BikeButler: A Personalized, Context-sensitive Bike Routing Tool using Open Data and VLM-based Analyses of Street View Imagery

Jared Hwang
Allen School of Computer Science
University of Washington, USA
jaredhwa@cs.washington.edu

John S. O'Meara
Issaquah High School
jsomeara@uw.edu

Zeyu Wang
Urban Design & Planning
University of Washington, USA
zeyuw1@uw.edu

Jasmine Zhang
Allen School of Computer Science
University of Washington, USA
jaszhang@cs.washington.edu

Jon E. Froehlich
Allen School of Computer Science
University of Washington, USA
jonf@cs.washington.edu

Figure 1: We introduce *BikeButler*, a personalized, context-sensitive bicycle route generation tool that enables users to generate, compare, virtually preview, and iteratively customize bike routes. Drawing on *OpenStreetMap*, open government data, and a custom *Vision Language Model* (VLM) analysis of street view imagery (SVI), BikeButler allows users to create profiles across eight bikeability factors, including *bike lanes*, *slope*, *vegetation*, and *surface quality* and to generate and compare context-weighted routes. See video demo.

Abstract

Urban cycling benefits personal wellbeing, public health, and global sustainability. While current tools such as Google and Apple Maps provide bike route recommendations, they do not account for a person's dynamic context (*e.g.*, commuting, recreation). We introduce *BikeButler*, a personalized, context-sensitive bicycle route generation tool that enables users to generate, compare, virtually preview, and iteratively customize bike routes via custom profiles that encode seven bikeability features, including *bike lane existence*, *slope*, *vegetation*, and *surface quality*—fusing data from *OpenStreetMap*, open government data, and a custom VLM-based analysis of Street View images. To design BikeButler, we employed a human-centered, iterative approach starting with formative interviews and culminating in a user study (*N=16*). Our findings demonstrate that bike routing preferences change as a function of context, that BikeButler enables users to quickly create and iterate context-sensitive routes,

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791292

and that generated routes differ significantly from Google Maps bike routing, reinforcing the importance of personalization.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; **Geographic visualization**.

Keywords

bikeability, route planning, computer vision, urban planning

ACM Reference Format:
Jared Hwang, John S. O'Meara, Zeyu Wang, Jasmine Zhang, and Jon E. Froehlich. 2026. BikeButler: A Personalized, Context-sensitive Bike Routing Tool using Open Data and VLM-based Analyses of Street View Imagery. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3772318.3791292

1 Introduction

Urban cycling is a cornerstone of sustainable transportation, offering significant benefits for individual health and well-being [34, 66, 73, 82] while contributing to broader public health and environmental goals [30, 71, 75, 113]. While current mapping tools such as *Google* and *Apple Maps* provide bike route recommendations, they do not account for a person's dynamic context: an early morning commute ride is fundamentally different from a recreational ride with the family. In this paper, we explore how people's bike route preferences change as a function of context and introduce and evaluate a novel bike routing tool called *BikeButler* to support context-sensitive decision making.

To design BikeButler, we draw on the long history of work in urban planning and transportation studies exploring *what* makes a city bikeable and *why* [16, 36, 43]. Cyclists prefer routes with explicit bicycle infrastructure such as bike lanes, trails, and greenways [57, 63, 77], low vehicular speed limits [77], limited topographical steepness [63], and low traffic density—and will even travel farther to use dedicated paths [64]. Similar to walking [87], bike route preferences extend beyond objective measures to subjective qualities, including vegetation [100], surface quality [77], and urban morphology, including the visual aesthetic of a street scene [100]. Despite this robust literature, there is little work exploring *how* these factors may differentially impact route choice as a function of context (*e.g.,* time-of-day, trip purpose, riding alone *vs.* others). Moreover, while mainstream mapping tools provide specific bike routing features, the route recommendations are opaque and not customizable. More dedicated tools such as *Strava* [2] and *RideWithGPS* [5] show popular routes in neighborhoods but the routes lack context and again are not parameterizable.

We introduce *BikeButler*, a personalized, context-sensitive bicycle route generation tool that enables users to generate, compare, virtually preview, and iteratively customize bike routes. BikeButler advances the state-of-the-art in three ways: first, to enable customization across both subjective and objective factors, we combine open government data, OpenStreetMap metadata, and a custom Vision Language Model (VLM)-based analysis of street view images (SVIs) enabling parameterization across eight factors including *elevation*, *bike lane availability*, and *vegetation.* Second, users can rapidly assess generated recommendations via color-coded route segmentations, elevation profiles, and SVI-based previews—supporting algorithmic transparency [86, 93]. Third, users can iteratively customize and compare route recommendations through an innovative segment-based voting mechanism or by changing factor weights in their profiles (*e.g.,* to emphasize pavement smoothness and bike lanes higher than route steepness and greenery).

To design BikeButler, we employed a human-centered, iterative approach starting with four formative interviews, one mid-fi evaluation, and culminating with a user study with 16 participants. We also performed a technical performance evaluation of the VLM's ability to score vegetation, surface quality, and bike lane width compared to two human labelers. We found that the VLM performed comparable to humans for vegetation and surface quality but not for characterizing bike lane width. For the user study, participants created a total of 187 personalized routes, reacting positively to key *BikeButler* features, including the ability to rapidly create, compare, and iterate on context-sensitive routes and to interactively assess route recommendations from SVI thumbnail previews, satellite imagery, and color-coded bikeability scores. We also compared participants' preferred BikeButler routes to default Google Maps bike routes, finding significant differences and reinforcing the importance of personalization.

In summary, our contributions are threefold: **(1) formative findings** across our semi-structured interviews establishing how cyclist route preferences shift depending on dynamic context, providing a new foundation for context-aware routing; **(2) a novel mixed-methods approach for parameterizable route assessment** that combines open, structured GIS data with a VLM-based analysis to score both objective and subjective qualities along a bike route and articulating opportunities and limitations therein; **(3) a new context-sensitive, bike routing tool** called BikeButler that provides personalized route recommendations, a novel interactive voting and previewing mechanism, and iterative, parameterization profiles. To support open science, all code is available on GitHub[1].

2 Related Work

We situate our work by synthesizing cross-disciplinary research on bike infrastructure preferences, bikeability indices and route optimization, and using computer vision (CV) to detect and qualify urban infrastructure.

2.1 Bikeability and Preferences

BikeButler draws on a rich body of prior work to surface and prioritize features that enhance bikeability [16, 50, 63, 74, 89]. Foremost is the existence of *dedicated cycling infrastructure*, such as bike lanes and trails/paths [57, 63, 90]. Mertens *et al.* found that separated cycle paths are the strongest determinant for perceived bikeability, and that "micro-environmental factors" related to safety such as speed limit and traffic density have a stronger impact when there is not a separated path [77]. In contrast, topography such as steep hills and elevation variability negatively impact bikeability [57, 63]. Lastly, features related to *comfort*, such as surface quality (*i.e.*, smooth pavement) [77] and street scene aesthetics [63, 90, 100] can positively impact perceived bikeability, with Van Holle *et al.* finding that

[1] https://github.com/makeabilitylab/BikeButler

vegetation was the most important environmental characteristic for street *invitingness* [100].

Preferences also change based on contexts; age is correlated with greater desire for low slope and more greenery [56], and traffic volumes impact leisure cyclists more than commuters [42]. Moreover, individuals have different effort tolerances to find fitting infrastructure; both Broach *et al.* [27] and Krizek *et al.* [64] found that cyclists were willing to travel farther to use a dedicated bike path. While the above work demonstrates the influence of contextual factors and the additional effort cyclists invest to reach preferable routes, the literature does not examine how dynamic contexts, such as a trip's purpose, affects route choice—our focus.

2.2 Bikeability Indices

To enable the study and comparison of urban cycling infrastructure, researchers have created *Bikeability Indicies* (BIs), which collate multiple bike-related factors into singular "*bikeability scores*" [18, 50, 57, 91, 110]. Typically, BIs are created either using infrastructure analysis of existing real-world conditions [50, 57] or via analyzing actual bike routes and trips by individuals [2, 26, 110]. For instance, Ito *et al.* use CV across SVIs to quantify 34 bikeability indicators (*e.g.*, buildings, potholes, cleanliness) and create a corresponding score [57] while Long *et al.* create an index using bike-share data [72]. A literature review of BI research [16] identified five key criteria for a high quality cycling network: safety, comfort, directness, coherence, and attractiveness [48, 107]. BikeButler introduces *personalized BIs* by enabling users to set dynamic weights across BI features via custom profiles and voting on suggested route segments.

2.3 Bicycle Route Optimization

While the above research provides robust measures for scoring urban bikeability and for understanding cyclist preferences therein, others have explored how to operationalize such findings into bike route recommendation tools [85, 89, 98]. For example, Preidhorsky *et al.* created an open-source map wiki, *Cyclopath* [85], where users could indicate road segments that were more or less bikeable, and Caggiani *et al.* surface several route options to users based on distance, air pollution, and safety [28]. Most similar to our work is Meng *et al.*, who specify three broad categories (accessibility, visual perceptibility, and cycling suitability), score roads based on those three categories using CV on Street View Images (SVIs), and generate routes of four types ("travel", "leisure", "commuting" and "bikeability") based on assigned weights in those three categories [76]. However, while demonstrating the effectiveness of using CV on SVIs and custom profile-weighted route generation, they did not create an *interactive tool* to generate, visualize, and iteratively refine routes, as we do with BikeButler. Moreover, we provide the first end-user study of such a system, identifying key preferences and additional desired contexts for future work.

Commercial tools also support bike routing and route-based comparisons. For example, Google Maps [46] and Apple Maps [6] show several options on search—often contextualized with bike lanes highlighted or with descriptions such as "*having bike lanes*" or being on "*shared paths*"—and allow selection between them. However, there is no ability for more fine-grained preference specification based on infrastructure or streetscape quality, as we provide in BikeButler. More expert tools, such as Strava [2] and RideWithGPS [5], utilize implicit (sensor-based) contributions of real cyclist routes and provide visualizations (*e.g.*, a heatmap of the most popular routes in a neighborhood). These apps generally also surface elevation data and sometimes bike lane data. However, the crowdsourced contributions are biased towards specific user populations (*e.g.*, Strava is aimed at serious cyclists), do not provide visibility into the *in-situ* bikeability along the route via SVI previews, and *also* do not provide for fine-grained feature based customization. BikeButler relies on open government data, OSM, and automated SVI analyses to provide visibility into the conditions of the route and allow for personalized route creation, but currently does not incorporate crowdsourced route information.

Figure 2: One participant shared an uncomfortable commute route they had "*found on OSM but didn't question it,*" since they trusted the software. Notably, there is no separation between cars and bikes. They later switched to a route they created manually using the Seattle official bike map [13].

2.4 AI for Infrastructure Detection

To assess bikeability, BikeButler uses a combination of structured GIS data and VLM-based analyses of street scenes. With advances in computer vision (CV), particularly in deep learning [62] and now VLMs (foundation models that can process visual and textual data) [33], researchers are increasingly applying automated analyses of SVI and satellite imagery to detect urban features, such as roads [55], sidewalks [38], signage [40, 69], buildings [103], and foliage [31]. Convolutional Neural Networks (CNNs) have been a primary source of success in this domain [57, 65, 76, 108], often by detecting and segmenting the features of interest and counting the number of pixels each feature contains [31, 45, 61, 68, 69, 105]—as in Ito *et al.* [57] and Meng *et al.* [76] for bikeability features (*e.g.*, buildings, road, trees) or Li *et al.* [68] for greenery and Kim *et al.* [61] for surface quality. However, the raw percentage of pixels of features may not fully capture how a *human* would perceive the bikeability of those conditions, as we attempt in BikeButler.

We thus turn to vision-capable foundation models, such as ChatGPT [10] and Gemini [11], which recent studies use to characterize urban infrastructure, given their ability to make zero or few-shot inferences [29, 59, 70, 95, 106, 111]. In particular, attempts to quantify

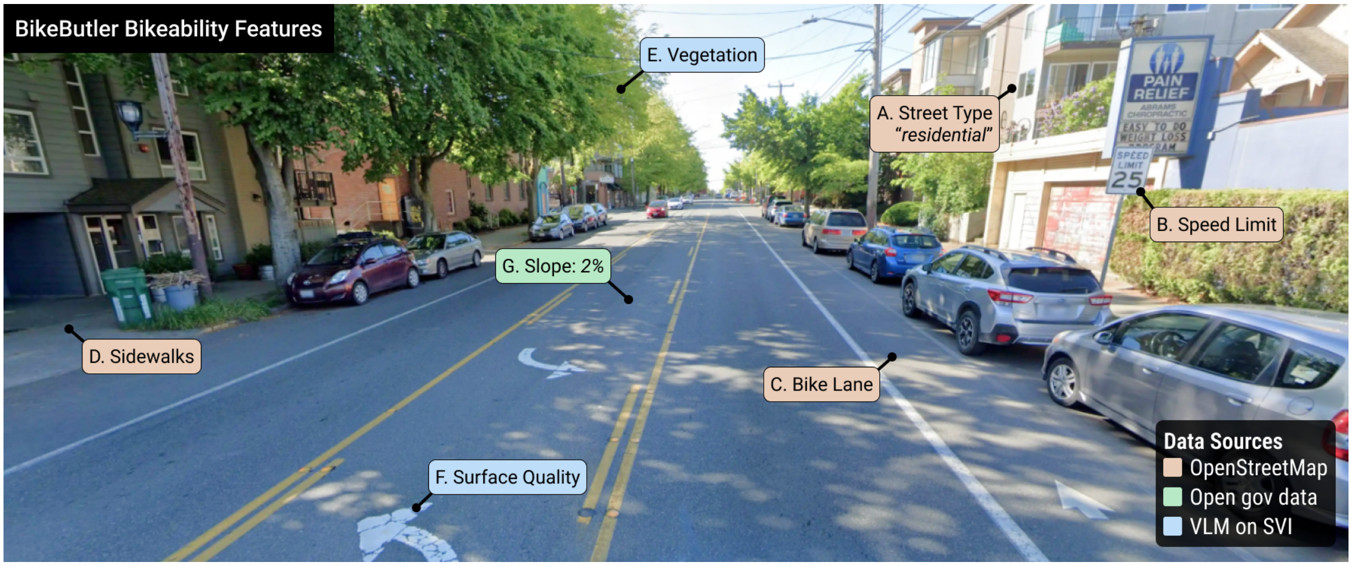


Figure 3: BikeButler identifies and scores seven bikeability features: (a) *street type,* (b) *speed limit,* (c) *bike lane presence,* (d) *sidewalk presence,* (e) *vegetation,* (f) *surface quality,* and (g) *slope* drawing on OSM and open government data as well as a Vision Language Model (VLM) applied to Street View Images (SVIs).

subjective qualities, like traffic safety [95] and walkability [29, 106] have shown promise and moderate to good alignment with human labels [29]. Specifically, Tang *et al.* found that VLMs performed more accurately when the features in question (*e.g.*, lane markings) were distinct and clearly visible, but performed worse with ambiguous thresholds (*e.g.*, vegetation, traffic density) [95]. Cai *et al.* found VLMs saw increased performance when provided expert-made rubrics to score by [29], which we provide in BikeButler. However, given the relative youth of foundation models, no such subjective assessments have been performed for *bikeability.*

3 Designing BikeButler

To design BikeButler, we pursued an iterative, human-centered design process, drawing on key literature (*e.g.*, [57, 60, 85]) and the results from four formative interview sessions. Participants were recruited via snowball sampling, had a range of bicycling expertise, and were all above 18. As a small study, our goal was to complement existing literature in bikeability by drilling down specifically into context-dependent route choice (*e.g.*, bike commuting *vs.* a family ride) and to solicit initial feedback to early-stage prototypes. P1 joined us twice: first for a semi-structured interview and then to provide feedback on a mid-fi prototype. Quotes have been lightly edited for concision.

