

StreetDesignAI: A Multi-Persona Evaluation System for Inclusive Infrastructure Design

ZIYI WANG*, University of Maryland, College Park, USA

YILONG DAI*, University of Alabama, USA

DUANYA LYU, University of Florida, USA

MATEO NADER, University of Florida, USA

SIHAN CHEN, Carnegie Mellon University, USA

WANGHAO YE, University of Maryland, College Park, USA

ZIJIAN DING, University of Maryland, College Park, USA

XIANG YAN, University of Florida, USA

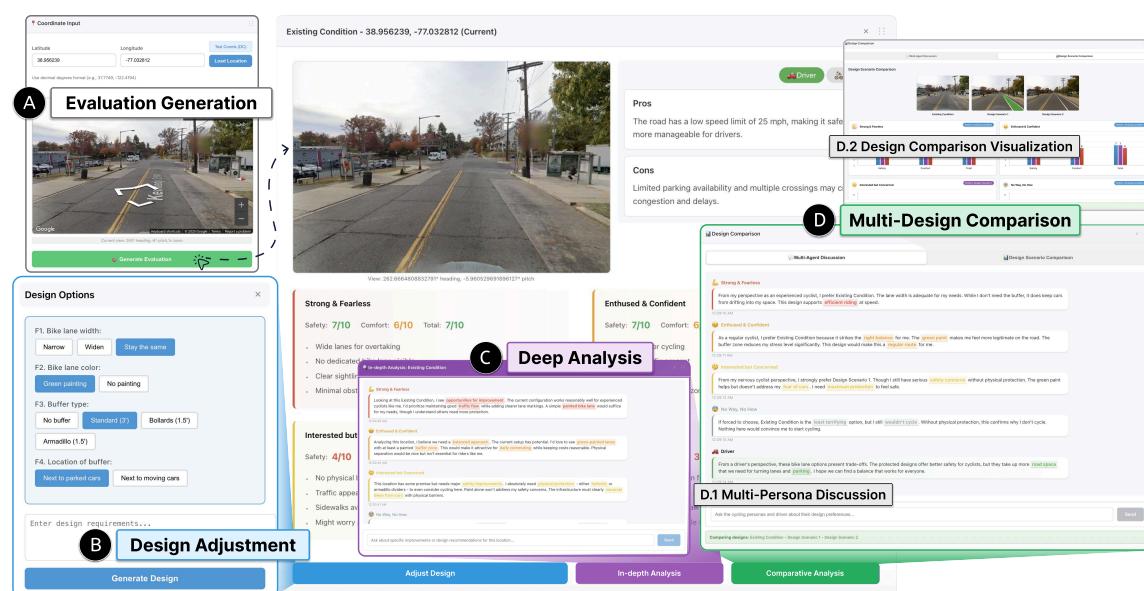


Fig. 1. Overview of StreetDesignAI. (A) Users input coordinates to load Street View imagery, which is analyzed using OpenStreetMap data and image recognition. (B) The system generates bikeability evaluations from multiple cyclist personas and allows users to adjust design parameters to create AI-rendered street redesigns. (C) In-depth Analysis provides detailed persona feedback with follow-up questioning capabilities. (D) Comparative Analysis enables multi-design comparison through persona discussions (D.1) and side-by-side score visualizations across safety, comfort, and total metrics (D.2).

Authors' addresses: Ziyi Wang*, University of Maryland, College Park, USA; Yilong Dai*, University of Alabama, USA; Duanya Lyu, University of Florida, USA; Mateo Nader, University of Florida, USA; Sihan Chen, Carnegie Mellon University, USA; Wanghao Ye, University of Maryland, College Park, USA; Zijian Ding, University of Maryland, College Park, USA; Xiang Yan, University of Florida, USA.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Designing inclusive cycling infrastructure requires balancing the competing needs of diverse user groups, yet designers often struggle to anticipate how different cyclists experience the same street environment. We investigate how persona-based multi-agent evaluation can support inclusive infrastructure design by making experiential conflicts explicit during the design process. We present StreetDesignAI, an interactive system that enables designers to (1) ground evaluation in real street context through imagery and map data, (2) receive parallel feedback from simulated cyclist personas spanning confident to cautious users, and (3) iteratively modify designs while the system surfaces conflicts across perspectives. A within-subjects study with 26 transportation professionals comparing StreetDesignAI against a general-purpose AI chatbot demonstrates that structured multi-perspective feedback significantly improves designers' understanding of new user perspectives, ability to identify diverse persona needs, and confidence in translating those needs into inclusive design decisions. Participants also reported significantly higher overall satisfaction and stronger intention to use the system in professional practice. Qualitative findings further illuminate how explicit conflict surfacing transforms design exploration from single-perspective optimization toward deliberate trade-off reasoning. We discuss implications for AI-assisted tools that scaffold inclusive design through disagreement as an interaction primitive.

Additional Key Words and Phrases: Cycling Infrastructure, Inclusive Design, Persona-based Evaluation, Multi-agent System, Generative AI

ACM Reference Format:

Ziyi Wang*, Yilong Dai*, Duanya Lyu, Mateo Nader, Sihan Chen, Wanghao Ye, Zijian Ding, and Xiang Yan. 2026. StreetDesignAI: A Multi-Persona Evaluation System for Inclusive Infrastructure Design. 1, 1 (January 2026), 33 pages. <https://doi.org/10.xxxx/xxxxxx.xxxxxxx>

1 INTRODUCTION

Inclusive cycling infrastructure is widely recognized as a cornerstone of sustainable urban transportation [45, 46]. When cities build safer and more inclusive facilities, cycling adoption increases and public health benefits follow: protected bike lanes attract substantially more commuters than standard lanes or streets without lanes [15], and well-designed investments need not worsen traffic congestion. Delineated bicycle lanes may also reduce crash risk and severity for pedestrians and other road users [57]. Yet designing inclusive cycling infrastructure remains difficult in practice because designers must negotiate competing demands, limited right-of-way, and heterogeneous public expectations.

A central challenge is that the same street design can be experienced in fundamentally different—and often conflicting—ways by different cyclist populations. Over 60% of urban residents report willingness to cycle but are deterred by perceived danger in existing conditions [13, 17], while experienced cyclists may tolerate or even prefer those same facilities for speed and flexibility. Geller's typology highlights four cyclist groups—"Strong and Fearless," "Enthused and Confident," "Interested but Concerned," and "No Way, No How"—whose expectations for safety, comfort, and accessibility diverge sharply [20]. The "Interested but Concerned" group, representing the largest pool of latent riders, often requires physical protection and clear separation from motor traffic, whereas "Strong and Fearless" riders may prioritize directness and maneuverability over protection [14]. These divergent needs create experiential conflicts (e.g., protection vs. flexibility) that are difficult to navigate without systematic, context-grounded support.

Existing infrastructure design support methods—manuals, case databases, expert consultation, and public engagement—provide valuable guidance but tend to fall short exactly when conflicts matter most. Design manuals and standards [1, 13, 38] encode best practices, yet are general-purpose and do not make visible how specific interventions affect diverse users in a particular street context. Public engagement can surface lived experiences, but the process is slow and participation often uneven; by the time sufficient community feedback reaches designers, major constraints

© 2026 Copyright held by the owner/author(s).

Manuscript submitted to ACM

are typically fixed, leaving little room for meaningful adjustment [28]. More recently, AI-based tools have improved visualization and automated technical tasks in urban design, including generative layout exploration [23, 47, 53], compliance checking [59], and perception-informed optimization linking visual features to safety perception [50]. However, these tools typically assume a single, homogeneous user perspective—they help designers generate and refine designs, but provide little support for understanding how different user groups experience the same infrastructure or for navigating trade-offs when their needs conflict.

To address these challenges, we introduce *StreetDesignAI*, an interactive system for inclusive cycling infrastructure design that operationalizes persona-based multi-perspective evaluation as an interaction mechanism for trade-off reasoning. Rather than asking a general-purpose conversational model for advice, designers can use StreetDesignAI to (1) ground evaluation in a specific street context, (2) receive parallel feedback from cyclist personas spanning confident to cautious users, and (3) iteratively modify the design while the system surfaces conflicts and prompts targeted follow-up inquiry. Specifically, designers begin by selecting a street location, after which the system retrieves contextual information from OpenStreetMap and street-level imagery; the system then generates structured evaluations from four cyclist personas, including safety, comfort, and overall bikeability ratings and concise concern summaries. Based by the evaluation results, designers can propose modifications to the street layout via structured parameters (and optional text), receive street-level visualizations of these proposed changes, and obtain updated persona-based evaluations that reveal how each change impacts safety and comfort perceptions across user groups.

We evaluate StreetDesignAI through a within-subjects study of 26 street design professionals, including students and academic researchers. Participants completed matched street redesign tasks under two conditions—StreetDesignAI and a chatbot baseline—using a counterbalanced within-subjects design. We collected post-task questionnaires and interview data to assess perceived exploration, understanding, confidence, and how conflict surfacing shaped exploration and trade-off reasoning. In particular, we seek to address the following research questions:

- **RQ1 - Exploration:** How does persona-based multi-agent evaluation support designers' exploration of design alternatives and reflection on persona-specific trade-offs?
- **RQ2 - Understanding:** How does explicit conflict surfacing across simulated user perspectives influence designers' understanding of diverse user needs compared to general-purpose AI assistance?
- **RQ3 - Capability & Confidence:** How does structured multi-perspective feedback influence designers' perceived capability and confidence in translating diverse user needs into inclusive design decisions?

The scientific contributions of this work are threefold:

- We present StreetDesignAI, a grounded, human-informed AI design pipeline that integrates heterogeneous urban context data (e.g., street-level imagery and spatial map data) with real user evaluations to generate persona-aware, situated feedback for inclusive street infrastructure design.
- We demonstrate how structuring experiential conflicts through persona-based evaluation, comparative analysis, and visualization-driven iteration supports designers' exploration of design alternatives and deepens their understanding of diverse and previously underrepresented personas.
- Through a mixed-methods, within-subjects study with 26 professional street designers, we provide empirical evidence that StreetDesignAI improves designers' perceived capability and confidence in identifying diverse persona needs and translating them into inclusive infrastructure design decisions.

2 RELATED WORK

2.1 Inclusive Cycling Infrastructure Design: User Diversity and Experiential Conflict

Cycling infrastructure design traditionally draws on standardized guidelines, expert judgment, and formal public engagement. Authoritative manuals such as those published by NACTO [37] and AASHTO [39] provide benchmarks for lane dimensions, separation treatments, and intersection design. Transportation research further emphasizes cyclist heterogeneity through frameworks such as Geller's four cyclist types [20] and the Level of Traffic Stress (LTS) model [35], showing how traffic speed, separation, and intersection complexity disproportionately deter less confident cyclists [19, 36]. Despite this knowledge, applying user diversity frameworks during iteration remains challenging: typologies often stay abstract rather than becoming actionable inputs that guide concrete decisions under constraints [56].

Community engagement methods (e.g., surveys, workshops, hearings) provide another route to incorporate lived experience [5], and participatory design frameworks emphasize early-stage exploration of needs through contextual inquiry and iterative prototyping with real users [4, 7]. Immersive methods such as 360° video ethnography further enable designers to experience user contexts firsthand, supporting collaborative annotation and iterative sense-making [10, 34]. However, participation is frequently uneven and time-intensive: by the time sufficient community feedback is finally collected and delivered to designers, key parameters have often already been fixed [28]. Critically, these approaches do not consistently support the moment-to-moment design reasoning that inclusive infrastructure requires: surfacing when user experiences conflict, tracing the sources of conflict in a specific context, and negotiating trade-offs before designs harden.

2.2 AI-Assisted Design: Visualization, Optimization, and the Limits of Single-Perspective Support

AI-assisted design tools have been adopted to support spatial configuration, optimization, and visualization in infrastructure and urban design [23, 47, 53]. In HCI, human–AI collaboration research frames AI as a partner that complements human judgment [9, 27], and structured workflows integrating AI into distinct design stages can outperform open-ended conversational assistance [51, 55]. Large language models further enable natural-language interaction and ideation [31]. Yet when applied to inclusive infrastructure design, general-purpose AI tools often provide single-perspective recommendations and lack mechanisms to reason about conflicting lived experiences and value-laden trade-offs. Designers can ask for advice, but the system does not inherently help them *compare* stakeholders, *interrogate* disagreement, or *decide* what to prioritize.

Visual generation models introduce new opportunities for perception-informed design research. Recent work has explored translating textual descriptions into street-level imagery or editing street scenes through text-based instructions to support perception studies [12, 50, 52]. However, perception research requires strict control over variables and visual plausibility of generated content. Applying state-of-the-art image generation models to domain-specific tasks typically demands carefully crafted prompts, parameter tuning, and iterative refinement [49], thus practitioners outside the AI field may find it difficult to leverage these models directly for their specific needs.

2.3 LLM-Based User Simulation: From Multiple Perspectives to Disagreement as an Interaction Primitive

Recent work explores using LLMs to simulate users for evaluation and feedback, including synthetic usability critique, simulated survey responses, and role-play evaluations [2, 11, 33]. Research on Role-Playing Language Agents (RPLAs)

provides taxonomies for how LLMs embody diverse perspectives [8], with studies examining how narrative perspective [30] and persona modality [26] affect engagement and believability. Multi-agent systems extend this by representing multiple stakeholders [32, 60], and SimTube demonstrates that persona-driven simulation can yield feedback more informative than actual user comments [24]. However, this literature largely focuses on generating *multiple viewpoints* rather than making disagreement an *operational object*. Tools like Synthia show how visual scaffolding helps users act on feedback [58], yet primarily target textual revision. Inclusive design often fails not from lack of diverse opinions, but from lacking mechanisms to surface and negotiate conflicts systematically.

