Do VLMs See What Sensors Feel? A Scalable Expert-Guided Design for Wheelchair Accessibility Assessment from Street View

Anonymous ACL submission

Abstract

Assessing built-environment interaction, such as wheelchair accessibility, is difficult because real-world mobility is shaped by distributed, context-dependent, and temporary barriers that are hard to capture at scale. To support scalable assessment, this paper examines whether vision-language models (VLMs) can identify accessibility barriers from Google Street View (GSV) imagery. We propose an expert-guided retrieval-augmented framework that combines GSV images, ADA-informed guidance, and expert-derived rubrics to evaluate accessibility dimensions. We collect a campus-scale dataset at the University of Florida, linking 407 unique GSV locations with GPS-derived wheelchair dwell behavior as a mobility-friction signal. Results show that VLM ratings are both negatively correlated and distributionally similar with dwell time, indicating partial but consistent alignment with a behavioral proxy for mobility friction. Visual cue analysis shows that certain environmental objects, such as curb ramps and crosswalks, are associated with higher VLM accessibility scores, while alignment remains limited for subtle surface conditions, transient obstructions, and viewpoint-dependent barriers. Overall, our findings show the potential of expert-guided VLMs for scalable accessibility assessment aligning with sensor-derived indicators of real-world wheelchair navigation.

1 Introduction

Wheelchair accessibility in urban systems is crucial for independent movement, autonomy, and social participation among people with mobility disabilities. Although the Americans with Disabilities Act (ADA) and related standards have improved baseline accessibility (U.S. Department of Justice, 2010), minimum compliance does not always ensure environments are comfortable, continuous, safe, or usable in everyday navigation (Kapsalis et al., 2024; Prescott et al., 2021). Wheelchair users still encounter uneven pavements (Kapsalis et al., 2024; Duvall et al., 2013), poorly designed curb ramps (Kapsalis et al., 2024; Frost and Bertocci, 2010; Wretstrand et al., 2010; Velho et al., 2016), obstructed pathways, steep slopes (D'Souza et al., 2019), surface discontinuities (Kapsalis et al., 2024; Bentzen et al., 2020), and other barriers that interrupt movement. Because these barriers are spatially distributed, context-dependent, and sometimes temporary, they are difficult to document through occasional assessments (Froehlich et al., 2019).

Because these barriers are difficult to capture systematically, current assessments rely heavily on manual site visits. For instance, university accessibility offices often cross-reference student schedules with physical audits of campus transit paths, a labor-intensive and unscalable process. To address this limitation, we use Google Street View as a scalable visual source for auditing outdoor environments. Prior studies suggest that VLMs can interpret Street View imagery for built-environment understanding, including urban perception and navigation-related reasoning (Zhang et al., 2025; Li et al., 2024b; Ro et al., 2025; Wang et al., 2026; Schumann et al., 2024). We therefore examine whether VLMs can evaluate accessibility by integrating visual cues, contextual reasoning, and domain-specific criteria.

A further model-level gap is whether VLMs can interpret Street View imagery in ways that correspond to people's physical experience. This is central to accessibility assessment, where the target is not only infrastructure appearance but the embodied interaction between wheelchair users and the environment. Although VLMs may identify problematic features in Street View imagery, it remains unclear whether visually inferred barriers align with **behavioral signals from real-world movement** such as prolonged dwelling of wheelchair users. These gaps motivate two research questions:

- **RQ1:** How well can VLMs identify wheelchair accessibility barriers from Street View imagery with and without expert-guided augmentation?
- **RQ2:** What geospatial patterns and visual cues explain correspondence and divergence between VLM and wheelchair sensor patterns?

To answer these questions, we propose an expert-guided system that uses VLMs to assess wheelchair accessibility barriers from Street View imagery and compare them with wheelchair sensor traces from the University of Florida campus. Our study makes three contributions: (i) a structured VLM pipeline for assessing barriers from Street View imagery; (ii) an expert-guided augmentation strategy integrating accessibility standards and expert-informed criteria; and (iii) a campus-scale benchmark linking Street View imagery with wheelchair sensor traces to identify where VLM-based assessments align with or diverge from sensor-derived mobility patterns. These contributions position our VLM pipeline as a scalable screening layer for accessibility assessment, helping urban planners and accessibility offices evaluate the built environment.

2 Related Work

We frame related work as follows: prior work tells us *what* to look for, but does not *scale*; Existing VLM studies scale but are not *grounded*. Our paper sits at the intersection.

2.1 Existing accessibility assessment methods

Manual Audits. ADA standards have improved physical accessibility but often fail to address the dynamic needs of users with mobility challenges. Field-audit approaches use ADA checklists to manually evaluate accessibility (New England ADA Center, 2014), providing explicit criteria for identifying barriers and remaining central to facility audits, sidewalk assessments, and campus reviews. However, they are labor-intensive, episodic, focused on minimum compliance rather than everyday wheelchair navigation, and not designed for end-user tools (Froehlich et al., 2019).

GIS/Crowdsourcing Platforms. GIS-based inventories and routing platforms like AccessMap evaluate accessibility through connected infrastructure networks (Li et al., 2018; Bolten and Caspi, 2019). To scale these efforts, crowdsourced systems like Project Sidewalk use Street View imagery and public participation (Hara et al., 2013). However, both face data bottlenecks: GIS models depend on meticulously maintained infrastructure data, while crowdsourcing relies on static labels that lack coverage or direct evidence of wheelchair users' lived experiences.

Sensors. Recent work explores wheelchair-mounted sensor systems to capture the embodied experience (Gatmaitan et al., 2025), such as inertial sensors/accelerometers (Gendle et al., 2012; Pinnock Branford et al., 2025), air quality sensors (Park et al., 2022), ego-centric cameras (Kutbi et al., 2020), and laser-enabled robotic arms (Zhong et al., 2019). By logging continuous data during transit, these sensors identify dynamic, real-world barriers (e.g., cracked pavement, sudden maneuvers, inaccessible doors) that static visual audits miss. While this grounds accessibility assessment in actual user mobility, it requires physical traversal within the environment, reinforcing the need for complementary and scalable methods.

2.2 VLMs for built-environment study

With the advance of large language models (LLMs), VLMs like LLaVA (Liu et al., 2023), Qwen-VL (Bai et al., 2023), and InternVL (Chen et al., 2024) have been developed to extend LLM capabilities by integrating visual perception with language reasoning, enabling advanced multimodal scene understanding and semantic environment analysis (Yang et al., 2024; Li et al., 2025b; Feng et al., 2025; Howell et al., 2026). In built-environment research, VLMs have been applied to Street View scene analysis, urban perception scoring, and navigation-related reasoning (Li et al., 2024a; Zhou et al., 2024; Wang et al., 2026), demonstrating an ability to interpret environmental context from visual input. More recent work shows that VLMs can also serve as world-model simulators: Street View-based reasoning has been used to estimate large-scale environmental events such as earthquake shaking experience (Li et al., 2025a), infer neighborhood health conditions (Du et al., 2025), and assess perceived walkability and safety from street-level imagery (Blečić et al., 2024; Tang et al., 2025). These results suggest that VLMs are well-positioned for scalable and explainable evaluation.

Despite these advances, applying VLMs to wheelchair accessibility assessment remains largely unexplored. Existing applications can hallucinate, lack grounding in lived-experience knowledge, and are rarely validated against physical sensors, which is consequential when assessments inform mobil-

ity decisions. To address these gaps, we develop a scalable pipeline that grounds VLM reasoning over Street View imagery in design standards and validates it against wheelchair mobility patterns.

3 Data and Methods

Figure 1 illustrates our overall workflow. We first introduce the pilot study area in Section 3.1. We then describe wheelchair-mounted GPS data collection in Section 3.2, which is used to derive real-world mobility-friction patterns. Section 3.3 explains how corresponding Street View images are retrieved along the same routes for scalable visual observations. Sensor data and Street View images capture complementary views of campus accessibility: lived wheelchair mobility and scalable visual context. Building on these paired data, Section 3.4 presents our expert-guided VLM pipeline for assessing potential accessibility barriers. Section 3.5 describes how model outputs are designed to interpret accessibility scores and reasoning. Finally, Section 3.6 explains how we compare VLM-derived assessments with GPS-derived mobility-friction patterns to evaluate alignment.

3.1 Study area

The study is conducted on the University of Florida (UF) campus in Gainesville, Florida, USA. UF provides a suitable setting for wheelchair accessibility assessment due to its heterogeneous built environment, including a mix of historic and modern infrastructure, varied sidewalk materials, curb ramps, and crosswalks. These varied route conditions frequently generate mobility barriers stemming from infrastructure design, surface quality, or temporary disruptions. Furthermore, as a major educational institution, the campus attracts a significant population of wheelchair users; it serves as a prime example of an environment that largely satisfies baseline ADA audits yet continues to present practical, everyday friction for wheelchair mobility.

To capture realistic accessibility experiences, we collect sensor data along high-traffic routes connecting key campus destinations (e.g., classrooms, libraries, and the student union). By focusing on everyday travel rather than artificially constructed scenarios, we can meaningfully evaluate whether Street View-based assessments align with actual wheelchair mobility patterns.

3.2 Wheelchair sensor data

We collect wheelchair-mounted GPS data across multiple campus trips using a smartphone attached to a manual wheelchair. The GPS streams record latitude and longitude during traversal and are used to reconstruct wheelchair trajectories and derive dwell behavior. In this study, dwell time serves as a grounded mobility-friction signal for comparing Street View-based VLM accessibility ratings with observed wheelchair movement.

We define dwell points as locations where the wheelchair remained within a localized area for an extended period rather than continuing along the route. Longer dwell time may indicate environmental or infrastructural conditions that slow or interrupt navigation, such as inaccessible entrances, blocked paths, uneven terrain, construction, congestion, or route-finding around inaccessible features. The processed dataset contains 1,253 dwell point candidates and 7,201 non-dwell points, from which we retained a 10% random sample of 720 non-dwell points for comparison. Details on dwell-point classification and contextual validation are provided in Appendices A and B.

