets serve fundamentally different purposes: ACDC provides ground truth for perception tasks, while DrivingGen evaluates video generation models.

5. Ithaca365: Dataset and driving perception under repeated and challenging weather conditions

URL: [View paper](#)

Brief Assessment

Ithaca365[67] focuses on repeated-route data collection for perception tasks (segmentation, detection, depth estimation) rather than evaluating generative world models. The datasets serve fundamentally different purposes: Ithaca365 provides ground-truth annotations for training perception systems, while DrivingGen evaluates video generation quality.

6. SID: Stereo Image Dataset for Autonomous Driving in Adverse Conditions

URL: [View paper](#)

Brief Assessment

SID[59] focuses on stereo image data collection for autonomous driving in adverse weather, not on benchmarking generative video world models. The datasets serve fundamentally different purposes: SID provides raw stereo images for perception algorithm development, while DrivingGen evaluates video generation models.

7. Bdd100k: A diverse driving dataset for heterogeneous multitask learning

URL: [View paper](#)

Prior Art Analysis

BDD100K[63] demonstrates that a diverse driving dataset covering varied weather conditions, times of day, and geographic regions was already established prior to DrivingGen. The candidate paper explicitly describes collecting 100k driving videos spanning varied weather (rain, snow, fog), times of day (dawn, day, night), and global geographic regions across multiple cities and countries. This directly challenges the novelty claim that DrivingGen was the first to introduce such diversity in a driving benchmark dataset.

Evidence

Evidence 1 - **Rationale:** Both papers claim to provide diverse driving datasets covering weather, geographic, and environmental variations. BDD100K[63] was published earlier and explicitly describes these same diversity dimensions. - **Original:** we present drivinggen, the first comprehensive benchmark for generative driving world models. drivinggen combines a diverse evaluation dataset-curated from both driving datasets and internet-scale video sources, spanning varied weather, time of day, geographic regions, and complex maneuvers - **Candidate:** we construct bdd100k, the largest driving video dataset with 100k videos and 10 tasks to evaluate the exciting progress of image recognition algorithms on autonomous driving. the dataset possesses geographic, environmental, and weather diversity, which is useful for training models that are less li...

Evidence 2 - **Rationale:** The candidate paper describes the same types of diversity (weather conditions, times of day, geographic coverage) that the original paper claims as novel contributions, demonstrating prior work exists. - **Original:** diverse driving dataset. we present a new evaluation dataset that captures diverse driving conditions and behaviors. unlike prior datasets biased toward sunny, daytime urban scenes, ours includes varied weather (rain, snow, fog, floods, sandstorms), times of day (dawn, day, night), global regions (n... - **Candidate:** in total, we have 100k driving videos (40 seconds each) collected from more than 50k rides, covering new y ork, san francisco bay area, and other regions as shown in figure 2. the dataset contains diverse scene types such as city streets, residential areas, and highways. furthermore, the videos were...

Evidence 3 - **Rationale:** BDD100K[63] explicitly addresses the same limitations (weather diversity, time of day balance, geographic coverage) that DrivingGen claims to be the first to solve, showing this was already accomplished in prior work. - **Original:** weather and time of day coverage is heavily skewed: datasets like nusenes (caesar et al., 2020) are dominated by clearweather, daytime driving, leaving rare but safety-critical conditions (night, snow, fog) underrepresented. 2) second, geographic coverage is limited, often confined to a few cities or... - **Candidate:** we have collected image-level annotation on six weather conditions, six scene types, and three distinct times of day, for each image. the videos contain large portions of extreme weather conditions, such as snow and rain. they also include a diverse number of different scenes across the world. nota...

8. Generalized predictive model for autonomous driving

URL: [View paper](#)

Brief Assessment

Generalized Predictive Model[14] focuses on building a video prediction model (GenAD) trained on OpenDV-2K dataset for generative world modeling, not on creating a comprehensive benchmark with diverse evaluation metrics for assessing driving world models across multiple dimensions like DrivingGen does.

9. A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook

URL: [View paper](#)

Brief Assessment

Datasets Survey[61] is a survey paper analyzing existing autonomous driving datasets from multiple perspectives. It does not introduce a new benchmark or evaluation dataset for generative driving world models, which is the focus of DrivingGen.