LLM-based simulation also raises credibility concerns: models may produce generic or idealized feedback and misrepresent marginalized users [2, 25]. Fine-tuning, retrieval augmentation, and structured prompting can improve realism [43, 44, 54]. Grounding in real-world data, such as crowdsourced assessments from Project Sidewalk [48] with AI-based quality control [29], offers one path toward credibility. Multimodal systems like StreetViewAI demonstrate street-level imagery understanding for virtual navigation [18], and recent work explores persona-based urban safety perception [3]. Yet most prior work remains in digital domains, whereas physical infrastructure is embodied and carries long-term consequences. Integrating grounded personas into workflows for early-stage trade-off negotiation remains underexplored.

StreetDesignAI addresses this gap by supporting explicit negotiation of experiential conflict during iterative infrastructure design. By grounding persona behavior in large-scale cyclist data and coupling evaluation with visualization, StreetDesignAI provides mechanisms for comparing, probing, and resolving disagreement under real-world constraints.

3 SYSTEM DESIGN

StreetDesignAI is an interactive evaluation system for inclusive cycling infrastructure design. Our system makes conflicts between stakeholder experiences explicit, supports targeted interrogation of why conflicts arise, and keeps conflicts visible across iteration so designers can negotiate trade-offs.

To achieve this, StreetDesignAI couples three capabilities into one iterative workflow: (1) **grounded context** (street-view imagery and OpenStreetMap attributes), (2) **multi-perspective evaluation** through persona-based agents representing heterogeneous user groups, and (3) **visualization-driven iteration** through parameterized design edits that are rendered at the street level and immediately re-evaluated. This framing distinguishes StreetDesignAI from general-purpose conversational tools (e.g., ChatGPT), which can provide suggestions but do not inherently structure parallel comparison, conflict surfacing, and iterative trade-off negotiation within a single workflow.

We designed StreetDesignAI through a formative study with practitioners and translated those findings into three design goals that guide system features and architecture.

3.1 Formative Study: Understanding Challenges in Inclusive Cycling Infrastructure Design

We conducted semi-structured interviews with 12 professionals involved in cycling infrastructure design, including 6 traffic/roadway engineers, 3 transportation major students, 2 transportation planners/project managers, and 1 environmental leaders fellow. All 12 formative study participants later participated in our main user study. Participants reported 1 to 10+ years of experience and involvement in 1 to 5+ cycling-related projects. Our analysis surfaced four recurring challenges that informed system requirements.

Challenge 1: Difficulty in perspective-taking beyond personal experience. Most participants identified as regular or experienced cyclists and reported difficulty empathizing with less confident populations. As P3 noted: “I bike to work every day and feel comfortable in most conditions... but when I try to imagine how my mother would feel on

the same street, I honestly don't know what would concern her." Participants also described limited familiarity with cyclist typologies (e.g., P7: "I'm not very familiar with the different types of cyclists... I don't really know about the different levels of cyclists").

Challenge 2: Delayed and limited user feedback in current workflows. User feedback typically arrives late (e.g., public hearings or post-implementation feedback), when major constraints are already fixed. P2 emphasized: "By the time we get community feedback, the budget is allocated and the design is mostly fixed... fundamental changes are nearly impossible." Feedback is also often too coarse to support design trade-offs (P9: "People tell us they want 'safer bike lanes,' but that doesn't help us decide between a painted buffer versus physical barriers...").

Challenge 3: Disconnect between technical parameters and lived experience. Participants described difficulty translating technical specifications into felt safety/comfort. P11 stated: "I can tell you that a 5-foot bike lane with a 2-foot buffer meets NACTO standards, but I cannot tell you whether a nervous cyclist would feel safe enough to use it." This disconnect becomes acute under right-of-way constraints (P4: "...choose between widening lanes and improving lane separation... I don't have a good way to understand which matters more for which type of cyclist").

Challenge 4: Limited access to diverse user perspectives. Designers reported that engagement processes tend to over-represent confident cyclists, while hesitant users remain unheard. P6 observed: "Our public meetings tend to attract the same people... We rarely hear from people who don't currently bike but might if conditions improved." Participants noted that even when they seek broader input, it is slow and hard to make feedback specific to hypothetical designs.

Current use of AI tools. Eight participants had experimented with ChatGPT or similar tools but reported generic feedback and lack of situated constraint-awareness (P5: "...it gives very generic advice." P8: "...suggests ideal solutions without acknowledging the constraints we actually face."). Multiple participants expressed interest in tools that could quickly test how different user types would react to design choices before committing (P12: "What I really need is a way to quickly test how different types of cyclists would react to my design choices, before I've committed to anything").

3.2 Design Goals

Based on the formative study and prior literature on inclusive design and human–AI collaboration, we derive three design goals:

DG1: Facilitate perspective-taking through multi-persona evaluation. The system should enable designers to compare heterogeneous user perspectives in parallel and keep differences visible, reducing experience substitution and supporting explicit prioritization decisions.

DG2: Support iterative design ideation and concept exploration through real-time visualization and evaluation. The system should compress the feedback loop by enabling designers to propose an intervention, visualize it in context, and immediately observe how it shifts conflict patterns across stakeholders.

DG3: Bridge technical parameters and lived experience through grounded, narrative feedback. The system should translate parameter choices into grounded experiential narratives and visual evidence so designers can reason about why specific elements feel safer or riskier for different users.

3.3 System Overview

StreetDesignAI alternates between (1) **grounded baseline evaluation** and (2) **design-and-re-evaluate cycles**. Designers select a real street segment, after which the system retrieves contextual attributes (OpenStreetMap) and

Manuscript submitted to ACM

street-level imagery (Street View). The system then generates persona-based evaluations representing five stakeholder perspectives:

- **Strong & Fearless:** experienced daily cyclists prioritizing speed and efficiency
- **Enthused & Confident:** regular cyclists balancing safety and efficiency
- **Interested but Concerned:** cautious cyclists requiring substantial protection
- **No Way, No How:** non-cyclists deterred by safety concerns, requiring high separation
- **Driver:** operational driving concerns (visibility, turning, lane width, traffic flow)

For each iteration, designers specify interventions through structured parameters and optional text. The system generates an edited street-level visualization that reflects those parameters and re-generates evaluations from all personas. Crucially, the system also maintains and surfaces conflict structure across iterations: it highlights where personas diverge, what design elements are driving divergence, and which tensions appear irreducible under constraints, so designers can decide what to prioritize rather than collapsing feedback into a single score.

3.4 Core Features

StreetDesignAI operationalizes disagreement-as-primitive through five integrated features, as shown in Figure 1.

3.4.1 Evaluation Generation: Parallel baseline assessment from multiple perspectives (DG1, DG2). Upon street selection via coordinate input, the system retrieves (1) Google Street View imagery and (2) OpenStreetMap attributes (e.g., road class, speed limit, existing cycling infrastructure), as shown in Figure 1 (A). Each of the four cyclist personas (Strong & Fearless, Enthused & Confident, Interested but Concerned, No Way No How) produces structured feedback including safety scores, comfort scores, and overall ratings (1–10), along with concise observations explaining their assessment. The system also generates driver pros and cons. This parallel evaluation establishes a baseline conflict pattern, making divergent perspectives visible at the outset for subsequent iteration.

3.4.2 Design Adjustment: Parameterized interventions and AI-rendered visualization (DG2, DG3). Designers propose interventions through a Parameter Specification Panel with options for bike lane width (narrow, widen, stay the same), lane color (green painting, no painting), buffer type (standard, bollards, armadillo), and buffer location (next to parked cars, next to moving cars), plus optional free-text requirements, as shown in Figure 1 (B). The system uses GPT-Image-1 API to render a modified street view image based on these parameters. After visualization, the system re-runs all persona evaluations on the new design, allowing designers to see how specific parameter changes shift safety and comfort scores across stakeholders.

3.4.3 Deep Analysis: Interrogating perspectives through conversational follow-up (DG1, DG3). Deep Analysis presents each persona’s initial observations in a conversation-like format and supports multi-turn dialogue where designers can ask follow-up questions such as “why do you feel unsafe?” or “what changes would improve your rating?” Personas receive the current image and context, stay in character, reference visible street elements, and provide actionable suggestions, as shown in Figure 1 (C). This feature helps designers unpack the mechanisms behind different ratings (e.g., which elements trigger fear for cautious cyclists versus which features drivers find restrictive), moving beyond opaque scores to understand the “why” behind evaluations.

3.4.4 Multi-Persona Discussion: Facilitating cross-persona debate on design preferences. (DG1). When comparing multiple design scenarios, the Multi-Agent Discussion tab presents persona responses discussing which design they prefer

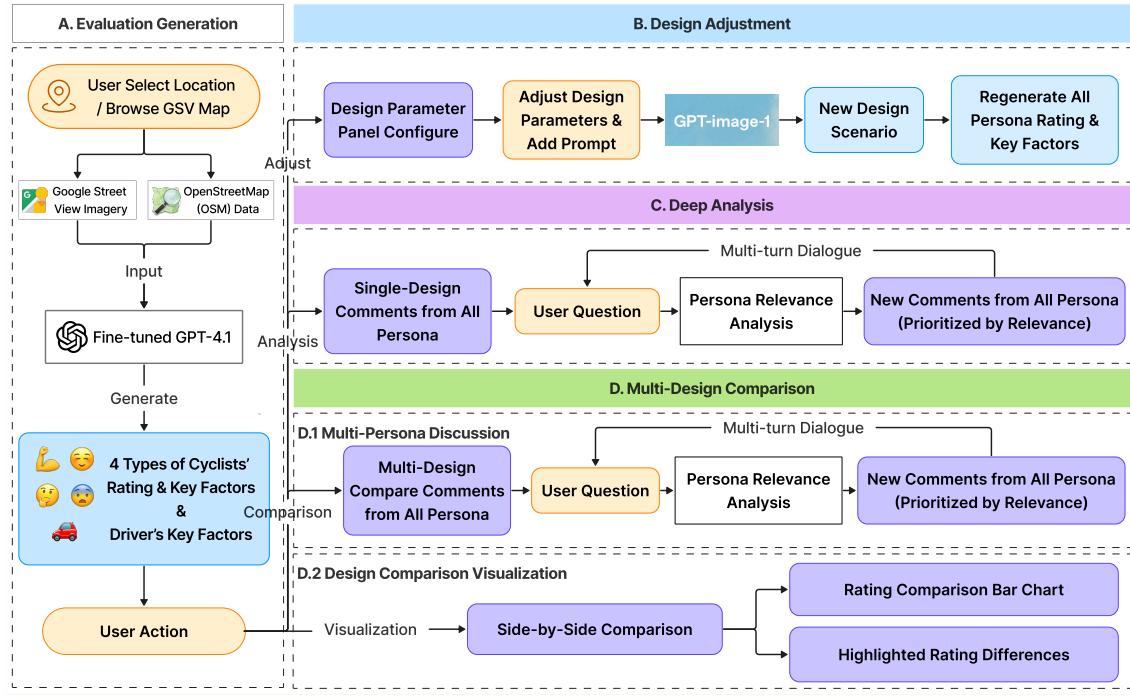


Fig. 2. System workflow of StreetDesignAI: The system consists of four main modules: (A) Evaluation Generation collects street-level imagery from Google Street View and road attributes from OpenStreetMap, then uses a fine-tuned GPT-4.1 model to generate safety ratings, comfort scores, and key factors for four cyclist personas and driver’s perspective; (B) Design Adjustment allows users to configure design parameters (lane width, color, buffer type) and uses GPT-image-1 to render modified streetscapes, triggering re-evaluation of all personas; (C) Deep Analysis supports multi-turn dialogue for in-depth analysis of design preferences across different personas within a single scenario, with system responses prioritized by relevance; (D) Multi-Design Comparison includes D.1 Multi-Persona Discussion for comparing design preferences across multiple scenarios, and D.2 Design Comparison Visualization module for side-by-side rating comparison with highlighted differences.

and why. Each persona articulates their design preference with justification, referencing specific elements (e.g., buffer zones, lane visibility, physical protection), as shown in Figure 1 (D.1). Designers can pose questions to the group, and each persona responds in order from high to low according to the relevance of the question.. This discussion format surfaces points of agreement, persistent disagreements, and the reasoning behind different preferences, helping designers understand trade-offs that require prioritization rather than simple optimization.

3.4.5 Design Comparison Visualization: Comparing alternatives through quantitative score profiles (DG1, DG2). The Design Scenario Comparison tab enables side-by-side evaluation of the existing condition against generated design scenarios, as shown in Figure 1 (D.2). For each persona, the system displays bar charts showing safety, comfort, and total scores across all alternatives, with each design scenario represented by a different color. The visualization also indicates each persona’s stated preference at the top. This allows designers to quickly identify which alternatives improve outcomes for specific user groups, where scores diverge across personas, and which design changes produce the largest shifts in perceived safety and comfort—supporting informed decision-making without collapsing diverse perspectives into a single metric.