Current tools & desired features. We began with questions about bike route choices and technology support tools like Google Maps. Participants emphasized key factors, including having visibility into the bikeability of a route and being able to prioritize bike infrastructure—affirming prior work—but also mentioned wanting options and flexibility. Towards current technology tools, all appreciated Google's convenience (including live navigation) and being able to see other cyclist activity on Strava, but felt there were gaps in what they were able to do. For example, P2 lamented that on Google Maps *"I don't typically see many options for biking... [compared to] driving,"* and P4 admitted that on Strava *"you can kind of get the lay of the land, but it doesn't really show you the conditions that you're riding on."*

Context-based preferences. All four participants emphasized that bike route preferences changed as a function of context—even beyond those found in literature, such as age [85], gender [109], and weather [97]. For example, P2 said *"It's difficult to say how I would prioritize things I want on a given day, because one day I might be doing groceries very casually, another day I might be rushing to work."* P3 preferred larger streets, saying *"I typically choose a busier road because... [I'll] always know where I am,"* and that big roads may be more likely to have shoulders where a bike would fit—though, when groupriding, they favor *"roads with bike lanes and less weaving, less [riding on] shoulders."*

Infrastructure features. In terms of infrastructure that supports safe and comfortable biking, our participants mentioned features consistent with prior work such as the width of the road and bike lanes, surrounding land use, pavement quality, road traffic volume, and presence of vegetation [57, 63, 77, 90, 100]. For example, P4 emphasized *"[I need] safety, which could be nice bike lanes, or calm or low traffic"* while P2 stated *"[if taking kids somewhere], I would want greenery or a place farther away from cars where everything is just moving slower."* Similar to prior work, participants mentioned being willing to bike further to achieve a route that fits their priorities: *"I will bike further afield depending on how flat it is"* (P1), consistent with [27, 64].

Mid-fi critique. P1 returned for a second session where we solicited feedback with a mid-fi interactive prototype. While positive overall (*"I like that it seems to be adaptive to your desires"*), they wanted more system transparency (*"I'm worried about changing [these preferences], can I go back?"*), the ability to compare route options (*"I want it to show me multiple routes at the same time"*),

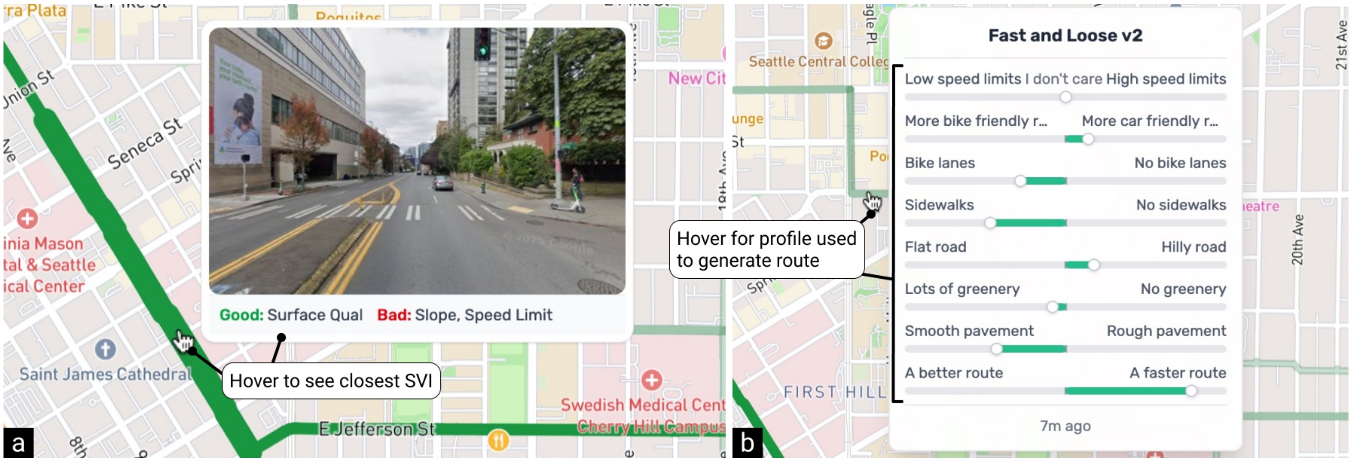


Figure 4: (a) Hovering over the current route will show the closest SVI to the mouse, as well as a text summary of which features are aligned and misaligned with the user's preferences. (b) Hovering over a previous route will show the profile used to create that route, as well as its underlying profile slider settings.

and preferred adjusting preference sliders *vs.* using the up/down votes on segments (*"[liking/disliking] is less fine grained and you don't understand what's happening behind the scenes"*).

3.1 Design Considerations

From the literature and our formative study sessions, we synthesize the following design considerations for a context-sensitive bike routing tool:

Transparency. Drawing on best practices in recommender systems [86, 93], enable users to understand *why* a route is recommended and interrogate the system to learn more.

Expressing preferences. Enable users to influence the route generation [86] and filter results by route type and desired infrastucture features.

Seamless comparison. Allow users to easily modify, compare and switch between routes, enabling them to iteratively refine their choices and discover their own preferences [81].

Best practices. Support key features common in other routing tools [2, 46], such as an elevation chart, seeing multiple routes, satellite view, and dragging the route to make small adjustments.

Key use cases. Finally, we distilled the following key use cases, including (1) commuting [53], (2) recreation [27], (3) utility (*e.g.*, grocery shopping) [52], (4) riding with other people [20], and (5) other, miscellaneous, transient contexts, such as weather [97], time of day [99], and mood.

4 The BikeButler System

Informed by the above design considerations, we iteratively created *BikeButler*, a personalized, context-sensitive bike routing and visualization tool, supporting context-level preference specification and route generation. As an initial prototype, we constructed BikeButler to support a single US city, Seattle, WA, a large metropolitan area with an expansive bicycle infrastructure network [13], varied topography, and reliable, high-density OSM bike data [83]; however, our underlying approach and algorithms should work in any city contingent on data availability. Below, we describe BikeButler's UI, data sources, and personalized routing algorithm. See also the video in *Supplementary Materials.*

4.1 User Interface and Interaction Design

BikeButler enables users to iteratively create and compare personalized bike routes using a custom-weighted node-network routing graph from a set bikeability profile. The primary UI (Figure 1) is composed of five key components: (1) the origin and destination (OD) search bar for generating routes and setting bikeability profiles; (2) the base map toggleable between a stylized view (default) and a satellite view; (3) overlaid route visualizations showing color-coded segments and up to two previously generated routes for the same OD pair; (4) a distance and elevation graph for the current and previous routes; and a (5) detailed view sidebar (on the right) that shows an SVI-based preview of each segment along with a profile-based summary. We provide key details below.

Generating and viewing routes. To generate a bike route, users first specify an OD pair and select or create a bikeability profile, which defaults to *All Around* or their last used selection. For the OD, users can either click directly on the map or type into the search bar (*e.g.*, *"Seattle Public Library"*, *"211 Alaskan Way"*), mirroring commercial tools. To set their bikeability profile, users can select one of three pre-generated profiles, including *All Around*, *Slow and Safe*, and *Fast and Loose*, or create their own via custom sliders (see Figure 6 and Figure 7). Clicking on the "*Get Bike Route*" button generates and visualizes the customized route, which is segmented and color-coded based on the user's selected profile—"*poor fit*" segments are in red, *neutral* in yellow, and "*good fit*" in green (Figure 1).

Inspecting a route. To enable route-related decision making and algorithmic transparency, BikeButler provides both high- and low-level route details, following Shneiderman's "*overview first then details on demand*" mantra [92]. First, users can mouseover route segments and immediately see the closest street scene (SVI) thumbnail at that location along with a bikeability score summary

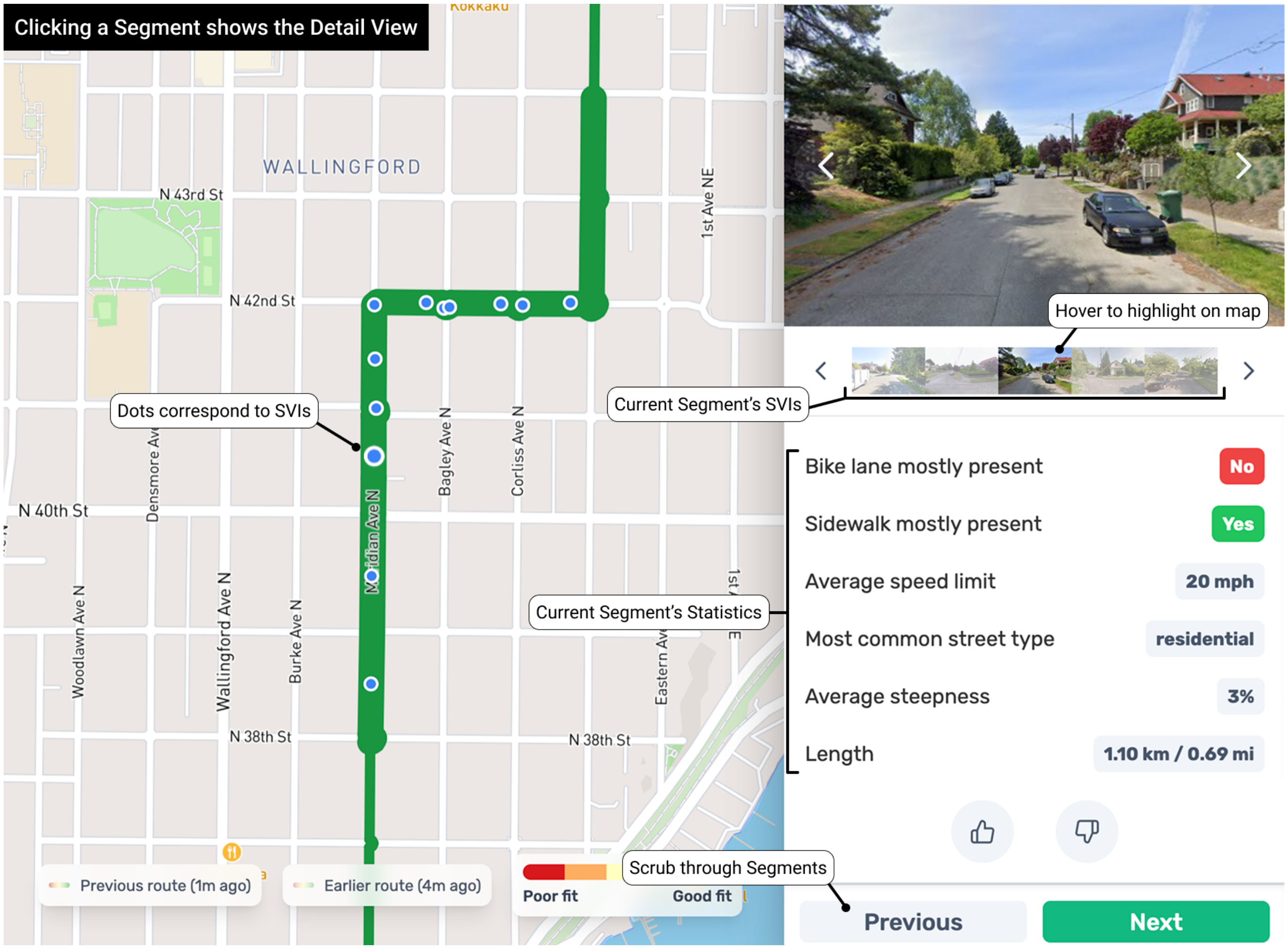


Figure 5: Clicking a segment on the map or on the sidebar opens the detailed view. The user can scroll through the SVIs along the segment, see details such as the average steepness, length, and average speed limit, and like and dislike the segment. Blue dots appear on the map indicating where each SVI is located, and hovering over the gallery will highlight its respective dot.

explaining the color coding (Figure 4a), enabling both a quantitative and user-oriented subjective assessment. In addition, the right sidebar shows an ordered segment-by-segment SVI-based preview of the route (Figure 1), inspired by [60]. Second, to drill down, users can click on a route segment either on the map or sidebar, which auto-zooms the segment into view, shows available SVI imagery located along those edges as blue dots along the segment, and provides detailed segment-related bikeability scoring data in the right sidebar (Figure 5), including *bike lane availability*, *sidewalk presence*, *average speed limit*, *street type*, *average steepness*, and *segment length*. Complementing the scoring assessments is a large SVI preview of the currently selected dot on the map and a small thumbnail gallery of all available SVI imagery along the segment, which is bidirectionally synchronized with the map (*e.g.,* hovering over a thumbnail changes the currently focused dot on the map and vice versa). Third, to further contextualize the route, users can examine and hover over the elevation chart on the bottom-left (Figure 1), which displays a topographical line graph (with dots demarcating each segment) along with the total distance and elevation gain/loss. The user can also toggle the basemap to a satellite view enabling birds-eye inspection.

Modifying and comparing routes. To modify a route, users can directly vote on a segment—by clicking on the thumbs up/down buttons on the segment preview cards—or by selecting or creating a new profile. For either interaction, BikeButler immediately shows a non-modal pop-up that says *"Your preferences have changed"* with a "*Regenerate Route*" button. The user can continue tweaking their profile or voting on other segments until clicking "regenerate." For the voting interactions, we alter the underlying sliders to become more similar/dissimilar to the liked/disliked segment.

