# LLM-Powered Digital Twins for Interactive Urban Mobility Simulation: Integrating SUMO with AI Agents

**Yiming Xu** *
The University of Texas at Austin
Austin, TX 78712
yiming.xu@utexas.edu

**Jihyung Park**
The University of Texas at Austin
Austin, TX 78712
jihyung803@utexas.edu

**Junfeng Jiao**
The University of Texas at Austin
Austin, TX 78712
jjiao@austin.utexas.edu

## Abstract

Urban mobility simulation platforms such as SUMO are widely used in transportation research but remain technically demanding for non-experts and disconnected from modern AI capabilities. This work presents an LLM-powered, web-based transportation digital twin that integrates SUMO with the OpenAI Agents SDK and the Model Context Protocol (MCP) to enable natural language interaction, dynamic scenario editing, and AI-assisted decision support. Our framework allows users to conversationally specify simulation tasks, which are translated by an LLM agent into SUMO configuration and TraCI commands. We extend the simulation environment with new editing features including road network modifications, traffic signal retiming, parking supply and pricing, and event simulation (e.g., incidents, work zones, special events). Guardrails and handoffs ensure safe execution and transparent auditing of simulation changes. By incorporating external tools through MCP, the platform further enriches scenarios with geographic and contextual data from services like Google Maps. We demonstrate the system using Austin, TX as a case study, showcasing how LLM-augmented digital twins can support planners, policymakers, and researchers in testing interventions, evaluating resilience, and exploring sustainable mobility strategies. Our results highlight how coupling LLM with transportation simulation makes digital twins more intuitive, interactive, and deployable, advancing AI-assisted urban planning and decision-making.

## 1 Introduction

Urban mobility systems are central to the functioning of cities, while they are increasingly strained by congestion, climate impacts, and rapid shifts in travel demand. Simulation platforms such as SUMO (Simulation of Urban Mobility) have become essential for evaluating transportation policies, designing traffic management strategies, and developing digital twins of urban environments. However, despite their technical maturity, these tools remain inaccessible to many non-expert stakeholders. Running simulations often requires domain expertise, familiarity with command-line workflows, and detailed knowledge of configuration files, limiting their use in participatory planning and rapid policy prototyping.

*Corresponding Author

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: UrbanAI.

At the same time, advances in Large Language Models (LLMs) have introduced new opportunities for human–AI collaboration. LLMs excel at translating natural language instructions into structured commands, coordinating tools, and mediating complex decision-making processes. However, their integration into urban simulation platforms has been limited, leaving open questions about how AI-driven interfaces might enhance the usability, interactivity, and trustworthiness of urban digital twins.

In this work, we present an LLM-powered, web-based digital twin for urban mobility, built on top of SUMO. Our system integrates the OpenAI Agents SDK with the Model Context Protocol (MCP) to enable conversational simulation control, scenario editing, and multi-agent orchestration. Users can interact with the platform by issuing natural language queries such as "simulate rush hour traffic with a lane closure on Congress Avenue in Austin", which will be automatically translated into SUMO configurations and executed through Traffic Control Interface (TraCI) API. We extend the simulation environment with new editing capabilities for road networks, traffic signals, parking, and events, making it possible to test interventions dynamically and evaluate their impacts on mobility and emissions.

Our contributions are threefold. First, we introduce a framework that couples LLM reasoning with a state-of-the-art urban mobility simulator, enabling natural language interaction, safe tool invocation, and integration with external data sources via MCP. Second, we develop and implement new scenario editing functionalities, including network, signal, parking, and event simulation, within a unified web interface. Third, we demonstrate the system in a case study of Austin, TX, highlighting how AI-augmented digital twins can lower barriers for planners and policymakers, support interactive experimentation, and foster more inclusive and responsive urban planning.

By bridging simulation, AI agents, and urban decision-making, this project advances the vision of intuitive, interactive, and deployable digital twins for sustainable and resilient cities.