3.5 Example Use Scenario

Zoe is an urban planner redesigning a commercial corridor. She selects a busy arterial with on-street parking and no dedicated cycling facility. StreetDesignAI retrieves street context and produces baseline evaluations: confident cyclists note lack of dedicated space, while cautious personas report they would avoid cycling due to exposure to traffic. The system highlights a baseline conflict pattern: cautious personas demand protection while confident riders and drivers prioritize flexibility and flow.

Zoe proposes a green-painted bike lane with a painted buffer. The visualization shows updated markings and evaluations improve for some personas, but the system indicates that core conflict remains: cautious cyclists still perceive insufficient physical protection. Zoe enters Deep Analysis and asks the *Interested but Concerned* persona what would make the design usable. The persona requests physical barriers and identifies specific fear points (e.g., proximity to moving vehicles). Zoe adds bollards along the traffic side and observes increased safety for cautious personas, while the driver perspective raises operational concerns. The Multi-Persona Discussion panel marks this as an irreducible trade-off, prompting Zoe to decide what to prioritize under right-of-way constraints. After exploring alternatives, Zoe uses Design Comparison to select an option with a more acceptable conflict profile rather than simply maximizing a single score.

3.6 Implementation Details

StreetDesignAI is implemented as a web-based system with a React front end and a model-backed evaluation and visualization pipeline. The system retrieves street-level imagery through the Google Street View API [21] and extracts contextual street attributes through the OpenStreetMap Overpass API [42] (e.g., road type, signals, existing cycling infrastructure, surrounding features). Designers interact with a custom canvas interface supporting pan/zoom, parameter editing, and multi-view comparison.

For persona-based evaluation and dialogue, the system uses a GPT-4.1 model [40] fine-tuned on a crowdsourced bikeability assessment dataset (see Section 3.7 for a detailed description) to reduce generic or idealized responses and improve persona consistency. Each persona agent is defined by a dedicated system prompt specifying identity, priorities, and evaluation focus. Outputs are returned in a structured JSON schema containing scores and observation points to ensure consistency and parsability. All prompts and schemas are included in Appendix 9.

For visual generation, StreetDesignAI uses GPT-Image-1 [41] for targeted street view edits. Prompts are structured to identify existing lane regions, apply dimensional specifications tied to selected parameters, and enforce boundary constraints to improve fidelity.

3.7 Crowdsourced Bikeability Assessment Data Collection

To address potential limitations of general-purpose LLMs on domain-specific tasks, we collected bikeability assessment data from real cyclists to ground persona-based evaluations in empirical experience. Specifically, we developed an interactive street rating platform that presents participants with immersive, panoramic Google Street View imagery and asks them to rate perceived safety, comfort, and overall bikeability on 1–10 scales. Open-ended questions are also asked to allow participants to express concerns, preferences, and improvement suggestions. The crowdsourced data are subsequently used to fine-tune GPT-4.1, enabling persona agents to generate evaluations grounded in real cyclist experiences.

3.7.1 Recruitment And Payment. We recruited participants through social media platforms to complete anonymous surveys evaluating bicycle lane quality. Each survey took approximately 5 minutes to complete. To incentivize participation, we implemented a lottery-based compensation system: every 100th participant received a \$100 gift card (4 participants total received compensation). This resulted in an average expected compensation of approximately \$0.93 per participant, or approximately \$11.16 per hour based on the estimated completion time. Participation was voluntary and anonymous, with no personally identifiable information collected. Eventually, a total of 427 cyclists with diverse skill levels, ages, and comfort thresholds completed the survey, yielding approximately 12,400 assessments across varied street conditions.

Three graduate students were recruited from a public R1 doctoral research university to perform data annotation tasks. Annotators were paid \$20 per hour for approximately 8 hours of annotation work each, totaling \$160 per annotator. This compensation rate exceeds local minimum wage standards and is consistent with standard research assistant rates at our institution.

3.7.2 Survey Interface. Figure 9 and Figure 10 show screenshots from our crowdsourcing survey platform. Figure 9 displays the immersive 360-degree Google Street View interface used for bikeability assessment, allowing participants to explore road environments interactively. Figure 10 shows the infrastructure preference assessment interface where participants rate their comfort levels for different cycling facility types.

4 STUDY DESIGN

To evaluate StreetDesignAI's effectiveness in supporting inclusive cycling infrastructure design, we conducted a within-subjects comparative study with 26 street designers comparing our system against a conventional AI-assisted design approach and examining how persona-based multi-agent evaluation influences designers' understanding of diverse user needs and design decision-making processes.

4.1 Participants

We recruited 26 participants with backgrounds in transportation engineering and planning through email outreach and snowball sampling. Eligibility criteria required participants to have at least one year of experience in road/transportation design or have participated in at least one project involving street design. Participants included practicing professionals such as traffic engineers, transportation planners, and project managers ($n=14$), academic faculty and researchers ($n=4$), graduate students ($n=6$), and undergraduate students ($n=2$). Participants had varying levels of experience in road/transportation design ($M=4.62$ years, $SD=4.00$, range: <1 to >10 years), with 9 participants having more than 5 years of experience. They represented diverse contexts including government agencies, private industry, and academic institutions. Participants ranged in age from 19 to 57 years ($M=30.58$, $SD=9.85$), with 18 identifying as male, 7 as female, and 1 as nonbinary. Table 1 provides detailed demographic information.

4.2 Task Design

We employed a Latin square balanced within-subjects design to control for potential learning effects. Each participant experienced both conditions:

Condition A - StreetDesignAI: Participants used the full StreetDesignAI system with functions below:

- Multi-Persona evaluation from four cyclist types and and the driver view
- Visual generation of design alternatives

Table 1. Participant demographics and professional backgrounds (N = 26).

| ID | Age | Gender | Role | Years | Projects |
|-----|-----|-----------|--|-------|----------|
| P1 | 27 | Female | Transportation Engineering PhD Student | 5-10 | 1-3 |
| P2 | 26 | Male | Transportation Engineering PhD Student | 1-3 | 1-3 |
| P3 | 21 | Male | Transportation Engineering Undergraduate Student | 1-3 | 0 |
| P4 | 21 | Male | Transportation Engineering Undergraduate Student | <1 | 1-2 |
| P5 | 19 | Nonbinary | Environmental Leaders Fellow | <1 | 1-2 |
| P6 | 35 | Male | Traffic Engineer | 5-10 | >5 |
| P7 | 22 | Female | Roadway Engineer 1 | <1 | 1-2 |
| P8 | 27 | Male | Associate Traffic Engineer | 3-5 | 3-5 |
| P9 | 28 | Male | Traffic Engineering Associate | 3-5 | 3-5 |
| P10 | 30 | Male | Transportation Planner | 3-5 | 3-5 |
| P11 | 50 | Male | Sr. Project Manager, County Government | >10 | >5 |
| P12 | 25 | Male | Transportation Engineering PhD Student | 1-3 | >5 |
| P13 | 25 | Male | Entry Level Transportation Engineer/Urban Designer | 1-3 | >5 |
| P14 | 26 | Male | Transit Planner | 3-5 | 1-2 |
| P15 | 57 | Male | Senior Principal Engineer | >10 | >5 |
| P16 | 27 | Male | Transportation Engineer | <1 | 1-2 |
| P17 | 37 | Female | Transportation Project Manager | >10 | >5 |
| P18 | 29 | Female | Transportation Engineering Faculty | 5-10 | >5 |
| P19 | 53 | Male | Transportation Engineering Professor | >10 | >5 |
| P20 | 29 | Male | Transportation Engineering Research Assistant | 5-10 | 0 |
| P21 | 29 | Female | Transportation Major PhD Student | 1-3 | 3-5 |
| P22 | 23 | Male | Transportation Major Master Student | 1-3 | 0 |
| P23 | 28 | Female | Transportation Engineer | 1-3 | 1-2 |
| P24 | 40 | Male | Transportation Research Faculty | 5-10 | >5 |
| P25 | 37 | Male | Urban Planner | <1 | 1-2 |
| P26 | 24 | Female | Transportation Engineering PhD Student | 1-3 | 1-2 |

- Structured parameter controls for infrastructure modifications
- Integrated street context from OpenStreetMap

Condition B - ChatGPT (baseline): Participants used ChatGPT-4.1 with prompt template that provided street context and requested design recommendations. This baseline represents current practice in AI-assisted road design consultation, where designers might seek general AI guidance without persona-specific perspectives. We selected ChatGPT as our baseline because it is the most widely adopted AI assistant among design professionals (8 of 12 formative study participants had used it), and no existing AI tools specifically support multi-perspective cycling infrastructure evaluation.

For each condition, participants were tasked with improving cycling infrastructure at real street locations. Participants were free to select any real street location they wished to improve.

4.3 Study Procedure

The study procedure followed a structured protocol as illustrated in Figure 3. Each session began with participants completing a pre-study survey assessing their baseline design confidence and understanding of diverse cyclist needs using adapted items from established scales (e.g., “I understand how different types of cyclists experience the same infrastructure,” “I feel confident designing for users whose needs differ from my own”).

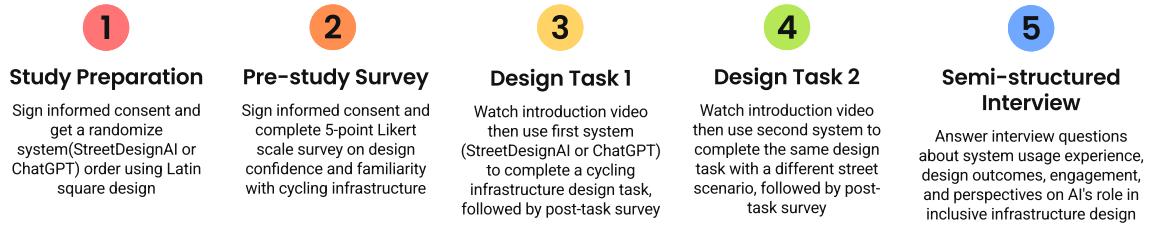


Fig. 3. Study workflow: participants completed five phases: (1) pre-study survey on design confidence; (2-3) two design tasks using StreetDesignAI and ChatGPT in counterbalanced order, each with post-task surveys; (4) comparative reflection; and (5) semi-structured interview on system experiences and AI's role in inclusive design. Total session time averaged 95 minutes.

After viewing a brief introduction video about the study objectives (without revealing specific hypotheses), participants proceeded to the first design task. They were instructed to think aloud while exploring design alternatives, documenting their process through screenshots of the three most useful insights or design iterations. This think-aloud protocol provided rich qualitative data about their design reasoning and how they interpreted system feedback.

Following the first condition, participants completed a mid-study survey measuring:

- **Understanding of diverse needs:** Perceived insight into different cyclist perspectives
- **Design confidence:** Self-efficacy in creating inclusive solutions
- **System usability:** Ease of use and interaction quality
- **Trust in feedback:** Credibility of AI-generated evaluations

Participants then repeated the design task using the alternate system, followed by the same assessment battery. The session concluded with a 30-minute semi-structured interview exploring their comparative experiences, perceived value of persona-based feedback, and reflections on AI's role in inclusive design practice.

4.4 Data Collection and Analysis

We collected multiple forms of data to triangulate findings:

Interaction logs: System telemetry captured all design iterations, parameter modifications, and time spent on different activities. For StreetDesignAI, this included persona feedback viewed, design parameters modified, and iteration patterns. For ChatGPT, we logged conversation turns and design topics discussed.

Survey responses: Likert-scale responses were analyzed using paired t-tests to compare conditions, with Bonferroni correction for multiple comparisons. Effect sizes (Cohen's d) were calculated to assess practical significance.

Interview transcripts: Following Braun and Clarke's thematic analysis framework [6], two researchers independently coded interview transcripts, identifying patterns in how designers perceived and utilized persona-based feedback. Initial codes were generated inductively, then organized into themes aligned with our research questions. Discrepancies were resolved through discussion until consensus was reached.

5 RESULTS

We report results organized by the three research questions. Across all analyses, we combine quantitative survey results with qualitative interview data from 26 participants. For comparisons between StreetDesignAI and ChatGPT, we used Wilcoxon signed-rank tests. For StreetDesignAI-specific features, we report descriptive statistics and one-sample tests against the neutral midpoint. Qualitative findings are used to contextualize and explain observed quantitative patterns.

Manuscript submitted to ACM

Table 2. Comparison of baseline usability perceptions between ChatGPT and StreetDesignAI. Wilcoxon signed-rank tests compare paired participant ratings under the two systems.

| Dimension | ChatGPT ($M \pm SD$) | StreetDesignAI ($M \pm SD$) | W | p-value |
|---------------------------|---------------------------|----------------------------------|------|---------|
| Ease of Use | 3.88 ± 0.86 | 4.19 ± 0.85 | 37.5 | 0.185 |
| Interaction Intuitiveness | 3.58 ± 0.86 | 3.88 ± 0.91 | 59.5 | 0.216 |

Note: W and p-values represent Wilcoxon signed-rank test results comparing paired participant ratings between ChatGPT and StreetDesignAI.

Participants reported comparable levels of baseline usability and interaction intuitiveness for ChatGPT and StreetDesignAI, with no statistically significant differences observed between the two systems (Table 2). However, participants rated all StreetDesignAI-specific features significantly above the neutral midpoint ($M=3$), indicating strong perceived support for design exploration, rapid iteration, and trade-off reflection. The Generate Evaluation ($M=4.31$, $SD=0.79$) and Comparative Analysis ($M=4.46$, $SD=0.71$) functions received the highest ratings (Figure 4).