3.3 Google Street View imagery

From the initial 8,443 collected locations, we perform a deduplication process to remove redundant Street View samples caused by the spatial resolution of Street View imagery. A 10-m distance threshold is applied to filter nearby duplicate locations. We further remove locations without valid Street View imagery, resulting in a final set of 407 unique locations. For each location, we extract Street View imagery from eight panoramic viewing angles to provide comprehensive visual coverage of the surrounding.

3.4 VLM pipeline design

We develop a retrieval-augmented multimodal reasoning framework for wheelchair accessibility assessment using Street View imagery (Figure 1). The pipeline first uses a VLM to identify accessibility-related environmental features (see Appendix F), including curb ramps, sidewalks, crosswalks, surface conditions, and obstacles. These features are then used to retrieve relevant accessibility guidance from a compact knowledge base distilled from ADA standards, UF campus design guidelines, PROWAG, and expert-derived heuristics.

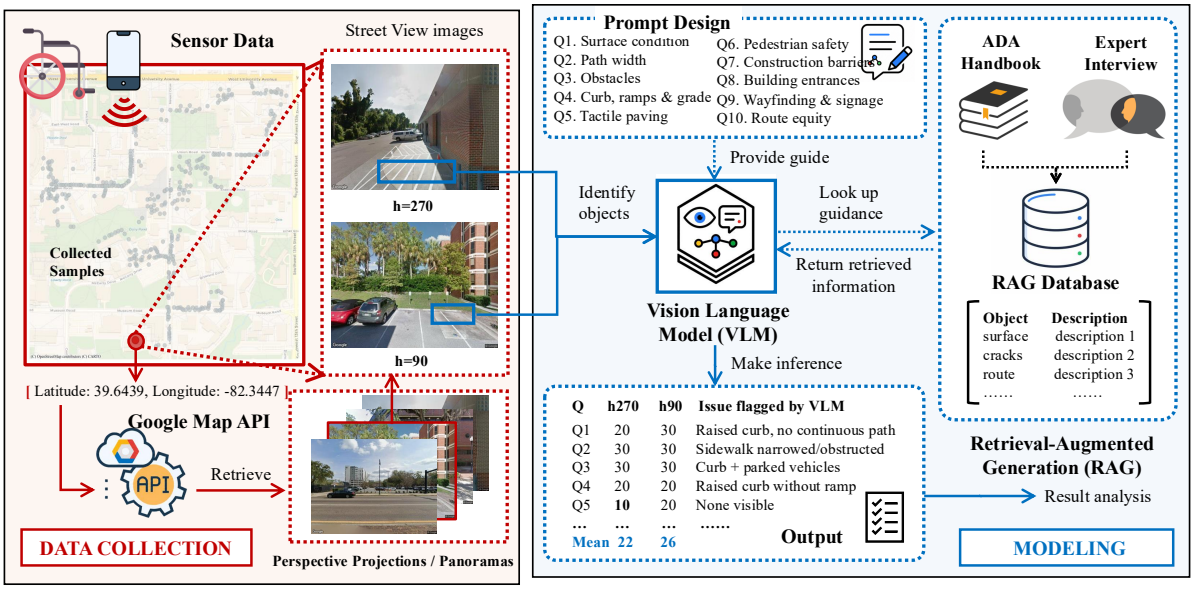

Figure 1: An illustration of research framework design.

We evaluate three prompting configurations: (i) a base VLM with a generic accessibility prompt, (ii) handbook augmentation with retrieved regulatory guidance, and (iii) expert-opinion augmentation incorporating rubric-based knowledge from wheelchair users and accessibility professionals. To improve grounding reliability, the framework constrains reasoning to observable visual evidence while discouraging unsupported assumptions about non-visible infrastructure. The model evaluates sidewalk continuity, curb transitions, pedestrian safety, and obstacles, and outputs a structured accessibility score ranging from 10 (very poor) to 50 (excellent) with concise justification. Details of prompt and RAG design are presented in Appendices I and J.

3.5 VLM output framework

We use a structured output schema to record per-image assessments of *surface condition*, *cross-slope*, *obstructions*, *curb-ramp presence and quality*, and *width adequacy*, together with an overall accessibility score and a free-text justification. This schema makes VLM outputs comparable across images, aggregable across locations, and auditable during error analysis (see Appendix J).

We then aggregate per-image judgments into segment- and route-level accessibility scores. This aggregation can account for image-level confidence, repeated viewpoints, and spatial redundancy, allowing VLM-derived assessments to be compared with sensor-derived mobility segments.

3.6 Evaluation

Agreement between sensor-derived mobility-friction patterns and VLM-based accessibility assessments is quantified using the Pearson correlation coefficient (r):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where x_i denotes the GPS-derived indicator of wheelchair mobility friction, which is dwelling time in seconds (as stated in Section 3.2), y_i represents the corresponding VLM-generated accessibility score, $\bar{x}$ and $\bar{y}$ are the sample means, and n is the total number of observations. Because longer dwell times reflect greater mobility friction, a strong negative correlation (r approaching -1) indicates high agreement between the observed mobility behavior and the VLM's multimodal accessibility reasoning. We compare r across different augmentation and multimodal input configurations to evaluate which setup yields accessibility assessments that align most with patterns observed in real-world wheelchair navigation.

We also report distributional agreement between dwell time and VLM-generated accessibility score using the Earth Mover's Distance (EMD), also known as the Wasserstein distance:

$$\begin{aligned}\mathrm{EMD}(X, Y) &= W_1(X, Y) \\ &= \inf_{\gamma \in \Pi(X,Y)} \int_{\mathbb{R}\times\mathbb{R}} |x - y| \, d\gamma(x, y),\end{aligned} \quad (2)$$

where X and Y denote the distributions of dwell time and VLM accessibility scores, respectively,

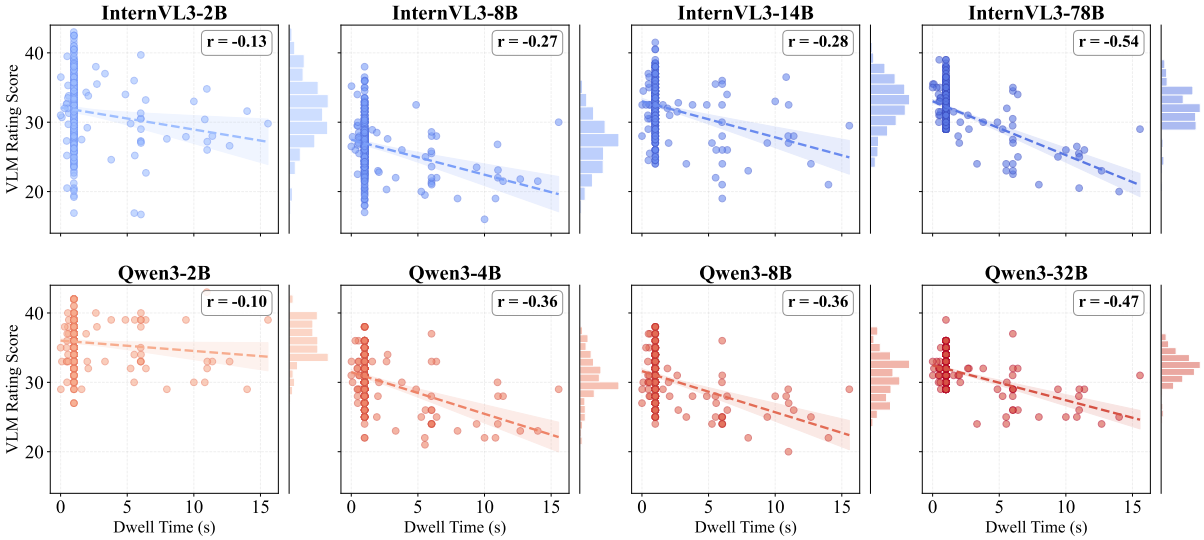

Figure 2: Correlation between wheelchair dwell time and VLM-based accessibility scores with RAG across 407 locations.

and $\Pi(X, Y)$ represents the set of all joint distributions with marginals X and Y. Lower EMD values indicate stronger distributional consistency between real-world mobility friction patterns and multimodal accessibility reasoning. Since mobility sensing data exhibit imbalanced and long-tailed distributions, we report both overall normalized EMD and quantile-based EMD, particularly Q4 with higher dwell time.

We conduct two complementary visual cue association analyses to examine semantic objects associated with VLM–sensor agreement and accessibility scoring patterns. For each object o, we compute rank differences between top- and bottom-ranked 10% regions under different criteria:

$$AMI_o = \text{rank}_{\text{align}}(o) - \text{rank}_{\text{misalign}}(o), \quad (3)$$

$$AII_o = \text{rank}_{\text{access}}(o) - \text{rank}_{\text{inaccess}}(o). \quad (4)$$

where $\text{rank}_{\text{align}}(o)$ and $\text{rank}_{\text{misalign}}(o)$ denote object frequency ranks in well- and poorly-aligned regions, while $\text{rank}_{\text{access}}(o)$ and $\text{rank}_{\text{inaccess}}(o)$ denote ranks in high- and low-accessibility regions. Smaller ranks indicate higher object frequency; thus, negative values indicate stronger association with aligned or highly accessible environments.

4 Results

4.1 Pipeline performance assessment

Quantitative Analysis. Figure 2 presents the pearson correlation, with trip-level breakdown results provided in Appendices Table 4. The results show a consistent negative correlation between dwelling time and VLM accessibility scores across both InternVL and Qwen model families. This ndicates that locations associated with prolonged wheelchair dwell times receive correspondingly lower VLM accessibility scores. These findings demonstrate the capability of VLMs to capture accessibility-related characteristics of the built environment from Street View imagery, aligning meaningfully with real-world mobility friction behavior.

Further, larger models generally demonstrate stronger agreement with GPS-derived mobility friction signals, with InternVL3-78B achieving the strongest correlation ($r = -0.54$) followed by Qwen3-32B ($r = -0.47$). In contrast, smaller variants such as InternVL3-2B and Qwen3-2B exhibit substantially weaker correlations ($r = -0.13$ and $r = -0.10$, respectively), indicating limited alignment with observed mobility friction patterns. The results suggest that greater model capacity improves a VLM's sensitivity to potential accessibility barriers reflected by GPS-derived dwell times.

Table 1: RAG versus non-RAG prompting performance comparison using r and EMD.

