Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rehearsing Urban Futures: Generative Agents and Daily Life Narratives

Leticia Izquierdo

leticiai@mit.edu

Massachusetts Institute of Technology

Jason Nawyn

Massachusetts Institute of Technology

Patricia Maes

Massachusetts Institute of Technology

Kent Larson

Massachusetts Institute of Technology

Article

Keywords:

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10884908/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

We describe an interactive approach to translating local, lived knowledge into actionable urban planning models. This combines two layers of generative simulation: a narrative layer that transforms resident testimony into daily-life scenarios, and a spatial layer that uses generative agents to simulate pedestrian activity. Participants can rehearse urban futures, observe the agents' natural-language reasoning, and directly contest outcomes that do not align with their lived experience. Through a large-scale installation in San Francisco and field evaluations with 30 residents across San Francisco and Boston, we analyzed interactions derived from planner interviews alongside resident agents built from participants' own testimony. Our findings highlight how narrative and agent interaction can reshape public deliberation by expanding residents' sense of involvement, and helping shift their proposals from incremental fixes (like lighting or benches) toward envisioning complex futures that current planning tools are not calibrated to evaluate. We also found that an agent's usefulness for elicitation depends less on its behavioral fidelity than its capacity to provoke correction: when an agent was visibly wrong, residents challenged it and provided precisely the nuanced knowledge the model lacked. Based on these results, we position the generative agent as a boundary object between a resident’s memory and a planner’s model, useful precisely because it can be argued with. Taken together, these results suggest that generative simulations can shift public participation from merely evaluating short-term interventions to actively rehearsing systemic urban futures.

1. Background

Both problems, illegible instruments and pre-determined future scenarios, have deep roots, and two research traditions have addressed them in part. The first approach is top-down. Modern urbanism began as the Industrial Revolution moved people from villages to cities, bringing overcrowding, poor sanitation, and epidemics such as cholera [2, 3]. Governments responded by imposing order from above: public health laws, housing ordinances, minimum street widths. This context birthed some of the earliest city mapping. In one famous example, maps were drawn by Dr. John Snow in 1854 to trace the source of cholera outbreaks in London to specific water pumps [4]. As cities grew, mapping shifted from tracking diseases to modeling urban processes. In 1964 Ira S. Lowry developed the Metropolis Model in Pittsburgh [5], combining economic theory with the gravity treatment of spatial interaction. This approach to predictive modeling reached its peak in the 1970s with Jay Forrester's Urban Dynamics at MIT [6]. Researchers simulated individuals with cellular-level behaviors in digital models to see whether complex growth patterns would emerge from simple rules, following in the footsteps of John von Neumann's *Theory of Self-Reproducing Automata* (1966) [7]. This birthed generative social science [8] and agent-based modelling [9].

But top-down urbanism did not always create communities that worked. Plans drawn from above could deliver housing, roads, and parks, yet they often missed what makes a neighborhood work: the everyday routines, relationships, and small practical realities that only residents know. The Pruitt-Igoe housing complex in St. Louis (1954–1972) is the classic example of this disconnect: it solved for sanitation but

failed to stimulate urban life. It lacked the commercial opportunities, the cultural venues, and practical amenities people needed to build a workable community [10]. In response to the failures of top-down urban renewal, a bottom-up approach emerged from the insights of activists such as Jane Jacobs. In *The Death and Life of Great American Cities* (1961) Jacobs saw street life dying as cities scrambled to make room for more automobiles [11]. In response, she advocated for a community-centric urbanism intended to prioritise human needs and relationships. Meanwhile, architectural theorist Christopher Alexander [12] argued that traditional planning rules, like "3 square meters of green space per person", failed to capture what actually makes spaces work. In *A Pattern Language* (1977), he collected over 250 patterns of public spaces that worked, gathered from people through their stories and experiences of community [12]. His most famous example is a newsstand next to a traffic light, which brings together people scanning headlines while waiting to cross the street.

The independent advantages of top-down and bottom-up urbanism were brought together in the work of N. John Habraken, a Dutch architectural theorist known for his 1961 manifesto, *Supports: An Alternative to Mass Housing* [13]. Ultimately he articulates a practical compromise: build the structural skeleton from above, then let residents fill it in from below, "you need this top-down structure to allow, then, the bottom-up infill."

The unifying idea that citizens and experts should design their own spaces, was eventually put into practice with tools like the Planning for Real™ platform (1977) by Tony Gibson [14]. These methods have gained such acceptance that international funding agencies such as the World Bank now require some form of direct community engagement to be included in successful proposals. An ecosystem of methods has emerged to facilitate community participation: interviews [15], which allow researchers to reconstruct the unwritten rules that govern street interaction; walk-alongs [16, 17], which let the environment itself bring up the topics that matter to the walker; urban sensors [18–20] and other quantitative techniques for collecting data at different levels of granularity [21–25]. The goal is to integrate across these information sources to increase explainability about what happens at street level, how spaces work, and what could be improved, according to residents' own experience of the city [26].

