TRAVELAGENT: GENERATIVE AGENTS IN THE BUILT ENVIRONMENT

A PREPRINT

Ariel Noyman[*]
MIT Media Lab
noyman@mit.edu

Kai Hu[*]
South China University of Technology
arhukai@mail.scut.edu.cn

Kent Larson
MIT Media Lab
kll@mit.edu

December 16, 2024

ABSTRACT

Understanding human behavior in the built environment is essential for designing functional and engaging urban places. Traditional methods, such as manual observations and surveys or simplified simulations, often fall short in capturing the complexity of real-world behaviors. We introduce **TravelAgent**, a simulation platform that utilizes Generative Agents to explore human experience, perception, and behavior in existing or planned environments. TravelAgents autonomously navigate and perform tasks in virtual 3D spaces, integrating visual cues, spatial memory, and prior knowledge for planning and decision-making. We demonstrate several use cases in which TravelAgent performs human-like activities in verity of environments, and conducted 100 simulations with 1898 steps across various interior and exterior environments and various agent archetypes, achieving a task completion rate of $\sim$75%. We perform spatial and lingual analysis of the agents' behavior, and show how they react, adapt, or fail to respond to changes in their surroundings. Our findings demonstrate emergent patterns of agent-environment interactions, including path optimization and goal-oriented navigation. TravelAgent offers a novel approach for simulating experience, perception, and behavior in urban design, architecture, and city planning, as a path for more sensible design of the built environments.

Keywords Urban Planning · Architectural Design · Human Behavior · Generative Agents · Simulation · Agent-Based Modeling

1 Introduction

1.1 Background

Human behavior in urban environments is central for the design of spaces that are functional, inclusive, and responsive to the diverse needs of their users **cite**. These behaviors shaped by a complex interplay of spatial configurations, visual stimuli, social dynamics, and individual preferences **cite**. This wide range of factors makes it challenging to asses how people will navigate, interact, and experience urban spaces, particularly in early stages of design**cite**.
In design processes, architects, urban planners, city officials and stakeholder often rely on a variety of tools and methods to evaluate the impact of design decisions on human behavior **cite**. These include manual observations, surveys, focus groups, and computational simulations, such as Agent-Based Models (ABMs)[Noyman, 2022, Chen and Cheng, 2010]. While these methods offer valuable insights, they often fall short of capturing the complexity of human experience in real-world settings **cite**. Manual observations are time-consuming and limited in scope, and surveys and focus groups may not fully represent the diversity of user experiences and preferences **cite**. Computational simulations, such as ABMs, provide a more scalable and controlled environment for studying human behavior, but often rely on simplified heuristics that do not fully capture the richness of human interactions[Batty, 2021].
Recent advances in machine learning and generative models have opened new possibilities for simulating human-like

[*]These authors contributed equally to this work.

behavior in urban environments[Luca et al., 2021]. Large Language Models (LLMs) and Generative Agents (GAs) have demonstrated the ability to generate text and perform complex tasks in response to natural language prompts[Dubey et al., 2024, Reid et al., 2024, Ray, 2023]. The primary advantage of generative agents lies in their ability to act as flexible and adaptive decision-makers in complex environments, and perform well in non-deterministic scenarios[Park et al., 2023, Kaiya et al., 2023]. These systems can simulate multi-step tasks, such as booking a flight or navigating a virtual environment, by chaining together a series of actions guided by an initial natural-language prompt **cite**.

1.2 TravelAgent: Generative Agents Simulation in Urban Environments

Here we introduce **TravelAgent**(TA), a simulation platform that leverages GA to simulate human behavior and experience in the built environment. TravelAgent integrates Large Language Models (LLM), image generation models, 3D models, and computer vision, to enable agents to navigate, observe, and interact with virtual environments using human-like decision-making processes. Unlike traditional simulations, which rely on predefined logic or fixed behavioral rules, TA incorporates cognitive principles such as spatial memory, visual perception, and context-based reasoning to perform autonomous decision making **cite**. TA is designed to easily simulate a wide range of scenarios, from pedestrian navigation tasks to open-ended exploratory activities, in both interior and exterior environments, and in different spatial and environmental conditions **cite**.
TA contributions are: (i) Human-like navigation capabilities, guided by vision, spatial memory, with or without prior knowledge of the agent's environment, (ii) Dynamic and adaptive decision-making processes, powered by a Chain-of-Thought system, (iii) An end-user, simple interface for simulating diverse behaviors, from routine navigation tasks to open-ended exploratory activities. By analyzing agent behaviors in a variety of simulated scenarios - ranging from large-scale urban-planning to micro-scale interior layouts - TA can provide designers and stakeholders with ways to evaluate and refine their designs and help identify potential risks, inefficiencies, and opportunities ahead of implementation.

1.3 Terminology

Agent-based models' (ABMs) are computational models that simulate the actions and interactions of agents in a given environment[Bonabeau, 2002]. ABMs are widely used in urban simulation to study complex systems, such as traffic flow, pedestrian dynamics, or infrastructural efficiency[Chen, 2012]. 'Generative Agents' (GA) are a new class of Large Language Model (LLM)-based agents, that preform multi-step tasks in response to natural language prompts **cite**. In this work, GA are introduced in the reasoning and decision-making part of TA, as they autonomously observe, navigate and interact with the environment. In this work, 'TravelAgent' (TA) is the platform that integrates GA with ABM to perform behavioral analysis of movement in the built environment.

1.4 Paper Organization

The remainder of this paper is organized as follows: Section 2 provides an overview of related work in spatial cognition, agent-based modeling, and pedestrian experience in urban design. Section 3 describes the methodology and implementation of the TravelAgent platform, including agent simulation, sensory inputs, and data collection. Section 4 presents several case studies of TravelAgent in different environments. Section 5 discusses the results and analysis of the case studies, and Section 6 concludes with a summary of the contributions, potential impact, and future directions of TravelAgent.

2 Related Work

Understanding human behavior in urban environments is a complex challenge that intersects multiple disciplines, including urban planning, spatial cognition, and computational modeling. This section reviews the key areas of research that inform the development of the TA platform, particularly focusing on urban experience analysis, spatial cognition, agent-based modeling, and the integration of computational models in simulating human behavior in the built environment.

2.1 Human Experience in Urban Design and City Planning

The importance of human-centered design principles, such as walkability, mixed-use development, and public space quality, has been widely recognized in shaping high-performance, vibrant cities **cite**. Recent research emphasizes the role of street-level analysis in urban design, highlighting the impact of pedestrian experience on urban vitality and social cohesion **cite**. The integration of new technologies has enabled designers and planners to visualize and evaluate

urban interventions in immersive ways, enhancing stakeholder engagement and decision-making **cite**.
"The View from the Road" by Kevin Lynch is a seminal work in human-centered city planning and the perceptual qualities of urban environments [Lynch, 1960]. Lynch introduces the idea of 'imageability,' which refers to the quality and organization of cityscape objects, so that it provides a coherent mental image of the city [Kim, 1999]. Similarly, Jane Jacobs's work highlighted the need for urban planners to consider the urban experience at the street level [Jacobs, 1961]. Gehl's research on public spaces and urban life further emphasizes the importance of human-scale and pedestrian-oriented design in creating vibrant and sustainable cities **cite**.
Today, Computer Aided Design and digital tools help designers, stakeholders, and decision-makers to reach a shared understanding of proposed urban interventions**cite**. Computerized visualizations can clarify the outcomes of urban design choices, including zoning regulations, building codes, and land-use allocations[Smith et al., 1998, Batty et al., 2000]. Nevertheless, digital visualizations cannot fully capture the nature of human behavior in urban environments. Often, these design aids are only accessible for a limited number of experts and stakeholders, whose perspectives may not fully represent the diverse needs and preferences of future users **cite**. Moreover, the static nature of these visualizations can overlook the dynamic and emergent behaviors that arise from exploration and interaction with the built environment **cite**. Modern technologies, such as virtual reality (VR) and augmented reality (AR), have the potential to bridge this gap by providing more immersive and interactive experiences for urban design and planning; however, these tools often require specialized equipment and expertise, and are limited to a single user or small group,limiting their accessibility and scalability [Fonseca et al., 2016].