Model Variant	Setting	r	Overall EMD	Q4 EMD
InternVL3-78B	Non-RAG	-0.40	0.592	0.421
InternVL3-78B	RAG	-0.54	0.530	0.337
Qwen3-VL-32B	Non-RAG	-0.40	0.517	0.345
Qwen3-VL-32B	RAG	-0.47	0.420	0.246

Qualitative Analysis. Figures 3 and 4 provide representative examples illustrating success and failure cases, while Figure 10 in Appendix presents additional qualitative examples illustrating the re-

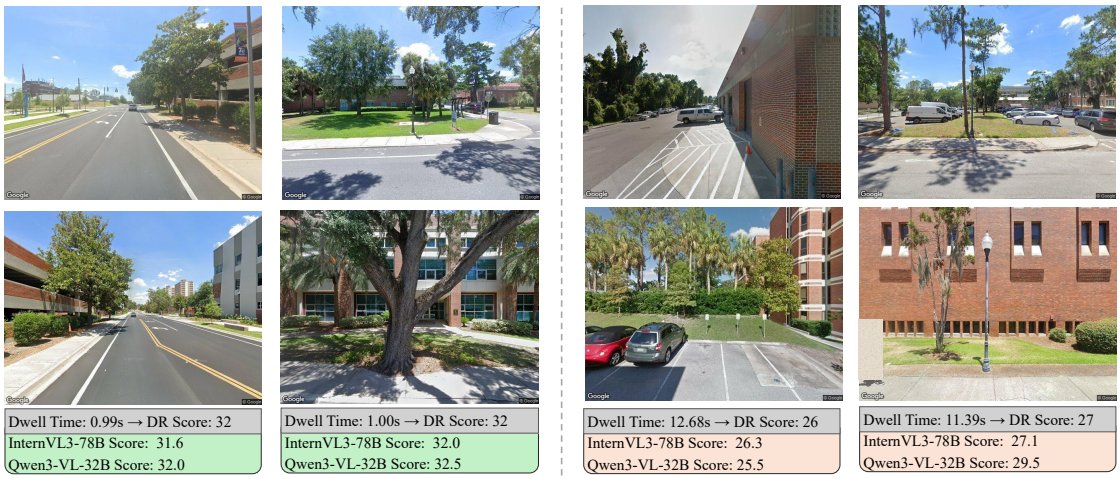

Figure 3: Success cases illustrating strong alignment between dwell time and VLM-derived accessibility scores. Columns 1–2 demonstrate accessible locations, while Columns 3–4 illustrate inaccessible locations. Dwell-derived reference (DR) scores are estimated using the fitted dwell–score correlation function.

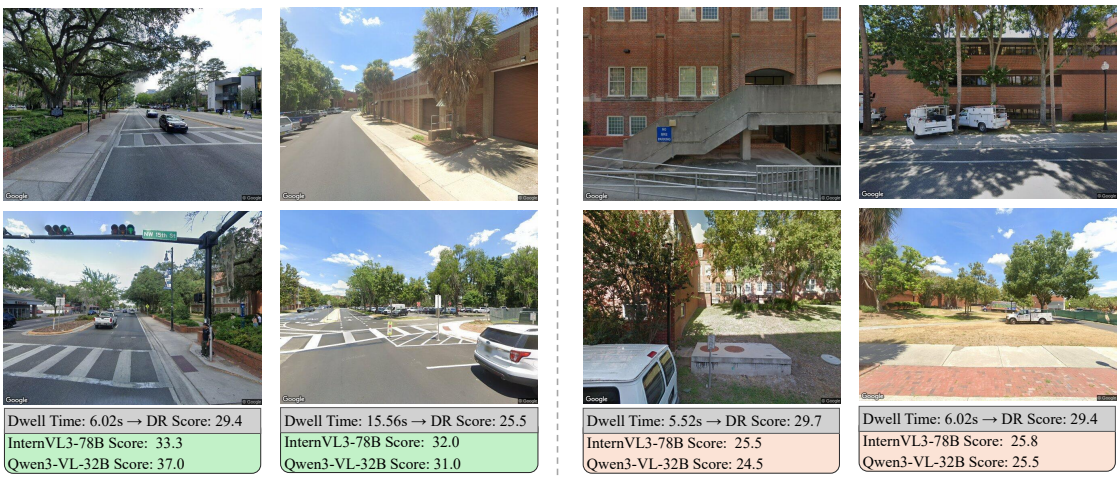

Figure 4: Failure cases illustrating weak alignment between dwell time and VLM-derived accessibility scores. Columns 1–2 demonstrate accessibility underestimation, while Columns 3–4 illustrate accessibility overestimation. Dwell-derived reference (DR) scores are estimated using the fitted dwell–score correlation function.

lationship between mobility sensing behavior and multimodal accessibility reasoning. The results demonstrate that both InternVL3-78B and Qwen3-VL-32B are sensitive to contextual pedestrian accessibility cues and partially reflect observed mobility behavior. In low dwell-time examples ($1.00s$), both models assign relatively higher accessibility scores (30–33), emphasizing smooth sidewalks, wider pathways, curb ramps, crosswalks, and fewer severe obstructions. In contrast, high dwell-time examples ($12.68s$) consistently receive lower scores (22–27), where both models identify fragmented pedestrian environments including raised curbs, missing curb ramps, constrained pathways, temporary barriers, and obstructed movement.

Figure 3 shows that the models effectively recognize accessibility-supportive pedestrian infrastructure and produce predictions closely aligned with GPS-derived mobility friction patterns. However, the failure cases in Figure 4 reveal limitations in complex environments involving ambiguous pedestrian continuity, viewpoint-dependent cues, transient obstructions, or subtle surface degradation, where models occasionally overestimate or underestimate accessibility. This shows that VLMs capture key accessibility-related environmental features while remaining sensitive to subtle pedestrian conditions and visual complexity.

4.2 Ablation study

RAG versus Non-RAG Reasoning. We first examine whether RAG improves alignment between VLM-based accessibility ratings and observed wheelchair mobility behavior. Compared with base prompting using generic accessibility instructions (Appendix E), RAG consistently yields stronger

negative correlations and lower EMD across all model variants. As shown in Table 1, RAG improves r from -0.40 to -0.54 for InternVL3-78B and from -0.40 to -0.47 for Qwen3-VL-32B, indicating stronger agreement between accessibility ratings and sensor-derived dwell time. RAG also reduces overall EMD from 0.592 to 0.530 for InternVL3-78B and from 0.517 to 0.420 for Qwen3-VL-32B, while decreasing Q4 EMD from 0.421 to 0.337 and from 0.345 to 0.246, respectively. The higher overall EMD values are partly influenced by imbalanced accessibility ratings, with most samples concentrated in moderately accessible regions. In contrast, Q4 EMD focuses on highly inaccessible locations and provides a more reliable measure under poor accessibility conditions, with lower values of approximately 0.25 to 0.35. Overall, these results suggest that grounding multimodal reasoning with retrieved ADA handbook knowledge and expert-derived guidance improves both correlation-based and distributional alignment between VLM accessibility assessment and GPS-derived wheelchair mobility friction.

Street View Perspective Projections versus Panoramas. We next examine how Street View input formats affects accessibility reasoning by comparing perspective projections (paired h=90° and h=270° perspectives) with full panoramas (Figure 5). Across both models, perspective projection inputs yield consistently stronger alignment with GPS-derived dwell time than panorama inputs. For example, InternVL3-78B achieves $r = -0.41$ with perspective projections compared to $r = -0.25$ with panorama inputs, and Qwen3-VL-32B shows a similar pattern ($r = -0.32$ vs. $r = -0.21$).

We hypothesize that this performance gap stems from both geometric and data limitations. Physically, the distortion of spherical projections can obscure fine-grained visual cues (e.g., surface texture, curb transitions, boundaries of obstructions), and panoramic views, which aggregate omnidirectional information, dilute attention compared to wide-angle configurations in perspective projection images that naturally align with a wheelchair user's forward route. Methodologically, VLMs are likely under-trained on panoramas compared to standard perspective-projection imagery

Multi-Question versus Single-Question. Table 2 shows that multi-question prompting substantially improves agreement between VLM accessibility predictions and mobility-derived behavioral signals compared to single-question prompting. InternVL3-78B improves from $r = -0.213$ under single-question prompting to $r = -0.54$ with ten-question reasoning, while Qwen3-32B improves from $r = -0.268$ to $r = -0.470$, with similar patterns observed across the EMD metric. The results suggest that decomposing wheelchair accessibility assessment into structured sub-questions enables more reliable multimodal reasoning and improves sensitivity to accessibility-related environmental cues.

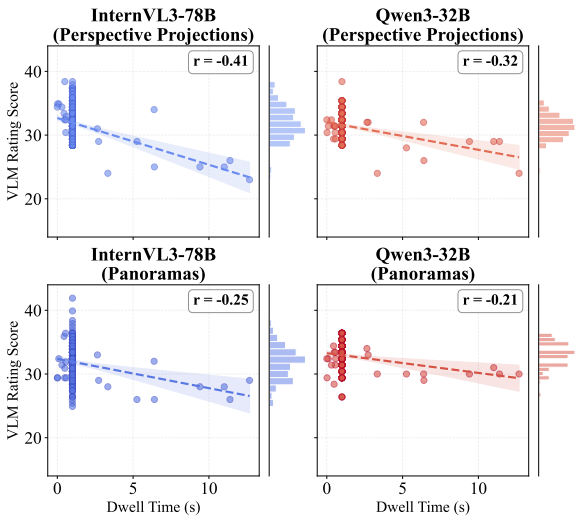

Figure 5: Correlation coefficients of VLM-generated ratings using 90-270 perspective projection (left) and panorama (right) Street View inputs across 246 sampled locations.

Table 2: Single- (SQ) versus multi- (MQ) question prompting performance measured by r and EMD.

Model Variant	Setting	r	Overall EMD	Q4 EMD
InternVL3-78B	SQ	-0.213	0.584	0.481
InternVL3-78B	MQ	-0.540	0.530	0.337
Qwen3-32B	SQ	-0.268	0.860	0.779
Qwen3-32B	MQ	-0.470	0.420	0.246

4.3 Geospatial and visual cue analysis

Geospatial Pattern. Quantitative agreement is evaluated at both measurement and trip levels through correlations between mobility sensing signals and VLM accessibility scores. Figure 6 further visualizes the campus-scale spatial relationship between GPS-derived dwell time and VLM accessibility score predictions. The models generally align with sensor observations in regions containing obvious accessibility barriers or clearly compliant pedestrian infrastructure. The spatial distributions suggest that VLMs capture potential accessibility barrier location patterns.