Once "*Regenerate Route*" is pressed, BikeButler uses the updated profile to generate a new route while showing the previous route semi-translucently, enabling comparison. If the user hovers over segments from the previous route, we no longer show the SVI preview card but instead the bikeability profile sliders that produced it (Figure 4b). Users can quickly switch between active routes by

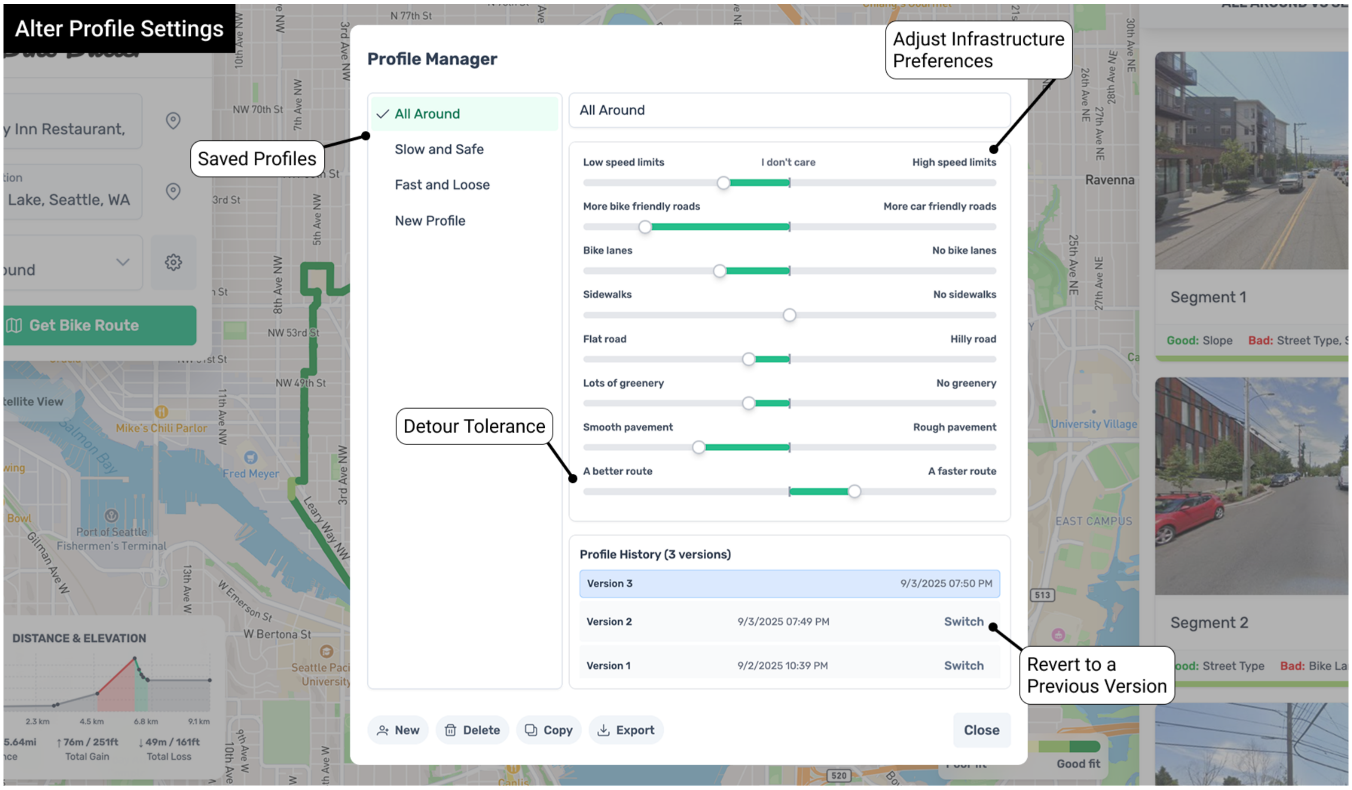


Figure 6: Clicking the settings dial opens the Profiles menu, where the user can create, copy, delete, and modify profiles. There are eight bikeability preference sliders, seven controlling desire/avoidance of streetscape qualities, and one indicating tolerance for detours. Changes and previous versions are saved so the user can easily switch back to a previous state.

clicking directly on them. To help track routes, we auto-version them and show a history of route creation at the bottom of the UI. To avoid overwhelming the user, we visualize up to three routes simultaneously for the same OD pair (active route as opaque plus two semi-translucent routes).

Creating a profile. At the core of BikeButler is the ability to create custom *bikeability profiles*. Users can select an existing profile by clicking the dropdown menu next to the OD field or create a new one via the profile settings hub (Figure 6). Each profile has eight sliders: seven for infrastructure features (*speed limit*, *street type*, *bike lane presence*, *sidewalk presence*, *slope*, *greenery*, *surface quality*), and one for detour tolerance. The left generally indicates higher "bike-friendliness." Section 4.3 describes how we encode these preferences into our context-sensitive routing algorithm.

4.2 Data Sources

BikeButler relies on a combination of bikeability features to generate context-sensitive routes; however, these features do not exist in single GIS databases or, in the case of subjective qualities such as *vegetation density* and *surface quality*, at all. Thus, BikeButler draws three sources of data: OSM metadata (for *speed limit*, *street type*, *bike lane presence*, and *sidewalk presence*), a custom VLM-based analysis of SVIs from Google Street View (GSV) (for *vegetation* and *surface quality*), and open government data (for *elevation*). See Figure 3 and Table 1.

OpenStreetMap. *OpenStreetMap* (OSM) is a collaborative, open-source geographic information system (GIS) that also increasingly provides structured data on pedestrian [8] and bike infrastructure, such as dedicated cycle lanes, paved trails, and even the existence of bike racks [80, 101]. As a crowdsourced database, information quality and data density can vary; however, studies show that OSM

Feature	Source	Min	Max	Neutral
Speed Limit	OSM	0	60	25
Street Type	OSM	See Section A.4		
Bike Lane	OSM	-1	1	0
Sidewalk	OSM	-1	1	0
Vegetation	SVI	1	5	3
Surface Quality	SVI	1	3	2
Slope	DEM	-10%*	10%	0

Table 1: For each infrastructure feature, we assume a minimum, maximum, and neutral value, allowing us to later normalize scores to the range [-1, 1]. *Slopes less than -8% are treated as positive (*i.e.*, we take the absolute value), as steep downhill slopes have also been shown to be dangerous [51, 96].

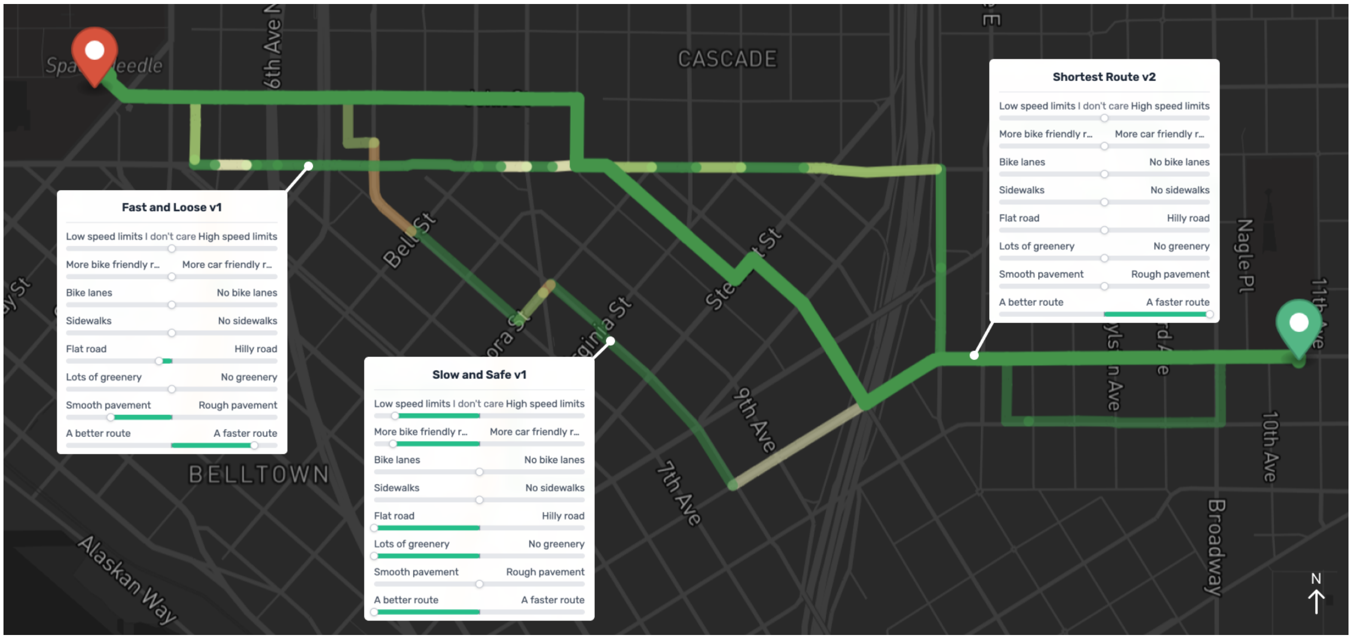


Figure 7: Altering profile settings can produce distinctly different routes with different priorities. Above is a shortest route, a slightly smoother route, and a more "bike-friendly" route. (Picture is color-modified for illustrative purposes).

quality can reach or exceed commercial systems [49, 79], including for bicycle-related data [41], our focus. In BikeButler, we use OSM to derive speed limits, bike and pedestrian infrastructure, and street type, as well as its road, bike, and footpath network for our routing algorithm. While OSM bike data quality varies [54, 114], our own informal experiments suggest Seattle's data is high quality and the local OSM community is diligent, reaffirming [83]. We fetch Seattle's bikeable network using `osmnx` [21] and the *OSM Overpass API* [9]. We remove all highways and any edge with the "bicycle" tag set to False, but include footpaths with the "bicycle" tag set to True. For the comprehensive `osmnx` filter, see Section A.3. For each network edge, we fetch the speed limit (`speed_limit`), street type (`highway`), bike lane presence (if the edge contains `bicycle`, `cycleway`, `bicycle_street`, or `cyclestreet`), and sidewalk presence (if the edge contains `sidewalk`).

Elevation Data. Because topography can significantly affect route choice [56, 63], BikeButler incorporates high-resolution elevation data drawn from Seattle's open government database [1]. Specifically, we use Seattle's 10-meter *Digital Elevation Model* (DEM) files to characterize the topology of each OSM-based street segment. To estimate directional slopes (positive indicating uphill), we fetch elevation data from segment endpoints.

Street View Images For the remaining features not present in structured databases (vegetation and surface quality), we preprocess and score GSV images using a Vision Language Model (VLM)—drawing inspiration from [57]—to characterize street scenes with text descriptions. These images are also used to show representative "first-person" bike preview thumbnails along a route, as in Figure 1, Figure 4a and Figure 5. A key challenge is in determining *which* SVIs best visually represent a bike route—serving both as useful images to the VLM as well as the end user for route previews. Thus, using the *Seattle Department of Transportation* (SDOT) street network [12], we generate sample points at critical decision points, including all intersections, complemented by 100 meter intervals along each street edge. For edges shorter than 100 meters, we take the midpoint. This produces 55,405 street edge sample points in Seattle. For each point, we query and download the closest available SVI within 50 meters from the GSV API [14], resulting in 51,997 unique images—seven sample points (0.013%) contained no SVI imagery. We then preprocess and save each panorama's vegetation and surface quality by running our VLM characterization pipeline on all downloaded SVIs: the front, back, and top-down views (simulated via gnomonic projection [23]) are extracted from each panorama, packaged with the street type and speed limit of that street, and passed to Gemini 2.5 Flash to score vegetation, surface quality, and bike lane existence on Likert-scales (*vegetation* from 1-5, 1 being "*barren*" and 5 "*lush*"; *surface quality* from 1-3, 1 being "*very poor*" and 3 "*excellent*"; *bike lane existence* where 1 is *present* and 0 *absent*; and *bike lane width* being *narrow*, *adequate*, or *wide*, inspired by [29]). See Section A.2 for the full prompt, and Section 5 for our evaluation.

4.3 BikeButler's Preference-Weighted Routing

To support personalized bike routing, BikeButler introduces a custom preference-weighted routing algorithm that scores each OSM-based street edge along our seven bikeability criteria and generates a route optimized for the user's profile preferences. We provide algorithmic details below; thresholds were learned through experimental iteration. See also our GitHub repository.

Preference vector. We encode the user's currently selected profile as an eight-dimensional preference vector along the seven

Diff	*Vegetation* R1 vs Gemini %	R2 vs Gemini %	R1 vs R2 %	(R1 & R2) vs. Gemini	*Surface Quality* R1 vs Gemini %	R2 vs Gemini %	R1 vs R2 %	(R1 & R2) vs. Gemini
-2	3.5%	5.0%	1.5%	3	0.0%	0.0%	3.0%	0
-1	23.0%	29.0%	11.5%	26	19.0%	9.5%	34.5%	10
0	58.0%	57.0%	58.5%	79	65.5%	58.0%	50.0%	76
1	15.0%	8.5%	27.0%	9	15.5%	31.0%	12.5%	14
2	0.5%	0.5%	1.5%	0	0.0%	1.5%	0.0%	0
				Sum: 117				Sum: 100

Table 2: Results of the comparison between R1, R2, and Gemini's scoring for Vegetation and Surface Quality. Raw percent overlap is similar between all three pairs, with Gemini agreeing between 58-66% of the time with at least one reviewer. Notably, R1 and R2 also have only about 50% agreement.

key bikeability criteria (*speed limit*, *street type*, *bike lane*, *sidewalk*, *slope*, *vegetation*, and *surface quality*) as well as detour tolerance (*i.e.*, the additional effort to reach a more fitting route). For each feature, preference is expressed as a float from [-5, +5], negative indicating *desire* and positive indicating *avoidance*. For example, a preference of -5 for "*speed limit*" indicates a strong desire for low speed limits. The last weight, detour tolerance, is a float from [0, 1] with 0 indicating maximum detour willingness and 1 no tolerance.

Street edge vectors. For each OSM edge, we similarly calculate and save a seven-dimensional vector encoding its bikeability score. Each score is normalized to the range [-1, 1], with negative indicating maximal "bikeability" and positive indicating minimal bikeability, matching the preference vector valence. For continuous numeric features (*speed limit*, *slope*, *vegetation*, and *surface quality*), a *max*, *min*, and *neutral* bikeability value is assumed and the range of the feature is normalized to [-1, 1]. For binary features (*bike lane*, *sidewalk*), presence is mapped to -1 and absence as +1. For street type (ordinal), each type is mapped to a numerical bikeability score from [-2, 2] based on our formative findings and the research team's own biking experience. For example, the OSM *"cycleway"* type is -2 and *"primary"* is +2; that range is then normalized to [-1, 1]. See Table 1 and Figure 3. Finally, for the VLM-based SVI features (*vegetation*, *surface quality*), we compute the average score of all SVIs along an edge. If none exist, the score is set to neutral (0).

Calculating edge weights. The weight for a given edge has two components: preference adherence (α) and edge length (β = *length*). We calculate α using a bonus/penalty framework starting with edge length as the base weight. For each feature, we define a "bonus window" sized by preference strength (smaller for stronger preferences) and positioned based on preference direction; when the edge's score for a feature is *within* the window, we apply a bonus, and when it is *outside* the window, we apply a penalty. The magnitude of the bonus/penalty scales proportionally to distance from the optimal (inversely, for bonuses). Finally, to then calculate α, if any feature has penalties, we disqualify that edge by setting $\alpha = 10 \cdot length \cdot \sum penalties$. Otherwise, $\alpha = (1 - \sum bonuses) \cdot length$. The final edge weight then combines both α and β, scaling by detour tolerance d: $weight = (1-d)\alpha + d\beta$. For more details, see Section A.1.

Routing and segmentation. With the road network and edge weights computed, we use Dijkstra's algorithm [35] to find the optimal path between the nodes closest to the origin and destination locations. To segment the routes, we split the route at points where the feature vector changes significantly. Here, we calculate cosine similarity between consecutive edge vectors (considering only nonzero preferences), grouping edges with similarity >0.5 with the previous segment and starting new segments when similarity $\leq$0.5. We then take a second pass to eliminate segments shorter than 20 meters by reassigning edges to the most similar neighboring segment. Then, for the aggregated segment statistics (*e.g.*, speed limit, *"mostly has bike lane"*, street type), we take the weighted mean (for street type, mode) based on length of each edge.

Segment voting. Finally, when the user "likes" a segment, we *add* the segment's feature scores to the preference vector, and for a "dislike," subtract. We then clamp the preference vector to [-5, 5]. Thus, when a segment is "liked", the preference vector is shifted towards similar segments, and the opposite when disliked.