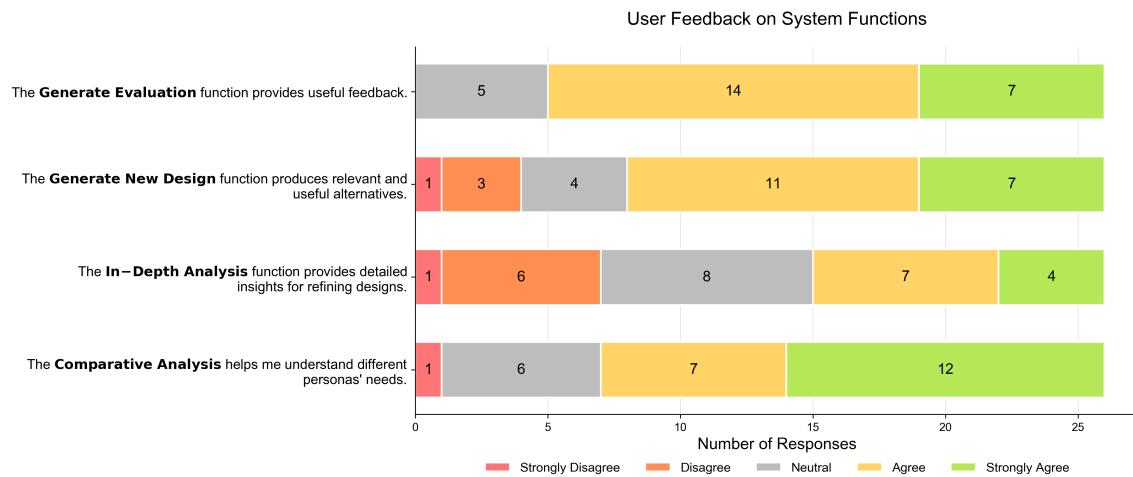


Fig. 4. Distribution of participant ratings for four key system functions (N=26). All functions rated above neutral midpoint.

Participants repeatedly emphasized that StreetDesignAI's structured interface and interaction model reduced friction in exploration. As P12 noted: "I think that was pretty intuitive. It gave you the image, what's wrong with it, and what these various users would probably think. The scoring system was clear and easy to understand." This structure contrasted with ChatGPT's text-heavy flow. P5 explained: "The interface on the website is much more intuitive than ChatGPT. It allows you to view each scenario more clearly as its own section instead of scrolling through text." Participants also highlighted the efficiency of button-based interactions. As P5 further elaborated: "Instead of filling out a text box, selecting buttons is much more efficient. The structured approach helps organize my thinking."

5.1 RQ1 - Exploration of Design Alternatives

5.1.1 Visual Generation and Iteration Speed. Rapid visualization enabled designers to explore alternatives early in the workflow. P6 observed that "within maybe 1-2 minutes, we can visualize changes from existing conditions to proposed

improvements. That speed is really valuable for early design exploration.” P25 similarly praised the generative capability: “The image generation is super good. Getting a realistic visualization that quickly is incredibly valuable for concept development.” This speed also supported practical constraints in professional contexts. P12 noted: “If I were in a rush to a meeting and needed to present design concepts, generating these images would be incredibly helpful. Much faster than Illustrator or Photoshop.”

Participants described the exploration value of visual generation as democratizing and communicative. P1 stated: “This image capability is extremely useful. Not everyone can draw, and even those who can can’t produce realistic renderings this fast.” P16 emphasized stakeholder communication: “the renderings even make the street look more beautiful...it helps sell the vision to stakeholders.” While acknowledging limitations, participants still found the feature useful for early-stage exploration. P5 offered: “Image generation has room for improvement, but the current capability is already useful for concept-level discussions.” and P18 added: “right-of-way constraints aren’t always reflected accurately, but for preliminary concepts, the visualizations are valuable.”

5.1.2 Persona Evaluation Quality. Evaluation and comparison features supported explicit reflection on persona-specific trade-offs. Participants valued the evaluation function’s fidelity to environmental cues and safety trade-offs. P3 noted: “the observations about cars are pretty good...how they affect drivers’ speed and the hazard of door zones when someone’s biking by.” and continued: “The observation about trees and environmental context is impressive. It does a good job understanding what’s actually in the environment.” P7 confirmed the plausibility of outputs: “the information it gives makes sense. It correctly identified that there’s no bike lane and that the pavement is uneven.” The speed advantage over manual work was salient. P1 stated: “the AI-generated pros and cons are impressive. It captures the key tradeoffs I would identify manually, but much faster.” P15 highlighted comfort-level details: “picked up on subtle details like pavement quality and sight lines that affect cyclist comfort.” Participants also noted occasional issues. P25 warned: “Sometimes it identified bike infrastructure that wasn’t there. The false positives could be confusing for users unfamiliar with the location.”

5.1.3 Comparative Analysis and Trade-off Reflection. Comparative analysis further strengthened trade-off reflection across alternatives. P10 appreciated presentation: “the color coding that highlights certain keywords makes it easy to quickly compare feedback across different design scenarios.” and also emphasized quantified deltas: “Seeing the safety score change from 3 to 5 after adding bollards gives clear, quantifiable feedback on design improvements.” P3 connected comparison with persona discussions: “I love the discussions between personas and the comparative analysis. Being able to see how each group responds to changes is very useful.” P2 found comparisons cognitively supportive: “having a basic idea in my head, this gives me a more intuitive comparison than I could do mentally. Visualizing the tradeoffs is powerful.” P16 described prioritization utility: “helped me understand which design elements had the biggest impact on different user groups. Very useful for prioritization.”

5.1.4 Design Parameter Choices. Analysis of 48 design scenarios across 26 sessions revealed consistent patterns in parameter selection (Figure 5). Green painted lanes were overwhelmingly preferred (85.4%, 41/48 designs) over unpainted alternatives (14.6%, 7/48). Physical separation through narrow bollards (35.4%, 17/48) and armadillo dividers (29.2%, 14/48) were collectively chosen in 64.6% of designs, exceeding standard painted buffers (27.1%, 13/48) and minimal buffering (8.3%, 4/48). Most participants selected wider lanes (47.9%, 23/48) over maintaining existing width (37.5%, 18/48) or narrowing (14.6%, 7/48). Buffers were more frequently placed adjacent to parked cars (62.5%, 30/48) than moving traffic (37.5%, 18/48).

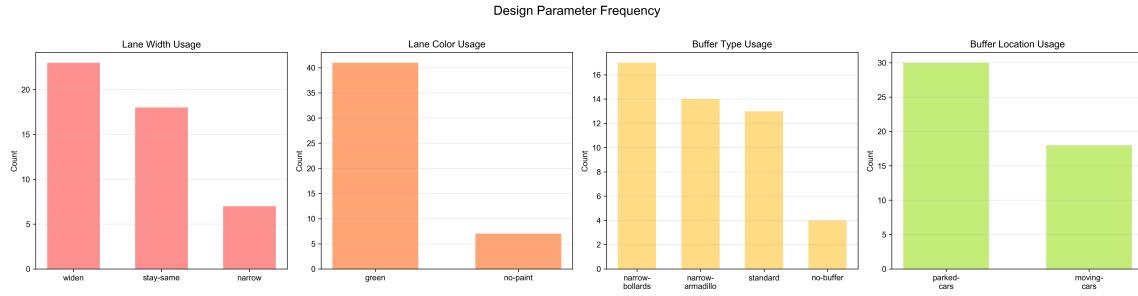


Fig. 5. Frequency of design parameter selections across 48 design scenarios: (a) lane width, (b) lane color, (c) buffer type, (d) buffer location.

The convergence on green paint (85.4%) and physical barriers (combined 64.6%) indicates that visibility and protection emerged as dominant design principles. As P12 explained during design exploration: “The green makes it unmistakably clear where cyclists should be, and the bollards give that physical reassurance that cars can’t drift into the lane.” The preference for parked-car-side buffering aligns with participants’ explicit concerns about dooring hazards, as noted by P8: “The buffer next to parked cars is essential. I’ve seen too many door-zone incidents to skimp on that protection.” The diversity of parameter combinations (48 design solutions) demonstrates that StreetDesignAI supported flexible exploration while revealing consistent strategic priorities.

5.1.5 In-Depth Analysis and Follow-up Interaction. The in-depth analysis enabled deeper exploration. P18 stated: “the feature I like most. Seeing how different risk perception groups respond to the same design is fascinating and useful.” P1 highlighted interactivity: “you can follow up with questions and dig deeper into specific concerns. The interactivity adds value.” P2 emphasized actionable guidance: “provides more practical, actionable information. When I asked about separation, it suggested specific buffer types and dimensions.” Participants also proposed refinements to increase realism and persona fidelity. P25 suggested: “the second agent could better consider the output of the first agent. More true back-and-forth dialogue would increase realism.” and P18 noted: “The ‘strong and fearless’ persona could be less focused on physical barriers since they’re supposed to be comfortable with less protection.”

5.2 RQ2 - Understanding of Diverse User Needs

Table 3. Comparison of perceived understanding and intention-to-use outcomes between ChatGPT and StreetDesignAI. Wilcoxon signed-rank tests compare paired participant ratings under the two systems.

| Dimension | ChatGPT ($M \pm SD$) | StreetDesignAI ($M \pm SD$) | W | p-value |
|---------------------------------------|---------------------------|----------------------------------|------|---------|
| Understanding new perspectives | 3.46 ± 0.99 | 4.04 ± 1.00 | 43.5 | 0.031* |
| Authenticity of persona feedback | 3.31 ± 1.01 | 3.77 ± 0.71 | 51.5 | 0.070 |
| Overall Satisfaction | 3.38 ± 0.85 | 3.85 ± 0.83 | 24.0 | 0.034* |
| Intention to use in professional work | 2.88 ± 1.34 | 3.62 ± 1.13 | 46.0 | 0.025* |

Note: * $p < 0.05$. W and p-values represent Wilcoxon signed-rank test results comparing paired participant ratings between ChatGPT and StreetDesignAI.

5.2.1 Comparative Understanding Outcomes. StreetDesignAI significantly outperformed ChatGPT in helping participants understand perspectives they had not previously considered in street design ($W=43.5, p = 0.031$), with mean

ratings increasing from 3.46 to 4.04 (Table 3). StreetDesignAI also led to significantly higher overall satisfaction (3.38 to 3.85, $W=24.0$, $p = 0.034$) and stronger intention to use the system in professional work (2.88 to 3.62, $W=46.0$, $p = 0.025$). While participants rated persona feedback from StreetDesignAI as more authentic than ChatGPT (3.31 to 3.77), this difference did not reach statistical significance ($p = 0.070$).

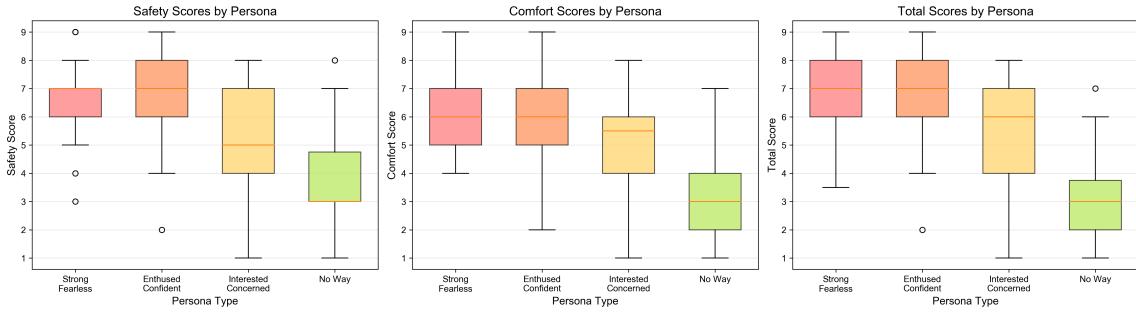


Fig. 6. Distribution of safety, comfort, and overall suitability scores across four cyclist personas (N=78 evaluations from 26 sessions).

5.2.2 Persona Score Distributions and Experiential Divergence. Analysis of 78 evaluations (30 existing conditions + 48 design scenarios) revealed systematic divergence in how different personas rated infrastructure (Figure 6). Strong & Fearless ($M=6.90$, $SD=1.08$) and Enthused & Confident ($M=6.91$, $SD=1.22$) cyclists rated scenarios similarly, with median overall scores around 7 and relatively compact distributions. In contrast, Interested but Concerned cyclists showed substantially lower scores ($M=5.58$, $SD=1.76$) with high variability, indicating that this group's experience is highly sensitive to design choices. No Way No How cyclists rated nearly all scenarios below 4 ($M=3.02$, $SD=1.23$), with minimal variation across scenarios. The 3.9-point gap between confident cyclists ($M\approx6.9$) and the most cautious group ($M=3.02$) quantifies the magnitude of experiential conflict that designers must navigate.

P18's design session exemplifies this divergence: on a Chicago street with existing bike lanes, Strong & Fearless rated safety at 9/10, while No Way No How rated it only 3/10—a 6-point gap on identical infrastructure. As P18 reflected after viewing such contrasts: “Seeing how different risk perception groups react to the same design really changed how I think about inclusive infrastructure. I used to design for the average cyclist, but there's no such thing.” The Interested but Concerned group's high variance ($SD=1.76$, the largest among all personas) reveals substantial opportunity for design interventions to improve their experience, positioning this population as a critical target for infrastructure improvements.

