Exhibit D – Research Project Requirement Template

Exploring Top-Down Visual Attention for Transportation Behavior Analysis: Walkability and Pedestrian Behaviors

Recipient/Grant (Contract) Number: The University of Texas at Austin; The City College of New York/Grant # 69A3552344815 and 69A3552348320

Center Name: National Center for Understanding Future Travel Behavior and Demand (TBD)

Research Priority: Improving Mobility of People and Goods

Principal Investigator(s): Zhigang Zhu, Alison Conway

Project Partners/Collaborators: Zihao Zhang, Landscape Design, Spitzer School of Architecture; Professor Xuan Wang, Machine Learning, CUNY City Tech

Research Project Funding: $180,000

Project Start and End Date: 06/01/2025 - 08/31/2026

Project Description: Employing the state-of-the-art research on attention and feedback mechanisms, especially with vision language models (VLMs), has not been fully explored previously for transportation behavior analysis, especially for the analysis of a variety of pedestrian behaviors related to sidewalks and streets. This project will expand our previous project exploring top-down visual attention for transportation behavior analysis into integrated research on both pedestrian behaviors and urban walkability by analyzing micro-level street scenes using deep learning models focusing primarily on VLMs. On one hand, walkability indices play an essential role in urban planning by providing guidelines to enhance street walkability for pedestrians. However, widely used indices such as Walk Score and the National Walkability Index only consider macro-level features. These include measures such as neighborhood population density, accessibility to amenities, proximity to transit stops, and the ratio of employment to housing. In response to these limitations, recent research increasingly incorporates semantic segmentation techniques in computer vision to extract micro-level features from street view images. These features are either used directly to predict walkability or combined with macro-level attributes to create new, more comprehensive indexes that capture both scales of influence. On the other hand, a great deal of pedestrian behaviors, particularly in urban settings like New York, would also be influenced by dynamic characteristics of the pedestrian environment. For example, the presence of other walkers, intersection cross traffic, delivery people, street entertainers, pan handlers, parked bikes and scooters, people walking pets, and other obstructions would be the dominant contributors to pedestrian behavior at the micro scale.

As many of these elements are beyond the obvious policy or design control, we aim to identify the interactions of these dynamic characteristics of pedestrians (as a behavior understanding task) with sidewalk and intersection designs (as a walkability analysis task) to derive actionable knowledge or information to improve the performance of both tasks and hence to influence design or policy for sidewalks and streets. By building on cognitive insights (e.g., that pedestrians look at key safety cues) and leveraging bio-inspired vision language model (VLM) architecture, the integration of top-down attention modeling with AI-driven image analysis promises more nuanced, human-centric metrics for walkable urban design and safer and more attractive pedestrian environments.

US DOT Priorities: This proposal addresses the following aspects of US DOT's transformation research priorities: (1). ***Integrated System of Systems***, including: Attributes of a Desired Transportation System of the Future (People-Centered, Data Driven and Intelligent) in terms of pedestrian safety in a metropolitan setting with complicated traffic situations; and Transportation Grand Challenge (Automation, AI and

Machine Learning) to derive information of both the actors (pedestrians) and the settings (sidewalks and streets). (2). ***Data-driven Insight***, especially to conduct exploratory research on transformational mobility data analytics with both images/videos and transportation-related metadata using large language models (LLMs) and VLMs. (3). ***New and Novel Technologies,*** especially to develop best practices for safe interaction of automated roadway vehicles with existing vehicles, roadside hardware, emergency responders, pedestrians, cyclists, and motorcyclists using the cutting-edge AI and LLMs.

Outputs: The outputs of the project include: (1) A ***comprehensive survey*** on the topic by the team on urban walkability and pedestrian behaviors using vision language models, which would be submitted to a journal. Unlike traditional walkability studies that rely on static built-environment metrics or recent big-data models that treat pedestrian flows as numerical patterns, this project investigates a top-down visual-attention framework, drawn from how people consciously scan their surroundings, and pairs it with the rich semantic insights of large vision language models to achieve deeper, high-level interpretations of urban scenes. (2) Preliminary results on using vision language models (VLMs) to analyze ***pedestrian behaviors*** by incorporating ***micro-level walkability*** of urban sidewalks, aiming to publish one or two conference papers. The visual attention mechanisms will be integrated in the VLMs by injecting domain-specific knowledge into the analysis of both image and text data. The domain knowledge includes dynamic transportation information, roadside hardware information, environmental conditions such as vegetation, location-based information (maps, events, tasks), as well as pedestrian behaviors and attention. (3) ***Multimodal pedestrian-scene interaction analytic tools*** for data labeling, analytics and visualization, aiming for a technical report and possibly a dataset. These include audio, visual and haptic features as well as accessibility functions that our team has studied for helping the navigation of people who are blind or have low vision. This will allow developers, engineers and users to access the multimodal walkability map of an urban scenes for interaction, interpretation and diagnosis. (4) ***A GitHub page*** will be created for each paper to be submitted to the premier vision conference CVPR or a transportation conference such as TRB, while posting on arXiv.

Outcomes/Impacts: The top-down approach implemented in this project integrates transportation-domain knowledge to the data-driven AI models, focusing primarily on vision language models (VLMs). The most immediate technical contribution of this project will be to promote task-specific deployment of these models to automate the capture of static and dynamic factors influencing micro-scale pedestrian behaviors. The comprehensive survey will provide insights to transportation researchers for the state-of-the-art AI/ML models and algorithms for pedestrian behavior analysis in the context of urban walkability using computer vision. This establishes a foundation and increases in the body of knowledge for applied researchers in transportation for real-world problems. In particular, this project has the potential to inform improvements to a number of transportation system applications including: (1) improved consideration of micro-scale features in pedestrian behavioral modeling; (2) development of new tools for pedestrian infrastructure design, performance monitoring, and asset management; (3) integration of additional micro-scale pedestrian factors into infrastructure and vehicle control systems; and (4) development of information-rich wayfinding tools. Ultimately, implementation of VLMs to inform these applications can help to (1) improve system safety by better aligning design and control systems with real pedestrian behaviors; (2) improve system user experiences by enabling consideration of micro-scale factors in wayfinding; and (3) provide efficient mechanisms for collecting pedestrian infrastructure and behavior data to inform planning, design, and project prioritization.

Final Research Report: A URL link to the final report will be provided upon completion of the project.