10. Augmented Cross Layer Refinement Network-Based Lane Detection in Adverse Weather Conditions

URL: [View paper](#)

Brief Assessment

Cross Layer Refinement[64] focuses on lane detection in adverse weather using image augmentation techniques, not on benchmark datasets for driving world models with diverse conditions.

Contribution 2: Novel multifaceted metrics for driving world models

Description: The authors propose a novel suite of evaluation metrics specifically designed for driving scenarios. These metrics comprehensively evaluate four dimensions: distribution-level measures for videos and trajectories, quality metrics accounting for perceptual and driving-specific factors, temporal consistency at scene and agent levels, and trajectory alignment measuring controllability.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Lidardm: Generative lidar simulation in a generated world

URL: [View paper](#)

Brief Assessment

LidarDM[53] focuses on LiDAR point cloud generation and evaluation, not on comprehensive video-based driving world model metrics. The candidate does not propose multifaceted evaluation metrics for assessing visual realism, trajectory plausibility, temporal coherence, and controllability in driving video generation as the original paper does.

2. Probing multimodal llms as world models for driving

URL: [View paper](#)

Brief Assessment

Probing Multimodal LLMs[51] focuses on evaluating multimodal LLMs' reasoning capabilities in driving scenarios (ego-motion, traffic detection, planning), not on proposing comprehensive evaluation metrics for generative driving world models that assess visual realism, trajectory plausibility, temporal coherence, and controllability as the original paper does.

3. World-in-world: World models in a closed-loop world

URL: [View paper](#)

Brief Assessment

World in World[56] focuses on closed-loop embodied task success across perception, navigation, and manipulation domains, not on driving-specific evaluation metrics. The candidate evaluates world models through task performance rather than the driving-specific dimensions (distribution-level measures, quality metrics for perceptual/driving factors, temporal consistency, trajectory alignment) proposed in the original paper.

4. Gigaworld-0: World models as data engine to empower embodied ai

URL: [View paper](#)

Brief Assessment

GigaWorld Zero[52] focuses on embodied AI and robotic manipulation scenarios, not autonomous driving. While it mentions evaluation metrics for world models, these are designed for robotic tasks (e.g., physical plausibility, geometric consistency) rather than driving-specific metrics like trajectory plausibility in traffic scenarios or driving-specific imaging factors.

5. Drivedreamer: Towards real-world-drive world models for autonomous driving

URL: [View paper](#)

Brief Assessment

DriveDreamer[4] focuses on controllable driving video generation using diffusion models and does not propose evaluation metrics for assessing driving world models across multiple dimensions like distribution, quality, temporal consistency, and trajectory alignment.

6. Geodrive: 3d geometry-informed driving world model with precise action control

URL: [View paper](#)

Brief Assessment

GeoDrive[54] focuses on 3D geometry-informed generation and action control, not on developing evaluation metrics. The paper evaluates its model using standard metrics (LPIPS, PSNR, SSIM, FID, FVD, ADE, FDE) rather than proposing novel evaluation frameworks for driving world models.

7. Act-bench: Towards action controllable world models for autonomous driving

URL: [View paper](#)

Brief Assessment

Act-Bench[57] focuses specifically on action controllability metrics (instruction-execution consistency and trajectory alignment), not the comprehensive four-dimensional evaluation suite (distribution, quality, temporal consistency, trajectory alignment) proposed in the original paper.

8. Panacea: Panoramic and controllable video generation for autonomous driving

URL: [View paper](#)

Brief Assessment

Panacea[55] focuses on video generation quality and controllability for autonomous driving, not on comprehensive evaluation metrics. While it uses standard metrics like FID and FVD, it does not propose a novel suite of evaluation metrics covering distribution, quality, temporal consistency, and trajectory alignment dimensions as described in the original contribution.

9. Seeing Clearly, Forgetting Deeply: Revisiting Fine-Tuned Video Generators for Driving Simulation

URL: [View paper](#)

Brief Assessment

Fine Tuned Generators[58] focuses on analyzing trade-offs in fine-tuning video generators (visual quality vs. dynamic understanding), not on proposing comprehensive evaluation metrics for driving world models. The paper uses existing metrics like FID, FVD, and object tracking for analysis rather than introducing a novel multifaceted evaluation framework.