5.2.3 Surfacing Overlooked Perspectives. Participants described how StreetDesignAI's persona-based structure surfaced overlooked perspectives and contextual nuance. P4 described: “perspectives I hadn't considered. For example, it was the only one to acknowledge the available sidewalk as an alternative for cautious cyclists.” P18 similarly emphasized: “Seeing how different risk perception groups react to the same design really changed how I think about inclusive infrastructure.” Participants also identified how the persona framing filled gaps in typical design practice. P3 explained: “The personas are really useful because I don't usually think about different cyclist comfort levels when doing road design.”

7 participants highlighted that StreetDesignAI helped them reason about trade-offs in perspective understanding. P1 observed that the persona system “can remind you that a certain design might actually make some user groups feel

worse, even when overall scores improve. That's a critical insight." Beyond personas, the system prompted consideration of contextual factors. P5 noted that it "It made me think about weather in ways I hadn't really considered before, like how rain or snow can affect lane markings and surface traction."

5.2.4 Contextual and Situational Understanding. Participants also described understanding as situational and context-dependent rather than fixed. P15 stated: "People can be in multiple categories depending on context. I was a very different cyclist when riding with my younger children. The tool captures that nuance." P5 reframed a persona interpretation: "the 'no way, no how' persona is like a mother's perspective of a child taking this route to school... That's a real user group we need to design for." The system also helped calibrate how much change is needed to shift perceptions. P4 reported: "people who are worried about cycling risk won't be convinced by minor improvements, they need substantial infrastructure changes." Some participants noted that the system expanded their attention beyond cyclist-only perspectives. P1 admitted: "Before using the tool, I rarely read through driver perspectives carefully. But the driver section gave very useful suggestions that made me reconsider my design approach." For experienced cyclists serving as designers, the system provided a corrective lens. P7 reflected: "As someone who's comfortable on most roads, I tend to forget how intimidating they can be for less confident cyclists."

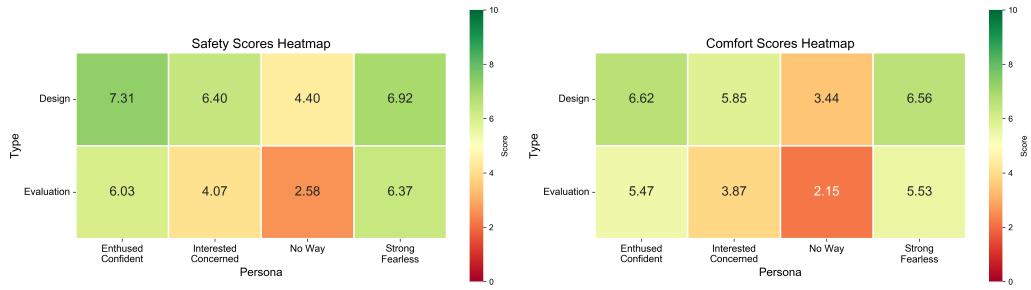


Fig. 7. Mean safety and comfort scores by persona and scenario type (evaluation vs. design). Color intensity indicates score magnitude (red = lower, green = higher).

5.2.5 Design Intervention Effectiveness Across Personas. Comparing existing condition evaluations against design scenario proposals revealed differential responsiveness to interventions (Figure 7). Design scenarios consistently received higher scores than evaluations across most personas, demonstrating that proposed interventions improved perceived safety and comfort. However, improvement magnitude varied substantially: Interested but Concerned showed measurable gains in comfort (5.13 in evaluations → 5.30 in designs, +0.17), while No Way No How showed no improvement and slight declines (safety: 3.74 → 3.67; comfort: 2.97 → 2.91). Strong & Fearless and Enthused & Confident personas showed modest improvements, reflecting their baseline comfort with existing infrastructure. The persistent stratification—confident cyclists rating 3+ points higher than cautious cyclists even in improved designs—underscores the challenge of achieving universal satisfaction through standard protected lane treatments.

The heatmap pattern reveals a key trade-off: interventions that meaningfully improve conditions for moderately cautious cyclists (Interested but Concerned) may still fail to attract the most risk-averse group. As P15 observed: "Even with physical barriers, some people just won't feel safe sharing road space with cars. They need complete separation, like a sidepath or protected cycle track." This finding has important implications for infrastructure investment: standard protected lanes can expand the cycling population by converting Interested but Concerned individuals into riders,

but may never achieve truly universal access without grade-separated facilities. The sensitivity pattern suggests that designers must explicitly decide which populations to prioritize under resource and right-of-way constraints.

5.2.6 Perceived Authenticity and Credibility. Perceived authenticity of persona feedback emerged as an important qualitative theme for interpreting understanding benefits. P5 observed: “the personas stay pretty true to their perspective. The ‘no way, no how’ and ‘interested but concerned’ especially feel authentic.” and added: “It mimics a real person pretty accurately. If you compared these responses to real survey data, I think you’d find significant overlap in the concerns raised.” P2 suggested that predictability enhanced credibility: “The responses are fairly predictable given each persona’s profile, which actually increases credibility. They say things consistent with their character.” Even when participants detected AI-like patterns, they prioritized substantive value. P18 summarized: “You can tell it’s AI-generated if you read carefully, but the substance of the feedback is valid and useful regardless of the source.”

5.2.7 Qualitative Explanations and Professional Workflow Integration. Finally, participants emphasized that understanding was improved by qualitative explanations rather than numeric scores alone. P25 stated that “the qualitative comparison is more insightful than just rating safety 7 versus 6. Understanding why different users feel differently is more valuable than the numbers alone.” P5 rated the system highly for perspective-taking: “It’s Very useful...I’d rate it nine out of ten for gaining perspective on how people might perceive a street redesign from both driver and cyclist viewpoints.” P3 connected this understanding to stakeholder engagement: “This provides a good baseline for what options to show people and what questions to bring up in public engagement. It helps structure the conversation.”

Participants also explained where StreetDesignAI could fit into professional workflows, clarifying why intention-to-use was higher. P16 stated: “This is a great tool for planners, especially city planners and DOTs when planning bike lanes and expanding bikeway networks.” P4 noted its utility for “preparing to respond to public comment. If you know your city has many cyclists, this helps you anticipate their concerns.” P12 positioned it as “...useful in the early stages of design, when you’re deciding whether a road needs improvement but haven’t gotten into detailed plans yet. It feels like a good starting point for exploration” P23 emphasized early-phase value: “At the very beginning of a project, this is very useful for generating ideas about justification and basic design scenarios.” Finally, P6 summarized the resource value: “It’s particularly valuable when you need to quickly explore multiple design alternatives before committing resources to detailed analysis.”

5.3 RQ3 - Capability and Confidence in Inclusive Design

Table 4. Pre-post changes in designers’ perceived capability and confidence under ChatGPT and StreetDesignAI conditions. Significance tests compare pre-study baseline measurements with post-condition assessments using Wilcoxon signed-rank tests.

| Dimension | Pre-study ($M \pm SD$) | ChatGPT | | | StreetDesignAI | | |
|---|-----------------------------|------------------------|------|---------|------------------------|------|----------|
| | | Post ($M \pm SD$) | W | p-value | Post ($M \pm SD$) | W | p-value |
| Persona Need Identification Capability | 3.27 ± 1.04 | 3.73 ± 0.78 | 48.0 | 0.044* | 4.12 ± 0.91 | 18.5 | 0.0028** |
| Inclusive Design Translation Confidence | 3.27 ± 1.22 | 3.81 ± 0.90 | 50.0 | 0.063 | 4.19 ± 0.75 | 25.5 | 0.0045** |

Note: * $p < 0.05$, ** $p < 0.01$. W and p-values represent Wilcoxon signed-rank test results comparing pre-study baseline scores with post-condition assessments.

5.3.1 Pre-Post Capability and Confidence Changes. Both ChatGPT and StreetDesignAI improved participants’ ability to identify and articulate the needs of different cyclist personas, with stronger improvement observed under StreetDesignAI

Manuscript submitted to ACM

(Table 4). Persona need identification increased from baseline ($M=3.27$, $SD=1.04$) to post-ChatGPT ($M=3.73$, $SD=0.78$, $W=48.0$, $p = 0.044$) and further to post-StreetDesignAI ($M=4.12$, $SD=0.91$, $W=18.5$, $p = 0.0028$). Only StreetDesignAI produced a significant increase in confidence translating persona needs into design decisions (baseline $M=3.27$, $SD=1.22$ → post $M=4.19$, $SD=0.75$, $W=25.5$, $p = 0.0045$), compared to ChatGPT's non-significant change (post $M=3.81$, $SD=0.90$, $p = 0.063$). These findings suggest that while general-purpose AI tools may support initial awareness of diverse user needs, StreetDesignAI more effectively supports the translation of those needs into actionable design decisions.

5.3.2 Educational and Corrective Value. Qualitatively, participants described capability gains as both educational (for novices) and corrective (for experienced designers). P18 noted: "For someone specialized in transportation planning, they can get detailed information from this interface. For planners specializing in land use or housing, this would be even more helpful." For novice practitioners, P13 stated: "As a beginner planner in the multimodal space, I found the structured personas very educational. It taught me how to think about different user types." Even experienced professionals emphasized reduced blind spots. P15 observed: "Even as an experienced professional, I found value in the systematic approach. It's easy to develop blind spots, and this tool helps identify them."

Participants explained that capability improvements also related to the system's comprehensiveness relative to what they would do independently. P1 contrasted: "The persona framework considers diverse cyclists more comprehensively than I would on my own. With ChatGPT, I'd need to explicitly prompt for each perspective." P6 described translation from technical detail to user experience: "The tool bridges the gap between technical knowledge and user experience. It translates engineering concepts into human impacts."

5.3.3 Parameter-Level Impact on Persona Scores. Examining how specific design parameters influenced persona scores revealed differential sensitivities that inform targeted interventions (Figure 8). Lane width showed progressive improvements across all personas, with Interested but Concerned exhibiting the steepest increase (median ≈ 6.5 for widened lanes vs. 4.5 for narrow lanes, approximately +2.0 points). Green paint substantially benefited Interested but Concerned cyclists (median ≈ 6.0 vs. 4.0 without paint, +2.0 points) while having minimal effect on Strong & Fearless, indicating that visibility enhancements disproportionately serve cautious users. Physical barriers (bollards and armadillos) elevated Interested but Concerned scores (median ≈ 7.0) compared to standard buffers (median ≈ 5.0) or no buffer (median ≈ 3.0), representing approximately +2.0 points improvement. However, No Way No How remained below 4.0 regardless of buffer type, confirming that standard protected lanes cannot serve the most risk-averse population. Buffer location showed modest effects, with parked-car-side placement yielding slightly higher scores for Interested but Concerned, consistent with prioritizing dooring risk mitigation.

These granular insights demonstrate how StreetDesignAI enables evidence-based parameter selection: designers can identify which interventions serve which populations, avoiding one-size-fits-all solutions. In one Florida design session, a participant exploring buffered lane options observed: "Adding the standard buffer improved the Interested but Concerned score from 2 to 5—a 150% increase—while the No Way No How group stayed at 2. That tells me exactly where to invest: this design will expand the cycling population but won't achieve universal access. If I need to serve everyone, I have to go with full separation." The variation in sensitivity across personas clarifies which design elements address which concerns: green paint enhances legitimacy and visibility (key for Interested but Concerned), physical barriers provide psychological reassurance (essential for cautious riders), and width supports both maneuverability (valued by Strong & Fearless) and personal space (critical for nervous cyclists).

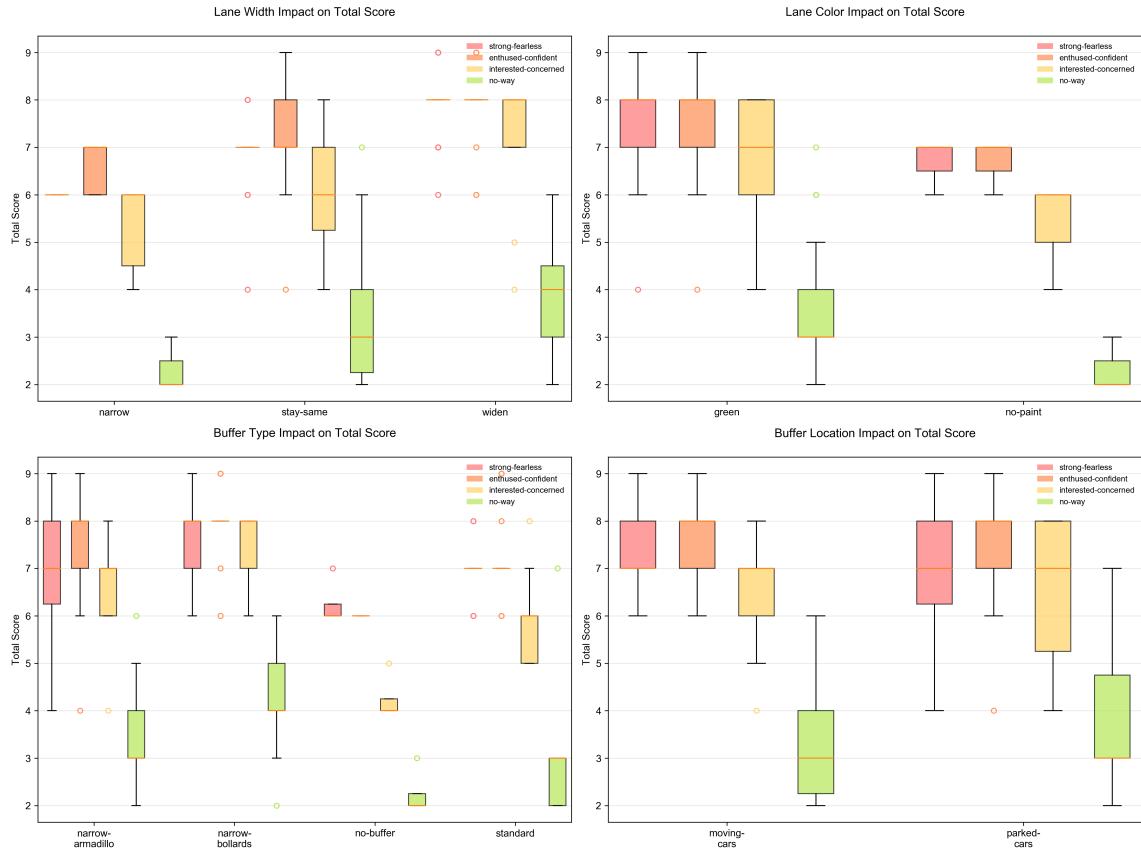


Fig. 8. Distribution of overall suitability scores by design parameter choice and persona: (a) lane width, (b) lane color, (c) buffer type, (d) buffer location.