2.2 Spatial Cognition and Human Navigation

Spatial cognition studies how humans perceive, interpret, and navigate physical spaces, which is critical for understanding movement patterns in urban environments[Montello, 1993, Golledge, 1999, Gath-Morad et al., 2024]. Research in this area emphasizes the importance of environmental cues, such as landmarks, signage, and spatial configurations, in facilitating way-finding and orientation [Javadi et al., 2017, Epstein et al., 2017]. Empirical studies have demonstrated that humans rely on a combination of egocentric (self-to-object) and allocentric (object-to-object) representations when navigating unfamiliar environments **cite**. Cognitive maps, mental representations of physical spaces, are constructed through experience and are essential for efficient navigation and spatial memory **cite**. In recent years, advances in neuroimaging and computational modeling have provided new insights into the neural mechanisms underlying spatial cognition, memory, and navigation **cite**. Virtual reality (VR) and augmented reality (AR) technologies have also been employed to study spatial cognition in controlled settings, providing valuable data for design and planing **cite**. Incorporating principles of spatial cognition into computational models can enhance the realism of simulated agents, allowing them to exhibit human-like navigation behaviors **cite**.

2.3 Agent-Based Modeling in Urban Simulation

Agent-Based Modeling (ABM) has been a cornerstone in urban simulation, enabling researchers to model the actions and interactions of autonomous agents within a defined environmentBankes [2002]. ABMs have been applied to investigate a variety of urban systems, including traffic flow analysisNguyen et al. [2021], pedestrian dynamicsFilomena and Verstegen [2021], and the evaluation of urban policies[Alfeo et al., 2019]. These models allow for the study of emergent behaviors resulting from simple interaction rules among agents, providing insights into complex urban systems. Despite their widespread use, traditional ABMs often rely on deterministic heuristics and predefined behavioral rulesRailsback and Grimm [2019], which can limit their ability to capture the richness and variability of human behavior **cite**. The simplification inherent in these models can overlook the cognitive processes and decision-making mechanisms that influence how individuals interact with urban spaces **cite**. Recent advancements have suggested enhancing ABMs by integrating machine learning techniques, such as reinforcement learning and neural networks[Kobayashi et al., 2023], to model more adaptive and intelligent agents. However, these approaches often require extensive data for training and validation, and can be computationally intensive, making them less accessible for practical urban design applications **cite**.

2.4 Generative Agents

The emergence of Large Language Models (LLMs) **cite**has opened new avenues for simulating human-like behavior through generative agents. These models leverage deep learning to process natural language inputs and generate contextually relevant responses, enabling more sophisticated interactions between agents and their environments[Chopra et al., 2024]. Generative agents have been utilized in various domains, including interactive storytelling **cite**, gaming **cite**, and social simulations **cite**. By integrating LLMs with perception modules, such as computer vision, agents can interpret and respond to complex environmental stimuli, making decisions that reflect human cognitive processes **cite**. In the context of urban simulation, combining LLMs with agent-based models allows for the creation of agents that can

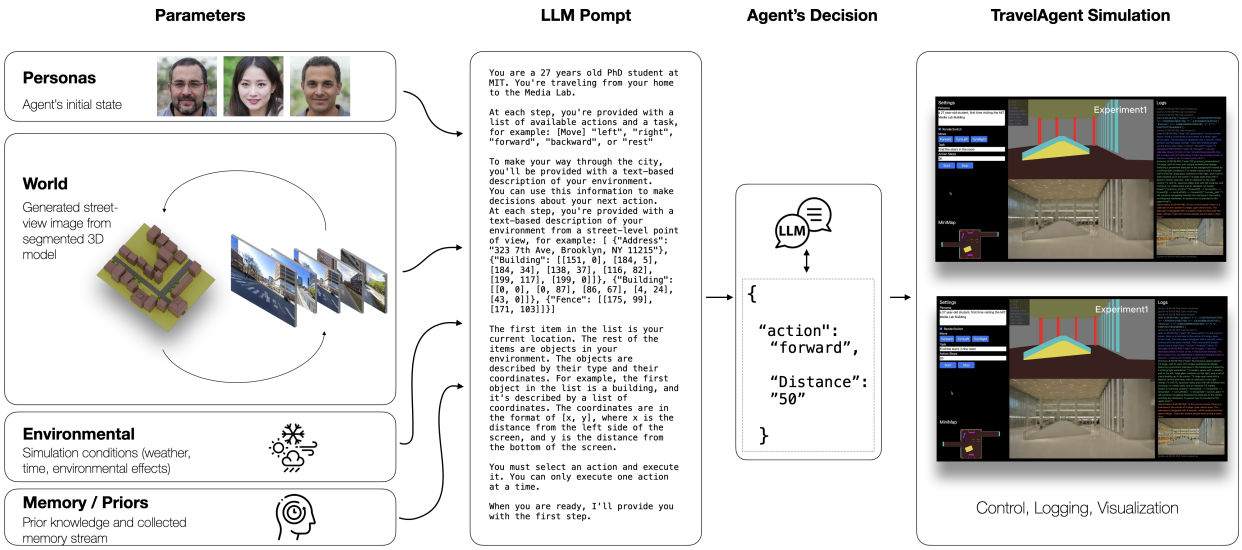

Figure 1: **cite**. The platform consists of Large Language Models (LLMs), computer vision models, and rapid image-generating diffusion models. Each experiment involves the creation of a 3D scene, generation of sensory inputs, processing of sensory inputs using an LLM and an Chain-of-Thought (CoT) framework, and post-experiment analysis.

navigate, perceive, and interact with virtual environments in a human-like manner[Atchade et al., 2024]. These agents can process visual inputs, maintain spatial memory, and make decisions based on prior knowledge and real-time observations, while flexibly adapting to changing conditions and unforeseen events **cite**.

Generative Agents (GA) represent a new class of agents that extend the capabilities of LLMs by incorporating reasoning, feedback, and memory to perform multi-step tasks autonomously **cite**. GA architectures typically include components such as a *Memory Stream* to record experiences, enabling long-term memory recall that factors in recency and relevance **cite**. A *Reflection* mechanism synthesizes memories into higher-level concepts, allowing agents to draw conclusions based on prior experiencesYao et al. [2023]. The *Planning* module translates reflections and environmental inputs into action plans, which are then executed through the *Action* phase[Huang et al., 2024]. The use of reasoning techniques like Chain-of-Thought (CoT) and Tree-of-Thought (ToT) allows GA to decompose tasks into manageable steps and adapt dynamically based on environmental feedback[Wei et al., 2023, Yao et al.]. These compound systems can independently tackle more complex tasks, acting as collaborative partners rather than relying solely on predefined rules **cite**.

Despite their capabilities, GA face several challenges. Efficient memory management is critical to avoid overload while maintaining relevant information[Hatalis et al., 2024]. Mimicking human processes such as reflection and experience remains a challenge due to current limitations of LLMs in understanding context and nuance[Mirzadeh et al., 2024]. Additionally, GA's ability to adapt to new environments and tasks is closely related to their orchestration and the quality of their reasoning mechanisms, making them sensitive to minor changes in setup and initialization **cite**.