4.4 Software Implementation

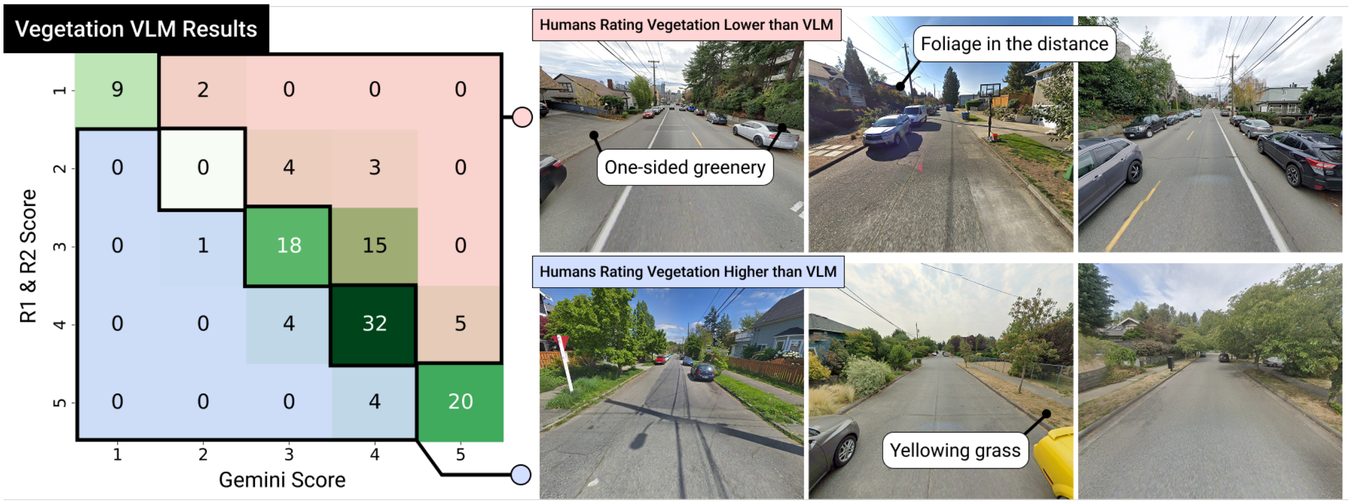
BikeButler is a web app implemented in *Nuxt* [7] with *Mapbox* [3] for the map and routing visualizations, and backend in *Python*, `osmnx` [21], and *Flask* [4]. We use the real-time *Google Places API* [15] with autocomplete for the origin-destination search. For route generation, we incorporate data from *OpenStreetMap* (OSM) [84], the *Seattle Department of Transportation* [1, 12], and *Google Street View (GSV)* [47]. For routing, we use `osmnx` with single_source_dijkstra [21], passing in our custom weight algorithm. We store the routing network, generate routes, and deploy BikeButler's backend on a dual Intel Xeon server (24 cores, 10GB allocated memory).

5 VLM Evaluation

Similar to Ito *et al.* [57] and Meng *et al.* [76], we use CV to automatically assess street scenes for bikeability; however, while both Ito and Meng employ custom-trained deep learning methods for semantic segmentation, we use a state-of-the-art VLM, *Gemini 2.5 Flash* [33], for its high-performance in image understanding. To examine a VLM's ability to score bikeability traits compared to a human, we performed a controlled technical performance evaluation. We assessed Gemini's ability to score *vegetation*, *surface quality*, and *bike lane existence* and *width* compared to two human labelers.

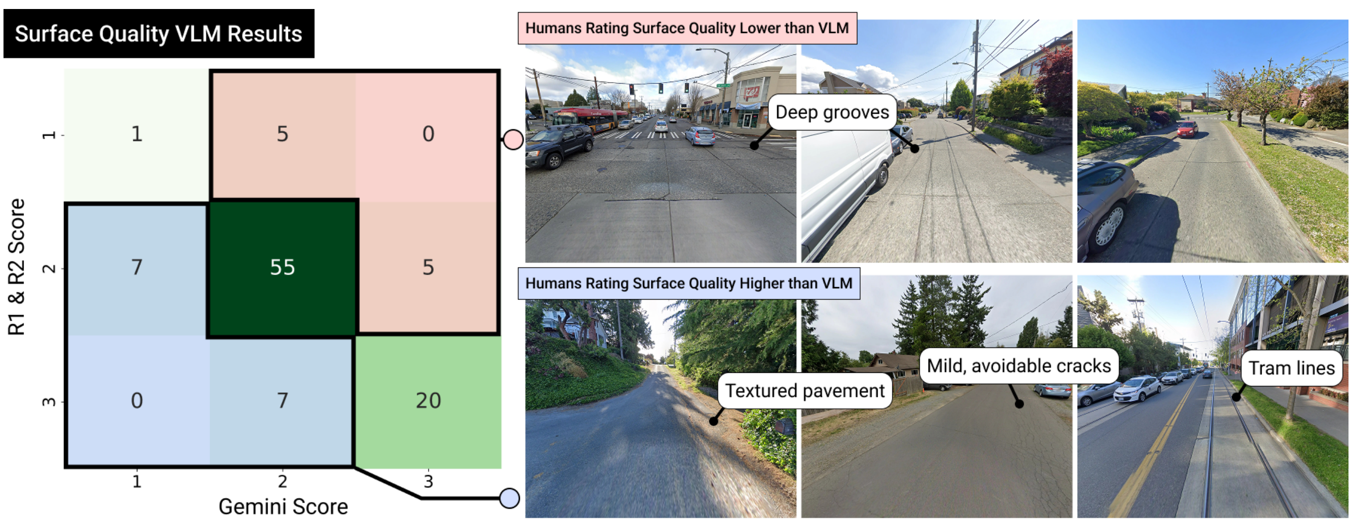
5.1 Method

To evaluate performance, we randomly sampled 200 panoramas from our total (*N*=51,997). For human labels, two members of the



R1 & R2 Score \ Gemini Score	1	2	3	4	5
1	9	2	0	0	0
2	0	0	4	3	0
3	0	1	18	15	0
4	0	0	4	32	5
5	0	0	0	4	20

Figure 8: Comparing Gemini's *vegetation* scoring performance (scale: 1-5) compared to only those 117 SVI images where R1 and R2 exactly agreed (N=117). A majority of scores (68%) aligned with the humans. Our qualitative analysis of errors identified where Gemini overestimated (*e.g.*, trees only on one side or minimal foliage) and underestimated (*e.g.*, yellowing grass, vegetation on periphery) compared to the human labelers.



R1 & R2 Score \ Gemini Score	1	2	3
1	1	5	0
2	7	55	5
3	0	7	20

Figure 9: Results of Gemini's performance in scoring *surface quality*, compared to the images where R1 and R2 agreed (N=100). Gemini overestimated the surface quality where there were deep grooves in the road, but underestimated when there were only minor cracks, textured pavement or tram lines. However, a majority of scores (76%) aligned with the humans.

research team (R1 and R2) manually scored each panorama. Both the humans and the VLM received the same prompt and data: the street type and speed limit, and the front, back, and top-down views extracted from the SVI panorama, and were instructed to score on the same scales and output, as in Section A.2. Because scoring street view scenes for bikeability is a *subjective* assessment, we do not just compare human labels to the VLM, but humans to each other. We calculate raw agreement between all three pairs (R1 & Gemini, R2 & Gemini, and R1 & R2) as well as the weighted Cohen's Kappa [32]. We also perform a qualitative analysis of where Gemini scored differently from humans.

5.2 Results

Overall, we found that Gemini was able to achieve moderate alignment with human labels for *vegetation* and *surface quality* but not for *bike lane existence* or *width*. Interestingly, the VLM achieved a similar level of alignment to each human as the two humans to each other. We expand on our findings below.

	Vegetation	Surface Quality	Bike Lane Width
R1 vs. Gemini	0.61	0.50	0.18
R2 vs. Gemini	0.55	0.27	0.24
R1 vs. R2	0.63	0.23	0.25

Table 3: Results of a Weighted Cohen's Kappa test between the three scorers of 200 SVI: R1, R2, and Gemini. Overall, the VLM achieved a similar level of alignment to each human as the two humans to each other.

Bike lane existence and width. As the worst performing category, Gemini only detected 37.5% of bike lanes R1 labeled and 25.6% that R2 labeled. Moreover, for width estimation, of the 16 bike lanes identified by R1 and 39 identified by R2, the VLM achieved a micro-F1 score of 0.06 for width categorization compared to R1 and 0.22 compared to R2.

Vegetation. In contrast, Gemini nearly matches human-level performance for *vegetation* scoring with exact rating agreement scores of 57% and 58% between R1 and R2. Of the 170 total disagreements (between either R1 or R2), 151 (88.8%) are only off by 1 (on the 5-point scale). As important context, the humans themselves agreed on only 117 out of 200 images (58.5%). See the weighted Cohen's Kappa [32] in Table 3. To more deeply examine differences between human and VLM scores, we created a confusion matrix of the 117 SVIs where both R1 and R2 agreed and compare it to VLM output (Figure 8, left). We then qualitatively examined all SVIs where the VLM either over- or under-estimated the human ratings (Figure 8, right). For VLM overestimates, most errors were due to greenery in the distance (*N=21*), on fringes but not dense (11), or only some grass, small bushes, or a single tree (6). For underestimates (*N*=9), four SVIs included yellowing grass and another four were largely composed of pavement but still had some greenery.

Surface Quality We perform the same analysis for surface quality (Table 3; Figure 9). In this case, we find a similar percent overlap between the three pairs. However, notably, Cohen's kappa between R2 and both other labelers is much lower (~0.25 compared to 0.50), and the two humans only agreed on 100 out of the 200 SVIs. Our complementary qualitative analysis (see Figure 9) identified both over- (*N*=10) and under-estimates (*N*=14). For overestimates, the dominating pattern (8) is deep grooves in the road. Underestimates are more mixed; four SVIs have textured pavement, five have mild avoidable cracks, and many have miscellaneous marks like shadows, grates, or tram lines.

Summary. In sum, Gemini performed comparable to two humans in rating *vegetation* and *surface quality* but not to identify *bike lanes* or *estimate widths*. Thus, we include VLM inferences from only the former two in BikeButler.

6 User Study Method

To evaluate BikeButler and explore the potential of a context-sensitive bicycle route planner, we conducted a two-part user study with 16 participants followed by a technical comparison between BikeButler and Google Maps using participant OD data. Towards our iterative design process, four of the 16 participants completed Part 1 (formative interviews) in an earlier session (Section 3) before returning for a lab-based evaluation of BikeButler (Part 2).

P	Age	Self-rated Expertise	Travel Freq	Primary Bike	Setting
1	25-34	Advanced Beginner	2-4/week	Road	In Person
2	25-34	Competent	<1/week	Hybrid	Remote
3	18-24	Proficient	>4/week	Road	In Person
4	25-34	Competent	>4/week	Hybrid	Remote
5	35-44	Expert	<1/week	E-bike	Remote
6	25-34	Proficient	2-4/week	Cruiser, City	Remote
7	25-34	Proficient	>4/week	Hybrid	Remote
8	18-24	Proficient	>4/week	Road	Remote
9	35-44	Proficient	2-4/week	Road, Mountain	Remote
10	35-44	Proficient	2-4/week	Road	In Person
11	35-44	Proficient	4/week	Road, E-bike	Remote
12	25-34	Competent	2-4/week	E-bike	Remote
13	25-34	Proficient	>4/week	E-bike	Remote
14	35-44	Competent	2-4/week	E-bike	Remote
15	45-64	Competent	2-4/week	Road	Remote
16	45-64	Proficient	>4/week	Road	Remote

Table 4: We recruited participants who are over 18, can speak English, and have cycled in an urban setting. Travel frequency includes personal and commuting trips. For self-rated expertise, we use the five-point Dreyfus model of skill acquisition [37] ("*Novice*," "*Advanced Beginner*," "*Competent*," "*Proficient*," "*Expert*").

6.1 Participants.

We recruited participants with varying cycling backgrounds via word-of-mouth and snowball sampling. Study advertisements linked to a screening survey that collected demographics, self-assessed biking expertise, bike-trip frequency, and primary type of bike (*e.g.*, road bike, e-bike). Participants had to be 18+ to participate, speak English, and have cycled in an urban setting. Fourteen participants were familiar with Seattle; thirteen lived there. See Table 4.

6.2 Procedure.

The study consisted of two parts: Part 1, a formative interview exploring how participants think about and plan bike routes across contexts, and Part 2, a lab-based study of BikeButler with three routing tasks. As noted above, one participant used a mid-fi prototype, after which we added comparing previous routes, saving profile versions, thumbnail text summaries, and the SVI map markers. All subsequent participants used the updated version. P16 did not complete the last task of Part 2 due to a hardware failure. Most interviews occurred via Zoom, with three conducted in person. Sessions lasted 60-90 minutes, and participants received US$25 per hour. A single researcher conducted all interviews. The full study protocol is in the *Supplementary Materials*.

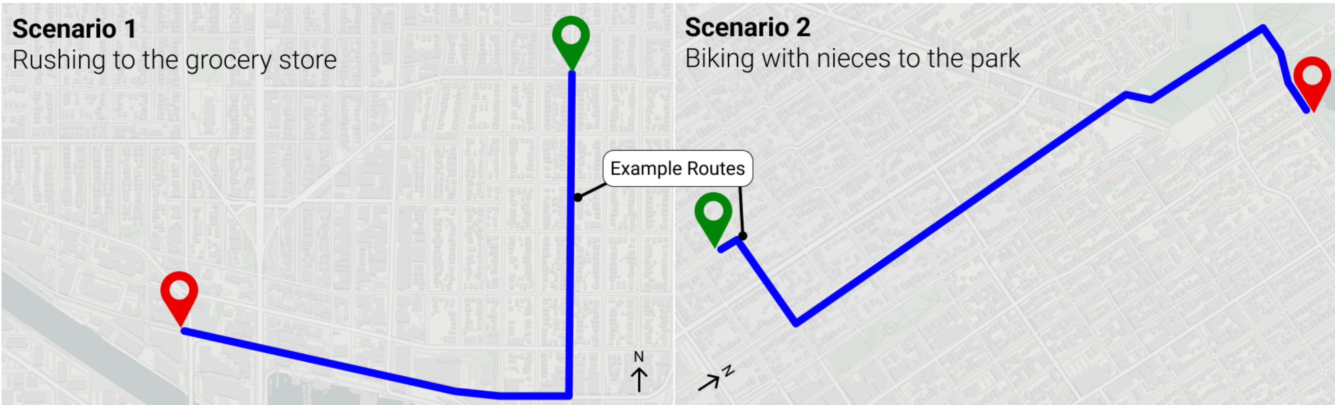


Figure 10: For Tasks 1 and 2, participants were provided origin/destination pairs and biking scenarios. They were asked to imagine "*going to the grocery store in a rush*" (Scenario 1) and "*bike with their two young nieces to the park*" (Scenario 2).

Part 1. Formative Interview. In Part 1, we asked participants about how they plan bike routes, prioritize bikeability features, and how these priorities may vary across contexts (*e.g.,* weather, ride purpose). We also inquired about their use of existing bike routing tools and their feature preferences and desires therein.

Part 2. BikeButler Tasks. Part 2 focused on three routing tasks: two scenario-based tasks and an open-ended task where participants could explore their own origin-destination pairs. Participants were encouraged to "think aloud" and ask questions. Before the first task, participants received a ~5 minute introduction to BikeButler, which included navigating the interface, generating an example route, and providing expert tips derived from our own usage, such as: how to use BikeButler's satellite view to supplement the SVI thumbnails and that increasing the "bike lane" slider may surface more arterial roads since smaller streets are unlikely to have bike lanes. Participants then received the BikeButler link, shared their screen, and were asked to familiarize themselves with the system and ask questions.

