Summary

Title: Generative AI Driven City-Scale Digital Twin for Virtual Testing and Benchmarking of Autonomous Vehicles

Authors: Henry Liu, Haowei Sun, Jiawei Wang

We propose a generative AI platform that automatically creates and maintains city-scale digital twins, allowing autonomous vehicles (AVs) to be validated virtually before public deployment. Instead of months of manual data collection and annotation, the system generates simulation-ready environments that update continuously with real-world changes and automatically expose AVs to diverse, safety-critical events. This approach reduces validation time from years to days while cutting costs and ensuring broader, more realistic safety assurance.

Today, AV validation depends on costly on-road testing and limited replay-based simulations that fail to capture the diversity of traffic and environments across cities. Creating HD maps often requires weeks of lidar scanning and manual editing, while rare safety-critical “corner cases” must be hand-crafted or replayed from on-road testing. Sensor simulation, still rooted in traditional graphics engines, leaves a significant sim-to-real performance gap. These manual and fragmented methods are slow, expensive, and fundamentally inadequate for ensuring comprehensive AV safety.

Our proposed platform addresses these bottlenecks with an end-to-end generative AI pipeline. An automated mapping module uses vision-language models to transform satellite imagery into centimeter-accurate, simulation-ready HD maps within hours, and keeps them up-to-date as cities evolve. A Naturalistic Driving Environment (NDE) reproduces human-like traffic behaviors from large-scale driving data, while the Naturalistic and Adversarial Driving Environment (NADE) injects rare but realistic adversities, such as sudden lane closures or unpredictable pedestrians, to stress-test AV stacks. A generative sensor simulation layer employ video foundation models to produce photorealistic multi-view camera data from generated trajectories, significantly narrowing the sim-to-real gap for perception modules or end-to-end AV systems.

Our previous studies have already validated some of the key components: HD maps with centimeter-level accuracy, peer-reviewed NDE and NADE frameworks in *Nature*, and camera-based simulations nearly indistinguishable from reality. The next challenge is scale—expanding from small areas to full cities. Over two years, we will extend the pipeline across multiple metropolitan areas, confirm that synthetic traffic matches real-world statistics, and show that NADE exposes safety-critical scenarios missed by conventional testing.

If successful, this technology will cut AV validation from years to days, allow transportation agencies to evaluate city-wide AV safety before deployment, and provide regulators and the public with evidence-based confidence in autonomous mobility, significantly advancing the safe and rapid rollout of next-generation transportation.