In the domain of urban simulation, these agents offer a promising approach to model more realistic human behaviors. By simulating multi-step decision-making processes and incorporating perceptual inputs, GA can provide insights into how individuals might interact with urban environments in ways that traditional ABMs cannot **cite**. This integration has the potential to enhance the design and evaluation of urban spaces by providing more nuanced and human-centric simulations.

3 Methodology

3.1 Agent Simulation

TravelAgent is composed of a Chain-of-Thought (CoT) architecture, computer vision, and fast image-generating models. Each TA experiment involves (i) a 3D scenes of interior or exterior built environment, (ii) sensory inputs such as realistic point-of-view images, segmentation maps, discovery maps, and memory, (iii) CoT to process sensory inputs, guiding

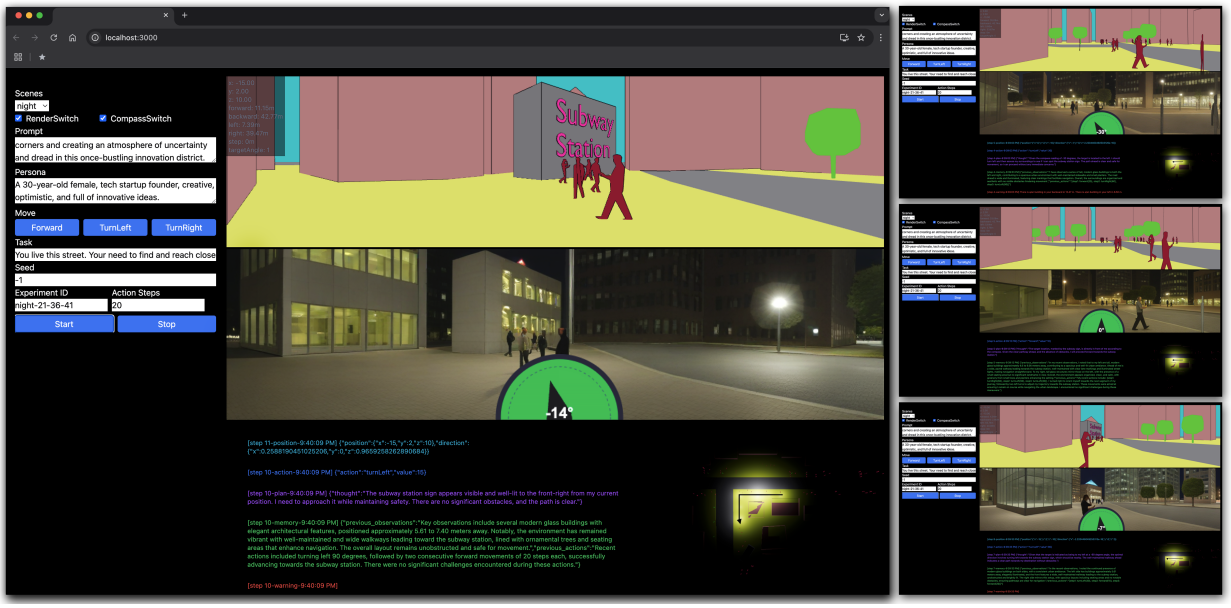

Figure 2: TravelAgent platform interface. The web app provides an end-to-end experimentation environment for testing and evaluating TravelAgents. Key elements include: on the left are the different inputs provided to the agent; on the bottom is the log of the Chain-of-Thought process: in orange are the agent's observations, in green are the agent's memories, in purple are the agent's plans, and in blue are the action/decision. On the top is the street-level view of the environment, with the agent's field of view highlighted in a circle. The baseline 3D environment is sufficient for guiding an image generation model to create eye-level images, as well as to provide depth estimation, and other sensory information for the CoT. Simulating in generative environments allows changing the scenario, agents' profiles, and tasks with a simple word change. The figure shows four step in of a simulation, where the agent is navigating a street to reach a subway station; the agent's progress is shown in the log, and the street-level view is updated at each step.

planning, decision-making, and memory management, and, (iv) post-experiment analysis and interactive dialogues with the agent. TA experiments are designed as follows:

Environment: The TA agent roams a 3D environment, created using common modeling software. The model can be crafted as very simple abstraction of the built environment, since an image generation diffusion model generates realistic street-level images from the basic 3D model. As described in Fig. 3, the 3d model is composed of colors and materials corresponding to real-world objects in the scene, such as walls or landscape objects, so that the class-guided image generation model can reference them to generate realistic-looking first-person images. The 3D environment is exported into a web interface using THREE.JS, which is then used to generate sensory inputs for the agents.

Agent Initialization: Agents are initialized with a set of parameters, which can range from simple properties to complex characters and environments. At minimum, these settings include the agent's starting location, orientation, environment description, number of allowed simulation steps, and task objectives. The agent's memory is initialized can be initialized as an empty set or already contain certain priors, and is updated as the agent explores and interacts with the environment. The task is defined as a set of objectives, such as reaching a destination, finding an object, or interacting with an entity, or free roaming. It is presented as a natural language prompt, which guides the agent's decision-making process; in later steps, the task is represented to the agent as part of the Chain-of-Thought process.

Simulation Steps: At each step of the simulation, the agent collects 'sensory inputs' from the environment, such as street-level images, compass, segmentation maps, discovery map, and memory3.2. The agent then processes these inputs using the CoT framework, first as observation, then as planning, and finally as action or decision. After making a decision, the agent moves in the environment using a key-value pair of interactions (i.e "move forward", "turn right", "finish") as well as a distance to move (i.e. "move 1 meter"). As shown in Fig.8, simulation steps might vary in length and impact; for example, a more 'certain' agent might decide to move forward 50 meters in one step, when another might only turn left to look for its goal. At the end of the step, the agent updates its memory, and repeats the process until the task is completed or the simulation ends.

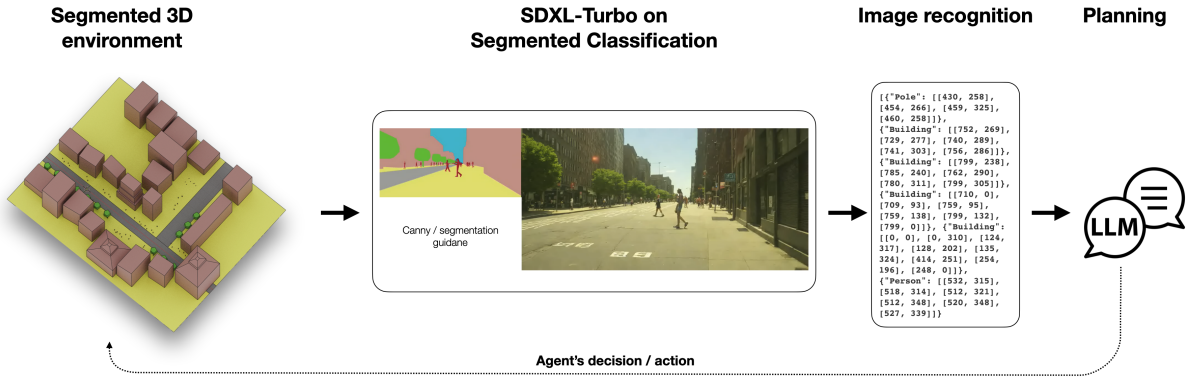

Figure 3: Eye-level image generation from 3D models. The image generation model uses a class or canny-guided fast diffusion model to generate realistic street-level images from the 3D model. The generated images then analyzed by the agent to infer the environment, objects, and spatial layout. The agent is provided with additional inference, such as depth estimation and collision warnings, to guide its decision-making process.