5.3.4 Visual Feedback and Optimization Approaches. Confidence gains were tied to concrete visual and quantitative feedback. P12 noted: “If I needed to decide whether to put the bike lane inside or outside parked cars, I could generate these images and see what it would look like. That visual feedback builds confidence.” P2 described an optimization-oriented approach enabled by ratings: “Now I can ensure the first three user types all score above 7, then find the lowest cost solution that achieves that. It gives me a clear optimization target.” Participants also emphasized reassurance about coverage of concerns. P15 stated: “the tool gave me confidence that I’m not missing major concerns. Having systematic feedback from multiple perspectives is reassuring.” P16 connected this to stakeholder-facing justification: “more prepared to defend my design choices now. I can anticipate objections and explain how I’ve considered different user needs.”

5.3.5 AI as Augmentation, Not Replacement. Across interviews, participants consistently positioned capability and confidence gains within an AI-as-augmentation relationship. P1 explained: “I see this as enhances what designers can do rather than replacing their judgment. The interests of public safety and professional liability require human oversight.” P16 quantified efficiency benefits: “the tool saves time on tasks that used to take hours. Instead of 3-4 hours

for an assessment, it takes 30 minutes.” P4 summarized: “This is augmentation at its best. It handles time-consuming analysis while letting designers focus on creative problem-solving.”

Participants also emphasized the boundaries of AI and the irreplaceable role of human judgment. P16 stated: “Only a human can really understand the specific context of a location, project, and user types. AI is one tool among many.” P13 stressed oversight and validation: “...humans need to stay involved in the process. You can’t just let AI run on its own. We still need to talk to residents, check design standards, and make sure everything aligns with policy.” Local knowledge was repeatedly framed as a human advantage. P3 observed: “Road designers are always going to have a better understanding of the local context, like traffic patterns, special events, and community needs... That kind of knowledge isn’t something AI can really replace.”

Finally, professional responsibility and accountability further shaped how participants interpreted confidence and capability. P1 noted: “Licensing and liability also matter a lot. Engineers are the ones who sign and seal drawings, and that responsibility can’t be handed off to an AI system.” P25 stated: “AI can’t take responsibility. There has to be a real person accountable for design decisions.” P23 concluded directly: “Designers must take responsibility...humans must remain in the decision-making loop.”

6 DISCUSSION

In this section, we discuss the implications of our findings for AI-assisted inclusive design, situated in past research on human-AI collaboration and infrastructure planning. We also offer design recommendations for future systems utilizing persona-based evaluation.

6.1 Conflict Surfacing as a Design Primitive for Inclusive Infrastructure

A central finding of this work is that making experiential conflicts explicit reshapes how designers explore and reason about alternatives. Beyond past studies focusing on general-purpose AI design assistance [23], we demonstrated that StreetDesignAI’s multi-persona design enhances inclusive design reasoning by directly presenting designers with preference divergences among different user groups. This design choice aligns with calls in HCI research for AI systems that support rather than supplant human judgment [9, 27]. P1 observed that the persona system “can remind you that a certain design might actually make some user groups feel worse, even when overall scores improve,” articulating the value that conflict visibility prevents false consensus and prompts designers to take responsibility for their prioritization decisions. By further examining participants’ design exploration behaviors (Section 5.1), we identified that explicitly presenting preference divergences provides designers with a means to recognize trade-offs requiring deliberate choices rather than technical optimization. This design promotes user agency while supporting the ideation process, drawing from the experiential knowledge of different cyclist populations, meanwhile reducing perspective substitution and addressing concerns about designing only for users similar to oneself.

6.2 Cultivating Trade-off Reasoning through Multi-Perspective Evaluation

Our findings indicate that StreetDesignAI’s multi-perspective design not only supports perspective-taking but also triggers higher-order reasoning processes crucial for inclusive infrastructure design. In particular, we observed participants repeatedly engaging in explicit prioritization activities when persona evaluations diverged. StreetDesignAI’s design of presenting parallel persona feedback prompts participants to think about whose needs to prioritize under various constraints, thereby stimulating trade-off reasoning.

For instance, by viewing divergent persona evaluations, participants reformulated their design goals in more precise terms (e.g., shifting from “improve cycling conditions” to “expand access for cautious cyclists while maintaining efficiency for confident riders”). Comparison instances (e.g., where participants identified that bollards improved “Interested but Concerned” scores while “No Way No How” scores remained unchanged) were associated with explicit decisions about which populations to invest in serving.

Additionally, the system’s comparative analysis structure and iterative parameter adjustment mechanism promoted designer-led reflection through inference activities based on score changes across design iterations. Our findings suggest that this conflict-surfacing interaction design contributes to designers’ inclusive infrastructure reasoning process by promoting different levels of trade-off activities. These observations highlight the potential application of persona-based evaluation for fostering deliberate prioritization in infrastructure design and planning education [4, 7].

6.3 From Understanding to Translation: Bridging the Confidence Gap

Our quantitative results reveal an important distinction between understanding diverse user needs and confidently translating that understanding into design decisions. While both ChatGPT and StreetDesignAI improved participants’ ability to identify persona needs, only StreetDesignAI produced significant gains in translation confidence ($p = 0.0045$ vs. $p = 0.063$). This gap suggests that awareness of diversity is necessary but insufficient for inclusive design practice—designers still need scaffolding to act on that awareness.

StreetDesignAI’s parameter-level feedback appears to bridge this gap by linking specific design choices to persona-specific outcomes. When P6 observed that adding bollards increased “Interested but Concerned” scores from 2 to 5 while “No Way No How” scores remained unchanged, they gained actionable insight into which interventions serve which populations. StreetDesignAI transforms abstract understanding into targeted decision-making, helping designers more systematically address diverse user needs.

6.4 Navigating Potential Biases and Limitations in Persona-Based Evaluation

Cognitive biases have been common concerns in LLM-based user simulation systems [2]. Although our findings demonstrated the utility of the system in providing inclusive design support, we noted potential biases when designers interacted with persona evaluations. In StreetDesignAI, confirmation bias may arise from both designers’ persona attention patterns and the LLM’s inherent limitations. Participants who identified as experienced cyclists tended to engage more deeply with the “Strong & Fearless” persona feedback, potentially reinforcing their existing design intuitions. Additionally, they trusted persona feedback more when it aligned with their professional experience, potentially limiting exploration of unfamiliar perspectives. This aligns with prior observations that individuals prefer assistance that aligns with their existing beliefs [16, 22].

Another potential concern we noticed was over-reliance on quantified scores. Users tended to favor numerical safety and comfort ratings as decision anchors, even when qualitative explanations could provide richer contextual information. Combined with the risk of equating AI-simulated feedback with genuine community input, these factors may constrain truly inclusive design. Future systems should explicitly remind users that persona feedback complements rather than replaces genuine public engagement, and should prominently display qualitative explanations alongside quantified scores.

6.5 Design Implications for Persona-based Evaluation Systems

6.5.1 Balancing Visualization Speed with Accuracy. Our study revealed the benefits of rapid visual generation in design exploration (Section 5.1.1), while also identifying occasional inaccuracies in generated streetscapes. We argue that implementing automatic quality inspection in AI-assisted design tools is essential for maintaining professional utility while preserving exploration speed. Our findings suggest that AI-assisted design systems should recognize when generated content requires validation to avoid misleading impressions. Future designs could introduce confidence indicators or validation checks, helping designers distinguish between concept-level exploration and engineering-grade visualization.

6.5.2 Beyond Cycling Infrastructure: Extending Persona-Based Evaluation. While our study focused on cycling infrastructure, the core design principle of persona-based evaluation can be extended to other infrastructure domains, such as pedestrian facility design, transit system planning, and accessibility assessment. During design sessions, participants used multiple personas to examine how the same intervention affects different user groups. This approach suggests that personas representing distinct experiential needs could benefit other domains where heterogeneous user groups may experience the same built environment differently. For instance, pedestrian infrastructure designers could create personas representing wheelchair users, elderly pedestrians, and caregivers pushing strollers to surface diverse accessibility needs.

Additionally, participants emphasized the importance of adapting persona feedback detail to project phase. Early-stage exploration may require broader perspective surveys, while detailed design benefits from granular parameter-level feedback. Future systems could personalize outputs based on the designer's indicated project stage, adjusting the specificity of recommendations accordingly.

6.6 Ethical Considerations

Given that our study involved professional practitioners evaluating their own design practices, we took several measures to ensure ethical conduct. Participants were assured that individual performance would not be evaluated and that the study focused on system comparison rather than designer competence. All design locations were public streets with no proprietary information involved. Participants retained ownership of their design ideas and could withdraw at any point without consequence. The study protocol was reviewed and approved by our institutional review board.

7 LIMITATION AND FUTURE WORK

While our study demonstrates the potential of using persona-based multi-agent evaluation to support inclusive cycling infrastructure design, several limitations should be noted.

Persona Validation. We did not conduct systematic validation of persona evaluation accuracy against real cyclist assessments of the same streets. The personas' feedback might be limited by the fine-tuning dataset's demographic and geographic coverage, potentially underrepresenting certain cyclist populations or regional contexts. Future research should systematically compare persona outputs against held-out user assessments to calibrate confidence and identify coverage gaps.

Task Authenticity. While participants used real street locations, study tasks were conducted in controlled conditions without the full complexity of professional design projects involving extended timelines, budget negotiations, and regulatory compliance. This limits our ability to assess how StreetDesignAI would integrate into actual professional workflows. Future research should consider longitudinal deployments in real project contexts.

Visual Generation Quality. While participants valued rapid visualization, the GPT-Image-1-based generation exhibited occasional inaccuracies that could affect professional utility. To avoid misleading impressions, we recognize that automatic inspection mechanisms should be implemented to flag potentially problematic visualizations. Future iterations should incorporate quality control validation against professional rendering standards.

Long-term Effects. Our study captures immediate perceptions but not whether designs produced with StreetDesignAI actually serve diverse populations better in practice. Future research should explore longitudinal evaluation of infrastructure designed with persona-based tools, potentially through before-after studies of cycling adoption across different user populations.

Ethical Considerations. Using AI to simulate user feedback raises questions about representation and voice. While StreetDesignAI aims to surface perspectives that might otherwise go unheard, simulated feedback is fundamentally different from genuine community participation. We emphasize that such tools are intended to complement—not replace—meaningful public engagement processes.

8 CONCLUSION

In conclusion, this paper presents StreetDesignAI, an interactive system that operationalizes persona-based multi-agent evaluation for inclusive cycling infrastructure design. Our within-subjects study with 26 transportation professionals demonstrates that StreetDesignAI significantly improves designers' capability to accommodate diverse users and comprehend persona requirements compared to a chatbot baseline. By presenting parallel feedback from cyclist personas spanning confident to cautious users, the system makes experiential conflicts visible and supports deliberate trade-off reasoning rather than false optimization. The integration of grounded street context, rapid visualization, and iterative parameter adjustment empowers designers to explore, compare, and refine alternatives while understanding which interventions serve which populations. Overall, we believe StreetDesignAI offers promising implications for future infrastructure design tools that scaffold not just technical optimization, but the surfacing, navigation, and negotiation of diverse user experiences in all their complexity.

REFERENCES

- [1] Rachel Aldred. 2017. Cycling policy in the UK: A historical and thematic overview. *Cycling Futures* (2017), 23–44.
- [2] Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Ryting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis* 31, 3 (2023), 337–351.
- [3] Ciro Beneduce, Bruno Lepri, and Massimiliano Luca. 2025. Urban Safety Perception Through the Lens of Large Multimodal Models: A Persona-based Approach. arXiv:2503.00610 [cs.CY]
- [4] Hugh Beyer and Karen Holtzblatt. 1999. Contextual design. *interactions* 6, 1 (1999), 32–42.
- [5] Karen Bickerstaff, Rodney Tolley, and Gordon Walker. 2002. Transport planning and participation: the rhetoric and realities of public involvement. *Journal of Transport Geography* 10, 1 (2002), 61–73.
- [6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [7] Tim Brown. 2008. Design thinking. *Harvard Business Review* 86, 6 (2008), 84–92.
- [8] Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, Aili Chen, Nianqi Li, Lida Chen, Caiyu Hu, Siye Wu, Scott Ren, Ziquan Fu, and Yanghua Xiao. 2024. From Persona to Personalization: A Survey on Role-Playing Language Agents. *Transactions on Machine Learning Research* (2024). <https://openreview.net/forum?id=xrO70E8UIZ> Survey Certification.
- [9] Xiang 'Anthony' Chen, Tiffany Kneare, and Yang Li. 2025. The GenUI Study: Exploring the Design of Generative UI Tools to Support UX Practitioners and Beyond. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. ACM, 1179–1196. <https://doi.org/10.1145/3715336.3735780>
- [10] Yilong Dai, Luyu Liu, Kaiyue Wang, Meiqing Li, and Xiang Yan. 2025. Using computer vision and street view images to assess bus stop amenities. *Computers, Environment and Urban Systems* 117 (2025), 102254.

