DataGenesis

Synthetic population
powered by AI

Synthetic population data for smarter, safer Transportation Systems

May 2025

Transportation agencies across the United States are grappling with persistent challenges in planning, forecasting, and safeguarding mobility systems. Chief among these are inaccuracies in travel demand models, insufficient data on pedestrian and cyclist movements, fragmented data ownership, and an overreliance on static models that are ill-suited to an era of dynamic change.

Federal Highway Administration (FHWA) studies have highlighted significant flaws in project-level traffic forecasts, which directly impact infrastructure investment decisions. These inaccuracies stem from **a lack of robust data, outdated modeling assumptions, and an inability to account for behavioral shifts or external shocks**. The Infrastructure Investment and Jobs Act (IIJA), specifically Section 11205, mandates improvements in travel demand modeling accuracy and the development of best practices for forecasting methodologies.

Moreover, current data ecosystems suffer from **jurisdictional silos and strict privacy regulations** (e.g., HIPAA, GDPR), which restrict access to personal and health-related information crucial to understanding transportation needs. Emergency responders, infrastructure planners, and transit agencies face delays and inefficiencies when data cannot be shared or contextualized.

Driving the Future: TxDOT's AI Vision for Smarter, Safer Mobility

The Texas Department of Transportation's **AI Strategic Plan (2025–2027)** charts a bold path to modernize transportation using artificial intelligence. Focused on safety, efficiency, and informed decision-making, the plan emphasizes ethical AI use through principles like transparency, security, and accountability. It aims to integrate AI across operations, from data management to real-time traffic control and emergency response.

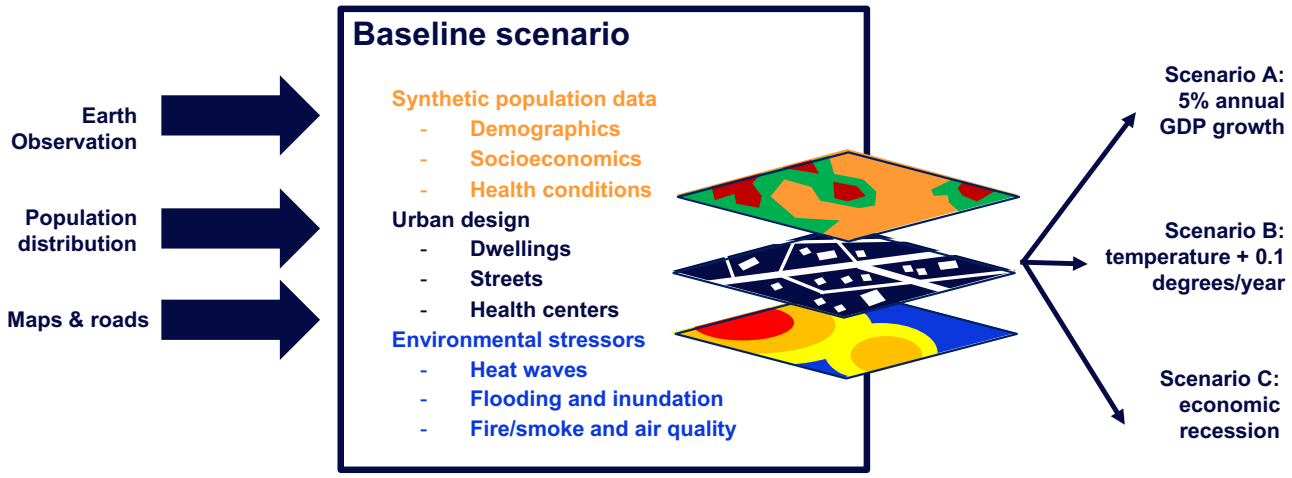
Core use cases include **crash hotspot detection**, **dynamic signal timing**, **automated PS&E reviews**, and **AI-driven emergency planning**—all closely aligned with Skymantics' DataGenesis platform. This tool offers synthetic populations, digital twins, and adaptive routing to improve travel demand modeling, infrastructure planning, and safety interventions, directly supporting TxDOT's goals.

The strategy also stresses building robust data foundations, expanding AI training, and forging innovation partnerships. Its focus on equity, multimodal systems, and predictive planning strongly complements the synthetic data-driven solutions presented in this white paper, paving the way for more resilient, people-centered mobility systems.

Finally, **most transportation modeling tools are parametric and deterministic**. They fail to support "what-if" scenarios involving climate change, migration, economic shocks, or new modes of mobility. As cities and rural areas evolve, transportation planning must become adaptive, predictive, and inclusive.

Towards Synthetic Digital Twins for transportation

DataGenesis, Skymantics' AI-powered synthetic population generator, addresses these systemic challenges by creating **Synthetic Digital Twins (SDTs).** An SDT is a geospatially enabled, multi-layer representation of reality that simulates the interplay between transportation infrastructure, environmental factors, and synthetic populations with rich demographic, socioeconomic, and behavioral attributes. Unlike aggregate statistics or anonymized datasets, DataGenesis produces statistically **accurate, yet entirely synthetic, microdata**. These synthetic agents are characterized by **more than 200 attributes**, including age, health status, income, mobility constraints, and housing context. Generated at scale, these populations can be used to simulate human behavior in response to changes in transportation networks or external conditions.