3.2 Sensory Inputs

TravelAgents interact with the environment through a variety of sensory inputs, including visual cues, spatial memory, and prior knowledge. These inputs are designed to mimic pedestrian eye-level perception and decision-making processes, without the usage of navigation tools or top-down maps. The following sensory inputs are provided to the agent at each step of the simulation:

Visual Perception: TA uses computer vision models to process visual information, such as object detection, scene segmentation, and depth estimation. We have explored a verity of image recognition models, such as YOLO, Mask2Former, and finally settled on OpenAI's GPT**cite**. Despite the advancements in image recognition, the agent's visual perception is still abstract, and the textual information sometimes lacks the nuance and context needed for complex decision-making. To address these shortcomings, depth and collision information is provided via ray-casting, in which agents emit rays from their viewpoint to estimate distances to nearby objects in the scene. This information is returned as class labels ('a wall', 'a tree') and 3D distance values ('front : 2 meters').

Discovery map: Despite the information provided by the visual perception model, in early experiments, we found that agents often return to previously visited locations, without recognizing them. To address this, we introduced a 'Discovery map' that displays a top-down view of the environment, including the agent's current location, orientation, and revealed areas. Every time the agent moves, the Discovery map updates to reflect the agent's new position and the areas it has explored, and is added to the CoT as a 2D bitmap image. Unlike traditional ABM, the Discovery map is not used for path-finding or navigation, but rather as a reference for the agent's spatial memory and decision-making process.

Compass: In TA tasks that occur in 'known' environments - where the agent supposedly visited before - a compass-like navigation cue provides directions and error indication. As shown in Fig. 2, the compass is shown at the bottom of the agent's field of view and is delivered to the CoT as part of the image inference. The compass provides general directions, such as 'North', 'South', 'East', and 'West', and is not tied to the agent's orientation; in other words, the agent may decide to move 'Forward' even if the compass indicates 'South'. In 'unknown' environments the compass is not shown, and the agent must rely on visual cues and memory to navigate.

Spatial Memory: TA maintains a spatial memory of the environment, which is updated at each step. The memory is a summation of the agent's past experiences, including visited locations, observed objects, and navigational cues. The memory is compressed and stored textually, representing the agent's experience or gained knowledge. As discussed in Section 5, the agent's memory is saved at each step, and can be used to research the agent's behavior, experience, and decision-making process or to initialize a new agent memory in future simulations.

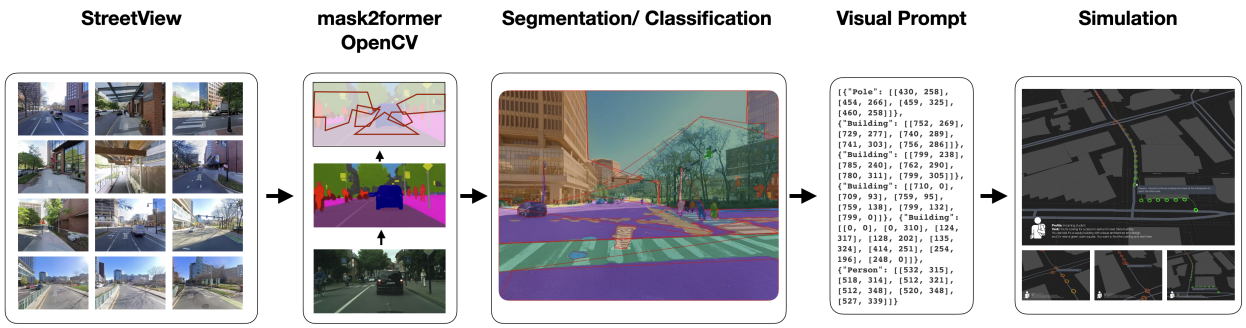

Figure 4: Visual Perception in 'Lunch Break' Experiment. The agent's visual perception is guided by Google Street View (GSV) images, which provide a first-person view of the environment. Mask2Former model is used to segment the image and identify objects, such as buildings, trees, and benches. An OpenCV convex hull algorithm is used to estimate the segments outlines, which are then used as a textual reference for the agent's navigation and decision-making process.

4 Case Studies

In order to streamline experimentation, we developed a web-based platform to enable users to create an evolute TAs in virtual environments. This platform allows users to easily integrate custom environments, define specific tasks and goals for the agents, and analyze their behaviors and decision-making processes. The platform was utilized to conduct a series of experiments in various settings, including both interior and exterior spaces, to assess the effectiveness of TA in simulating human behavior. This section presents the experiments conducted; the results and analysis are discussed in Section 5.

4.1 Early Experiments

Exterior - Lunch Break in Kendall Square: An early experiment tested the agent's ability to navigate and make decisions in real-world urban environments. The experiment involved simulating a student search for a lunch break spot in Kendall Square, Cambridge, MA. The agent was tasked with finding a place to eat lunch with shade and seating near their lab. Since no map or route was provided, the agent's prompt suggested they were unfamiliar with the area (a student visiting campus), and relied on visual cues and memory to navigate. The agent was provided with Google Street View (GSV) images of the area to simulate its point of view. At each step, the agent was shown a new GSV image and asked to make a decision based on the visual information, its memory, and the task requirements.
Unlike the following experiments, the usage of GSV images provided the agent with a more realistic view of the urban environment, but also introduced certain challenges. For example, the agent's visual perception was limited to GSV images, which are mainly captured around to the center of the road and do not provide a full pedestrian view of the street. Moreover, navigation cues such as depth, distance, and collision warnings were more challenging to provide, as the GSV images do not include this information. Lastly, GSV data cannot be easily altered to present different scenarios or environments, such as different seasons, times of day, or weather conditions, which limited the experiment's flexibility and scalability.
Interior - Laboratory Exploration: A following experiment involved simulating a researcher's exploration of a laboratory space. Unlike the previous experiment, the lab environment was designed using a flexible 3D model, with specific objects and features, such as elevators, stairs, and lab equipment. The agent was tasked with finding a specific public area in the lab, in order to meet with a colleague. Given the confined nature of the lab environment, the agent's decision-making process was more constrained than in the previous experiment. The agent relied heavily on visual cues, such as room layouts, and object placement, to navigate the space and locate the public area.

4.2 Main Experiment: 'Train Station'

The initial experiments were conducted in loosely controlled environments and did not incorporate most of the Chain-of-Thought (CoT) design. A key takeaway from these early studies was that agents should be less influenced by the initial conditions or the current step of the simulation, and more driven by the overall task and the agent's goals, in concert with their memories, observations, and planning. To test this hypothesis, we conducted an experiment simulating a common activity as part of a daily commute in an urban area. The agent was tasked with navigating city streets to reach

a local train station. As before, the agent was provided with no maps and did not follow any shortest path or navigation algorithm.

4.2.1 Experiment Design

The experiment was structured around a consistent urban environment and task, but varied across different scenarios, personas, and initial conditions. The environment comprised an urban scene featuring a mix of buildings, streets, and pedestrians, with the task being to navigate to a nearby subway station as part of a daily commute. Each agent received a natural language prompt detailing the agent's persona, time of day, weather conditions, and task requirements. These environmental specifications also informed the prompts for the street-image generative model. Table 6.2 summarize the scenarios. The personas and scenarios were combined to create a matrix of 100 experiments. Table 1 illustrates the associations between season, location, time, persona, and scene, resulting in five distinct scenarios.
In all scenarios, the agent's primary objective was to navigate the streets and reach a designated subway station. Upon successful completion of this task, the agent was informed that the train service was unavailable and was subsequently assigned an additional task from a predefined set of sub-tasks:

- 'If the train is unavailable, find an alternative way to get to work.'
- 'Buy coffee before work if there's time.'
- 'Interact with a friend across the street if encountered.'