## 2 Related Work

### 2.1 Urban Mobility Simulation and Digital Twins

Microscopic traffic simulators such as SUMO (Lopez et al., 2018) and MATSim (Horni et al., 2016) are widely used for evaluating traffic operations, emissions, and mobility policies. More recently, the concept of urban digital twins has gained momentum, linking real-time data with simulation models to create interactive representations of urban systems (Derenick et al., 2021; Madnick et al., 2023). Digital twin initiatives have been demonstrated in contexts such as Singapore's Virtual Singapore and various European smart city projects, highlighting their potential to support urban planning and emergency preparedness. However, these platforms typically require expert knowledge of simulation tools and data pipelines, limiting their accessibility for broader stakeholder engagement.

### 2.2 AI for Urban Systems

Machine learning and reinforcement learning approaches have been extensively applied to traffic signal control (Wei et al., 2019; Chu et al., 2020), routing and demand forecasting (Ke et al., 2017; Zhang et al., 2020), and shared mobility optimization (Jiang et al., 2021). These studies demonstrate the ability of AI models to improve efficiency and robustness in urban mobility. Yet, most AI applications in this domain are task-specific, trained on fixed datasets or constrained environments, and rarely integrated with simulation platforms in ways that allow interactive experimentation or human-in-the-loop decision-making.

### 2.3 Large Language Models for Decision Support

Large language models (LLMs) such as GPT-4 have demonstrated strong capabilities in instruction following, tool use, and multi-agent coordination (OpenAI, 2024). Recent research has explored LLMs as interfaces for data analysis (Liu et al., 2023), simulation control (Wang et al., 2024), and collaborative problem-solving (Park et al., 2023). In the urban context, early work has examined LLMs for policy summarization, citizen engagement, and urban analytics (Xu et al., 2024). However, integrating LLMs directly with mobility simulators and digital twins remains largely unexplored.

### 2.4 Gap and Our Contribution

While digital twins provide rich simulation environments and AI methods enhance optimization, there remains a disconnect between simulation engines, AI reasoning, and human accessibility. Existing platforms either lack intuitive interfaces for non-experts or fail to integrate AI agents capable of orchestrating complex urban simulation workflows. Our work addresses this gap by combining SUMO, web technologies, and LLM agents into a unified system that enables natural language interaction, scenario editing, and AI-assisted decision support for urban mobility planning.

## 3 System Architecture

Our system integrates the SUMO microscopic traffic simulator with a web interface and an LLM-powered agent layer to enable interactive urban mobility simulations. The architecture (Figure 1) is organized into five layers: simulation engine, backend services, LLM orchestration, external tool integration, and frontend visualization.

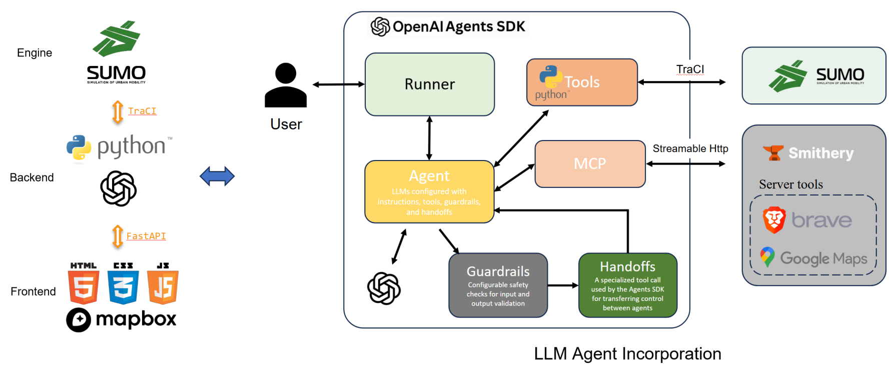


Figure 1: System architecture of the LLM-powered SUMO-Web digital twin

At the core of the platform is SUMO, which provides microscopic traffic simulation of road networks, vehicle movements, and control strategies. Simulations are executed via TraCI, SUMO's TCP-based API, enabling real-time monitoring and manipulation of the simulation state (e.g., adjusting vehicle routes, modifying signals, or querying performance metrics).