After the introduction, participants began the three routing tasks, starting with two scenario-based tasks before an open-ended exploration. For the scenario tasks, participants were all given the same origin-destination pair and told to find the best route that fit the context, for their preferences. Scenario 1 asked participants to "*imagine you are going to the grocery store in a rush*" while Scenario 2 asked participants to "*imagine taking their two young nieces, aged 8 and 10, to the park (they are confident bikers, but still young)*," with an unavoidable hill climb (Figure 10). For the third task, participants were instructed to start with a familiar route and interactively explore options. For the two participants unfamiliar with Seattle, we provided a list of candidate origins and destinations, including grocery stores and tourist sites. Sessions concluded with a short interview reflecting on BikeButler and opportunities for improvement.

6.3 Analysis.

We recorded and transcribed all interview sessions. For analysis, we used a combination of deductive and inductive coding [24]. We created an initial codebook based on the interview protocol, and one primary researcher coded all interviews, updating the codebook as necessary. Another researcher then peer-checked three randomly sampled coded transcripts and met with the primary researcher to discuss disagreements [78]. The codebook was then updated and the transcripts were spot-checked accordingly. The codebook is included in *Supplementary Materials.*

7 User Study Results

We describe key findings related to context-dependent route choice, bike infrastructure preferences, and reactions to and usage of BikeButler. Recall that four participants completed Part 1 as part of our co-design process (see Section 3), and then later returned for Part 2 once the prototype was finished. All other participants (*N*=12) completed Part 1 and 2 in a single session with the finalized prototype. We include only the 12 new participant findings for Part 1 below and all but P1 (*N*=15) for Part 2 (as P1 used an earlier prototype).

7.1 Part 1: Formative Interview Findings

We highlight findings related to context, bike feature prioritization, and current routing tools emphasizing recurrent themes from our co-design sessions as well as new emergent themes.

Context-based preferences. A key focus of our work is in exploring how cycling route choice is influenced by dynamic context. Echoing our co-design participants, bike route choice is a complex, nuanced, and multi-faceted decision making process that involves time availability, skill level, ride purpose, weather, street scene aesthetics, bike infrastructure, and more. Beyond, participants emphasized the influence of biking partners, especially children, as captured by P15: *"I'm more comfortable taking risk when it's just me, but [with kids] we put a lot more stake in safety, navigability, and visibility"* and P9 highlighted that each child is different *"because some kids are crazy and some are cautious."* Others mentioned social influences beyond children, as P7 stated *"my girlfriend bikes with me and she's uncomfortable on the road, so we go on nice greenways."* Additional factors included lighting availability (*"I prefer busier streets at night because they're [well-lit],"* P13), physical activity tolerance (*"It's unpleasant if [I show up to work tired], but if [I'm coming home] it doesn't matter as much,"* P9), and even mood (*"I switch routes based on my mood,"* P10).

Bike infrastructure. Reaffirming prior work and our co-design sessions, participants most desired dedicated bike infrastructure, while hills, surface quality, and visual novelty also played a role. For example, P14 said *"bike lanes are good, so long as they're not the super narrow ones because those get really busy,"* P7 said *"if I have children on the back of my bike, I'd be worried about getting put on roads that are really potholey,"* and P10 commented *"I would often go for novelty to see different places, like a bakery along the way."* However, other desires were not consistent, suggesting the need for personalization. For a recreational ride, P13 was *"more likely to want to climb a big hill because there's a cool view at the top."* Participants also disagreed about residential streets and alleyways, some saying they felt safer and some saying the opposite, as encapsulated by P12: *"you might end up having [side streets] to yourself, but some [cars] speed through there even if they're not supposed to."*

Current routing tools and desired features. For bike routing, most participants preferred Google Maps, appreciating its ease-of-use and turn-by-turn navigation (*"I'll put it into Google Maps, make sure it makes sense, then just go,"* P14), though lamented its lack of customization, minimal algorithm transparency, and overemphasis on direct routes. P7 noted *"it will sometimes just give you the most direct route and streets that I know I don't want to go on"* and P6 observed an overemphasis on bike lanes *vs.* nearby safer streets, *"they'll give you where there is a bike lane, but you would be better served to just go on some parallel neighborhood street."* Some participants supplemented Google Maps with tools like Strava or RideWithGPS to see route popularity, with P7 finding *"understanding where people are biking is really helpful,"* and P10 echoing *"if I'm doing recreational riding, I'll spend a long time on RideWithGPS and dragging around and [picking] connections."* However P9 noted potential user biases: *"you have to mentally weight the heat map versus what kind of user Strava has"* (P9). For desired features, participants wanted customizability, real-time route conditions, and ways to input local expertise, with P15 desiring options to *"avoid roads with speed limits of 25 or higher, avoid construction, maximize protected bike lanes or off-road trails,"* P13 emphasizing personal knowledge: *"I'm not usually looking for 'does this road have a bike lane' because I usually already know that kind of thing."* P3 wanted insight on key decision points (*e.g.*, intersections) along a route, *"to know exactly which way to go."*

7.2 Part 2: BikeButler Findings

In Part 2, participants used BikeButler across three tasks and engaged in a debrief interview. Overall, participants responded positively to BikeButler, particularly the ability to interactively customize and compare routes and then rapidly assess them via SVI thumbnail previews, satellite imagery, and elevation graphs. As P14 said, *"I would absolutely use this especially going somewhere new,"* and P2 said *"I would use it, it takes the guessing out of using Google."* BikeButler's features enhanced user's understanding of routes, local geography, and tradeoffs between key factors such as car traffic, bike lanes, and distance. In total, participants created 187 routes across the three tasks (12.5 routes per participant), demonstrating BikeButler's participant's engagement with personalization features. Key criticisms included the opaqueness of profile sliders' weighting, our road segmentation approach, and the lack of direct manipulation of routes (*i.e.,* unlike Google Maps, you can not yet click on and drag route points in BikeButler). We describe key findings related to understanding routes, customizable preferences, and the time investment to use BikeButler below.

Learning the tool. When first familiarizing themselves with BikeButler, participants began by entering their own origin destination pairs and examining the initial route by panning on the map, mousing over segments, and reading the elevation chart. For those participants familiar with Seattle, they entered an OD pair for a route they knew well and compared their own lived experience to BikeButler's recommendation, for example, P15 started by going to *"the waterfront, we bike there all the time."* They then tried different profile settings, mostly by switching between the pre-made profiles but some participants tried the sliders. Overall, first impressions were positive: P9 exclaimed, *"this is far cooler than I had imagined."*

Tasks 1 and 2: Role-play scenarios For the first two tasks, participants were given role-play scenarios along with origin-destination pairs, asked to find satisfactory routes, and to explain their reasoning. Overall, participants created an average of 3.3 (*median*=3) and 4.2 (*median*=4) routes, respectively across Tasks 1 and 2. For Task 1, which emphasized "rushing," participants generally emphasized profile features such as shorter routes and having bike lanes *vs.* Task 2, which included children, and consequently preferentially weighted less hills, smooth pavement, and bike friendly roads.

To create and compare routes, participants generally began by building an initial scenario-relevant profile and generating a route, as P9 said: *"[I'll see] what it gives me to start with and see what I feel like I need to change."* To examine recommended routes, participants first observed the color-codings, focusing most on the red segments (*e.g.*, *"why are you bad?"* P14). Users would then use a combination of the SVIs, the satellite view, and the elevation chart to understand the route and identify what parts they did not like and why. Most commonly, users would open the detail view, click through SVI previews, then check the elevation chart (*"yeah, 3% is fine,"* P9). Then, they would alter their preferences based on what they did not like (*"let's see if I emphasize surface quality a bit more, maybe that will change things,"* P8), and repeat until finding a satisfying route (*"so it's having me go down 45th, which I agree. This is what I would do,"* P4). Some participants ended the task with caveats—P9 said they would *"[go] with this but just shoot across"* a one-block detour BikeButler suggested. In Task 2, participants took longer (4.18 min *vs.* 3.33 min), perhaps to be more careful (*e.g.*, *"I'm a bit torn because this is more efficient, but maybe being on the safe side with kids is better..."* P2).

Task 3: Open-ended exploration For the open-ended exploration, participants generated an average of 5.2 routes (*median*=5.5) by choosing an origin and destination, creating a profile with some starter settings, then iterating as in Tasks 1 and 2. As a demonstration of how BikeButler effectively generates different routes for participants, we display six participants' routes in Figure 11. Participants explored commutes (*N*=4), recreational trips (*N*=5), errands (*N*=2), tourism (*N*=2), and trips to restaurants (*N*=2). Most tried to recreate the route they normally take by *"fiddling with the sliders"* (P9), either for affirmation (*"good to know I have a good route,"* P4), or for route discovery (*"I want to try that out and see if I like it better than the route that I've been taking,"* P5). This led to some exciting moments—like P6, *"what if I optimize for greenery...*

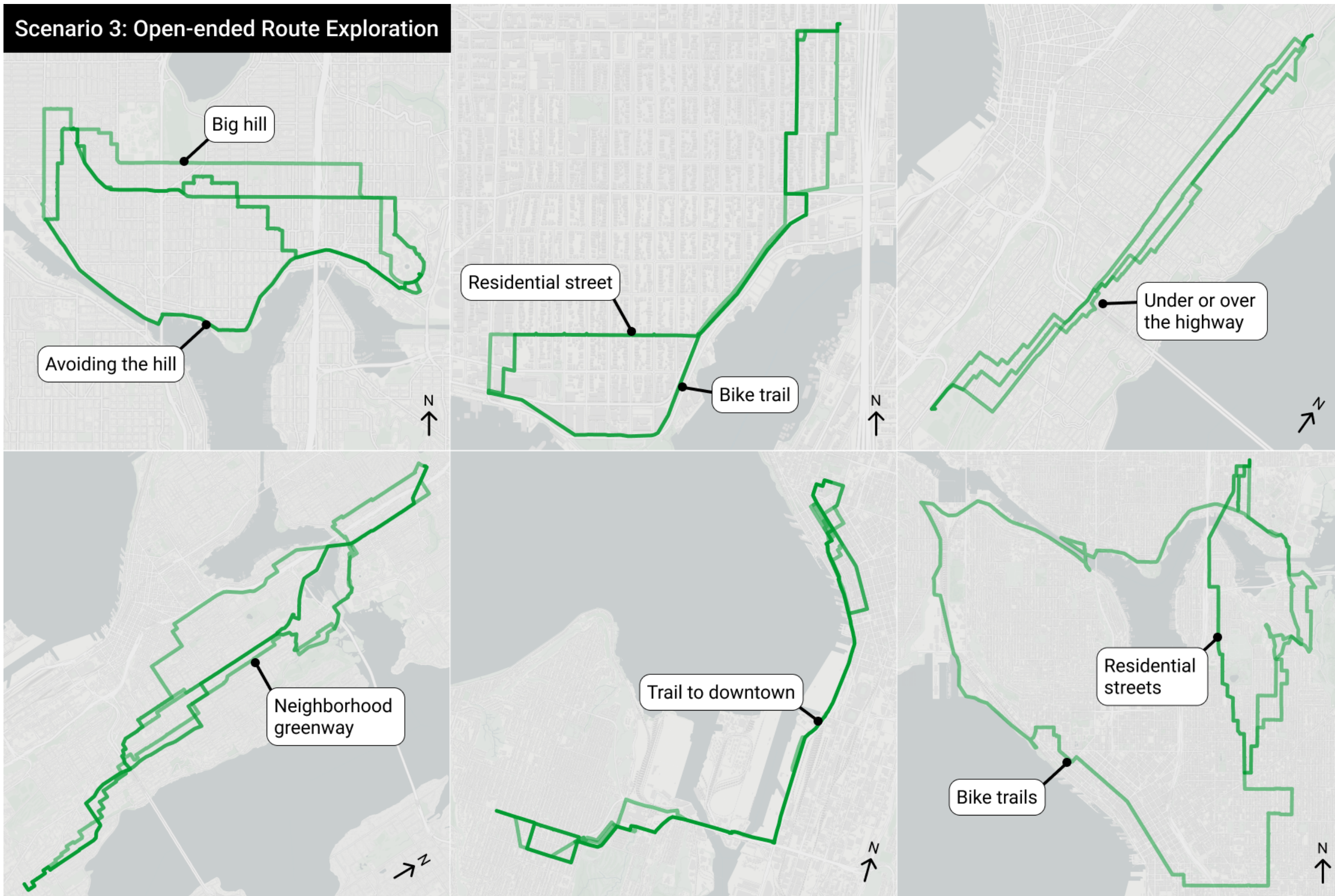


Figure 11: In Scenario 3, participants were asked to pick a route they were familiar with and explore different options via BikeButler. Many participants focused on finding a way to get from their origin to destination according to some goal they had in mind. We display six participants' generated routes, to demonstrate the ability to fine-tune and iterate using BikeButler.

there, that's it, that's my commute"—but others were disappointed (*"it would be cool if you could drag so it would tell you that your route is better or worse for these reasons,"* P9). When exploring route options, participants followed the same approach as Tasks 1 and 2 by identifying points they disliked, then iterating on their preferences.

7.2.1 *Overarching Themes.* We close with some overarching themes.

Understanding route recommendations. To assess and understand route recommendations, participants appreciated the SVI previews, color-coded segments, score summaries, elevation chart, and satellite views. For example, P12 said *"having the street view photos is helpful, it gives you a sense of what to expect,"* and P10 appreciated the seamless integration and responsiveness of seeing SVI previews via mouse hover *"[like] peripheral vision kind of way."* Participants used the SVIs in concert with other data to further their understanding of the route and geography, *"I don't know of another tool that has [all of this information] in one place"* (P13) and *"it's nice to have a reference as to why [the route] wasn't good"* (P14). Emphasizing how this information can even impact the *decision* to bike, P15 said *"[its] knowing what to anticipate... when I can't find enough information on Google [I won't bike]."* However, some expressed concerns about data accuracy and segmentation logic. P10 noted that a segment labeled as "*mostly having a bike lane*" did not in several SVIs, causing distrust in the data. Other participants noted that the slope was averaged and that*"sometimes one segment seems to have many different environments"* (P12).