These sub-tasks were designed to evaluate the agent's adaptability to unforeseen circumstances and its decision-making capabilities when presented with new information. This approach also tested the agent's ability to manage an expanded context window, a key aspect in the application of Large Language Models (LLMs) for dynamic and complex environments**cite**.

4.2.2 Agent Initation and Navigation Cues

In addition to the generated street-view and the discovery map 3.2, the agent is provided with a compass-like navigation cue that offers general directions and error indications, as detailed in 3.2. This compass acts as a reference for prior knowledge of the location, simulating the agent's memory from previous visits. The agent's initial prompt further reinforces this familiarity: *'You live on this street. It is morning, and you are on your way to work. You need to reach the nearby subway station. To reach the station, head down the street, then turn left, and it is on your left. It has a large 'Subway' sign on it.'*

5 Results and Discussion

In this section we present the results of the case studies, focusing on the 'Daily Commute' experiment. We discuss the agent's behavior, decision-making process, and the opportunities and challenges of using TravelAgent.
In the initial experiments (see 4.1) the agent's decisions were primarily influenced by immediate visual stimuli and physical elements like walls, trees, and benches. The agent's path was largely determined by early choices, such as direction at a fork, leading to a deterministic behavior pattern. This suggests a strong influence of initial conditions, with limited adaptation to new information. To address this, subsequent simulations incorporated dynamic decision-making, allowing the agent to re-evaluate choices based on updated sensory inputs and environmental cues, enhancing its adaptability.
In the primary experiment, 'Subway Station', we conducted 100 simulations encompassing 1898 steps. Approximately 76% of the agents successfully completed their tasks within the designated step count. The remaining agents failed to achieve their objectives due to various factors, such as encountering obstacles (e.g., pedestrians or buildings) or making continuous turns that led to disorientation. This section provides a comprehensive evaluation of the agents' behavior and decision-making processes through spatial analysis, thematic and topical modeling, and sentiment analysis of the agents' cognitive outputs and observations.
importantly, reaching the goal of finding the subway station was not the main objective of these experiments. Instead, TA is designed to inspect the agent's internal observations and decisions so that the legibility and coherence of the environmental design could be evaluated based on the agent's experience.

5.1 Spatial Analysis

An evaluation of both successful and unsuccessful navigation paths across different scenarios in the 'Subway Station' experiment revealed distinct spatial patterns, as shown in Fig. 5 (bottom). Agents in the 'Night' scenario exhibited the

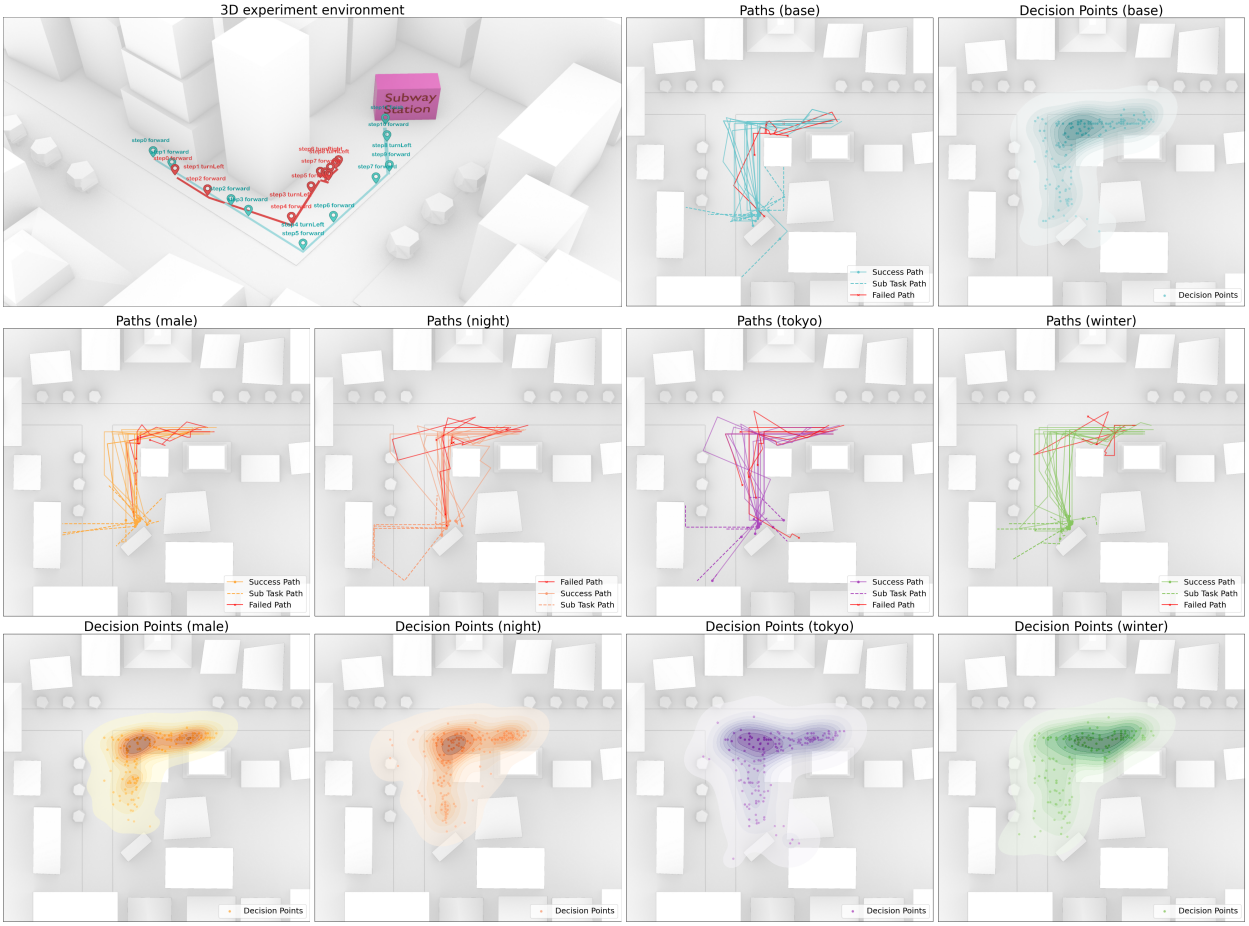

Figure 5: Spatial Evaluation of 'Train Station' Experiment. For each scenario, we evaluate the agent's successful and failed paths; If the agent reaches and recognizes the subway station, the path is considered successful, and the agent is given a new task. The top left figure overlays a successful and a failed path of the 'Base' scenario. The top-right figure shows the paths of all agents in the 'Base' scenario, and the aggregation of all decision points in this scenario. The bottom figure shows the paths of all agents in all other scenarios. Notably, the agents in the 'Night' scenario have the most failed and inconsistent paths, which is also reflected in more sparse decision points aggregation. Conversely, the agents in the 'Winter' scenario have more consistent paths, with a clear aggregation of decision points early in the simulation.

highest frequency of navigation failures and the most inconsistent path trajectories. This is reflected in a more dispersed aggregation of decision points, suggesting difficulties in spatial orientation and decision-making under reduced visibility conditions. Conversely, agents in the 'Winter' scenario demonstrated more consistent and successful navigation paths, with decision points clustered tightly and earlier in the simulation. Correlating with the agents' top terms in their observations, the agents in the 'Winter' scenario focused on the presence of snow, ice, and cold weather, which may have influenced their navigation strategies and decision-making processes.
These findings highlight the significant influence of environmental variables on agent navigation and pedestrian-level decision-making.