Visual Cue Association Analysis. Figures 7 and

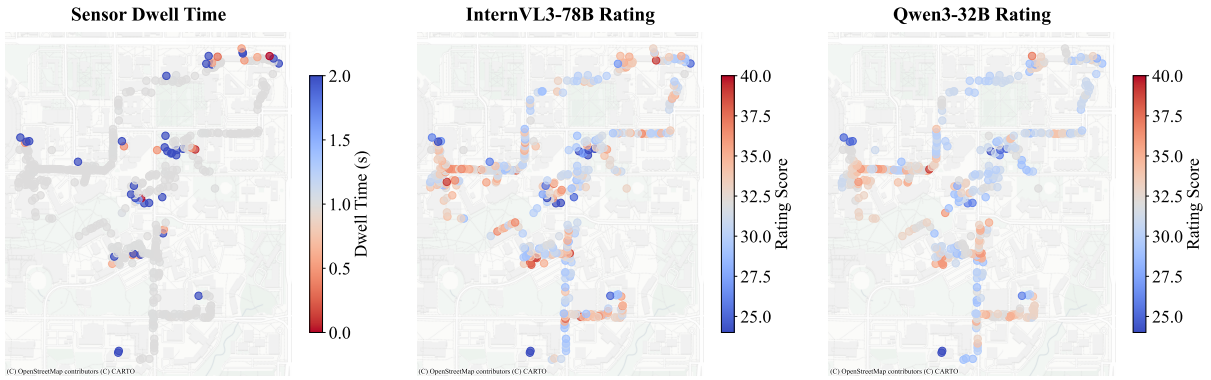

Figure 6: Spatial distribution comparison between measured dwell time and VLM accessibility scores.

8 present complementary rank-difference analyses for VLM–sensor agreement by AMI and accessibility scoring patterns by AII. Across both models, infrastructures that supports physical accessibility such as *crosswalks*, *curb ramps*, *tactile paving*, and *accessible signage* are associated with stronger agreement and higher accessibility scores, whereas degraded pedestrian conditions including *cracks*, *gaps*, *loose surfaces*, *stairs*, *parked vehicles*, and *obstructions* are associated with lower accessibility scores and greater prediction disagreement. Interestingly, InternVL-78B demonstrates greater sensitivity to structural discontinuities, while Qwen3-32B emphasizes surface-level obstacles and pathway constraints.

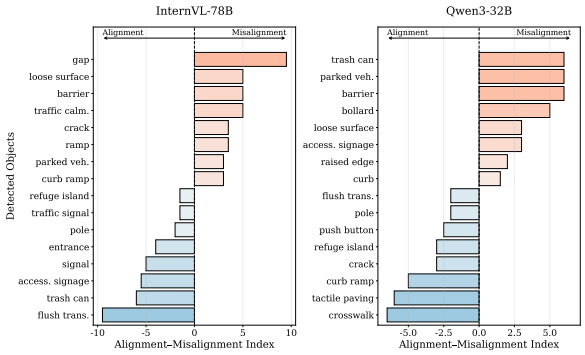

Figure 7: Detected object AMI between well-fitted and poorly-fitted accessibility predictions. Positive AMI indicate semantic objects disproportionately associated with poorly fitted locations (misalignment), whereas negative values indicate stronger association with well-fitted locations (alignment).

5 Discussion

Our results show that VLM-based Street View analysis can meaningfully approximate, but not replace, sensor- and lived-experience-based accessibility assessment. The strongest configuration (RAG-augmented InternVL3-78B) reaches $r = -0.54$ with sensor-derived dwell time, positioning the pipeline as a *scalable tool* that flags where on-the-ground audits are most needed.

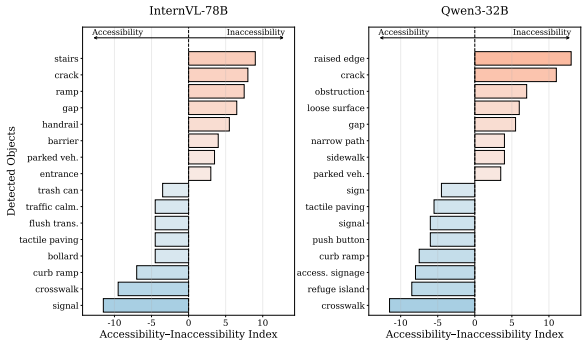

Figure 8: Detected object AII between high- and low-accessibility predictions. Positive values indicate semantic objects disproportionately associated with lower-rated and less accessible environments, whereas negative values indicate stronger association with highly accessible environments.

The ablation study clarifies which components matter. RAG grounding in ADA handbooks and expert rubrics consistently improves alignment with sensor signals, with the largest gains in InternVL3-78B, suggesting that handbook and expert-rubric grounding can augment models' knowledge and indicate a design pattern for civic-AI applications. Street view perspective projections further outperform panoramas, indicating that simple multi-angle sampling offers a better cost–accuracy tradeoff than omnidirectional processing.

The visual-cue and geospatial analyses reveal where VLM reasoning is reliable or inconsistent. Models align with sensors around salient barriers and compliant infrastructure, but diverge under subtle surface degradation and transient obstructions. AMI and AII further show that crosswalks and tactile paving support reliable reasoning, while fragmented environments and parked vehicles remain challenging. Overall, the expert-grounded VLM pipeline provides a scalable screening tool to prioritize locations for human review rather than replace accessibility audits.

Limitations

Several limitations should be noted. First, our study is a single-campus pilot conducted at UF; while the campus offers heterogeneous accessibility conditions, multi-campus or multi-city generalization remains untested. Second, Street View imagery is subject to temporal staleness relative to live conditions, particularly for transient barriers such as construction, congestion, and seasonal obstructions, which limits agreement with real-time sensor traces. Third, sensor coverage is constrained to traversable routes, meaning that fully inaccessible features (e.g., entrances without ramps that block traversal entirely) are under-represented in the behavioral ground truth. Fourth, the expert rubric reflects input from a relatively small expert pool, and broader engagement with wheelchair users and accessibility professionals would strengthen the lived-experience grounding. Finally, dwell time serves as a behavioral proxy rather than a direct accessibility assessment label, and may also reflect non-barrier factors such as waiting, resting, or pedestrian congestion.

Future work could address these limitations in several directions: (i) extending validation to multiple campuses and urban contexts to assess generalization; (ii) integrating live imagery sources such as mobile capture and dashcam data to reduce temporal staleness; (iii) tightening human-in-the-loop verification with wheelchair users and accessibility professionals; and (iv) feeding pipeline outputs into campus planning and maintenance workflows to evaluate downstream impact. A broader goal is to support routinely updated, equity-relevant accessibility maps for urban decision-makers.

Ethical Considerations

This work aims to support accessibility assessment at scale, with the goal of benefiting people with mobility impairments by identifying barriers in the built environment. However, several ethical considerations merit attention.

Privacy. Street View imagery is publicly available but may incidentally capture identifiable individuals. We rely solely on pre-existing imagery and do not collect or store personally identifiable information. Wheelchair dwell behavior is derived from aggregated GPS traces and is not linked to specific individuals.

Bias and Fairness. VLMs may reflect biases present in their training data, potentially leading to inconsistent performance across different environmental contexts, geographic regions, or barrier types underrepresented in training corpora. Practitioners should be cautious about deploying such models in high-stakes accessibility audits without human expert validation, particularly for subtle or context-dependent barriers.

Limitations of Automated Assessment. Our framework is intended as a scalable screening tool, not a substitute for on-the-ground professional accessibility evaluation or ADA compliance audits. Automated ratings should be interpreted as signals that prioritize locations for further review, not as definitive assessments of accessibility.

Data Collection. The wheelchair GPS data used in this study was collected at the University of Florida under appropriate institutional oversight. Participants' data were anonymized prior to analysis. No vulnerable populations were targeted or adversely affected by this research.

Intended Use. We release our findings to support urban planners, accessibility advocates, and researchers in identifying mobility barriers at scale. We caution against using this framework to make legal compliance determinations or to replace disability community input in infrastructure decision-making process.

Statement of AI Assistant Use

AI tools have been used only for minor editing, code refinement, proofreading, and language polishing. All research questions, methodological design, data collection procedures, accessibility assessment pipeline, sensor-based evaluation strategy, statistical analyses, and interpretation of results were developed and verified by the authors.

References

Hashem Abu Tariah, Nasser Ghasham, Mohammad Alolayan, Bandar Alahmadi, and Abdullah Alqarni. 2018. Wheelchair accessibility of mosques in riyadh. *Work*, 60(3):385–391.

Architectural and Transportation Barriers Compliance Board. 2023. Accessibility guidelines for pedestrian facilities in the public right-of-way. Federal Register, 88 FR 53604. Accessed May 21, 2026.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, and 1 others. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Billie Louise Bentzen, Alan C Scott, Robert Wall Emerson, and Janet M Barlow. 2020. Effect of tactile walking surface indicators on travelers with mobility disabilities. *Transportation research record*, 2674(7):410–419.

Ivan Blečić, Valeria Saiu, and Giuseppe A. Trunfio. 2024. Enhancing urban walkability assessment with multimodal large language models. In *International Conference on Computational Science and Its Applications*, pages 394–411. Springer.

Nicholas Bolten and Anat Caspi. 2019. Accessmap website demonstration: Individualized, accessible pedestrian trip planning at scale. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 676–678. ACM.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, and 1 others. 2024. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24185–24198.

Hongru Du, Yang Zhao, Jianan Zhao, Shaochong Xu, Xihong Lin, Yiran Chen, Lauren M Gardner, and Hao 'Frank' Yang. 2025. Advancing real-time infectious disease forecasting using large language models. *Nature Computational Science*, 5(6):467–480.

Jonathan Duvall, Rory Cooper, Eric Sinagra, Dianna Stuckey, Joshua Brown, and Jonathan Pearlman. 2013. Development of surface roughness standards for pathways used by wheelchairs. *Transportation research record*, 2387(1):149–156.