10. Worldsimbench: Towards video generation models as world simulators

URL: [View paper](#)

Brief Assessment

WorldSimBench[37] focuses on evaluating world simulators across open-ended embodied environments, autonomous driving, and robot manipulation with hierarchical dimensions (visual quality, condition consistency, embodiment). The original paper's metrics are specifically designed for driving scenarios with distribution-level measures, quality metrics for perceptual and driving-specific factors, temporal consistency at scene and agent levels, and trajectory alignment for controllability—a different evaluation framework than WorldSimBench's approach.

Contribution 3: Comprehensive evaluation of 14 state-of-the-art models

Description: The authors conduct extensive benchmarking of 14 generative world models spanning general video models, physics-based models, and driving-specific models. This evaluation reveals important insights about trade-offs between visual quality and physical consistency, providing the first comprehensive comparison in the driving domain.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Drive&Gen: Co-Evaluating End-to-End Driving and Video Generation Models

URL: [View paper](#)

Brief Assessment

Drive and Gen[69] focuses on co-evaluating video generation models and end-to-end driving planners through behavioral metrics, not on comprehensive benchmarking of 14 generative world models across visual quality and physical consistency dimensions as in the original paper.

2. Vista: A generalizable driving world model with high fidelity and versatile controllability

URL: [View paper](#)

Brief Assessment

Vista[68] focuses on developing a single generalizable driving world model rather than benchmarking multiple models. The paper does not present a comprehensive evaluation framework comparing 14 different generative models across multiple dimensions.

Appendix: Text Similarity Detection

Textual similarity detection checked 22 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Vista: A generalizable driving world model with high fidelity and versatile controllability

Detected in: Contribution: contribution_3

△ **Note:** This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] DrivingGen: A Comprehensive Benchmark for Generative Video World Models in Autonomous Driving [View paper](#)
- [1] Drivearena: A closed-loop generative simulation platform for autonomous driving [View paper](#)
- [2] Muvo: A multimodal generative world model for autonomous driving with geometric representations [View paper](#)
- [3] Drivedreamer-2: Llm-enhanced world models for diverse driving video generation [View paper](#)
- [4] Drivedreamer: Towards real-world-drive world models for autonomous driving [View paper](#)
- [5] A survey of world models for autonomous driving [View paper](#)
- [6] Glad: A streaming scene generator for autonomous driving [View paper](#)
- [7] Gaia-2: A controllable multi-view generative world model for autonomous driving [View paper](#)
- [8] Exploring the interplay between video generation and world models in autonomous driving: A survey [View paper](#)
- [9] Occworld: Learning a 3d occupancy world model for autonomous driving [View paper](#)
- [10] Textual explanations for self-driving vehicles [View paper](#)
- [11] DrivingWorld: Constructing world model for autonomous driving via video GPT [View paper](#)
- [12] MagicDrive-V2: High-resolution long video generation for autonomous driving with adaptive control [View paper](#)
- [13] Think2drive: Efficient reinforcement learning by thinking with latent world model for autonomous driving (in carla-v2) [View paper](#)
- [14] Generalized predictive model for autonomous driving [View paper](#)
- [15] Cogen: 3d consistent video generation via adaptive conditioning for autonomous driving [View paper](#)
- [16] Dreamforge: Motion-aware autoregressive video generation for multi-view driving scenes [View paper](#)
- [17] Prospective role of foundation models in advancing autonomous vehicles [View paper](#)
- [18] Driveworld: 4d pre-trained scene understanding via world models for autonomous driving [View paper](#)
- [19] Worldreamer: Towards general world models for video generation via predicting masked tokens [View paper](#)
- [20] PosePilot: Steering Camera Pose for Generative World Models with Self-supervised Depth [View paper](#)
- [21] Adawn: Adaptive world model based planning for autonomous driving [View paper](#)
- [22] DrivingDiffusion: layout-guided multi-view driving scenarios video generation with latent diffusion model [View paper](#)
- [23] Mitigating covariate shift in imitation learning for autonomous vehicles using latent space generative world models [View paper](#)
- [24] A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU [View paper](#)
- [25] Geosim: Realistic video simulation via geometry-aware composition for self-driving [View paper](#)