We build a set of Python services (FastAPI) that wrap TraCI commands and expose them as web-accessible endpoints. These services also manage file generation for SUMO inputs (e.g., network, demand, configuration). Legacy scripts such as osmWebWizard.py are incorporated into this layer, automating the process of converting OpenStreetMap (OSM) data into SUMO networks. The backend further supports scenario editing functions for road networks, signals, parking supply, and events, which are stored as versioned patches to maintain auditability.

Above the backend, we integrate the OpenAI Agents SDK to provide an LLM-based orchestration layer. The LLM acts as a conversational controller, translating natural language instructions (e.g., "close a lane on 5th Street and simulate the morning peak") into structured backend API calls. The Agents SDK enables: 1) Tool invocation: LLMs call SUMO-control functions as tools; 2) Guardrails: Schema validation and safety checks ensure only valid simulation commands are executed; 3) Handoffs: Tasks can be distributed across multiple agents (e.g., one specialized in emissions, another in routing) for modular workflows.

To extend beyond SUMO's internal models, we integrate external services using the Model Context Protocol (MCP). This enables the LLM assistant to enrich simulations with contextual information—for example, retrieving real-time traffic incidents from Google Maps, querying datasets through Smithery, or leveraging web search via Brave. These capabilities allow scenario configurations to reflect both synthetic and real-world conditions.

The user interacts with the platform through a web interface built with HTML, CSS, and JavaScript, with Mapbox providing geospatial visualization. Users can inspect networks, traffic flows, parking demand, and event effects directly in the browser. The frontend also integrates the LLM assistant as a chat interface, enabling human–AI collaboration where users describe desired interventions in natural language, and the assistant manages translation, execution, and result reporting.

## 4 Functionalities

While the initial SUMO-Web platform focused primarily on importing OpenStreetMap (OSM) networks, generating demand files, and running simulations through a web interface, the present work significantly extends its capabilities. We introduce four categories of interactive scenario editing functions that allow users to design, modify, and evaluate interventions dynamically. These functions transform the platform from a static simulator into a comprehensive digital twin testbed.

Users can add, remove, or modify road segments and lanes directly through the web interface or via natural language commands mediated by the LLM assistant. Supported operations include adjusting the number of lanes, changing lane attributes (e.g., speed limits, allowed modes), and redefining geometries. These edits are automatically converted into SUMO-compatible network patch files and validated with guardrails to prevent invalid topologies (e.g., disconnected edges).

The platform allows real-time editing of traffic control logic at intersections. Users can specify fixed-time signal plans (cycle length, offsets, phase durations), introduce protected movements, or test adaptive control strategies. The LLM assistant translates user instructions into XML signal definitions, which are reloaded into the running SUMO simulation through TraCI. This feature enables rapid experimentation with signal coordination policies at both corridor and network scales.

To capture parking dynamics—often neglected in microscopic simulators—we implement modules for configuring on-street and off-street parking supply, pricing, and search behavior. Parking demand is modeled as an extension of trip generation, while search strategies (e.g., greedy, probabilistic, or time-budgeted) affect vehicle trajectories and congestion. Users can test policies such as adjusting parking fees or altering garage capacities, with direct visualization of occupancy levels and spillover effects.

The system supports injection of exogenous events that disrupt traffic conditions, including incidents, work zones, special events, and weather-related capacity changes. Each event is parameterized by location, severity, and duration, and can be introduced mid-simulation via the LLM interface. This functionality enables exploration of resilience and response strategies, such as rerouting during accidents or evaluating demand surges during concerts and sporting events.