Synthetic Digital Twin conceptual architecture

Synthetic Digital Twins empower stakeholders to:

- **Simulate travel demand changes** due to demographic shifts, economic trends, or land use policies.
- **Model discrete transportation scenarios** such as sheltering logistics or commuting time.
- **Forecast vehicle/pedestrian flows** and vulnerable road user exposure in safety-critical contexts.
- **Perform coverage analyses** that identify underserved communities without compromising privacy.

By integrating synthetic data with geospatial layers and environmental events (e.g., economic growth, migrations, natural disasters), DataGenesis supports adaptive planning, risk assessment, and infrastructure resilience. Planners can **generate alternative futures, evaluate policy interventions, and prioritize investments** with unparalleled flexibility.

Use cases in transportation applications

a. Travel demand forecasting

DataGenesis enhances traditional forecasting by simulating future demand across diverse population groups, accounting for dynamic factors such as aging, household composition, and commuting patterns. This aligns with FHWA's goals to improve model fidelity and incorporate uncertainty ranges into forecasts.

b. Infrastructure planning

DOTs can use Synthetic Digital Twins to evaluate the impact of road expansions, public transit projects, or housing developments on future traffic volumes, commuting time, and modal splits. Agent-based simulations help identify choke points, plan for efficient access, and assess long-term viability across scenarios.

c. Safety modeling

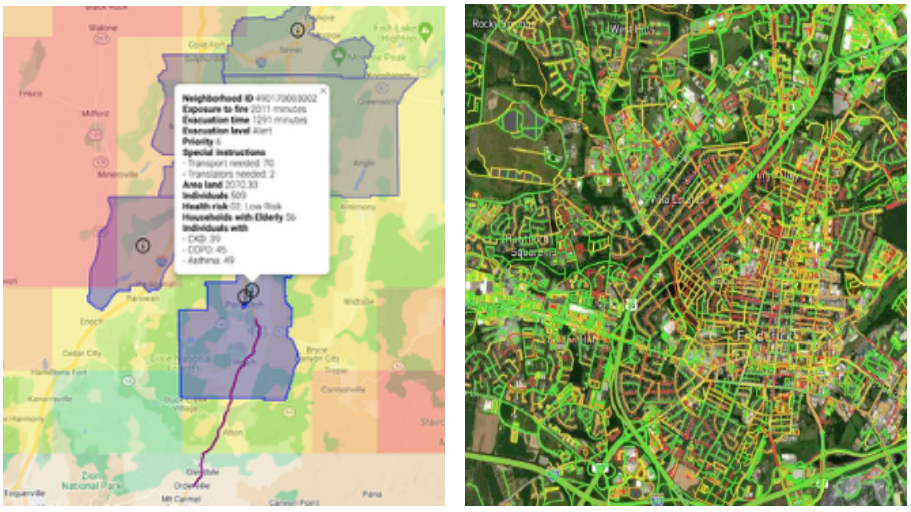
DataGenesis generates Realistic Artificial Datasets (RADs) including risk exposure datasets for road users, including private and commercial drivers, pedestrians, and cyclists. These datasets help calibrate safety models, estimate crash modification factors, and inform the design of safer multimodal corridors.

d. Emergency management and evacuation planning

During disaster scenarios, DataGenesis models population density, health status, and mobility needs to optimize evacuation routes, prioritize at-risk groups, and coordinate resource allocation. This can be applied to urban and rural contexts where routes become constrained due to proximity to a disaster.

e. Commuter/traveler behavior simulation

Synthetic populations can be modeled in terms of psychographics and expected commuter/traveler behavior based on socioeconomic level, access to transportation modes, or household demographics. This supports discrete event simulations to inform targeted interventions in communities.



Example DataGenesis applications: a) Vulnerability-based evacuation planning b) Daytime pedestrian volume estimation per residential area

DataGenesis features

DataGenesis enables powerful decision support for transportation leaders, equipping them to plan smarter, act faster, and deliver safer, more equitable mobility systems for the future.

- **Generative AI engine:** Synthesizes realistic, privacy-safe populations from over 140 authoritative public data sources, ensuring comprehensive demographic, health, and socioeconomic representation.
- **Multivariable correlation preservation:** Maintains meaningful interdependencies between variables to support nuanced modeling of life events such as aging, relocation, family dynamics, employment, and disease progression.
- **Temporal modeling:** Enables advanced scenario planning over time, allowing users to test future outcomes based on economic shifts, climate trends, or policy interventions.
- **Geospatial precision:** Generates data at highly granular resolutions (e.g., Census block group), making it fully compatible with GIS systems and spatial analytics platforms.
- **Validation suite:** Employs rigorous QA tools, including rule-based logic, statistical testing, and expert review cycles, to ensure the realism, consistency, and analytical fitness of generated datasets.
- **Advanced routing logic:** Powers evacuation and routing simulations with intelligent algorithms that optimize based on capacity, urgency, vulnerability, and system constraints.
- **Multi-modal and scenario-aware:** Accounts for pedestrian, vehicle, and transit modalities under dynamic constraints like road closures, infrastructure failure, or vulnerable populations with limited mobility.
- **Integration ready:** Offers robust APIs and data export formats to integrate synthetic populations into use cases ranging from infrastructure planning and freight logistics to first responder simulation and emergency preparedness.

DataGenesis
by skymantics

SBA 8(a)-certified Disadvantaged Small Business

We provide AI, data and geospatial solutions in
Government, healthcare and financial industries

www.skymantics.com

info@skymantics.com