- [26] ReSim: Reliable World Simulation for Autonomous Driving [View paper](#)
- [27] RlGf: Reinforcement learning with geometric feedback for autonomous driving video generation [View paper](#)
- [28] Epona: Autoregressive Diffusion World Model for Autonomous Driving [View paper](#)
- [29] Data-driven Diffusion Models for Enhancing Safety in Autonomous Vehicle Traffic Simulations [View paper](#)
- [30] Understanding world or predicting future? a comprehensive survey of world models [View paper](#)
- [31] GAIA-1: A Generative World Model for Autonomous Driving [View paper](#)
- [32] Assessing Quality Metrics for Neural Reality Gap Input Mitigation in Autonomous Driving Testing [View paper](#)
- [33] DiffRoad: Realistic and Diverse Road Scenario Generation for Autonomous Vehicle Testing [View paper](#)
- [34] Drivinggen: Efficient safety-critical driving video generation with latent diffusion models [View paper](#)
- [35] Enhance sample efficiency and robustness of end-to-end urban autonomous driving via semantic masked world model [View paper](#)
- [36] Unleashing generalization of end-to-end autonomous driving with controllable long video generation [View paper](#)
- [37] Worldsimbench: Towards video generation models as world simulators [View paper](#)
- [38] Unimlvg: Unified framework for multi-view long video generation with comprehensive control capabilities for autonomous driving [View paper](#)
- [39] Vehiclesim: realistic and 3D-aware video editing with one image for autonomous driving [View paper](#)
- [40] Foundation Models in Autonomous Driving: A Survey on Scenario Generation and Scenario Analysis [View paper](#)
- [41] Simworld: A unified benchmark for simulator-conditioned scene generation via world model [View paper](#)
- [42] MindDrive: An All-in-One Framework Bridging World Models and Vision-Language Model for End-to-End Autonomous Driving [View paper](#)
- [43] GEM: A Generalizable Ego-Vision Multimodal World Model for Fine-Grained Ego-Motion, Object Dynamics, and Scene Composition Control [View paper](#)
- [44] Generative Diffusion-Based Contract Design for Efficient AI Twin Migration in Vehicular Embodied AI Networks [View paper](#)
- [45] TrajectoryE prediction using video generation in autonomous driving [View paper](#)
- [46] End-to-End Driving with Online Trajectory Evaluation via BEV World Model [View paper](#)
- [47] Gendds: Generating diverse driving video scenarios with prompt-to-video generative model [View paper](#)
- [48] Low-Rank Sparse Generative Adversarial Unsupervised Domain Adaptation for Multitarget Traffic Scene Semantic Segmentation [View paper](#)
- [49] Planning with Adaptive World Models for Autonomous Driving [View paper](#)
- [50] MaskGWM: A Generalizable Driving World Model with Video Mask Reconstruction [View paper](#)
- [51] Probing multimodal llms as world models for driving [View paper](#)
- [52] Gigaworld-0: World models as data engine to empower embodied ai [View paper](#)
- [53] Lidardm: Generative lidar simulation in a generated world [View paper](#)
- [54] Geodrive: 3d geometry-informed driving world model with precise action control [View paper](#)
- [55] Panacea: Panoramic and controllable video generation for autonomous driving [View paper](#)
- [56] World-in-world: World models in a closed-loop world [View paper](#)
- [57] Act-bench: Towards action controllable world models for autonomous driving [View paper](#)
- [58] Seeing Clearly, Forgetting Deeply: Revisiting Fine-Tuned Video Generators for Driving Simulation [View paper](#)
- [59] SID: Stereo Image Dataset for Autonomous Driving in Adverse Conditions [View paper](#)
- [60] S2R-Bench: A Sim-to-Real Evaluation Benchmark for Autonomous Driving [View paper](#)
- [61] A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook [View paper](#)
- [62] Towards a Transitional Weather Scene Recognition Approach for Autonomous Vehicles [View paper](#)
- [63] Bdd100k: A diverse driving dataset for heterogeneous multitask learning [View paper](#)
- [64] Augmented Cross Layer Refinement Network-Based Lane Detection in Adverse Weather Conditions [View paper](#)
- [65] One million scenes for autonomous driving: Once dataset [View paper](#)
- [66] ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding [View paper](#)
- [67] Ithaca365: Dataset and driving perception under repeated and challenging weather conditions [View paper](#)
- [68] Vista: A generalizable driving world model with high fidelity and versatile controllability [View paper](#)
- [69] Drive&Gen: Co-Evaluating End-to-End Driving and Video Generation Models [View paper](#)