Customizing routes Participants appreciated being able to rapidly experiment with profiles and compare results: *"I really liked being able to shift the settings and really quickly see how it suggested a new route"* (P15), *"I liked being able to choose different options for the route, and there were factors that I hadn't thought about, like sidewalks"* (P5), and *"it just gave me so much freedom"* (P8). Still, participants wanted to better understand the effect of profiles and liking/disliking segments on route reaction. Most participants used sliders: P13 said, *"it feels more straightforward to set the sliders for the entire route, vs. its unclear what disliking a segment will do."* Others, like P7, felt the opposite: *"I don't have a great sense of what I'm trying to do with the sliders."*

Desired Features. In terms of desired features, the most common was to directly "drag" routes, to make small changes without altering the profile, followed by requesting additional bikeability

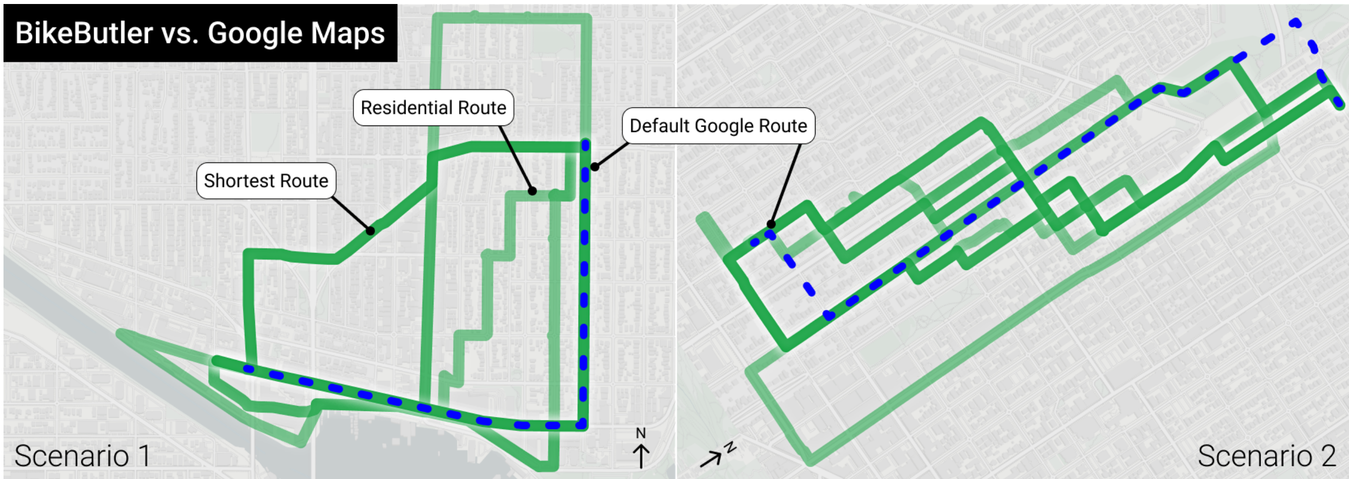


Figure 12: To examine how context-sensitive routes may differ from commercial tools, we compared participants' final routes in Scenarios 1 and 2 between BikeButler (translucent green; deeper shades correspond to more overlap) and Google Maps (blue). BikeButler produced 114 routes, which were downselected by participants to the final unique 22 shown above while Google produced only four.

factors like *intersection/turn* awareness, *lighting*, and *traffic volume*. For example, P14 said *"the annoying part is it has a whole bunch of left turns"* while P11 emphasized: *"that's great, but this has so many turns and I don't like to look at my phone when biking... I prefer to make turns at controlled intersections."* Finally, participants emphasized that for BikeButler to be truly useful, it would need mobile integration and live GPS navigation (akin to Google Maps).

7.3 Comparing Google Maps Routes to BikeButler

To examine how BikeButler's context-sensitive routing algorithm may differ from commercial tools, we compared participants' routes in Scenarios 1 and 2 between BikeButler and Google Maps (Figure 12). In total, BikeButler produced 50 Scenario 1 routes and 64 Scenario 2 routes, which were downselected by participants to a final unique eight in Scenario 1 and 14 in Scenario 2 shown in Figure 12. Google produced only one in Scenario 1 and three in Scenario 2. To enable bikeability comparisons, we fed the default Google Maps routes into our street-edge bike scoring algorithm.

We identified three key differences: first, while Scenario 2 BikeButler routes were longer (by 8%), they were less steep (by 8%). Second, interestingly, Google appears to bias towards dedicated bike infrastructure: our participants ultimately selected final routes with fewer bike lanes (28% less in Scenario 1 and 54% less in Scenario 2) but more residential streets (1.3x in Scenario 1; 2.6x in Scenario 2). Third, BikeButler routes had lower speed limits, better vegetation, and better surface quality—by 11%, 6%, and 12%, respectively in Scenario 1, and 30%, 4%, and 9% in Scenario 2.

8 Discussion

BikeButler explores the potential for a context-sensitive personalized routing tool to enable dynamic customization of bike routes, recognizing that preferences vary based on individual and context. Results from our user study indicate that individuals value being able to seek out routes with specific bikeability features, and appreciate at-a-glance summarizations and *first-person* views of the route, ultimately reducing uncertainty. We discuss limitations, design implications for future bike routing and recommender tools, and future opportunities for the BikeButler system.

8.1 Limitations

Our work has five primary limitations. First, we performed our study with only 16 participants and a single US city; while our participants had diverse bicycling experiences, future work should expand a US focus. Second, and relatedly, BikeButler is reliant on a multitude of datasets—OSM, open government data, and Google Street View. Without good coverage and high-quality data [39, 54, 114], BikeButler will not perform. Third, while our work demonstrates the feasibility of using VLM-based analyses of street scenes for bikeability and achieved human-level performance compared to human labelers, more work is necessary to understand how inference errors impact routing and opportunities for improvement. Fourth, study sessions were limited to ~60 minutes and three tasks; some participants commented that they needed more time to truly understand the impact of profile changes. Moreover, participants did not actually bike recommended routes and compare them to BikeButler's assessments. Future work should examine BikeButler's use via a deployment study. The tool is currently live at https://bikebutler.cs.washington.edu/.

8.2 Scaling BikeButler

BikeButler relies on three widely available sources of data—OSM, open government elevation data, and Google Street View (GSV)—and should function anywhere these data exist. However, utility is contingent on data quality. As mentioned, GSV and OpenStreetMap's reliability and currency is variable [39, 54, 114]. Large cities (such as Seattle) may have broader and more up-to-date coverage, while rural areas may suffer [104]. In these cases, more work and a deployment study is necessary to determine the impact of poor data-quality on BikeButler.

8.3 Implications for Design

Extending and reaffirming the Design Considerations outlined in Section 3.1, from our results we enumerate five key recommendations for interactive tools and recommender systems at large, containing implications beyond just bike-routing tools:

- **Preference awareness.** Tools should recognize that individuals and contexts bear varying preferences [63, 64, 77], and allow for fine-grained personalization accordingly.
- **Seamless iteration.** In BikeButler, users selected routes through a process of preference specification, generation, examination, and alteration. Tools should enable such an iteration loop for maximal flexibility and discovery [81], and allow users to make finegrained changes without significantly altering underlying preferences or suggestion generation algorithms.
- **Data availability and transparency.** Users of BikeButler used every source of data available to them (*e.g.*, SVIs, satellite view, elevation chart, OSM data) to make informed decisions and learn about a route's bikeability. Where one data source was unreliable, they easily referred to another, filling in the gap. Interactive tools should support *multimodal* data where possible, and in a transparent fashion, supporting trust and preparedness [86, 93].
- **Human-in-the-loop.** Tools should empower users to provide insight on the data quality at specific points according to their own experiences, providing live feedback both for developers and other users.
- **Multimodal use cases.** Cyclists examine routes both at home, and on the road. Tools should recognize and optimize for multimodal use cases, acknowledging that the UI priorities may differ, *e.g.,x* when live-navigating *vs.* route planning.

8.4 The Future of Context-sensitive Bike Routing Tools

As an initial prototype, we limited the range of BikeButler to Seattle; but our long-term vision is to have high-quality personalized bike routing everywhere. While BikeButler would technically function anywhere contingent on OSM and SVI availability, we see vast potential for greater street-scene understanding from VLMs, utilization of multimodal data sources, and live-routing awareness.

VLM based streetscape analysis. In our work, and previous studies using VLMs for *subjective* visual analysis [29, 95, 106], VLMs have comparatively strong performance in zero or few-shot scenarios where there are no existing datasets—however, when compared to human labels, they produce only moderate alignment. We attempted a single prompt and a single agent; however, prompt engineering [102], ensemble voting [44], or prompts that utilize output from classical models [57] or other foundation models like Segment Anything [62] for more quantitative input to the VLM could boost performance. Furthermore, we demonstrated that although Gemini was able to reach similar alignment to a human as the humans to each other, the *humans themselves* also had only moderate alignment. This suggests that attempting to score subjective qualities may be misguided, and that further work with robust measures and more individuals is required for greater understanding of how to apply VLMs effectively.

Multimodal data. BikeButler represents a step towards *context-aware* agents that utilize *multimodal* datasets to reason. We utilized government datasets, crowdsourced maps and geotags, and AI on first-person images to give users a more holistic view of the bikeability landscape compared to current tools. We focused on eight features, drawn from literature [57, 63, 77]. However, other factors have been shown to influence bikeability that were not within our data sources: time of day [94], light [25], and weather [17] can be sourced from the user's position. Construction and traffic volumes [42] could come from user-reports or sensor-based live reporting [26]. Historical data, like crime and traffic incidents [26] could influence perceived safety. Providing these data to the user would allow them to more holistically evaluate a route, and inherent redundancies (*e.g.*, satellite and SVIs) would allow for one to fill the gaps in another.

Biking and beyond. BikeButler demonstrates the value of context-based personalized routing for cycling. However, our approach and algorithm could be applied to other transport modes: walking preferences influenced by light [25], noise [88], and cleanliness [22]; indoor navigation considering accessibility [58, 67]; driving [112] to avoid intersections or local streets [19]. While each would require their own data sources and analysis methods, the principle of context-based preferences has broad application.

9 Conclusion

Through mixed-methods research, including a 16 participant user study and developing BikeButler, we contribute:

(1) The first work that concretely establishes context-dependent bike route preferences.
(2) Design guidelines for future bicycle routing tools.
(3) An evaluation of a state-of-the-art VLM's ability to qualify bike infrastructure.
(4) The first context-sensitive personalized bike route generation and visualization tool.

Our work highlights the potential for context-sensitive bike routing tools to encourage greater bicycle adoption by increasing visibility into a region's perceived bikeability according to an individual's own preferences, and we release BikeButler's code for open science. We showcase the need for such tools and demonstrate its feasibility at a large scale.

Acknowledgments

We thank all study participants, Michael Saugstad for his assistance with our study's qualitative analysis, as well as Cynthia Dong, Jerry Cao, and Richard Li for their guidance in our study and experiment design. This work was supported by NSF grant #2411222.

References

[1] 2001. 10-Meter DEM Files in the Seattle 1° by 2° Quadrangle. https://gis.ess.washington.edu/data/raster/tenmeter/byquad/seattle/index.html.

[2] 2009. Strava | Running, Cycling & Hiking App - Train, Track & Share. https://www.strava.com/.

[3] 2010. API Docs. https://docs.mapbox.com//api/guides/.

[4] 2010. Welcome to Flask — Flask Documentation (3.1.x). https://flask.palletsprojects.com/en/stable/.

[5] 2011. Ride with GPS: The Most Trusted Platform for Cycling Routes and GPS Navigation | Ride with GPS. https://ridewithgps.com/new-home-experiment?lang=en.

[6] 2012. Maps. https://www.apple.com/maps/.

[7] 2016. Nuxt: The Progressive Web Framework. https://nuxt.com/.

[8] 2016. OpenSidewalks – The Taskar Center for Accessible Technology.

[9] 2022. API - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/API.

[10] 2023. ChatGPT. https://chatgpt.com/?locale=en-US.

[11] 2023. Gemini Developer API | Gemma Open Models. https://ai.google.dev/.

[12] 2024. Seattle Streets | Seattle GeoData. https://data-seattlecitygis.opendata.arcgis.com/datasets/seattle-streets/explore?location=47.609317%2C-122.329964%2C11.82.

[13] 2025. Bike Web Map - Transportation | Seattle.Gov. https://www.seattle.gov/transportation/projects-and-programs/programs/bike-program/bike-web-map.

[14] 2025. Google Maps Platform Documentation | Street View Static API. https://developers.google.com/maps/documentation/streetview.

[15] 2025. Overview | Places API. https://developers.google.com/maps/documentation/places/web-service/overview.

[16] Tufail Ahmed, Ali Pirdavani, Geert Wets, and Davy Janssens. 2024. Bicycle Infrastructure Design Principles in Urban Bikeability Indices: A Systematic Review. *Sustainability* 16, 6 (Jan. 2024), 2545. doi:10.3390/su16062545

[17] Ran An, Renee Zahnow, Dorina Pojani, and Jonathan Corcoran. 2019. Weather and Cycling in New York: The Case of Citibike. *Journal of Transport Geography* 77 (May 2019), 97–112. doi:10.1016/j.jtrangeo.2019.04.016

[18] Julián Arellana, María Saltarín, Ana Margarita Larrañaga, Virginia I. González, and César Augusto Henao. 2020. Developing an Urban Bikeability Index for Different Types of Cyclists as a Tool to Prioritise Bicycle Infrastructure Investments. *Transportation Research Part A: Policy and Practice* 139 (Sept. 2020), 310–334. doi:10.1016/j.tra.2020.07.010

[19] Theo Arentze, Tao Feng, Harry Timmermans, and Jops Robroeks. 2012. Context-Dependent Influence of Road Attributes and Pricing Policies on Route Choice Behavior of Truck Drivers: Results of a Conjoint Choice Experiment. *Transportation* 39, 6 (Nov. 2012), 1173–1188. doi:10.1007/s11116-012-9391-z

[20] Roger Beecham and Jo Wood. 2014. Characterising Group-Cycling Journeys Using Interactive Graphics. *Transportation Research Part C: Emerging Technologies* 47 (Oct. 2014), 194–206. doi:10.1016/j.trc.2014.03.007

[21] Geoff Boeing. 2025. Modeling and Analyzing Urban Networks and Amenities With OSMnx. *Geographical Analysis* (May 2025), gean.70009. doi:10.1111/gean.70009

[22] Hieronymus C. Borst, Henk M. E. Miedema, Sanne I. de Vries, Jamie M. A. Graham, and Jef E. F. van Dongen. 2008. Relationships between Street Characteristics and Perceived Attractiveness for Walking Reported by Elderly People. *Journal of Environmental Psychology* 28, 4 (Dec. 2008), 353–361. doi:10.1016/j.jenvp.2008.02.010

[23] A. D. Bradley. 1940. The Gnomonic Projection of the Sphere. *The American Mathematical Monthly* (Dec. 1940).

[24] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa

[25] Joseph Breda, Daniel Campos Zamora, Shwetak Patel, and Jon E. Froehlich. 2025. NightLight: Passively Mapping Nighttime Sidewalk Light Data for Improved Pedestrian Routing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3706598.3714299

[26] Joseph Breda, Keyu Chen, Thomas Plötz, and Shwetak Patel. 2025. ProxiCycle: Passively Mapping Cyclist Safety Using Smart Handlebars for Near-Miss Detection. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. doi:10.1145/3706598.3713325

[27] Joseph Broach, Jennifer Dill, and John Gliebe. 2012. Where Do Cyclists Ride? A Route Choice Model Developed with Revealed Preference GPS Data. *Transportation Research Part A: Policy and Practice* 46, 10 (Dec. 2012), 1730–1740. doi:10.1016/j.tra.2012.07.005

[28] Leonardo Caggiani, Rosalia Camporeale, and Michele Ottomanelli. 2017. A Real Time Multi-Objective Cyclists Route Choice Model for a Bike-Sharing Mobile Application. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. 645–650. doi:10.1109/MTITS.2017.8005593

[29] Chenyi Cai, Kosuke Kuriyama, Youlong Gu, Filip Biljecki, and Pieter Herthogs. 2025. Can a Large Language Model Assess Urban Design Quality? Evaluating Walkability Metrics Across Expertise Levels. arXiv:2504.21040 [cs] doi:10.48550/arXiv.2504.21040

[30] Wu Chen, Trine Agervig Carstensen, Ranran Wang, Sybil Derrible, David Rojas Rueda, Mark J. Nieuwenhuijsen, and Gang Liu. 2022. Historical Patterns and Sustainability Implications of Worldwide Bicycle Ownership and Use. *Communications Earth & Environment* 3, 1 (Aug. 2022), 171. doi:10.1038/s43247-022-00497-4

[31] Kwanghun Choi, Wontaek Lim, Byungwoo Chang, Jinah Jeong, Inyoo Kim, Chan-Ryul Park, and Dongwook W. Ko. 2022. An Automatic Approach for Tree Species Detection and Profile Estimation of Urban Street Trees Using Deep Learning and Google Street View Images. *ISPRS Journal of Photogrammetry and Remote Sensing* 190 (Aug. 2022), 165–180. doi:10.1016/j.isprsjprs.2022.06.004

[32] Jacob Cohen. 1968. Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin* 70, 4 (1968), 213–220. doi:10.1037/h0026256

[33] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, and et al. 2025. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities. arXiv:2507.06261 [cs] doi:10.48550/arXiv.2507.06261

[34] Jeroen Johan de Hartog, Hanna Boogaard, Hans Nijland, and Gerard Hoek. 2010. Do the Health Benefits of Cycling Outweigh the Risks? *Environmental Health Perspectives* 118, 8 (Aug. 2010), 1109–1116. doi:10.1289/ehp.0901747

[35] E. W. Dijkstra. 2022. A Note on Two Problems in Connexion with Graphs. In *Edsger Wybe Dijkstra: His Life,Work, and Legacy* (1 ed.). Vol. 45. Association for Computing Machinery, New York, NY, USA, 287–290.