Clive D'Souza, Victor L Paquet, James A Lenker, and Edward Steinfeld. 2019. Self-reported difficulty and preferences of wheeled mobility device users for simulated low-floor bus boarding, interior circulation and disembarking. *Disability and Rehabilitation: Assistive Technology*, 14(2):109–121.

Federal Highway Administration. 2006. Lesson 20: Traffic calming. University Course on Bicycle and Pedestrian Transportation FHWA-HRT-05-123, U.S. Department of Transportation, Federal Highway Administration. Archived publication.

Jie Feng, Shengyuan Wang, Tianhui Liu, Yanxin Xi, and Yong Li. 2025. Urbanllava: A multi-modal large language model for urban intelligence. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6209–6219.

Jon E. Froehlich, Anke M. Brock, Anat Caspi, João Guerreiro, Kotaro Hara, Reuben Kirkham, Johannes Schöning, and Benjamin Tannert. 2019. Grand challenges in accessible maps. *interactions*, 26(2):78–81.

Karen L Frost and Gina Bertocci. 2010. Retrospective review of adverse incidents involving passengers seated in wheeled mobility devices while traveling in large accessible transit vehicles. *Medical Engineering & Physics*, 32(3):230–236.

Isabelle Gatmaitan, Yiwen Dong, Shabboo Valipoor, and Vivian W. H. Wong. 2025. Poster abstract: Dynamic mobility barrier detection using wheelchair-mounted sensors. In *Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '25, page 310–311, New York, NY, USA. Association for Computing Machinery.

Shawn C Gendle, Mark Richardson, James Leeper, L Brent Hardin, J Matt Green, and Phillip A Bishop. 2012. Wheelchair-mounted accelerometers for measurement of physical activity. *Disability and Rehabilitation: Assistive Technology*, 7(2):139–148.

Kotaro Hara, Victoria Le, and Jon E. Froehlich. 2013. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 631–640. ACM.

Anthony Howell, Nancy Wu, Sharmistha Bagchi-Sen, Yushim Kim, and Qian Chayn Sun. 2026. Multimodal large language models, street view images and urban policy-intelligence: recovering the sustainability effects of redlining. *npj Urban Sustainability*.

Efthimis Kapsalis, Nils Jaeger, and Jonathan Hale. 2024. Disabled-by-design: effects of inaccessible urban public spaces on users of mobility assistive devices–a systematic review. *Disability and Rehabilitation: Assistive Technology*, 19(3):604–622.

Mohammed Kutbi, Xiaoxue Du, Yizhe Chang, Bo Sun, Nikolaos Agadakos, Haoxiang Li, Gang Hua, and Philippos Mordohai. 2020. Usability studies of an egocentric vision-based robotic wheelchair. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1):1–23.

Ivan Chew Boon Leong and Susan E Higgins. 2010. Public library services for wheelchair-bound young people in singapore. *Public Library Quarterly*, 29(3):210–229.

Hanyan Li, Jack Cebe, Sara Khoeini, Yanzhi Ann Xu, Chelsea Dyess, and Randall Guensler. 2018. A semi-automated method to generate gis-based sidewalk networks for asset management and pedestrian accessibility assessment. *Transportation Research Record*, 2672(44):1–9.

Heng Li, Minghan Li, Zhi-Qi Cheng, Yifei Dong, Yuxuan Zhou, Jun-Yan He, Qi Dai, Teruko Mitamura, and Alexander G Hauptmann. 2024a. Human-aware vision-and-language navigation: Bridging simulation to reality with dynamic human interactions. *Advances in Neural Information Processing Systems*, 37:119411–119442.

Ling Li, Yu Ye, Yao Zhou, Bingchuan Jiang, and Wei Zeng. 2024b. Georeasoner: Geo-localization with reasoning in street views using a large vision-language model. *arXiv preprint arXiv:2406.18572*.

Lingyao Li, Dawei Li, Zhenhui Ou, Xiaoran Xu, Jingxiao Liu, Zihui Ma, Runlong Yu, and Min Deng. 2025a. Llms as world models: Data-driven and human-centered pre-event simulation for disaster impact assessment. *arXiv preprint arXiv:2506.06355*.

Lingyao Li, Runlong Yu, Qikai Hu, Bowei Li, Min Deng, Yang Zhou, and Xiaowei Jia. 2025b. From pixels to places: A systematic benchmark for evaluating image geolocalization ability in large language models. *arXiv preprint arXiv:2508.01608*.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916.

New England ADA Center. 2014. Ada checklist for existing facilities. Developed by the Institute for Human Centered Design.

June Young Park, Jessica Eisma, Sarah Rose, and Mikila Salazar. 2022. Campus as a living lab: Discovering the comfort of wheelchair users in the pedestrian network by experiential learning with high school students.

Kathylee Pinnock Branford, Meegan G Van Straaten, Omid Jahanian, Melissa MB Morrow, and Stephen M Cain. 2025. An inertial sensor-based comprehensive analysis of manual wheelchair user mobility during daily life in people with sci. *PloS one*, 20(9):e0323050.

Mike Prescott, William C Miller, Jaimie Borisoff, Polly Tan, Nova Garside, Robert Feick, and W Ben Mortenson. 2021. An exploration of the navigational behaviours of people who use wheeled mobility devices in unfamiliar pedestrian environments. *Journal of Transport & Health*, 20:100975.

Juneyoung Ro, Namwoo Kim, and Yoonjin Yoon. 2025. How well do vision-language models understand cities? a comparative study on spatial reasoning from street-view images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6476–6485.

Raphael Schumann, Wanrong Zhu, Weixi Feng, Tsu-Jui Fu, Stefan Riezler, and William Yang Wang. 2024. Velma: Verbalization embodiment of llm agents for vision and language navigation in street view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18924–18933.

Yihong Tang, Ao Qu, Xujing Yu, Weipeng Deng, Jun Ma, Jinhua Zhao, and Lijun Sun. 2025. From street views to urban science: Discovering road safety factors with multimodal large language models. *arXiv preprint arXiv:2506.02242*.

U.S. Department of Justice. 2010. *2010 ADA Standards for Accessible Design*. U.S. Department of Justice. Accessed May 21, 2026.

R Velho, C Holloway, A Symonds, and B Balmer. 2016. The effect of transport accessibility on the social inclusion of wheelchair users: A mixed method analysis. social inclusion; 4 (3): 24-35.

Junjue Wang, Weihao Xuan, Heli Qi, Zihang Chen, Hongruixuan Chen, Zhuo Zheng, Junshi Xia, Yanfei Zhong, and Naoto Yokoya. 2026. Cityvlm: Towards sustainable urban development via multi-view coordinated vision–language model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 232:62–74.

Anders Wretstrand, Per-Olof Bylund, Jan Petzäll, and Torbjörn Falkmer. 2010. Injuries in special transport services—situations and risk levels involving wheelchair users. *Medical engineering & physics*, 32(3):248–253.

Jihan Yang, Runyu Ding, Ellis Brown, Xiaojuan Qi, and Saining Xie. 2024. V-irl: Grounding virtual intelligence in real life. In *European conference on computer vision*, pages 36–55. Springer.

Jiaxin Zhang, Yunqin Li, Tomohiro Fukuda, and Bowen Wang. 2025. Urban safety perception assessments via integrating multimodal large language models with street view images. *Cities*, 165:106122.

Ming Zhong, Yanqiang Zhang, Xi Yang, Yufeng Yao, Junlong Guo, Yaping Wang, and Yaxin Liu. 2019. Assistive grasping based on laser-point detection with application to wheelchair-mounted robotic arms. *Sensors*, 19(2):303.

Gengze Zhou, Yicong Hong, and Qi Wu. 2024. Navgpt: Explicit reasoning in vision-and-language navigation with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 7641–7649.

6 Appendices

A Details on study area and sensor processing

This appendix provides additional detail on the study setting, sensor processing workflow, and supplementary analyses used to support the main results. We first describe how the wheelchair GPS data were processed into dwell-based mobility-friction signals. We then provide the equations used in the preprocessing pipeline, the experimental environment, trip-level correlation results, the rule-based RAG lookup table, and the accessibility question design. The full prompt templates are placed at the end of the appendix for readability.

The UF campus includes academic buildings, libraries, dining facilities, recreational spaces, and major pedestrian connectors that generate substantial daily foot traffic. Route conditions vary across locations, including sidewalk continuity, terrain transitions, curb ramps, building entrances, crosswalk design, and temporary disruptions such as construction or pedestrian congestion. Unlike controlled accessibility audits, university campuses are dynamic mobility environments where infrastructure quality, pedestrian activity, and route usability jointly shape navigation experiences.

Figure 9 summarizes the full sensor-processing workflow. The figure shows how raw Sensor Logger trip files are cleaned, temporally ordered, converted into movement features, and grouped into dwell candidate areas for comparison with VLM-derived accessibility scores.

First, wheelchair sensor data are collected across routes designed to reflect common and functionally important mobility patterns. These trips capture navigation through sidewalks, building entrances, ramps, crosswalks, elevators, high-traffic walkways, classroom buildings, and other mobility-relevant infrastructure.

Dwell points are then classified according to relative stopping duration. Points within the interquartile range are classified as normal dwell time, points above the 75th percentile as long dwell time, and points above the 95th percentile as very long dwell time. Because dwell time is not a direct accessibility label, we treat it as a mobility-friction signal requiring contextual interpretation. Long dwell events may reflect physical barriers, but they may also result from waiting at crosswalks, resting, adjusting equipment, or pedestrian congestion. To reduce this ambiguity, a sample of dwell points is reviewed using field knowledge, including photographs and field notes, and annotated when points could be linked to observed built-environment conditions such as inaccessible entrances, rough surfaces, construction barriers, crowded walkways, and route deviations.

B Data processing equations

This section gives the equations used to derive movement and dwell features from GPS trajectories. These equations are included to make the sensor-processing pipeline transparent and reproducible. For each trip, GPS points were ordered by timestamp. Let point i have timestamp t_i, latitude ϕ_i, and longitude λ_i, where latitude and longitude are converted to radians for distance calculations.