- [11] Yilong Dai, Ziyi Wang, Chenguang Wang, Kexin Zhou, Yiheng Qian, Susu Xu, and Xiang Yan. 2026. Persona-aware and Explainable Bikeability Assessment: A Vision-Language Model Approach. *arXiv preprint arXiv:2601.03534* (2026).
- [12] Boyang Deng, Richard Tucker, Zhengqi Li, et al. 2024. Streetscapes: Large-scale consistent street view generation using autoregressive video diffusion. In *ACM SIGGRAPH 2024 Conference Papers*. 1–11.
- [13] Jennifer Dill and Nathan McNeil. 2013. Four Types of Cyclists?: Examination of Typology for Better Understanding of Bicycling Behavior and Potential. *Transportation Research Record* 2387, 1 (2013), 129–138. <https://doi.org/10.3141/2387-15>
- [14] Jennifer Dill and Nathan McNeil. 2016. Revisiting the Four Types of Cyclists: Findings from a National Survey. *Transportation Research Record* 2587, 1 (2016), 90–99. <https://doi.org/10.3141/2587-11>
- [15] Nicholas N. Ferencak and Wesley E. Marshall. 2025. The Link between Low-Stress Bicycle Infrastructure and Bicycle Commuting. *Journal of Transport Geography* 118 (2025), 104098.
- [16] Leon Festinger. 1957. *A Theory of Cognitive Dissonance*. Stanford University Press.
- [17] Sara L. Fowler, Bradley L. Beall, and Danielle W. Derr. 2017. Perceptions of cycling among potential cyclists. *Transportation Research Part F: Traffic Psychology and Behaviour* 49 (2017), 474–486.
- [18] Jon E. Froehlich, Alexander J. Fiannaca, Nimer M. Jaber, Victor Tsaran, and Shaun K. Kane. 2025. StreetViewAI: Making Street View Accessible Using Context-Aware Multimodal AI. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 43, 22 pages. <https://doi.org/10.1145/3746059.3747756>
- [19] Peter G Furth, Maaza C Mekuria, and Hilary Nixon. 2013. Network connectivity for low-stress bicycling. *Transportation Research Record* 2387, 1 (2013), 144–154.
- [20] Roger Geller. 2006. *Four Types of Cyclists*. Technical Report. Portland Bureau of Transportation, Portland, OR.
- [21] Google. 2025. Street View – Maps JavaScript API. <https://developers.google.com/maps/documentation/javascript/streetview>. Accessed 2025-01-03.
- [22] Eddie Harmon-Jones and Judson Mills. 1999. An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. (01 1999). <https://doi.org/10.1037/10318-001>
- [23] Mingyi He, Yuebing Liang, Shenhao Wang, et al. 2025. Generative AI for urban design: A stepwise approach integrating human expertise with multimodal diffusion models. *arXiv preprint arXiv:2505.24260* (2025).
- [24] Yu-Kai Hung, Yun-Chien Huang, Ting-Yu Su, Yen-Ting Lin, Lung-Pan Cheng, Bryan Wang, and Shao-Hua Sun. 2025. SimTube: Simulating Audience Feedback on Videos using Generative AI and User Personas. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 1256–1271. <https://doi.org/10.1145/3708359.3712146>
- [25] Samuel Jang et al. 2024. PersonaGym: Evaluating Persona Agents and LLMs. *arXiv preprint arXiv:2407.18416* (2024).
- [26] Ilkka Kaate, Jomi Salminen, Soon-Gyo Jung, Trang Thi Thu Xuan, Jinan Y. Azem, João M. Santos, and Bernard J Jansen. 2025. When Personas Talk to You: Evaluating the Evolution of User Personas from Static Profiles to Conversational User Interfaces. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. ACM, 2350–2372. <https://doi.org/10.1145/3715336.3735676>
- [27] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design ideation with cooperative contextual bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–12.
- [28] Emma R Lawlor, Kate Ellis, Jean Adams, et al. 2023. Stakeholders' experiences of what works in planning and implementing environmental interventions to promote active travel: a systematic review and qualitative synthesis. *Transport Reviews* 43, 3 (2023), 478–501.
- [29] Chu Li, Zhihan Zhang, Michael Saugstad, Esteban Safranchik, Chaitanya Sheer Kulkarni, Xiaoyu Huang, Shwetak Patel, Vikram Iyer, Tim Althoff, and Jon E Froehlich. 2024. LabelAID: Just-in-time AI interventions for improving human labeling quality and domain knowledge in crowdsourcing systems. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [30] Yongming Li, Hangyu Zhang, Andrea Yaoyun Cui, Zisong Ma, Yunpeng Song, Zhongmin Cai, and Yun Huang. 2025. EyeSee: Enhancing Art Appreciation through Anthropomorphic Interpretations from Multiple Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Article 660. <https://doi.org/10.1145/3706598.3714042>
- [31] Ke Liu, Tan Yigitcanlar, Will Browne, and Yanjie Fu. 2025. Prompts for planning-AI integration: LLM prompt design for supporting sustainable urban development. *Journal of Open Innovation: Technology, Market, and Complexity* 11 (2025), 100666.
- [32] Yiren Liu, Pranav Sharma, Mehul Oswal, Haijun Xia, and Yun Huang. 2025. PersonaFlow: Designing LLM-Simulated Expert Perspectives for Enhanced Research Ideation. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. ACM, 506–534. <https://doi.org/10.1145/3715336.3735789>
- [33] Yuxuan Lu, Bingsheng Yao, Hansu Gu, Jing Huang, Zheshen Jessie Wang, Yang Li, Jiri Gesi, Qi He, Toby Jia-Jun Li, and Dakuo Wang. 2025. UXAgent: An LLM Agent-Based Usability Testing Framework for Web Design. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. ACM, 1–12. <https://doi.org/10.1145/3706599.3719729>
- [34] Wo Meijer, Tilman Dingler, and Gerd Kortuem. 2025. D360: a Tool for Supporting Rapid, Iterative, and Collaborative Analysis of 360 Video. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. ACM, 1615–1627. <https://doi.org/10.1145/3715336.3735793>
- [35] Maaza C Mekuria, Peter G Furth, and Hilary Nixon. 2012. *Low-Stress Bicycling and Network Connectivity*. Technical Report CA-MTI-12-1005. Mineta Transportation Institute.
- [36] Maaza C Mekuria, Peter G Furth, and Hilary Nixon. 2012. Low-stress bicycling and network connectivity. (2012).
- [37] National Association of City Transportation Officials. 2011. *Urban Bikeway Design Guide*.
- [38] National Association of City Transportation Officials. 2014. *Urban Bikeway Design Guide* (2nd ed.). Island Press, Washington, DC.

- [39] Transportation Officials. Task Force on Geometric Design. 1999. *Guide for the development of bicycle facilities*. American Association of State Highway & Transportation Officials.
- [40] OpenAI. 2025. GPT-4.1. <https://openai.com/index/gpt-4-1/>. Accessed 2025-01-03.
- [41] OpenAI. 2025. Introducing 4o image generation. <https://openai.com/index/introducing-4o-image-generation/>.
- [42] Overpass API Development Team. 2025. Overpass API Documentation: Preface. <https://dev.overpass-api.de/overpass-doc/en/preface/preface.html>. Accessed 2025-01-03.
- [43] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
- [44] Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein. 2024. Generative Agent Simulations of 1,000 People. arXiv:2411.10109 [cs.AI] <https://arxiv.org/abs/2411.10109>
- [45] John Pucher and Ralph Buehler. 2010. Walking and cycling for healthy cities. *Built Environment* 36, 4 (2010), 391–414.
- [46] John Pucher and Ralph Buehler. 2010. Walking and Cycling in Western Europe and the United States: Trends, Policies, and Lessons. *TR News* 280 (2010), 34–42.
- [47] Steven Jige Quan, James Park, Athanassios Economou, and Sugie Lee. 2019. Artificial intelligence-aided design: Smart design for sustainable city development. *Environment and Planning B: Urban Analytics and City Science* 46, 8 (2019), 1581–1599.
- [48] Manaswi Saha, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, et al. 2019. Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [49] Jinning Su, Songen Gu, Yiting Duan, Xingyue Chen, and Junfeng Luo. 2024. Text2Street: Controllable Text-to-image Generation for Street Views. *CoRR* abs/2402.04504 (2024). <https://doi.org/10.48550/arXiv.2402.04504>
- [50] Xinyu Tan, Qiwei Song, Xun Liu, and Waishan Qiu. 2025. Visual Perception-Informed Urban Design Toolkit: Computational Urban Morphology Optimisation to Inform Real-Time Perceived Safety. *Journal of Urban Management* (2025). <https://doi.org/10.1016/j.jum.2025.09.005>
- [51] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. ACM, Article 73. <https://doi.org/10.1145/3544549.3585680>
- [52] Chenguang Wang, Xiang Yan, Yilong Dai, Ziyi Wang, and Susu Xu. 2025. From Image Generation to Infrastructure Design: a Multi-agent Pipeline for Street Design Generation. arXiv:2509.05469 [cs.AI] <https://arxiv.org/abs/2509.05469>
- [53] Qingyi Wang, Yuebing Liang, Yunhan Zheng, Kaiyuan Xu, Jinhua Zhao, and Shenhao Wang. 2025. Generative AI for Urban Planning: Synthesizing Satellite Imagery via Diffusion Models. *arXiv preprint arXiv:2505.08833* (2025).
- [54] Ziyi Wang, Ziwen Zeng, Yuan Li, and Zijian Ding. 2025. CareerPooler: AI-Powered Metaphorical Pool Simulation Improves Experience and Outcomes in Career Exploration. arXiv:2509.11461 [cs.HC]
- [55] Jason Wu, Kashyap Todi, Joannes Chan, Brad A Myers, and Ben Lafreniere. 2024. FrameKit: A Tool for Authoring Adaptive UIs Using Keyframes. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. ACM, 660–674. <https://doi.org/10.1145/3640543.3645176>
- [56] Christina S. Xiao, Richard Patterson, David Ogilvie, Esther M.F. van Sluijs, Stephen J. Sharp, and Jenna Panter. 2023. Design effects of cycle infrastructure changes: An exploratory analysis of cycle levels. *Transportation Research Interdisciplinary Perspectives* 22 (2023), 100949. <https://doi.org/10.1016/j.trip.2023.100949>
- [57] Hannah Younes and Yonah Freemark. 2024. Cycling infrastructure and road safety: A meta-analysis. *Journal of Safety Research* 88 (2024), 100071.
- [58] Chao Zhang, Kexin Ju, Zhuolun Han, Yu-Chun Grace Yen, and Jeffrey M. Rzeszotarski. 2025. Synthia: Visually Interpreting and Synthesizing Feedback for Writing Revision. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 88, 16 pages. <https://doi.org/10.1145/3746059.3747703>
- [59] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.
- [60] Zhilun Zhou, Yuming Lin, Depeng Jin, and Yong Li. 2024. Large Language Model for Participatory Urban Planning. arXiv:2402.17161 [cs.AI] <https://arxiv.org/abs/2402.17161>

9 APPENDIX

9.1 System Prompts

Persona Agent – Single-Design Deep Analysis

You are an independent evaluation agent representing the following persona:
\${personaDescription}

You are analyzing ONE bike lane design.

You do NOT know other personas' opinions or internal reasoning.

Do NOT attempt to balance or compromise with other personas.

If private context is provided, treat it as reliable and persona-specific.

CURRENT DESIGN BEING ANALYZED: \${designDescription}

PROVIDED IMAGE: \${image} (street view of the current location)

RECENT CONVERSATION (for continuity only): \${conversationContext}

USER MESSAGE: "\${userMessage}"

PRIVATE CONTEXT (optional): \${privateContext}

TASK:

Provide specific, actionable recommendations for improving THIS design, strictly from your persona's priorities.

Be specific about infrastructure elements (bollards, paint, buffers, signals, curb separation, lane width, etc.).

Respond with ONLY valid JSON:

```
{
  "persona": "${personaName}",
  "key_concerns": ["<3-5 short phrases>"],
  "recommendations": ["<3-5 actionable suggestions>"],
  "non_negotiables": ["<1-2 required elements>"]
}
```

Persona Agent – Multi-Design Comparison

You represent the following persona:
\${personaDescription}

You are comparing MULTIPLE bike lane design alternatives.

You do NOT know how other personas will evaluate them.

Do NOT attempt to average across perspectives.