[36] Jennifer Dill and Theresa Carr. 2003. Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. *Transportation Research Record* 1828, 1 (Jan. 2003), 116–123. doi:10.3141/1828-14

[37] Stuart E. Dreyfus and Hubert L. Dreyfus. 1980. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. ORC802 (Feb. 1980).

[38] Michael Duan, Shosuke Kiami, Logan Milandin, Johnson Kuang, Michael Saugstad, Maryam Hosseini, and Jon E. Froehlich. 2022. Scaling Crowd+AI Sidewalk Accessibility Assessments: Initial Experiments Examining Label Quality and Cross-city Training on Performance. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3517428.3550381

[39] Zicheng Fan, Chen-Chieh Feng, and Filip Biljecki. 2025. Coverage and Bias of Street View Imagery in Mapping the Urban Environment. *Computers, Environment and Urban Systems* 117 (April 2025), 102253. doi:10.1016/j.compenvurbsys.2025.102253

[40] Chiung-Yao Fang, Sei-Wang Chen, and Chiou-Shann Fuh. 2003. Road-Sign Detection and Tracking. *IEEE Transactions on Vehicular Technology* 52, 5 (Sept. 2003), 1329–1341. doi:10.1109/TVT.2003.810999

[41] Colin Ferster, Jaimy Fischer, Kevin Manaugh, Trisalyn Nelson, and Meghan Winters. 2020. Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation* 14, 1 (2020), 64–73. arXiv:https://doi.org/10.1080/15568318.2018.1519746 doi:10.1080/15568318.2018.1519746

[42] Charlie E Foster, Jenna R Panter, and Nicholas J Wareham. 2011. Assessing the Impact of Road Traffic on Cycling for Leisure and Cycling to Work. *The International Journal of Behavioral Nutrition and Physical Activity* 8 (June 2011), 61. doi:10.1186/1479-5868-8-61

[43] Simon D.S. Fraser and Karen Lock. 2011. Cycling for Transport and Public Health: A Systematic Review of the Effect of the Environment on Cycling. *European Journal of Public Health* 21, 6 (Dec. 2011), 738–743. doi:10.1093/eurpub/ckq145

[44] M. A. Ganaie, Minghui Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan. 2022. Ensemble Deep Learning: A Review. *Engineering Applications of Artificial Intelligence* 115 (Oct. 2022), 105151. doi:10.1016/j.engappai.2022.105151

[45] Fang-Ying Gong, Zhao-Cheng Zeng, Fan Zhang, Xiaojiang Li, Edward Ng, and Leslie K. Norford. 2018. Mapping Sky, Tree, and Building View Factors of Street Canyons in a High-Density Urban Environment. *Building and Environment* 134 (April 2018), 155–167. doi:10.1016/j.buildenv.2018.02.042

[46] Google. 2005. Google Maps. https://www.google.com/maps

[47] Google. 2007. Explore Street View and add your own 360 images to Google Maps. https://www.google.com/streetview/

[48] R. Groot. 2007. *Design Manual for Bicycle Traffic.* Number 25.

[49] Mordechai Haklay. 2010. How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design* 37, 4 (Aug. 2010), 682–703. doi:10.1068/b35097

[50] Michael Hardinghaus, Simon Nieland, Marius Lehne, and Jan Weschke. 2021. More than Bike Lanes—A Multifactorial Index of Urban Bikeability. *Sustainability* 13, 21 (Jan. 2021), 11584. doi:10.3390/su132111584

[51] M. Anne Harris, Conor C. O. Reynolds, Meghan Winters, Peter A. Cripton, Hui Shen, Mary L. Chipman, Michael D. Cusimano, Shelina Babul, Jeffrey R. Brubacher, Steven M. Friedman, Garth Hunte, Melody Monro, Lee Vernich, and Kay Teschke. 2013. Comparing the Effects of Infrastructure on Bicycling Injury at Intersections and Non-Intersections Using a Case–Crossover Design. *Injury Prevention* 19, 5 (Oct. 2013), 303–310. doi:10.1136/injuryprev-2012-040561

[52] Kristiann C. Heesch and Shannon Sahlqvist. 2013. Key Influences on Motivations for Utility Cycling (Cycling for Transport to and from Places). *Health Promotion Journal of Australia* 24, 3 (2013), 227–233. doi:10.1071/HE13062

[53] Eva Heinen, Bert van Wee, and Kees Maat. 2010. Commuting by Bicycle: An Overview of the Literature. *Transport Reviews* 30, 1 (Jan. 2010), 59–96. doi:10.1080/01441640903187001

[54] Marco Helbich, Chritoph Amelunxen, Pascal Neis, and Alexander Zipf. 2012. Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. *Proceedings of GI_Forum* 4 (2012), 24.

[55] Maryam Hosseini, Andres Sevtsuk, Fabio Miranda, Roberto M. Cesar, and Claudio T. Silva. 2023. Mapping the Walk: A Scalable Computer Vision Approach for Generating Sidewalk Network Datasets from Aerial Imagery. *Computers, Environment and Urban Systems* 101 (April 2023), 101950. doi:10.1016/j.compenvurbsys.2023.101950

[56] Koichi Ito, Prateek Bansal, and Filip Biljecki. 2024. Examining the Causal Impacts of the Built Environment on Cycling Activities Using Time-Series Street View Imagery. *Transportation Research Part A: Policy and Practice* 190 (Dec. 2024), 104286. doi:10.1016/j.tra.2024.104286

[57] Koichi Ito and Filip Biljecki. 2021. Assessing Bikeability with Street View Imagery and Computer Vision. *Transportation Research Part C: Emerging Technologies* 132 (Nov. 2021), 103371. doi:10.1016/j.trc.2021.103371

[58] Watthanasak Jeamwatthanachai, Mike Wald, and Gary Wills. 2019. Indoor Navigation by Blind People: Behaviors and Challenges in Unfamiliar Spaces and Buildings. *British Journal of Visual Impairment* 37, 2 (May 2019), 140–153. doi:10.1177/0264619619833723

[59] Natchapon Jongwiriyanurak, Zichao Zeng, June Moh Goo, Xinglei Wang, Ilya Ilyankou, Kerkritt Sriroongvikrai, Nicola Christie, Meihui Wang, Huanfa Chen, and James Haworth. 2025. V-RoAst: Visual Road Assessment. Can VLM Be a Road Safety Assessor Using the iRAP Standard? arXiv:2408.10872 [cs] doi:10.48550/arXiv.2408.10872

[60] Pushpak Karnick, David Cline, Stefan Jeschke, Anshuman Razdan, and Peter Wonka. 2010. Route Visualization Using Detail Lenses. *IEEE Transactions on Visualization and Computer Graphics* 16, 2 (March 2010), 235–247. doi:10.1109/TVCG.2009.65

[61] Young-Mok Kim, Young-Gil Kim, Seung-Yong Son, Soo-Yeon Lim, Bong-Yeol Choi, and Doo-Hyun Choi. 2022. Review of Recent Automated Pothole-Detection Methods. *Applied Sciences* 12, 11 (Jan. 2022), 5320. doi:10.3390/app12115320

[62] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs] doi:10.48550/arXiv.2304.02643

[63] Patricia Jasmin Krenn, Pekka Oja, and Sylvia Titze. 2015. Development of a Bikeability Index to Assess the Bicycle-Friendliness of Urban Environments. *Open Journal of Civil Engineering* 5, 4 (Nov. 2015), 451–459. doi:10.4236/ojce.2015.54045

[64] Kevin J. Krizek, Ahmed El-Geneidy, and Kristin Thompson. 2007. A Detailed Analysis of How an Urban Trail System Affects Cyclists' Travel. *Transportation* 34, 5 (Sept. 2007), 611–624. doi:10.1007/s11116-007-9130-z

[65] Minchu Kulkarni, Chu Li, Jaye Jungmin Ahn, Katrina Oi Yau Ma, Zhihan Zhang, Michael Saugstad, Kevin Wu, Yochai Eisenberg, Valerie Novack, Brent Chamberlain, and Jon E. Froehlich. 2023. BusStopCV: A Real-time AI Assistant for Labeling Bus Stop Accessibility Features in Streetscape Imagery. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3597638.3614481

[66] Louise-Ann Leyland, Ben Spencer, Nick Beale, Tim Jones, and Carien M. van Reekum. 2019. The Effect of Cycling on Cognitive Function and Well-Being in Older Adults. *PLOS ONE* 14, 2 (Feb. 2019), e0211779. doi:10.1371/journal.pone.0211779

[67] Chu Li, Rock Yuren Pang, Delphine Labbé, Yochai Eisenberg, Maryam Hosseini, and Jon E. Froehlich. 2025. Accessibility for Whom? Perceptions of Mobility Barriers Across Disability Groups and Implications for Designing Personalized Maps. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3706598.3713421

[68] Xiaojiang Li, Carlo Ratti, and Ian Seiferling. 2018. Quantifying the Shade Provision of Street Trees in Urban Landscape: A Case Study in Boston, USA, Using Google Street View. *Landscape and Urban Planning* 169 (Jan. 2018), 81–91. doi:10.1016/j.landurbplan.2017.08.011

[69] Tianjiao Liang, Hong Bao, Weiguo Pan, and Feng Pan. 2022. Traffic Sign Detection via Improved Sparse R-CNN for Autonomous Vehicles. *Journal of Advanced Transportation* 2022, 1 (2022), 3825532. doi:10.1155/2022/3825532

[70] Xiucheng Liang, Jinheng Xie, Tianhong Zhao, Rudi Stouffs, and Filip Biljecki. 2025. OpenFACADES: An Open Framework for Architectural Caption and Attribute Data Enrichment via Street View Imagery. arXiv:2504.02866 [cs] doi:10.48550/arXiv.2504.02866

[71] Graeme Lindsay, Alexandra Macmillan, and Alistair Woodward. 2011. Moving Urban Trips from Cars to Bicycles: Impact on Health and Emissions. *Australian and New Zealand Journal of Public Health* 35, 1 (Feb. 2011), 54–60. doi:10.1111/j.1753-6405.2010.00621.x

[72] Ying Long and Jianting Zhao. 2020. What Makes a City Bikeable? A Study of Intercity and Intracity Patterns of Bicycle Ridership Using Mobike Big Data Records. *Built Environment* 46, 1 (Feb. 2020), 55–75. doi:10.2148/benv.46.1.55

[73] Liang Ma, Runing Ye, and Hongyu Wang. 2021. Exploring the Causal Effects of Bicycling for Transportation on Mental Health. *Transportation Research Part D: Transport and Environment* 93 (April 2021), 102773. doi:10.1016/j.trd.2021.102773

[74] Nathan McNeil. 2011. Bikeability and the 20-Min Neighborhood: How Infrastructure and Destinations Influence Bicycle Accessibility. *Transportation Research Record* 2247, 1 (Jan. 2011), 53–63. doi:10.3141/2247-07

[75] Michael McQueen, John MacArthur, and Christopher Cherry. 2020. The E-Bike Potential: Estimating Regional e-Bike Impacts on Greenhouse Gas Emissions. *Transportation Research Part D: Transport and Environment* 87 (Oct. 2020), 102482. doi:10.1016/j.trd.2020.102482

[76] Siyuan Meng and Hao Zheng. 2023. A Personalized Bikeability-Based Cycling Route Recommendation Method with Machine Learning. *International Journal of Applied Earth Observation and Geoinformation* 121 (July 2023), 103373. doi:10.1016/j.jag.2023.103373

[77] Lieze Mertens, Delfien Van Dyck, Ariane Ghekiere, Ilse De Bourdeaudhuij, Benedicte Deforche, Nico Van de Weghe, and Jelle Van Cauwenberg. 2016. Which Environmental Factors Most Strongly Influence a Street's Appeal for Bicycle Transport among Adults? A Conjoint Study Using Manipulated Photographs. *International Journal of Health Geographics* 15, 1 (Sept. 2016), 31. doi:10.1186/s12942-016-0058-4

[78] Janice M. Morse. 2015. Critical Analysis of Strategies for Determining Rigor in Qualitative Inquiry. *Qualitative Health Research* 25, 9 (Sept. 2015), 1212–1222. doi:10.1177/1049732315588501

[79] Pascal Neis, Dennis Zielstra, and Alexander Zipf. 2012. The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet* 4, 1 (March 2012), 1–21. doi:10.3390/fi4010001

[80] Trisalyn Nelson, Colin Ferster, Karen Laberee, Daniel Fuller, and Meghan Winters. 2021. Crowdsourced data for bicycling research and practice. *Transport Reviews* 41, 1 (2021), 97–114. arXiv:https://doi.org/10.1080/01441647.2020.1806943 doi:10.1080/01441647.2020.1806943

[81] Don Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

[82] P. Oja, S. Titze, A. Bauman, B. de Geus, P. Krenn, B. Reger-Nash, and T. Kohlberger. 2011. Health Benefits of Cycling: A Systematic Review. *Scandinavian Journal of Medicine & Science in Sports* 21, 4 (2011), 496–509. doi:10.1111/j.1600-0838.2011.01299.x

[83] Kazi Shahrukh Omar, Gustavo Moreira, Daniel Hodczak, Maryam Hosseini, and Fabio Miranda. 2022. Crowdsourcing and Sidewalk Data: A Preliminary Study on the Trustworthiness of OpenStreetMap Data in the US. arXiv:2210.02350 [cs] doi:10.48550/arXiv.2210.02350

[84] OpenStreetMap contributors. 2017. OpenStreetMap. https://www.openstreetmap.org

[85] Reid Priedhorsky, Mikhil Masli, and Loren Terveen. 2010. Eliciting and Focusing Geographic Volunteer Work. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. ACM, Savannah Georgia USA, 61–70. doi:10.1145/1718918.1718931

[86] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art. *User Modeling and User-Adapted Interaction* 22, 4 (Oct. 2012), 317–355. doi:10.1007/s11257-011-9115-7

[87] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (Santiago, Chile) *(HT '14)*. Association for Computing Machinery, New York, NY, USA, 116–125. doi:10.1145/2631775.2631799

[88] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The Shortest Path to Happiness: Recommending Beautiful, Quiet, and Happy Routes in the City. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT '14)*. Association for Computing Machinery, New York, NY, USA, 116–125. doi:10.1145/2631775.2631799