5.2 Term Frequency Analysis

To further understand the agent's behavior and decision-making process, we conducted a thematic analysis of the agent's cognitive outputs and environmental observations. We extracted the agent's internal representations from the simulation logs and applied various natural language processing (NLP) techniques, including tokenization, lemmatization, and n-gram analysis **cite**. Subsequently, we utilized term frequency-inverse document frequency (TF-IDF) vectorization to identify the most prevalent features in the agent's outputs. These features were then clustered to show the key topics

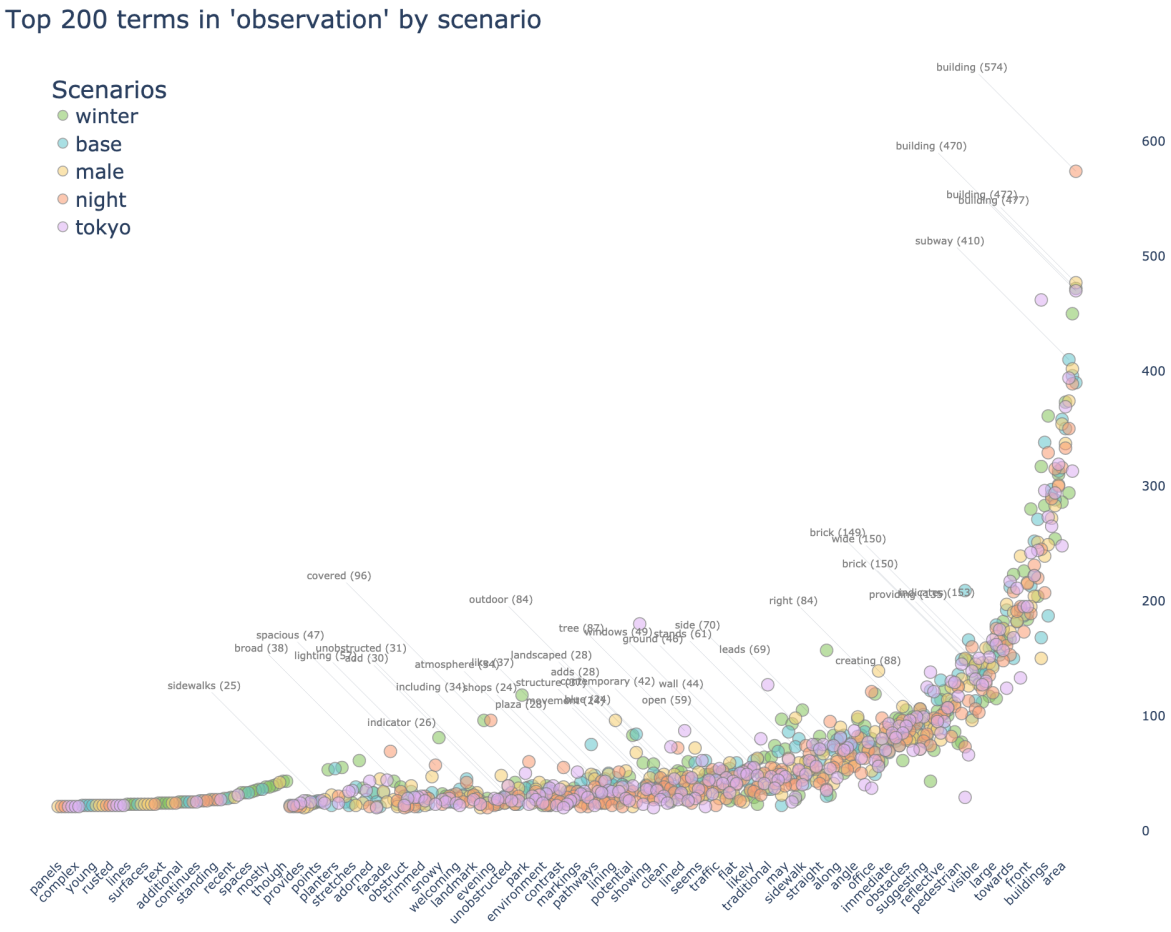

Figure 6: Most common words in the agent's 'plan' stream. The words are clustered by scenario. Naturally, the most prevalent words are related to the agent's task, the target, its action, and the immediate environment.

that emerged from the agent's decision-making process.
We evaluated 1,898 unique steps in the simulation logs. From these, we extracted the following word counts: Planning - 91,340, Memory: 156,663, Observations: 219,742. The results visualized in Fig. 6, illustrate the most prevalent terms in the agent's 'plan' stream, clustered by scenario. Top features are mostly related to the agent's immediate surrounding environment and its task, such as 'target', 'action', 'move', and 'street'. In all scenarios, the agent reflects on the shape, form, and materiality of its environment ('bricks', 'glass', 'flat', 'building'), as well as the presence of obstacles and navigational cues ('obstacle').

5.3 Topical Modeling

To examine the cognitive processes underlying the agent's behavior, we conducted topic modeling on the observation and planning streams. Key topics were extracted using LLM clustering and semantic analysis, categorizing terms into navigation, visibility, movement, obstacles, and urban environment. Latent Dirichlet allocation (LDA) was applied to the agent's logs to identify semantic themes.
As depicted in Fig. 7, the analysis revealed a predominance of terms related to urban features and navigation, with less emphasis on movement, obstacles, and visibility. The topic model indicated a strong focus on direct path-following and target-seeking behaviors, with limited exploration or route optimization. The Natural Language Toolkit (NLTK) and GPT-4 were utilized for textual analysis and initial semantic categorization.
The distribution of topics suggests a goal-oriented focus, with the agent primarily concentrating on reaching its target. This focus appears to limit consideration of alternative routes and exploratory behaviors. As shown in Fig. 5, the 3D environment included potential shortcuts and detours to assess the agent's adaptability. Despite these features, the agent

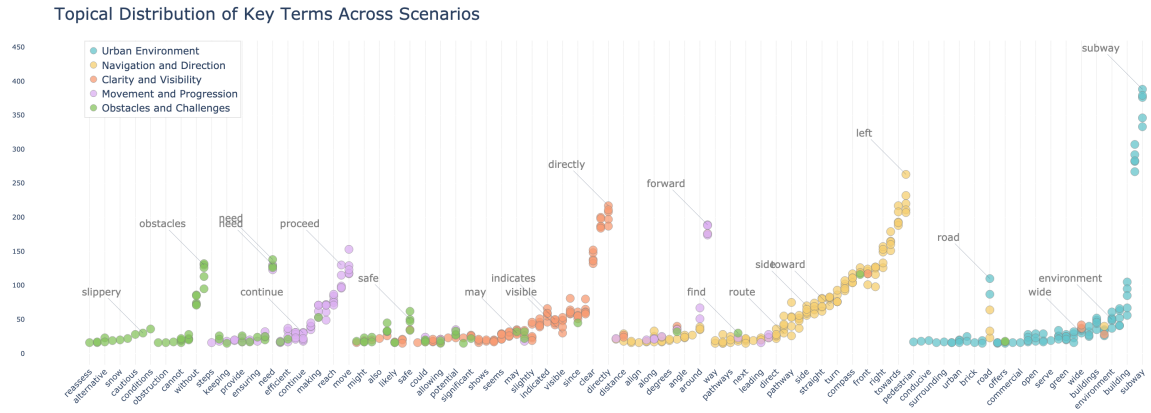

Figure 7: Topical modelling of agent's observation and plan across scenarios. The key terms in the agent's observation and plan are clustered into five major topics: navigation, visibility, movement, obstacles, and urban environment. First, an NLP method is used to extract the main topics from the entire corpus of the agent's plan and observation. Then, each of the top terms is assigned a topic based on its semantic content, and the distribution of topics is visualized across scenarios.

consistently followed the main road, indicating a deterministic decision-making process.
The agent's behavior may be influenced by its initial conditions, prompting reliance on prior knowledge and compass rather than adapting to new opportunities. Further research is required to explore the agent's adaptability and decision-making in more dynamic and open-ended environments.