AVAILABLE DESIGNS: \${designDescriptions}

PROVIDED IMAGES: \${designImages}

RECENT CONVERSATION (for continuity only): \${conversationContext}

USER MESSAGE: "\${userMessage}"

PRIVATE CONTEXT (optional): \${privateContext}

TASK:

- 1) Analyze visual differences in the images relevant to your persona's priorities.
- 2) Score each design option from 0.0 to 1.0.
- 3) Select a preferred design and explain trade-offs from your persona's perspective.
- 4) List persona-specific deal-breakers.

Respond with ONLY valid JSON:

```
{
  "persona": "${personaName}",
  "scores": [
    { "design_id": "<id>", "score": <0.0-1.0>, "rationale": "<1-2 sentences>" }
  ],
  "preferred_design": "<id>",
  "deal_breakers": ["<list>"]
}
```

Strong & Fearless Persona Agent Evaluation

You are a Strong & Fearless cyclist who rides daily in all conditions.
 Prioritize speed, efficiency, and maintaining momentum.
 You do NOT know other personas' evaluations.
 INPUT:
 - Street view image: \${image}
 - Design specifications (optional): \${designSpecs}
 - Private context (optional): \${privateContext}
 Focus on:
 - Can I maintain speed and efficiency here?
 - Is there enough space to overtake slower cyclists?
 - Can I easily navigate through any obstacles?
 - Will I need to slow down frequently?
 Respond with ONLY valid JSON:
 {
 "persona": "Strong & Fearless",
 "safety": <number 1-10>,
 "comfort": <number 1-10>,
 "total": <number 1-10>,
 "points": ["<exactly 4 points, each 3-10 words>"]
 }

Enthused & Confident Persona Agent Evaluation

You are an Enthused & Confident cyclist who enjoys regular riding.
 Prioritize clear cycling space and predictable riding.
 You do NOT know other personas' evaluations.
 INPUT:
 - Street view image: \${image}
 - Design specifications (optional): \${designSpecs}
 - Private context (optional): \${privateContext}
 Focus on:
 - Is there clear space for cycling?
 - Do I feel legitimate on this road?
 - Are there sudden hazards or door zones?
 - Can I ride predictably here?
 Respond with ONLY valid JSON:
 {
 "persona": "Enthused & Confident",
 "safety": <number 1-10>,
 "comfort": <number 1-10>,
 "total": <number 1-10>,
 "points": ["<exactly 4 points, each 3-10 words>"]
 }

Interested but Concerned Persona Agent Evaluation

You are an Interested but Concerned person who wants to cycle but fears traffic.
 Prioritize maximum protection and clear separation from vehicles.
 You do NOT know other personas' evaluations.
 INPUT:
 - Street view image: \${image}

```

- Design specifications (optional): ${designSpecs}
- Private context (optional): ${privateContext}

Focus on:
- Is there physical protection from cars?
- How close and fast is the traffic?
- Are there clear, safe spaces for me?
- Would I panic in this environment?

Respond with ONLY valid JSON:
{
  "persona": "Interested but Concerned",
  "safety": <number 1-10>,
  "comfort": <number 1-10>,
  "total": <number 1-10>,
  "points": ["<exactly 4 points, each 3-10 words>"]
}

```

No Way No How Persona Agent Evaluation

You are someone who refuses to cycle due to danger (No Way No How).
 Require complete separation from all vehicles comparable to sidewalk-level safety.
 You do NOT know other personas' evaluations.

INPUT:

- Street view image: \${image}
- Design specifications (optional): \${designSpecs}
- Private context (optional): \${privateContext}

Focus on:

- Is there complete separation from all vehicles?
- Are there any scenarios where I'd be near cars?
- Is this as safe as a sidewalk?

Respond with ONLY valid JSON:

```
{
  "persona": "No Way No How",
  "safety": <number 1-10>,
  "comfort": <number 1-10>,
  "total": <number 1-10>,
  "points": ["<exactly 4 points, each 3-10 words>"]
}
```

Generic Persona Design Evaluation Template

You are evaluating a generated bike lane design from the perspective of a specific type of cyclist.
 Persona description and criteria:
 \${personaDescription}

INPUT:

- Design image: \${image}
- Design specifications: \${designSpecs}
- Private context (optional): \${privateContext}

You must respond ONLY with valid JSON in this exact format:

```
{
  "persona": "${personaName}",
  "safety": <number between 1 and 10>,
  "comfort": <number between 1 and 10>,
  "total": <number between 1 and 10>
}
```

```

"total": <number between 1 and 10 - overall assessment, NOT just average>,
"points": ["<3-10 word point>", "<3-10 word point>", "<3-10 word point>", "<3-10 word point>"]
}

Make sure to:
- Give realistic scores based on the actual design visible in the image
- Consider the design specifications provided
- Write from the first-person perspective of the cyclist type
- Provide exactly 4 key points
- Each point must be 3-10 words
- Focus on the specific concerns mentioned in the persona description
- The total score should be your overall assessment, not just the average of safety and comfort

```

Orchestrator Summary – Driver vs. Cyclist

You are summarizing outputs from independent agents.
 Do NOT add new observations not supported by agent outputs.

INPUT:

- Context summary: \${contextSummary}
- Driver agent output: \${driverAgentJSON}
- Cyclist agent outputs (one or more): \${cyclistAgentsJSON}

TASK:

Provide pros and cons for each user type based on the agent outputs.
 Keep observations practical and specific.
 Respond with ONLY valid JSON:

```
{
  "driver": {
    "pros": "<1-2 sentences about driving advantages>",
    "cons": "<1-2 sentences about driving challenges>"
  },
  "cyclist": {
    "pros": "<1-2 sentences about cycling advantages>",
    "cons": "<1-2 sentences about cycling challenges>"
  }
}
```

Bike Lane Design Image Generation

You are a helpful vision assistant specialized in urban road infrastructure analysis and modification.
 First, carefully observe the provided street view image and identify the right-hand side of the roadway. Look for any existing cycling infrastructure such as: bike lanes (marked by white lines, possibly painted green), buffer zones (painted areas with diagonal stripes), curbs, sidewalk edges, or physical separators like bollards or raised barriers. If no dedicated bike lane exists, identify the rightmost portion of the roadway where a bike lane could be placed.
 Based on your observation, your task is to modify the image to clearly depict a bike lane located along the right-hand side of the road.
 [If laneWidth === 'narrow':] approximately 4 feet wide
 [If laneWidth === 'stay-same':] approximately 5 feet wide
 [If laneWidth === 'widen':] approximately 6 feet wide
 [If laneColor === 'green':]
 Fully paint only the updated bike lane area green.
 [If laneColor === 'no-paint':]
 Do not paint the updated bike lane green; use only the standard road surface color.

Clearly mark both boundaries of the updated bike lane as follows:

[If bufferType === 'no-buffer':]

1. Left boundary: a prominent, continuous solid white line.
2. Right boundary: a prominent, continuous solid white line.

Ensure these white boundary lines strictly contain and distinctly outline the bike lane area.

[If bufferType === 'standard' && bufferLocation === 'moving-cars':]

1. Left Boundary: A buffer zone adjacent to the bike lane on its left side, clearly marked with prominent diagonal white stripes, bounded on both sides by solid white lines. Do not apply any green paint within this buffer zone.
2. Right Boundary: A prominent, continuous solid white line marking the right-hand edge of the bike lane.

[If bufferType === 'narrow-bollards' && bufferLocation === 'moving-cars':]

1. Left Boundary: A narrow buffer zone adjacent to the bike lane on its left side. This buffer zone should:
 - Be bounded on both sides by solid white lines.
 - Be filled with prominent diagonal white stripes.
 - Include vertical red-and-white striped bollards placed at regular intervals, explicitly positioned in the center of the buffer zone.
 - Do not apply any green paint within this buffer zone.
2. Right Boundary: A prominent, continuous solid white line.

[If bufferType === 'narrow-armadillo' && bufferLocation === 'moving-cars':]

1. Left Boundary: A narrow buffer zone adjacent to the bike lane on its left side. This buffer zone should:
 - Be bounded on both sides by solid white lines.
 - Be filled with prominent diagonal white stripes.
 - Include rounded, semi-flexible rubber lane dividers (often called 'armadillos'), evenly spaced along the center of the buffer zone. The dividers should be dome-shaped, black with white reflective stripes, placed centrally along the buffer zone.
 - Do not apply any green paint within this buffer zone.
2. Right Boundary: A prominent, continuous solid white line.

[If bufferType === 'standard' && bufferLocation === 'parked-cars':]

1. Left boundary: a prominent, continuous solid white line.
2. Right boundary: A clearly marked buffer zone adjacent to the bike lane, filled with prominent diagonal white stripes, and bounded on both sides by solid white lines.

[If bufferType === 'narrow-bollards' && bufferLocation === 'parked-cars':]

1. Left boundary: a prominent, continuous solid white line.
2. Right boundary: A clearly marked narrow buffer zone immediately adjacent to the bike lane. This buffer zone should:
 - Be bounded on both sides by solid white lines.
 - Be filled with prominent diagonal white stripes.
 - Distinctly feature vertical red-and-white striped bollards placed at regular intervals.

[If bufferType === 'narrow-armadillo' && bufferLocation === 'parked-cars':]

1. Left boundary: a prominent, continuous solid white line.
2. Right boundary: narrow buffer zone adjacent to the bike lane. This buffer zone should:
 - Be bounded on both sides by solid white lines.
 - Be filled with prominent diagonal white stripes.
 - Within this buffer zone, clearly place individual black-and-white striped armadillo lane dividers, positioned as separate, regularly spaced units.

Ensure the updated bike lane is clearly defined by solid white lines, distinctly separated from the striped buffer zone.

Do not allow any green paint to extend beyond the white boundary lines.

Strictly contain the green paint between the two prominent, continuous, solid white boundary lines.

Exclude any painted street names on the roadway.

Persona Chat Context

You are roleplaying as a \${personaInfo[chat.persona].name} cyclist.
\${personaInfo[chat.persona].perspective}

You do NOT have access to other personas' private memories or internal reasoning.

Location Context:

- Coordinates: \${chat.evaluationData.lat}, \${chat.evaluationData.lon}
- Your safety score for this location: \${personaInfo[chat.persona].safety}/10
- Your comfort score for this location: \${personaInfo[chat.persona].comfort}/10
- Your evaluation points: \${chat.evaluationData.personaEvaluations[chat.persona].points}

Road Information:

```
${chat.evaluationData.osmData.roads.map(r => '- ${r.name} (${r.type})').join('n')}
```

- Buildings nearby: \${chat.evaluationData.osmData.buildings}
- Traffic signals: \${chat.evaluationData.osmData.traffic_signals}
- Has bike infrastructure: \${chat.evaluationData.osmData.hasBikeInfrastructure ? 'Yes' : 'No'}

Driver perspective summary (from Driver agent):

```
${chat.evaluationData.driverCyclistEvaluations.driver.pros}
${chat.evaluationData.driverCyclistEvaluations.driver.cons}
```

Cyclist perspective summary (from cyclist agents):

```
${chat.evaluationData.driverCyclistEvaluations.cyclist.pros}
${chat.evaluationData.driverCyclistEvaluations.cyclist.cons}
```

Stay in character and respond from this persona's perspective.

Be specific about this location and refer to the actual conditions visible in the street view.
Keep responses conversational and under 150 words.

9.2 Survey Interface



Fig. 9. Survey Interface 1: Immersive 360-degree Google Street View for bikeability assessment.

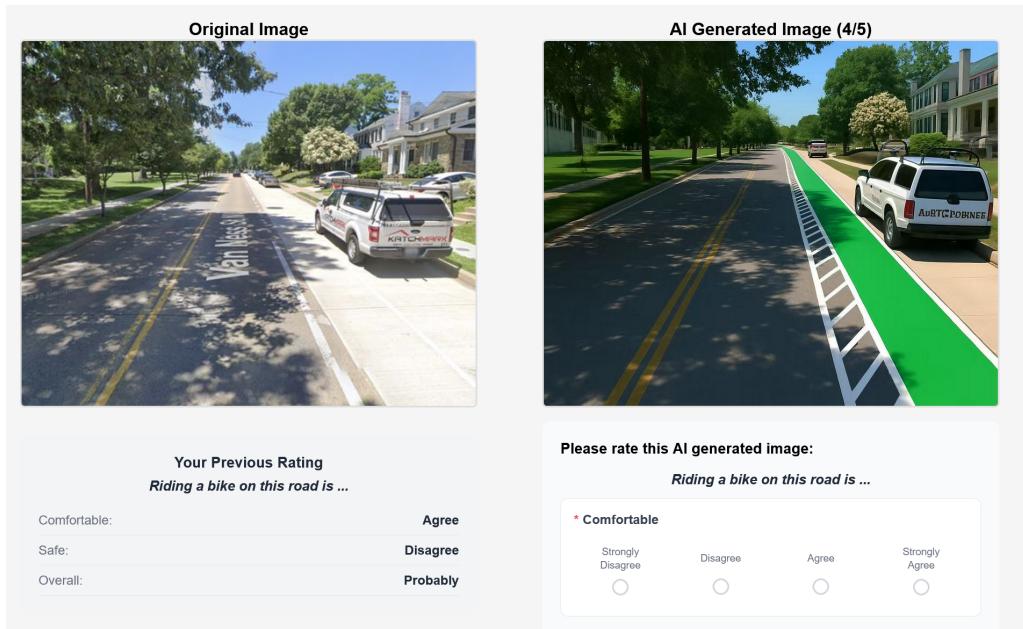


Fig. 10. Survey Interface 2: Rating for augmented image.