Time Gap. The time difference between consecutive GPS points was calculated as:

$$\Delta t_i = t_{i+1} - t_i.$$

Haversine Distance. The distance between consecutive GPS points was calculated using the haversine formula:

$$\Delta\phi_i = \phi_{i+1} - \phi_i,$$
$$\Delta\lambda_i = \lambda_{i+1} - \lambda_i,$$
$$a_i = \sin^2\left(\frac{\Delta\phi_i}{2}\right) + \cos(\phi_i)\cos(\phi_{i+1})\sin^2\left(\frac{\Delta\lambda_i}{2}\right),$$
$$d_i = 2R\arcsin\left(\sqrt{a_i}\right).$$

where $R = 6{,}371{,}000$ meters is the approximate radius of the Earth, and d_i is the distance from point i to point $i+1$ in meters.

GPS-Estimated Speed. Estimated GPS speed was calculated as:

$$v_i = \frac{d_i}{\Delta t_i},$$

where v_i is estimated speed in meters per second.

Dwell Candidate Rule. Points were labeled as dwell candidates when estimated speed was less than or equal to 0.7 meters per second:

$$\text{DwellCandidate}_i = \begin{cases} 1, & v_i \le 0.7, \\ 0, & v_i > 0.7. \end{cases}$$

Spatial Binning. Latitude and longitude were rounded into spatial bins using a grid size of $g = 0.0001$:

$$\text{lat_bin}_i = \text{round}\left(\frac{\text{lat}_i}{g}\right) \times g,$$
$$\text{lon_bin}_i = \text{round}\left(\frac{\text{lon}_i}{g}\right) \times g.$$

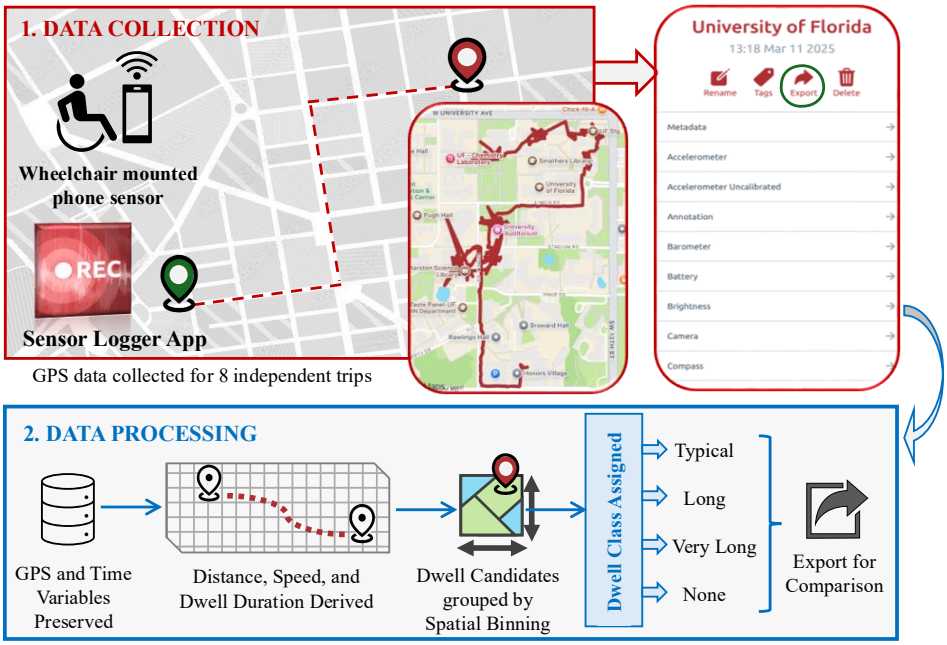

Figure 9: Data collection and processing workflow. Sensor Logger trip files were cleaned to preserve GPS and time fields, concatenated with trip identifiers, and sorted temporally. Movement features were calculated from consecutive GPS points, including time gap, haversine distance, and estimated speed. Dwell candidate points were grouped into spatial bins to identify dwell areas, which were classified by duration and exported for VLM comparison. Equations for the movement and dwell calculations are provided in Appendix B.

Total Dwell Time. For each trip and spatial bin group j, total dwell time was calculated as:

$$D_j = \sum_{i \in j} \Delta t_i,$$

where j represents a trip/bin group containing dwell-candidate points.

Dwell Classification. Dwell areas were classified using the 75th and 95th percentile thresholds of total dwell time:

$$\text{Dwell Class}_j = \begin{cases} \text{Very long}, & D_j \geq Q_{95}, \\ \text{Long}, & Q_{75} \leq D_j < Q_{95}, \\ \text{Typical}, & Q_{25} \leq D_j < Q_{75}. \end{cases}$$

C Experimental environment

We evaluate multiple open-source VLMs including InternVL3-2B, InternVL3-8B, InternVL3-14B, InternVL3-78B, Qwen3-VL-2B, Qwen3-VL-4B, Qwen3-VL-8B, and Qwen3-VL-32B. These models span parameter scales from 2B to 78B parameters, enabling systematic comparison across lightweight, medium-scale, and large-scale multimodal architectures.

All experiments are conducted on NVIDIA DGX B200 SuperPOD infrastructure consisting of NVIDIA Blackwell B200 GPUs with 180 GB GPU memory per device. The computing nodes are interconnected through NVIDIA Quantum-2 InfiniBand networking with 400 Gb/s bandwidth to support distributed large-scale VLM inference. GPU resources are managed through the SLURM workload manager under CUDA 12 environments.

The implementation is developed using Python 3.10 with PyTorch, HuggingFace Transformers, Accelerate, and DeepSpeed for distributed multimodal inference and large-model execution. Unless otherwise specified, all experiments use the default model configurations and inference settings provided by the official implementations of InternVL3 and Qwen3-VL to ensure reproducibility and fair comparison across model variants.

All models are evaluated under a unified prompting and evaluation framework. Greedy decoding is adopted during inference to ensure deterministic and reproducible accessibility score generation. Input images are processed using the default visual preprocessing pipelines associated with each model architecture. Prompting parameters, including temperature, top-p, and maximum generation length, follow the default inference configurations of the corresponding repositories unless modifications are required for structured output formatting.

The retrieval-augmented prompting framework uses fixed ADA-informed accessibility guidance and expert-derived evaluation rubrics across all experiments to maintain consistency between model

variants. Comparative experiments evaluate the effects of model scaling, RAG, and prompting strategies on multimodal environmental understanding, accessibility barrier detection, and alignment with GPS-derived wheelchair dwell behavior.

D Qualitative examples

Figure 10 provides additional qualitative examples that complement the main result, showing how the VLMs translate visible street-view cues into wheelchair accessibility scores and reasoning, and how these judgments correspond to sensor-derived dwell behavior. The figure also illustrates how InternVL3-78B and Qwen3-VL-32B generate structured ratings and reasoning across multiple accessibility dimensions.

E Baseline correlation results

Figure 11 shows the correlation between wheelchair dwell time and VLM accessibility scores across InternVL and Qwen model variants. Overall, the baseline results show a consistent negative relationship between dwell time and predicted accessibility: locations with longer wheelchair dwell time tend to receive lower accessibility ratings. However, the strength of this relationship differs across model families and model sizes. These results also show that larger models are generally more sensitive to visual cues, achieving higher performance. At the same time, the baseline correlations remain moderate even for the strongest models.

Compared with the RAG-enhanced results reported in the main text, the baseline setting produces weaker alignment for the strongest models. For example, InternVL3-78B improves from $r = -0.40$ in the baseline setting to $r = -0.54$ with RAG, while Qwen3-32B improves from $r = -0.40$ to $r = -0.47$. These gains suggest that retrieved accessibility guidance helps the models interpret visible features in more wheelchair-relevant ways. This difference also indicates that retrieved accessibility guidance helps the models connect visible environmental features, such as curb transitions, surface quality, pathway continuity, and obstructions, to wheelchair-relevant accessibility judgments more reliably.

F Rule-based lookup table for RAG

Before applying the accessibility questions, we use a rule-based lookup step to connect visual features detected from GSV imagery with the relevant accessibility concepts in the RAG database. This lookup table acts as a lightweight grounding layer: it helps the model retrieve guidance related to the visible scene without requiring the model to infer the full regulatory category from scratch. Table 3 lists the mapping between common detected visual features and the accessibility definition categories used for retrieval.

Table 3: Rule-based lookup table used for RAG. Detected visual accessibility features are semantically mapped to ADA-informed categories for retrieval-guided reasoning.

Detected Visual Feature	Mapped Accessibility Definition Category
sidewalk	walkable surface; accessible route
crack	surface condition; cracks, gaps, or openings
gap	cracks, gaps, or openings
raised edge	raised edges; height differences; abrupt edges
loose surface	loose or mixed surfaces; surface condition
narrow path	path width; clear passage; maneuvering space
obstruction	obstacle or obstruction; encroaching walkable path
pole	obstacle or obstruction
sign / bollard	street furniture or roadside obstruction
trash can / parked vehicle	obstacle or obstruction
curb	curb; raised edge or level difference
curb ramp	curb cut; ramp
ramp	ramp; broken ramp surface
flush transition	flush or blended transition
tactile paving	detectable warning surface
crosswalk	crossing or pedestrian crossing support
pedestrian signal	pedestrian signal
push button	crossing control button
refuge island	refuge island
traffic calming	traffic calming features
temporary barrier	temporary barrier
building entrance	accessible route
stairs	height difference or abrupt edge
handrail	ramp accessibility support
accessible signage	platform edge or decision point

G Trip-level correlation analysis

The main results report overall correlations across all locations. To examine whether the observed alignment is consistent across individual wheelchair trips, Table 4 reports trip-level correlation coefficients between dwell time and VLM accessibility scores for each model variant. Although the strongest models generally maintain negative correlations across several trips, the magnitude varies substantially due to between-trip het-

Figure 10: Representative visual examples from low and high dwell-time locations. Both models assess built environment wheelchair accessibility on a 10–50 scale from inaccessible to accessible environments, with concise multimodal justification.