All edits are stored as versioned scenario patches, enabling undo/redo operations and comparative evaluation against baseline conditions. Guardrails enforce domain constraints (e.g., speed limits within realistic bounds, signal phases covering all movements), ensuring that user-driven or LLM-suggested interventions remain valid. This promotes both trustworthiness and reproducibility in experimentation.

## 5 Conclusion

This paper presented an LLM-powered digital twin for urban mobility, integrating the SUMO microscopic traffic simulator with a web-based interface, an LLM orchestration layer, and external data services via the Model Context Protocol. By combining natural language interaction with real-time simulation control, our framework lowers technical barriers to mobility modeling and supports human–AI collaboration in urban planning. We introduced a suite of new scenario editing functions—including road network modification, traffic signal retiming, parking simulation, and event injection—that transform SUMO-Web from a static interface into a comprehensive experimentation environment. Our work contributes to the growing vision of AI-augmented digital twins as tools for sustainable and resilient urban decision-making. By coupling LLM reasoning with simulation engines, we show how planners and policymakers can engage with complex transportation systems more intuitively and iteratively.

Looking forward, future research will explore multimodal interfaces (speech, AR/VR), tighter integration with real-time urban data streams, and reinforcement learning agents for proactive optimization. We see this as a step toward building deployable, intelligent digital twins that bridge the gap between advanced AI methods and the practical challenges of urban governance.

## 6 Code and Demonstration

Project code and demonstrations are available at: `https://github.com/xuyimingxym/SUMO_LLM_Agent`.

## References


[1] Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P. & Wiessner, M. (2018) Microscopic traffic simulation using SUMO. In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582. IEEE.

[2] Horni, A., Nagel, K. & Axhausen, K.W. (2016) *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press.

[3] Derenick, J.C., Vashisth, D., Fong, M. & Banerjee, S. (2021) Digital twins for smart cities: Concepts, applications, and challenges. *Smart Cities*, 4(1), pp. 88–106. MDPI.

[4] Madnick, S., Fonseca, B., Khatri, V. & Lee, D. (2023) Digital twins in urban infrastructure: A review and research agenda. *Computers, Environment and Urban Systems*, 102, 101927. Elsevier.

[5] Wei, H., Xu, N., Zhang, H., Zheng, G., Zang, X.L., Xu, K. & Li, Z. (2019) Colight: Learning network-level cooperation for traffic signal control. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1913–1922. ACM.

[6] Chu, T., Wang, J., Codecà, L. & Li, Z. (2020) Multi-agent deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(4), pp. 3413–3420. AAAI Press.

[7] Ke, J., Zheng, H., Yang, H. & Chen, X.M. (2017) Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 85, pp. 591–608. Elsevier.

[8] Zhang, J., Zheng, Y. & Qi, D. (2020) Deep learning models for traffic forecasting: Review and case studies. *Transportation Research Part C: Emerging Technologies*, 111, pp. 61–79. Elsevier.

[9] Jiang, S., Chen, X., Yin, Y. & Zhu, H. (2021) Optimization of shared mobility services with demand learning. *Transportation Research Part B: Methodological*, 146, pp. 221–242. Elsevier.

[10] Liu, J., Zheng, Z., Wang, X. & Li, Z. (2023) Is ChatGPT a good data analyst? A preliminary study. *arXiv preprint arXiv:2301.13852*.

[11] Wang, X., Li, R., Gupta, R. & Zhao, L. (2024) Large language models as simulation controllers. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*. PMLR.

[12] Park, J.S., O'Brien, C.J., Cai, C.J., Morris, M.R., Liang, P. & Bernstein, M.S. (2023) Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pp. 1–22. ACM.

[13] Xu, Y., Jiao, J. & Ma, Z. (2024) LLM-augmented urban analytics for planning and citizen engagement. *Computers, Environment and Urban Systems*, 107, 102021. Elsevier.

[14] OpenAI. (2024) GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*.