5.4 Sentiment Analysis

To evaluate the agent's emotional state in both successful and failed paths, we conducted sentiment analysis using the Natural Language Toolkit's (NLTK) VADER sentiment analyzer**cite**. The agent's 'thought' and 'observation' streams were classified into three categories: positive, neutral, and negative. The results, visualized in Fig. 8, indicate a distinction between successful and failed paths. In both cases, the agent's sentiment is mostly positive, and occasional negative sentiment can be observed. However, failed paths exhibit clusters of both negative sentiments and iterative actions ('search' steps), indicating the agent's 'frustration' and 'confusion' when unable to find its path. In contrast, successful paths show a more positive sentiment, with fewer search steps and a clearer path to the target.
As described in 4.2.1, the agent is given an additional task upon reaching its goal. Similar to failed paths, in additional tasks, the agent's sentiment is more negative and 'search' oriented, as it is required to adapt to new information and make additional decisions. This suggests that the agent's adaptability and decision-making process are influenced by the complexity of the task and can influence the agent's 'emotional' state. Further research is needed to explore the agent's emotional responses and adaptability in more dynamic and complex environments.

6 Discussion

This paper presents TravelAgent (TA), a novel platform for simulating human behavior in urban environments. TA integrates generative agents, computer vision, and Chain-of-Thought (CoT) frameworks to model human-like decision-making processes, experiences, and interactions in virtual environments. The platform enables users to create and evaluate TA in various scenarios, providing insights into agent-environment behavior and urban design. Experiments conducted in this study demonstrate the effectiveness of TA in simulating behavior and assessing user interactions in urban environments. The results of the case studies highlight the platform's potential for evaluating design decisions, predicting user interactions, and informing urban design and architecture. This section discusses the implications of TA for urban design and architecture, as well as the opportunities and challenges of using generative agents in simulating human behavior.

6.1 Implications for Urban Design and Architecture

The findings from the TA simulations have several potential implications for urban design and architecture, particularly in the areas of way-finding, environmental legibility, and user experience. Below, we discuss three key implications:

Figure 8: Sentiment analysis of the agent's thoughts and observations using NLTK's VADER sentiment analyzer. Sentiments are classified into three categories: positive (blue), neutral (gray), and negative (light pink). The darkest red represents the "Finish" step, where the agent is given an additional randomized task. Notably, in failed paths, there are clear clusters of 'search' actions, i.e. the agent aims to reroute itself or find a new path by looking left and right repeatedly. This then leads to a negative sentiment, as the agent is unable to find its path. In successful paths, the agent's sentiment is more positive, and search steps are less frequent.

1. way-finding and Navigation: The ability of TA to simulate human-like navigation and decision-making processes can provide insights into how people interact with urban environments. By analyzing the paths taken by agents and their decision points, designers can identify potential way-finding challenges and optimize the layout of streets, signage, and landmarks to improve navigability. For example, the clustering of decision points in the 'Winter' scenario suggests that certain environmental cues, such as clear sightlines and distinctive landmarks, can significantly aid navigation. Incorporating these elements into urban design can enhance way-finding and reduce the cognitive load on pedestrians. Similarly, TA can be adapted to investigate indoor design and way-finding in complex buildings, such as hospitals, airports, and shopping malls. By simulating human behavior in these environments, designers can assess the effectiveness of different way-finding strategies, such as signage, lighting, and spatial layout, and identify areas for improvement. Unlike traditional ABM, the usage of Generative Agents in these simulations can provide a more nuanced understanding of user interactions and preferences, enabling designers to create more intuitive and user-friendly environments.
2. Environmental Legibility: Environmental legibility refers to the ease with which people can understand and navigate a space **cite**. As discussed in 5, TA simulations highlight the importance of visual cues, spatial memory, and prior knowledge in shaping human behavior. By examining the agents' observations and plans, designers can assess the legibility of different design options and make informed decisions to enhance the clarity and coherence of urban spaces. For example, the prevalence of terms related to urban features and navigation in the agent's cognitive outputs suggests that clear landmarks, visual cues, and spatial organization can improve environmental legibility.
3. Experience and Safety: Sentiment analysis can provide insights into the emotional responses of users to different environments. By identifying areas where agents experience frustration or confusion, designers can address potential safety concerns and improve the overall experience of users. For example, the clustering of negative sentiments and search steps in failed paths might suggest that certain areas of the environment can be more confusing or challenging

to navigate than others. By simulating a verity of agents with different profiles and needs, designers can evaluate the impact of design decisions on user experience, safety, and well-being, and make informed choices to create more inclusive and accessible environments.

6.2 Limitations and Future Work

While the TA platform demonstrates potential in simulating human behavior in urban environments, several limitations currently constrain its effectiveness. This section outlines key limitations and suggests future directions for enhancing the platform's capabilities.

Environmental Complexity and Dynamics: The current implementation of TA is confined to relatively simple environments with limited goals and interactions. To more accurately reflect real-world urban settings, future work should propose environmental complexity by incorporating a wider dynamic elements and social interactions. This includes integrating varying building types, dynamic weather conditions, and fluctuating pedestrian and vehicular traffic. Additionally, enhancing the platform to support multi-agent interactions would enable the study of social dynamics and collective behaviors, such as crowd movements during public events and pedestrian flows in busy streets. Incorporating agent-to-agent communication, group dynamics, and flocking behaviors can better mirror the complexity of human interactions in built environments.

Computational Efficiency: The present TA framework demands substantial computational resources to generate sensory inputs, process agent behaviors, and perform Chain-of-Thought (CoT) analysis. Each simulation step involves generating street-level images, segmentation maps, depth estimations, CoT reasoning, and memory management, together resulting in significant processing time. Future efforts should focus on optimizing the platform's performance to reduce simulation time and resource requirements. This may involve paralleling processes and offloading some of the agent's decision-making to traditional ABM logic.

Agent Diversity and Personalization: Currently, TA agents may not fully capture the diversity of user groups and their unique needs. Future work should aim to develop more sophisticated agent profiles that reflect varying demographic characteristics such as age, mobility, cultural background, and personal preferences. For instance, simulating agents representing elderly users, individuals with disabilities, or children would enable designers to evaluate the accessibility and inclusivity of urban spaces.

Validation and Real-World Integration: While TA provides valuable insights into human-like interactions within urban environments, further validation is necessary to assess its accuracy and reliability. Future research should involve comparing TA simulations with real-world observations and traditional agent-based models, as well as conducting user studies to evaluate its predictive capabilities. Incorporating real-time data feeds, such as traffic and pedestrian patterns, could also enhance the realism and applicability of the simulations.

Applications in Urban Safety and Policy Planning: The perception of safety in urban environments is influenced by factors like lighting, visibility, and social cues **cite**. TA can be utilized to simulate and assess different design options for improving urban safety and security. Additionally, simulating the impact of policy interventions—such as changes in zoning regulations, transportation infrastructure, and public space design—can help policymakers anticipate potential outcomes and make more informed decisions that address the needs of urban residents.

By addressing these opportunities, future TravelAgent platforms can evolve into an indispensable tool for advancing urban design and human behavior simulation. Similar to other CAD and BIM tools, TA can become an integral part of the design process, enabling designers to evaluate and optimize their designs based on human-centric criteria and agent-driven user feedback.

References

Ariel Noyman. *CityScope: An Urban Modeling and Simulation Platform*. PhD thesis, Massachusetts Institute of Technology, 2022.

Bo Chen and Harry H. Cheng. A Review of the Applications of Agent Technology in Traffic and Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):485–497, June 2010. ISSN 1524-9050, 1558-0016. doi:10.1109/TITS.2010.2048313.