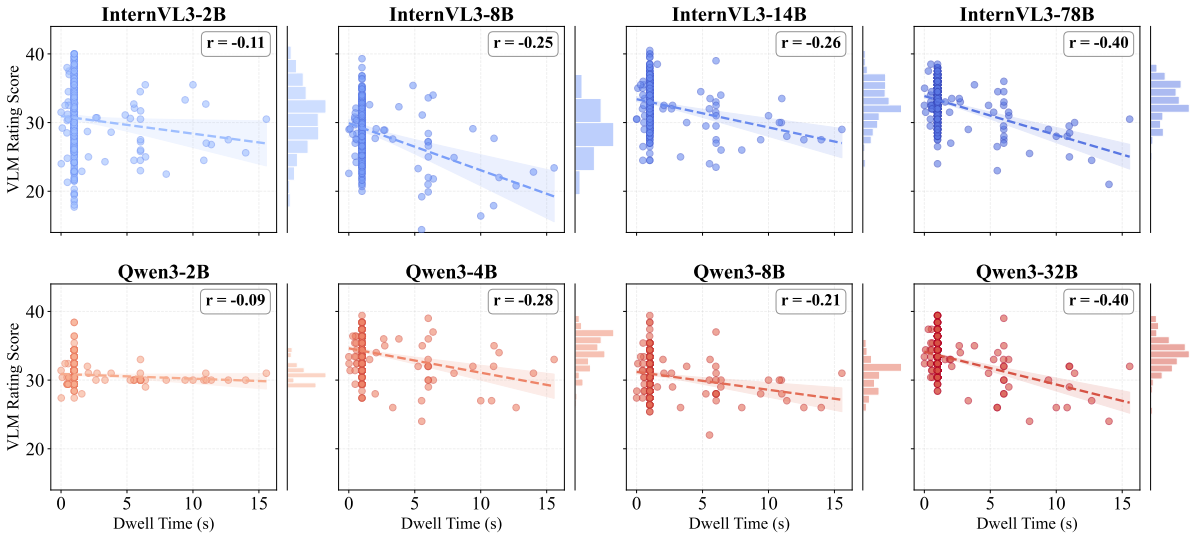

Figure 11: Correlation between wheelchair dwell time and VLM-based accessibility ratings under the baseline setting.

erogeneity, including differences in route length, visual complexity, matched GSV image density, and the local causes of dwell behavior such as resting, waiting, or pedestrian congestion. Therefore, pooled correlations should be interpreted as aggregate patterns rather than stable per-route estimates, and robust deployment would require larger and more balanced route-level sampling.

H Accessibility questions in prompt

Table 5 presents the accessibility questions used in the multi-question prompt. These questions were designed to cover the major dimensions of wheelchair navigation visible in street-view imagery, including surface condition, path geometry, obstacles, curb transitions, tactile paving, pedestrian crossing support, temporary barriers, entrances, signage, and route equity. The third column shows how each question is grounded in prior accessibility literature and accessibility guidance.

I RAG feature definitions

Table 6 provides the accessibility definitions used in the retrieval database. These definitions help translate detected visual features into accessibility-relevant concepts, allowing the model to reason with domain guidance while still grounding its answer in visible evidence from GSV imagery.

Table 4: Trip-level correlation coefficients across InternVL and Qwen model variants. Negative values indicate that longer dwell time is associated with lower VLM-derived accessibility scores. The total column reports the overall correlation across all evaluated trip locations.

Model	Trip ID								Total
	T1	T2	T3	T4	T5	T6	T7	T8	
InternVL-2B	0.003	-0.390	-0.273	-0.039	0.061	-0.208	-0.152	-0.166	-0.134
InternVL-8B	-0.305	-0.153	-0.337	0.104	-0.651	-0.320	0.002	-0.284	-0.266
InternVL-14B	-0.413	-0.148	-0.519	-0.121	0.136	-0.272	-0.088	-0.422	-0.285
InternVL-78B	-0.559	-0.248	-0.583	-0.155	-0.649	-0.423	0.067	-0.677	-0.542
Qwen-2B	0.096	0.325	-0.338	0.124	-0.421	-0.303	0.228	-0.056	-0.102
Qwen-4B	-0.404	-0.098	-0.549	-0.076	-0.501	-0.258	-0.066	-0.426	-0.363
Qwen-8B	-0.108	-0.148	-0.412	-0.132	-0.353	-0.321	0.443	-0.481	-0.364
Qwen-32B	0.018	-0.075	-0.629	-0.114	-0.836	-0.319	-0.213	-0.543	-0.401

Table 5: LLM accessibility questions and literature support matrix.

Question	LLM Question	Literature Support
Q1: Surface Condition	Do walkable surfaces appear smooth, continuous, and stable, or are there visible cracks, uneven slabs, patches, gaps, raised edges, loose materials, or mixed surfaces such as gravel, cobblestone, brick, mulch, or grass that could make movement difficult for mobility-aid users?	Prior work identifies rough, uneven, and sloped pathways as key outdoor accessibility barriers, and wide or frequent cracks can expose wheelchair users to harmful whole-body vibration (Kapsalis et al., 2024; Duvall et al., 2013). PROWAG also requires pedestrian access route surfaces to be stable, firm, slip resistant, and free of abrupt level changes (Architectural and Transportation Barriers Compliance Board, 2023).
Q2: Path Width & Geometry	Does the route appear visually open and continuous, or does it appear narrowed, pinched, sharply turning, zig-zagged, crowded, or constrained in a way that could make navigation difficult for a wheelchair, scooter, walker, or crutch user?	ADA guidance requires accessible routes to maintain continuous clear width, with additional clearance for turns and narrow obstructions (U.S. Department of Justice, 2010). PROWAG similarly emphasizes clear width and passing space for pedestrian access routes (Architectural and Transportation Barriers Compliance Board, 2023).
Q3: Obstacles	Are there visible objects, vegetation, signs, poles, bollards, furniture, trash cans, debris, parked vehicles, or other obstacles encroaching on the walkable path or reducing clear passage?	ADA standards address protruding objects and state that they cannot reduce the required clear width on accessible routes (U.S. Department of Justice, 2010). PROWAG also specifies requirements for protruding objects and vertical clearance along pedestrian circulation paths (Architectural and Transportation Barriers Compliance Board, 2023).
Q4: Curb Cuts, Ramps & Grade	At visible crossings or level changes, are curb cuts, ramps, or flush transitions present? Are there visible steps, raised curbs, abrupt edges, gaps, broken ramp surfaces, or other level changes that may create a barrier?	Prior work identifies curb ramps and ramp slopes as major accessibility concerns, noting that excessive cross-slopes and ramp gradients can increase strain and compromise independent mobility (Kapsalis et al., 2024; Frost and Bertocci, 2010; Wretstrand et al., 2010; Velho et al., 2016). Even slopes within permissible ranges can cause difficulty for some users (D'Souza et al., 2019). ADA and Access Board guidance also specify ramp width, slope, and accessible-route requirements (U.S. Department of Justice, 2010).
Q5: Tactile Paving	Is tactile paving or detectable warning surfacing visible at crossings, curb cuts, platform edges, or decision points? If present, does it appear unobstructed, continuous, and in usable condition, or does it appear worn, blocked, broken, missing, or misaligned?	Tactile paving can support pedestrians with visual impairments but may also create fatigue or instability for mobility-aid users, requiring evaluation of both presence and condition (Kapsalis et al., 2024; Bentzen et al., 2020). PROWAG requires detectable warning surfaces at curb ramps, refuge islands, platforms, and related transition points (Architectural and Transportation Barriers Compliance Board, 2023).

Continued on next page

Table 5: LLM accessibility questions and literature support matrix (continued).

Question	LLM Question	Literature Support
Q6: Pedestrian Crossing Safety	Are visible pedestrian crossing supports present, such as marked crosswalks, curb cuts, pedestrian signals, push-button controls, refuge islands, signage, or traffic-calming features? Do controls or signs appear high, blocked, set back, or difficult to see or reach from a seated position?	PROWAG requires accessible pedestrian signals and push buttons where pedestrian signals are provided, including audible or vibrotactile indications and reachable controls (Architectural and Transportation Barriers Compliance Board, 2023). Prior work also notes that high road signs and crosswalk buttons can hinder orientation and access for mobility-aid users, who often sit lower than ambulatory pedestrians (Kapsalis et al., 2024; Prescott et al., 2021).
Q7: Construction or Temporary Barriers	Are there visible temporary barriers such as construction fencing, cones, scaffolding, closures, equipment, or temporary signage blocking or narrowing the pedestrian path? If so, is a clearly marked alternate pedestrian route visible and continuous?	PROWAG requires an alternate pedestrian access route when a pedestrian circulation path is temporarily inaccessible due to construction, maintenance, closure, or similar conditions (Architectural and Transportation Barriers Compliance Board, 2023).
Q8: Building Entrances & Access Features	Do visible building entrances appear level, ramp-connected, or otherwise reachable without stairs? Are there visible steps, raised thresholds, narrow-looking doorways, heavy/manual doors, missing ramps, blocked entrances, or unclear accessible entrance signage?	Entrance features are often among the least accessible elements for mobility-aid users (Kapsalis et al., 2024). ADA standards require accessible entrances and routes and specify minimum clear width for door openings (U.S. Department of Justice, 2010). Prior studies also identify high handles, heavy manual doors, and difficult-to-modify ramps as entrance-specific barriers (Abu Tariah et al., 2018; Leong and Higgins, 2010).
Q9: Wayfinding & Signage	Is accessibility-related signage visible, such as wheelchair symbols, accessible entrance signs, elevator signs, route arrows, detour signs, or crossing instructions? Is the signage visible, unobstructed, and positioned where pedestrians could plausibly see it?	ADA standards require directional and identification signs, including International Symbol of Accessibility signage for accessible facilities and routes (U.S. Department of Justice, 2010). ADA also specifies tactile and visual sign requirements for permanent rooms and spaces.
Q10: Route Equity	From the visible scene, does the accessible path appear to follow the same general route as the main pedestrian path, or is there evidence that wheelchair or mobility-aid users must take a separated, indirect, hidden, back-of-building, or unclear route?	Prior work emphasizes that inaccessible or inflexible public-space elements can exclude mobility-aid users or force them into unsafe and inconvenient spatial situations, with impacts on health and quality of life (Kapsalis et al., 2024). ADA's path-of-travel concept frames accessibility as a continuous, unobstructed connection among sidewalks, streets, entrances, and facility spaces, rather than the presence of an isolated ramp (U.S. Department of Justice, 2010).

