Enhancing Autonomous Vehicle Testing with Diffusion-Based Camera and LiDAR Sensor Simulation

BUSINESS PROBLEM

Waymo invests significant resources in collecting test and validation miles to safely expand its autonomous ride-hailing service into new territories. This process is both costly and time-consuming. The business needs a more scalable, cost-efficient method to test critical driving scenarios. Generative AI, particularly diffusion models, provides an opportunity to reduce the need for physical testing by simulating realistic, multimodal sensor data—including RGB camera images, LiDAR depth maps, and 3D point clouds—thereby lowering operational expenses and expediting expansion efforts.

APPROACH

This work leverages and extends stable diffusion models to simulate 360-degree driving scenes with synchronized camera RGB and LiDAR sensor data. The approach involves encoding data into the model's latent space, training a multimodal diffusion pipeline for generating consistent, multi-view data, and evaluating the results using realism, depth accuracy, and sensor alignment metrics.

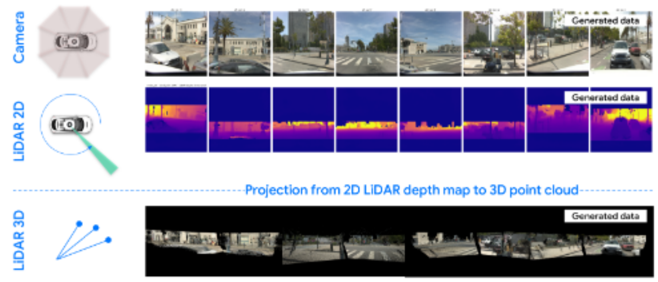
DATA SOURCES

The project utilized multimodal sensor datasets captured from autonomous vehicle testing in urban environments. These datasets were internally sourced and reflect real-world driving scenarios, providing a rich foundation for training and evaluating generative models.

Data Types and Format

The data included high-resolution RGB images from multiple vehicle-mounted cameras, 2D LiDAR depth maps, and 3D point clouds. Formats used were consistent with machine learning pipelines.

Methodology: A multimodal generative AI model simulates eight camera and 360' LiDAR sensor signals for new driving scenes



The trained latent stable diffusion model can enhance simulation realism for testing of autonomous vehicles. This could in-turn lead to a reduction of required on-road testing miles saving operational costs.

Author: Linn Bieske

IMPACT

This work has the potential to enhance Waymo's simulation capabilities by enabling the generation of realistic, multimodal driving environments. By producing synchronized, high-fidelity RGB and LiDAR sensor data, the developed model could support more efficient iteration, targeted scenario exploration, and robustness evaluation of autonomous systems. The project may contribute to foundational generative AI research within the organization and facilitate internal knowledge sharing through well-documented pipelines and presentations. Looking ahead, the approach could inform future efforts aimed at real-time simulation and integration of additional sensor modalities. If further developed and validated, this direction might help reduce operational costs, improve scalability when entering new territories, and diversify training data sources for autonomous vehicle testing.

DRIVERS	The autonomous vehicle industry demands scalable, safe, and cost-effective methods for testing and validation. As companies aim to expand ride-hailing services into new territories, the need to reduce reliance on physical test miles—while maintaining scenario coverage and sensor fidelity—was a key catalyst for exploring generative simulation techniques using AI.
BARRIERS	Key challenges included aligning camera RGB and LiDAR data in the latent space, managing the computational complexity of training multimodal diffusion models, and ensuring the realism and consistency of generated sensor outputs. Limited off-the-shelf tools for such multimodal generative modeling required extensive custom development.
ENABLERS	The project was supported by a collaborative research environment, access to large-scale proprietary sensor data, and mentorship from experts in perception and simulation. Integration with ongoing internal research efforts also enabled feedback loops and alignment with strategic goals.
ACTIONS	To implement the solution, we curated training datasets, developed a multimodal latent diffusion model, and integrated a variational autoencoder (VAE) for encoding LiDAR data. We iteratively trained and evaluated the model on simulation quality metrics and documented the findings to support reproducibility and knowledge transfer.
INNOVATION	This work introduces a transformative simulation approach by unifying camera and LiDAR data in a shared latent space using a stable diffusion framework. It enables synchronized, high-fidelity 360° scene generation, bridging generative image synthesis and sensor emulation. This marks a leap beyond traditional tools and sets the stage for scalable, data-driven autonomy development.
IMPROVEMENT	The model demonstrated the ability to generate high-fidelity, spatially consistent sensor data across camera and LiDAR modalities. This approach has the potential to reduce reliance on costly physical test miles and significantly broaden the range and diversity of driving scenarios available for simulation and validation.
BEST PRACTICES	Success in similar projects depends on robust data preprocessing, thoughtful architectural integration of sensor modalities, clear performance metrics, and iterative feedback from domain experts. Documentation and close collaboration with internal teams significantly enhance adoption and scalability.
OTHER APPLICATIONS	Beyond autonomous vehicle testing, this generative simulation framework could be applied to synthetic data generation for robotics, AR/VR environments, smart city planning, or any domain requiring multimodal sensor fusion. It also lays the groundwork for real-time generative systems in hardware-in-the-loop testing environments.