Michael Batty. Multiple models. *Environment and Planning B: Urban Analytics and City Science*, 48(8):2129–2132, 2021. ISSN 2399-8083, 2399-8091. doi:10.1177/23998083211051139.

Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)*, 55(1):1–44, 2021.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Machel Reid, Nikolay Savinov, Denis Teplyashin, Dmitry Lepikhin, Timothy Lillicrap, Jean-baptiste Alayrac, Radu Soricut, Angeliki Lazaridou, Orhan Firat, Julian Schrittwieser, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

Partha Pratim Ray. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3:121–154, 2023. ISSN 2667-3452. doi:10.1016/j.iotcps.2023.04.003.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22, 2023.

Zhao Kaiya, Michelangelo Naim, Jovana Kondic, Manuel Cortes, Jiaxin Ge, Shuying Luo, Guangyu Robert Yang, and Andrew Ahn. Lyfe Agents: Generative agents for low-cost real-time social interactions, 2023.

Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl_3):7280–7287, May 2002. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.082080899.

Liang Chen. Agent-based modeling in urban and architectural research: A brief literature review. *Frontiers of Architectural Research*, 1(2):166–177, 2012. ISSN 2095-2635. doi:10.1016/j.foar.2012.03.003.

Kevin Lynch. *The image of the city*. MIT press, 1960.

Young Ook Kim. *Spatial Configuration, Spatial Cognition and Spatial Behaviour: the role of architectural intelligibility in shaping spatial experience*. University of London, University College London (United Kingdom), 1999.

Jane Jacobs. *The death and life of great American cities*. Random House LLC, 1961.

Andy Smith, Martin Dodge, and Simon Doyle. *Visual communication in urban planning and urban design*. University College London, Centre for Advanced Spatial Analysis (CASA), 1998.

Michael Batty, David Chapman, Steve Evans, Mordechai Haklay, Stefan Kueppers, Naru Shiode, Andy Smith, and Paul M Torrens. Visualizing the city: communicating urban design to planners and decision-makers. 2000.

David Fonseca, Francesc Valls, Ernest Redondo, and Sergi Villagrasa. Informal interactions in 3d education: Citizenship participation and assessment of virtual urban proposals. *Computers in Human Behavior*, 55:504–518, 2016.

Daniel R Montello. Scale and multiple psychologies of space. *Spatial information theory a theoretical basis for GIS*, pages 312–321, 1993.

Reginald G Golledge. Human wayfinding and cognitive maps. *Wayfinding behavior: Cognitive mapping and other spatial processes*, pages 5–45, 1999.

Michal Gath-Morad, Jascha Grübel, Koen Steemers, Kerstin Sailer, Lola Ben-Alon, Christoph Hölscher, and Leonel Aguilar. The role of strategic visibility in shaping wayfinding behavior in multilevel buildings. *Scientific Reports*, 14 (1):3735, 2024.

Amir-Homayoun Javadi, Beata Emo, Luke R Howard, and et al. Hippocampal and prefrontal processing of network topology to simulate the future. *Nature Communications*, 8(1):1–11, 2017. doi:10.1038/s41467-017-00112-0.

Russell A Epstein, Eva Z Patai, Joshua B Julian, and Hugo J Spiers. The cognitive map in humans: Spatial navigation and beyond. *Nature neuroscience*, 20(11):1504–1513, 2017. doi:10.1038/nn.4656.

Steven C. Bankes. Agent-based modeling: A revolution? *Proceedings of the National Academy of Sciences*, 99 (suppl_3):7199–7200, May 2002. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.072081299.

Johannes Nguyen, Simon T. Powers, Neil Urquhart, Thomas Farrenkopf, and Michael Guckert. An overview of agent-based traffic simulators. *Transportation Research Interdisciplinary Perspectives*, 12:100486, December 2021. ISSN 25901982. doi:10.1016/j.trip.2021.100486.

Gabriele Filomena and Judith A. Verstegen. Modelling the effect of landmarks on pedestrian dynamics in urban environments. *Computers, Environment and Urban Systems*, 86:101573, March 2021. ISSN 01989715. doi:10.1016/j.compenvurbsys.2020.101573.

Antonio Luca Alfeo, Eduardo Castelló Ferrer, Yago Lizarribar Carrillo, Arnaud Grignard, Luis Alonso Pastor, Dylan T. Sleeper, Mario G. C. A. Cimino, Bruno Lepri, Gigliola Vaglini, Kent Larson, Marco Dorigo, and Alex 'Sandy' Pentland. Urban Swarms: A new approach for autonomous waste management. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4233–4240, May 2019. doi:10.1109/ICRA.2019.8794020.

Steven F Railsback and Volker Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2019.

Akihiro Kobayashi, Naoto Takeda, Yudai Yamazaki, and Daisuke Kamisaka. Modeling and generating human mobility trajectories using transformer with day encoding. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, pages 7–10, 2023.

Ayush Chopra, Shashank Kumar, Nurullah Giray-Kuru, Ramesh Raskar, and Arnau Quera-Bofarull. On the limits of agency in agent-based models, October 2024.

Parfait Atchade, Adrian Mora, Luis Alonso-Pastor, Arnaud Grignard, Ariel Noyman, Leticia Izquierdo, Carlo Adornetto, Kai Hu, Fernando Fernandez, Hossein Rahnama, Margaret Church, Markus ElKatsha, and Kent Larson. Humanized Agent-based Models: A Framework, August 2024.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing Reasoning and Acting in Language Models, March 2023.

Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. Understanding the planning of LLM agents: A survey, February 2024.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, January 2023.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of Thoughts: Deliberate Problem Solving with Large Language Models.

Kostas Hatalis, Despina Christou, Joshua Myers, Steven Jones, Keith Lambert, Adam Amos-Binks, Zohreh Dannenhauer, and Dustin Dannenhauer. Memory Matters: The Need to Improve Long-Term Memory in LLM-Agents. *Proceedings of the AAAI Symposium Series*, 2(1):277–280, January 2024. ISSN 2994-4317. doi:10.1609/aaaiss.v2i1.27688.

Iman Mirzadeh, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, and Mehrdad Farajtabar. GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models, October 2024.

Appendix

Appendix A: Scenario Descriptions

- *Scenario 1 - 'Base':* A bright summer morning in Kendall Square, Cambridge, MA. Modern glass buildings and bustling streets with pedestrians, cyclists, and outdoor cafés creating a lively and vibrant atmosphere.
- *Scenario 2 - 'Winter':* A snowy winter morning in Kendall Square, Cambridge, MA. Heavy snow blankets the streets and modern buildings, with bundled-up pedestrians, and a quiet stillness filling the air.
- *Scenario 3 -'Tokyo':* A bright summer morning in Tokyo, Japan. Mix of traditional wooden buildings and modern structures. Bustling streets filled with pedestrians, cyclists, and outdoor tea houses creating a vibrant atmosphere.
- *Scenario 4 - 'Night':* A quite summer nighttime street scape in Kendall Square, Cambridge, MA. Streets are softly illuminated by warm streetlights and the gentle glow of modern office buildings with large glass facades.
- *Scenario 5 - 'Persona':* The agent's persona was varied between a 30-year-old female and a male around the same age. This was meant to examine wether the agent's gender influenced its decision-making process.

Appendix B: Scenario Matrix

Season	Location	Time	Persona	Scene
Summer	Kendall Square, Boston	Morning	30-year-old female/male	Scene 1
Winter	Kendall Square, Boston	Morning	30-year-old female/male	Scene 2
Summer	Tokyo	Morning	30-year-old female/male	Scene 3
Summer	Kendall Square, Boston	Night	30-year-old female/male	Scene 4

Table 1: Associations between Season, Location, Time, Persona, and Scene