[89] Giulia Reggiani, Tim van Oijen, Homayoun Hamedmoghadam, Winnie Daamen, Hai L. Vu, and Serge Hoogendoorn. 2022. Understanding Bikeability: A Methodology to Assess Urban Networks. *Transportation* 49, 3 (June 2022), 897–925. doi:10.1007/s11116-021-10198-0

[90] Tomás Rossetti, Verónica Saud, and Ricardo Hurtubia. 2019. I Want to Ride It Where I like: Measuring Design Preferences in Cycling Infrastructure. *Transportation* 46, 3 (June 2019), 697–718. doi:10.1007/s11116-017-9830-y

[91] Jonas Schmid-Querg, Andreas Keler, and Georgios Grigoropoulos. 2021. The Munich Bikeability Index: A Practical Approach for Measuring Urban Bikeability. *Sustainability* 13, 1 (Jan. 2021), 428. doi:10.3390/su13010428

[92] Ben Shneiderman. 2003. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*, BENJAMIN B. BEDERSON and BEN SHNEIDERMAN (Eds.). Morgan Kaufmann, San Francisco, 364–371. doi:10.1016/B978-155860915-0/50046-9

[93] Rashmi Sinha and Kirsten Swearingen. 2002. The Role of Transparency in Recommender Systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. Association for Computing Machinery, New York, NY, USA, 830–831. doi:10.1145/506443.506619

[94] John M Sullivan and Michael J Flannagan. 2002. The Role of Ambient Light Level in Fatal Crashes: Inferences from Daylight Saving Time Transitions. *Accident Analysis & Prevention* 34, 4 (July 2002), 487–498. doi:10.1016/S0001-4575(01)00046-X

[95] 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:2506.02242 [cs] doi:10.48550/arXiv.2506.02242

[96] Kay Teschke, M. Anne Harris, Conor C. O. Reynolds, Meghan Winters, Shelina Babul, Mary Chipman, Michael D. Cusimano, Jeff R. Brubacher, Garth Hunte, Steven M. Friedman, Melody Monro, Hui Shen, Lee Vernich, and Peter A. Cripton. 2012. Route Infrastructure and the Risk of Injuries to Bicyclists: A Case-Crossover Study. *American Journal of Public Health* 102, 12 (Dec. 2012), 2336–2343. doi:10.2105/AJPH.2012.300762

[97] Tom Thomas, Rinus Jaarsma, and Bas Tutert. 2013. Exploring Temporal Fluctuations of Daily Cycling Demand on Dutch Cycle Paths: The Influence of Weather on Cycling. *Transportation* 40, 1 (Jan. 2013), 1–22. doi:10.1007/s11116-012-9398-5

[98] Francesco Maria Turno, Aleksejs Vesjolijs, Jelena Kijonoka, and Irina Yatskiv Jackiva. 2024. Context-Aware Personalized Route Recommendations for Bicycle Users Using Large-Scale Mobility Data. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*. 449–454. doi:10.1109/COMPSAC61105.2024.00068

[99] Jim Uttley, Steve Fotios, and Robin Lovelace. 2020. Road Lighting Density and Brightness Linked with Increased Cycling Rates After-Dark. *PLOS ONE* 15, 5 (May 2020), e0233105. doi:10.1371/journal.pone.0233105

[100] Veerle Van Holle, Jelle Van Cauwenberg, Benedicte Deforche, Liesbet Goubert, Lea Maes, Jack Nasar, Nico Van de Weghe, Jo Salmon, and Ilse De Bourdeaudhuij. 2014. Environmental Invitingness for Transport-Related Cycling in Middle-Aged Adults: A Proof of Concept Study Using Photographs. *Transportation Research Part A: Policy and Practice* 69 (Nov. 2014), 432–446. doi:10.1016/j.tra.2014.09.009

[101] Ane Rahbek Vierø, Anastassia Vybornova, and Michael Szell. 2024. BikeDNA: A tool for bicycle infrastructure data and network assessment. *Environment and Planning B: Urban Analytics and City Science* 51, 2 (2024), 512–528. arXiv:https://doi.org/10.1177/23998083231184471 doi:10.1177/23998083231184471

[102] Jiaqi Wang, Zhengliang Liu, Lin Zhao, Zihao Wu, Chong Ma, Sigang Yu, Haixing Dai, Qiushi Yang, Yiheng Liu, Songyao Zhang, Enze Shi, Yi Pan, Tuo Zhang, Dajiang Zhu, Xiang Li, Xi Jiang, Bao Ge, Yixuan Yuan, Dinggang Shen, Tianming Liu, and Shu Zhang. 2023. Review of Large Vision Models and Visual Prompt Engineering. *Meta-Radiology* 1, 3 (Nov. 2023), 100047. doi:10.1016/j.metrad.2023.100047

[103] Zeyu Wang, Koichi Ito, and Filip Biljecki. 2024. Assessing the Equity and Evolution of Urban Visual Perceptual Quality with Time Series Street View Imagery. *Cities* 145 (Feb. 2024), 104704. doi:10.1016/j.cities.2023.104704

[104] Zeyu Wang, Yingchao Jian, Adam Visokay, Don MacKenzie, and Jon E. Froehlich. 2025. Street View for Whom? An Initial Examination of Google Street View's Urban Coverage and Socioeconomic Indicators in the US.. In *The 4th ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications*. Minneapolis, MN, USA. doi:10.1145/3764919.3770883

[105] Zeyu Wang, Yingjie Liu, and Don MacKenzie. 2026. Visual clustering of urban environments: Associations with household and travel characteristics. *Journal of Transport Geography* 131 (2026), 104546.

[106] Musab Wedyan, Yu-Chen Yeh, Fatemeh Saeidi-Rizi, Tai-Quan Peng, and Chun-Yen Chang. 2025. Urban Walkability through Different Lenses: A Comparative Study of GPT-4o and Human Perceptions. *PLOS ONE* 20, 4 (April 2025), e0322078. doi:10.1371/journal.pone.0322078

[107] Simone Weikl and Patricia Mayer. 2023. Data-Driven Quality Assessment of Cycling Networks. *Frontiers in Future Transportation* 4 (March 2023). doi:10.3389/ffutr.2023.1127742

[108] Galen Weld, Esther Jang, Anthony Li, Aileen Zeng, Kurtis Heimerl, and Jon E. Froehlich. 2019. Deep Learning for Automatically Detecting Sidewalk Accessibility Problems Using Streetscape Imagery. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 196–209. doi:10.1145/3308561.3353798

[109] Meghan Winters and Kay Teschke. 2010. Route Preferences among Adults in the near Market for Bicycling: Findings of the Cycling in Cities Study. *American journal of health promotion: AJHP* 25, 1 (2010), 40–47. doi:10.4278/ajhp.081006-QUAN-236

[110] Meghan Winters, Kay Teschke, Michael Brauer, and Daniel Fuller. 2016. Bike Score®: Associations between Urban Bikeability and Cycling Behavior in 24 Cities. *International Journal of Behavioral Nutrition and Physical Activity* 13, 1 (Feb. 2016), 18. doi:10.1186/s12966-016-0339-0

[111] Zhenhua Xu, Yan Bai, Yujia Zhang, Zhuoling Li, Fei Xia, Kwan-Yee K. Wong, Jianqiang Wang, and Hengshuang Zhao. 2025. DriveGPT4-V2: Harnessing Large Language Model Capabilities for Enhanced Closed-Loop Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17261–17270.

[112] Bin Yang, Chenjuan Guo, Yu Ma, and Christian S. Jensen. 2015. Toward Personalized, Context-Aware Routing. *The VLDB Journal* 24, 2 (April 2015), 297–318. doi:10.1007/s00778-015-0378-1

[113] Yongping Zhang and Zhifu Mi. 2018. Environmental Benefits of Bike Sharing: A Big Data-Based Analysis. *Applied Energy* 220 (June 2018), 296–301. doi:10.1016/j.apenergy.2018.03.101

[114] Shudan Zheng and Jianghua Zheng. 2014. Assessing the Completeness and Positional Accuracy of OpenStreetMap in China. In *Thematic Cartography for the Society*, Temenoujka Bandrova, Milan Konecny, and Sisi Zlatanova (Eds.). Springer International Publishing, Cham, 171–189. doi:10.1007/978-3-319-08180-9_14

A Appendix

A.1 Calculating Edge Weights

Given the preference vector and a directed multigraph road network with scored edge vectors, we can calculate a dynamic weight for each edge representing a optimal similarity match, informed by [76]. The weight consists of two components: a preference adherence term (α) and an edge length term ($\beta = length$). We calculate α using a bonus/penalty framework that starts with the edge length as the base weight. For each feature, we define a "bonus window", with its size based on the user's preference strength, and apply a bonus to the weight if the edge's score is within the window, and a penalty if it is outside. This effectively *decreases* the total weight if a feature is strongly aligned with the preference, and *increases* the total weight if it is misaligned.

Since edge vector values range from -1 to 1, the bonus window positioning depends on preference direction: negative preferences create windows starting at -1, while positive preferences create windows ending at +1. The window size is determined by preference magnitude—stronger preferences yield smaller windows, while neutral preferences yield the largest windows. For example, a preference of -5 will have a window size of 0.4, making the bonus range [-1, -0.6]. A neutral preference of 0 will have a window size of 2, making the bonus range [-1, 1] (the entire possible range). Features with values inside the bonus window contribute bonuses to the weight; those outside contribute penalties. The magnitude of each bonus or penalty scales with how close the edge's feature value is to the optimal value. For instance, with a negative preference, an edge scoring -1 (optimal) receives 100% of its maximum possible bonus, while an edge scoring -0.1 receives only 10%. The maximum bonus per feature is determined by the number of nonzero features in the preference vector, such that if all features achieve their optimal scores, the total bonus reaches 100% of the edge length (*aka*, maximum bonus percent per feature = $\frac{100\%}{\text{\# nonzero preferences}}$).

We calculate α by observing the total bonuses and penalties. If there are penalties, that means that at least one feature failed to match the stated preferences, and we therefore disqualify the edge by setting α to the edge length times the penalties, multiplied by 10 as an additional avoidance factor:

$$\alpha = 10 \cdot length \cdot \sum penalties$$

Otherwise, the edge is valid for consideration, and α is equal to the length times the bonuses:

$$\alpha = (1 - \sum bonuses) \cdot length$$

Finally, the total weight of an edge is the sum of α and β, scaled according to the tolerance for detours (d):

$$weight = (1 - d)\alpha + d\beta$$

A.2 Gemini Prompt

```
You are an expert urban planner and cycling infrastructure
    analyst. Your task is to assess the bikeability of a
    specific street segment based on visual and data inputs.
    You must analyze all provided information to generate a
    structured assessment.
*Your Goal:* Evaluate the safety, comfort, and overall
    experience for a person cycling on this street.
*Inputs You Will Receive:*
1.  *[Image 1: Front View]* A street-level image looking
    forward along the direction of travel.
2.  *[Image 2: Back View]* A street-level image looking
    backward from the same point.
3.  *[Image 3: Top-Down View]* A projected, top-down view of
    the street segment.
4.  *[Data: Speed Limit]* The posted or legal speed limit
    for vehicles (e.g., "30 mph" or "50 km/h"). May be
    missing/null.
5.  *[Data: Street Archetype]* The general classification of
    the street from OpenStreetMap (e.g., "residential,"
    "tertiary," "commercial"). May be missing/null.
*Your Task:*
Based on all the provided inputs, carefully evaluate the
    following factors and provide your output in the
    specified JSON format.
---
### *Analysis Categories & Rating Scales*
*1. Vegetation:*
  - *Scale:* 1 to 5
  - *Definition:* Rate the amount of green space, tree
      canopy, and pleasant foliage.
    - *1:* Barren. No trees or greenery.
    - *3:* Some. Occasional trees or grass verges.
    - *5:* Lush. Significant, continuous tree canopy or
        well-maintained green spaces.
*2. Surface Quality:*
  - *Scale:* 1 to 3
  - *Definition:* Rate the smoothness and condition of the
      road surface where a cyclist would ride.
    - *1:* Very Poor. Numerous potholes, large cracks, or
        extremely uneven surfaces that would be dangerous or
        very uncomfortable.
    - *2:* Fair. Some cracks, patches, or imperfections, but
        generally navigable.
    - *3:* Excellent. Smooth, clean, and well-maintained
        asphalt or concrete.
*3. Bike Lane:*
  - *Existence (select one):* Yes or No. Look for any
      dedicated space for cyclists marked on the pavement
      (e.g., painted lines, symbols).
  - *Estimated Width (if "Yes"):*
    - Narrow: Appears less than 1.5m / 5ft wide.
    - Adequate: Appears 1.5m - 2.1m / 5-7ft wide.
    - Wide: Appears wider than 2.1m / 7ft.
    - If No bike lane exists, this must be N/A.
*4. Sidewalks:*
  - *Existence (select one):* None, One Side Only, Both Sides
  - *Condition (overall):* Poor, Fair, Good
*5. Overall Bikeability Score:*
  - *Scale:* 1 to 10
  - *Definition:* A holistic score combining all factors.
      Consider vehicle speed, traffic volume (inferred from
      street type and width), presence and quality of bike
      lanes, and overall comfort.
    - *1-3:* Hostile. Unsafe for most people. High
        speed/volume traffic with no dedicated space for
        bikes.
    - *4-6:* Usable with Caution. Tolerable for confident
        cyclists but stressful. May have a bike lane on a
        busy road or be a slow, shared street.
    - *7-8:* Comfortable. A good, safe experience for many
        cyclists. Features good bike lanes or is on a very
        low-speed, low-traffic street.
    - *9-10:* Excellent. "All Ages & Abilities" standard.
        Physically protected, low-stress environment, ideal
        for all users.
*6. Bikeability Summary:*
  - A concise, 1-2 sentence summary explaining the overall
      score. Mention the most significant positive and
      negative factors.
```

A.3 OSM Filter

Here we include the osmnx filter used to fetch the Seattle bikeable road network.

```
custom_filter = (
    '["highway"]' # Must be a road
    '["highway"!~"abandoned|bus_guideway|construction
        |corridor|elevator|escalator|motor|no|planned
        |platform|proposed|raceway|razed|steps]'
        # Must not be one of these types
    '["area"!~"yes"]'# Must not be an area.
    '["access"!~"private"]'
        # Must not be private access.
    '["bicycle"!~"no"]'
        # The "bicycle" tag must not be "no".
    '["service"!~"private"]'
        # Must not be a private service road.
)
G = ox.graph_from_bbox(
    (self.topleft_lon, self.botright_lat,
    self.botright_lon, self.topleft_lat),
    custom_filter=custom_filter, retain_all=False)
```

A.4 Highway Types

Here we list the mappings from the OSM tag highway to a range [-2, 2]. The range is then normalized to [-1, 1] in the edge vector.

```
highway_types = {
    'motorway': 2,
    'trunk': 2,
    'primary': 2,
    'secondary': 1,
```

```
    'tertiary': 0.5,
    'unclassified': 0,
    'residential': -1,
    'motorway_link': 2,
    'trunk_link': 2,
    'primary_link': 2,
    'secondary_link': 1,
    'tertiary_link': 0.5,
    'living_street': -2,
    'service': -2,
    'pedestrian': -2,
    'track': -2,
    'road': 1,
    'footway': -2,
    'bridleway': -2,
    'path': -2,
    'sidewalk': -1,
    'crossing': -2,
    'traffic_island': 0,
    'cycleway': -2,
}
```