Table 6: RAG feature definition table.

Accessibility Element	Definition
Walkable surface / walking surface	A pedestrian surface that is part of a route. Under the ADA Standards, walking surfaces that are part of an accessible route must comply with floor/ground surface requirements, including stability, firmness, and slip resistance (U.S. Department of Justice, 2010, §403).
Surface condition	The visible state of the walking surface, including whether it appears stable, firm, continuous, broken, uneven, loose, or interrupted. ADA floor and ground surfaces must be "stable, firm, and slip resistant," (U.S. Department of Justice, 2010, §1002.4.1) and PROWAG applies the same requirement to pedestrian access route surfaces (Architectural and Transportation Barriers Compliance Board, 2023).
Cracks / gaps / openings	Openings or gaps in a walking surface can affect accessibility. ADA standards state that openings in floor or ground surfaces generally must not allow passage of a sphere more than 1/2 inch in diameter, and elongated openings must be oriented perpendicular to the dominant direction of travel (U.S. Department of Justice, 2010, §302.3).
Raised edges / height differences / changes in level / raised curb / abrupt edge	A change in level is a vertical or abrupt difference between walking surfaces. ADA standards specify how small vertical or beveled changes may be treated, and changes above 1/2 inch must be ramped (U.S. Department of Justice, 2010, §303.4).
Loose or mixed surfaces	Surfaces such as gravel, mulch, grass, brick, cobblestone, or other irregular materials should be treated as potential surface-condition concerns because accessible floor/ground surfaces must be stable, firm, and slip resistant (U.S. Department of Justice, 2010, §403).

Continued on next page

Table 6: RAG feature definition table (continued).

Accessibility Element	Definition
Route / accessible route	Accessible routes shall consist of one or more of the following components: walking surfaces with a running slope not steeper than 1:20, doorways, ramps, curb ramps excluding the flared sides, elevators, and platform lifts (U.S. Department of Justice, 2010, §402.2).
Path width / clear width / clear passage	The unobstructed clear width of the walkable path available for movement. An object is considered to reduce clear passage when it extends into, blocks, or narrows the usable route (U.S. Department of Justice, 2010, §§403.5.1, 307.5).
Tight turns / maneuvering space	Turning or maneuvering space refers to the space needed for wheelchair users or other mobility-device users to turn and navigate. ADA turning space can be circular or T-shaped, with specified minimum dimensions (U.S. Department of Justice, 2010, §403.5.2).
Wheelchair	a manually-operated or power-driven device designed primarily for use by an individual with a mobility disability for the main purpose of indoor or of both indoor and outdoor locomotion (ADA, 2020).
Obstacle / obstruction	Any object, vegetation, sign, pole, furniture, vehicle, debris, or other encroachment that protrudes into, narrows, blocks, or reduces the clear usable width of the accessible/pedestrian route (U.S. Department of Justice, 2010, §§307, 307.5, and 403.5.1).
Encroaching on the walkable path	Encroaching: When an object, vegetation, sign, pole, furniture, parked vehicle, trash can, debris, or other feature extends into the walkable path and reduces the clear space available for movement (U.S. Department of Justice, 2010, §§307, 403.5.1).
Signs / poles / bollards / furniture / trash cans / debris / parked vehicles	These should be treated as possible obstructions when they visibly reduce clear width, block a curb ramp, obstruct a crossing, or protrude into the pedestrian circulation path.
Curb	A raised feature along the side of a street that delineates the edge of the roadway or pedestrian circulation path (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3C).
Curb cut / curb ramp	A sloped connection that is cut through or built up to a curb. Curb ramps may be perpendicular or parallel to the curb or to the street they serve or be a combination thereof (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3C).
Ramp	A sloped walking surface with a running slope steeper than 1:20 (5.0%) that accomplishes a change in level and is not part of a pedestrian circulation path that follows the roadway grade. A curb ramp is not a ramp (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3R).
Flush transition / Blended transition	A wraparound connection at a corner, or a flush connection where there is no curb to cut through, other than a curb ramp (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3B).
Grade	PROWAG defines grade by reference to running slope, meaning the slope parallel to the direction of pedestrian travel. For this project, the LLM should not estimate grade quantitatively (Architectural and Transportation Barriers Compliance Board, 2023, §§R104.3G, R104.3R).
Broken ramp surface	A ramp surface with visible cracks, missing material, unevenness, or surface deterioration. ADA ramp surfaces must comply with floor/ground surface requirements, and changes in level other than permitted slope conditions are not allowed on ramp runs (U.S. Department of Justice, 2010, §405).
Tactile paving / Tenji blocks / Detectable warning surface	A standardized surface feature built in or applied to pedestrian circulation paths and other pedestrian facilities to warn of hazards. These surfaces have a raised dome pattern called truncated domes. ADA standards require detectable warnings to consist of truncated domes and specify dome size, spacing, and visual contrast (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3D) (U.S. Department of Justice, 2010, §705).
Crossing / crosswalk	That part of a roadway that is located at an intersection included within the connections of the lateral lines of the pedestrian circulation paths on opposite sides of the highway measured from the curbs, or in the absence of curbs, from the edges of the traversable roadway, and in the absence of a pedestrian circulation path on one side of the roadway, the part of a roadway included within the extension of the lateral lines of the pedestrian circulation path at right angles to the center line; or at any portion of a roadway at an intersection or elsewhere distinctly indicated as a pedestrian crossing by pavement marking lines on the surface. Crosswalks at intersections may be marked or unmarked (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3C).
Platform Edge / Decision points	The boarding edge of a transit platform, a location where a pedestrian must choose a route, crossing, entrance, detour, or direction (U.S. Department of Justice, 2010, §705.2).
Pedestrian crossing support	Visible features that support pedestrian crossing, such as crosswalk markings, curb ramps, pedestrian signals, push buttons, refuge islands, signage, or traffic-calming features.

Continued on next page

Table 6: RAG feature definition table (continued).

Accessibility Element	Definition
Pedestrian signal	A signal device used to direct pedestrian traffic at a crosswalk. PROWAG defines a pedestrian signal head as a device containing the walking-person and upraised-hand symbols (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3P).
Push-button crossing control	A button to activate a device or signal timing for pedestrians, bicyclists, or others crossing a roadway (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3P).
Refuge Island	A defined area in the direction of pedestrian travel located between traffic lanes for pedestrian refuge within a median, splitter island, or channelizing island (Architectural and Transportation Barriers Compliance Board, 2023, §R104.3P).
Traffic-calming features	Physical measures intended to reduce negative effects of motor vehicle use, alter driver behavior, and improve conditions for nonmotorized street users. Examples include speed bumps, raised crosswalks, and intersections; among others (Federal Highway Administration, 2006).
Temporary barrier	A temporary obstruction or closure that blocks, narrows, or redirects a pedestrian circulation path. This may include construction fencing, cones, scaffolding, closures, equipment, or temporary signage.

J Prompt design

We place the full prompt templates at the end of the appendix to keep the main supplementary results and tables easier to read. The RAG multi-question prompt integrates retrieved accessibility guidance with visible environmental evidence and asks the model to produce structured accessibility scores with concise justifications. We also include a single-question prompt used to examine performance under a simplified assessment setting.

J.1 RAG Multi-Question Prompt Template

```
SYSTEM_PROMPT = """
You are an accessibility assessment expert specialized in wheelchair navigation and pedestrian
infrastructure evaluation. Analyze visible pedestrian infrastructure, environmental conditions,
and retrieved accessibility guidance to provide a reliable wheelchair accessibility assessment
for manual wheelchair users.
"""

ACCESSIBILITY_PROMPT = """
Task: {QUESTION_SET['task']}
View: {IMAGE}

The image captures pedestrian infrastructure and surrounding environmental conditions relevant
to wheelchair accessibility.

Detected Environmental Features
  • Detected keywords: {detected_keywords}

Retrieved Accessibility Guidance

    {retrieval_block}

Assessment Information
  • Question ID: {question_item['id']}
  • Category: {question_item['category']}
  • Question: {question_item['question']}

Accessibility Rating Scale
  • 10 = Very poor accessibility
  • 20 = Poor accessibility
  • 30 = Moderate accessibility
  • 40 = Good accessibility
  • 50 = Excellent accessibility

Assessment Guidelines
  • Use retrieved guidance only when supported by visible evidence.
  • Do not assume inaccessible or occluded infrastructure exists.
  • Base all reasoning strictly on observable environmental conditions.
  • Evaluate sidewalk continuity, curb transitions, obstacles, navigation safety, and pedestrian
    accessibility.
```

Provide a concise justification grounded only in visible evidence.

Output Format

```
    {"score": <1020|30|40|50>,| "justification": "<fewer than 20 words>"}
```

"""

J.2 Single Question Prompt Template

SYSTEM_PROMPT = """
You are an accessibility assessment expert specialized in wheelchair navigation and pedestrian infrastructure evaluation. Analyze visible pedestrian infrastructure and environmental conditions to provide a reliable accessibility assessment for manual wheelchair users.
"""

ACCESSIBILITY_PROMPT = """
Task: built_environment_accessibility_and_vibration_assessment
View: {IMAGE}

The image captures pedestrian infrastructure and surrounding environmental conditions relevant to wheelchair accessibility.

Assessment Information

- Question ID: ONE_QUESTION
- Category: GENERAL_CONDITION
- Question: Evaluate the pedestrian pathway considering surface condition, usability, and navigability. Provide an overall rating of how smoothly a manual wheelchair user could navigate the environment.

Accessibility Rating Scale

- 10 = Very poor accessibility
- 20 = Poor accessibility
- 30 = Moderate accessibility
- 40 = Good accessibility
- 50 = Excellent accessibility

Assessment Guidelines

- Base the assessment strictly on visible evidence.
- Do not assume inaccessible or occluded infrastructure exists.
- Evaluate sidewalk continuity, surface smoothness, curb transitions, obstacles, navigation safety, and overall usability.
- Consider overall wheelchair comfort and navigability.

Provide a concise justification grounded only in visible evidence.

Output Format

```
    {"score": <1020|30|40|50>,| "justification": "<fewer than 20 words>"}
```

"""