Nevertheless, an implicit critique has often accompanied these well-intentioned approaches: do they actually work? Does anything change, or is this all lip service? Community-led methods also share a structural limit to their scale. How many people can you actually reach? Is the sample truly representative? In response to these critiques, researchers have attempted to embed bottom-up methods within top-down processes to recruit the advantages of scalability without sacrificing individual-scale insights. One example of this approach involves agent-based simulation, a computational technique that recreates complex systems by analyzing and reproducing the independent actions of individual, autonomous entities known as agents. Izquierdo et al. (2021), for example, built agent profiles from interviews with the population of Lomas del Centinela, in Mexico, and tested possible outcomes of community interventions based on the collective behavior of these agents [27]. However, these models remained limited by the rules written into them: the action space of a simulation was whatever had been coded by hand; the representative profiles were few; the behavioral repertoire of each

agent was limited; and the agentic rules were fixed at the start of the simulation, so no learning could occur. Practically, this means they could only ever evaluate interventions within the range of prior observation.

This fundamental limitation is inherent to all closed systems, but it can be worked around with clever engineering that draws from the much larger reasoning space encoded in massive data models like LLMs. A relevant example of this approach is Social Simulacra and Generative agents: Interactive simulacra of human behavior. by Park et al. (2022) [28, 29], where believable social behavior emerges in a small virtual city without any programmed rules. In their latest work on scaling this approach, these researchers conducted one thousand two-hour interviews to create one agent per person [30]. The agents were then able to anticipate participants' answers two weeks later with eighty-five percent accuracy.

So, would a synthetic participant ever be a legitimate part of an urban planning intervention? Three objections would likely arise. First, flattening: agents trained on averages smooth away the very differences of gender, age, class, neighborhood, that planning exists to negotiate [31]. Second, the illusion of consensus: if you simulate deliberation, you can manufacture agreement that no real community ever gave [32]. Third, representation without presence: an agent speaks for a person who is not in the room, exactly the failure mode participatory planning was invented to fix [33]. Community-scale simulation requires deeper social grounding, and humans must stay in the process.

Michael Batty, a Professor of Planning at University College London, offers a reasonable perspective on this: think of generative models not as prediction, but as rehearsal for collective imagination and decision-making. Across his work—from his 2001 paper with Paul Torrens [34] to *The Computable City* [5]—Batty argues that computational models are decision-making tools for uncertainty, not predictors of fixed outcomes. As he notes in *Inventing Future Cities* [36], models exist to actively invent desired futures rather than discover predetermined ones.

The way these models are represented computationally is difficult to understand for anyone who did not build them. HCI researchers have worked on this problem for decades. The root of communicating complex city plans in a tangible, accessible manner traces back to Patrick Geddes and his pioneering 1892 'Outlook Tower'—a six-story civic observatory in Edinburgh designed as a visual laboratory to help citizens understand their city within its wider regional context [37, 38]. In the 1970s, Nicholas Negroponte, then director of the Architecture Machine Group (AMG) at MIT, proposed having interactive conversations with machines so that humans and machines could iterate collaboratively on architectural designs [39]. Chris Schmandt, who would later establish the Speech Interface Group at the MIT Media Lab, moved the concept into an immersive environment with the Put-That-There interface [40, 41], demonstrating how humans could communicate more naturally with computers by combining voice recognition and hand gestures simultaneously. Around the same time, another AMG researcher, Andrew Lippman (now director of Viral Communications at the Media Lab), launched the Aspen Movie Map project [42, 43], which filmed every street in Aspen, allowing you to virtually drive through the city on a

touchscreen. This research legacy inspired more recent work at the MIT Media Lab, such as the mixed-reality CityScopeAR and The Rehearsal Machine [44, 45], which positions the participant at street level to evaluate a future vision from within, and keep iterating on it until it works.

To our knowledge, no existing tool combines all of these objectives at once—keeping real residents' voices at the center of the simulation, and making its outputs legible enough for those same residents to deliberate over.

2. Neighborhood Platform

Our exploration began with the most direct record available of what a neighborhood wants: interviews with its residents. We leveraged interviews about a Family Zoning proposal collected by the San Francisco Planning Department's outreach programs along the Geary–Park Presidio Boulevard corridor in the Inner Richmond in 2025 (Fig. 1). The resulting Neighborhood Platform therefore builds on testimony from a real planning process that had already begun, but was not being operationalized beyond summary reports. From those transcripts we extracted micro-stories: narrative fragments describing the daily frictions of displacement pressure, transit gaps, and school commutes. A large language model (Claude 3.7 Sonnet, Anthropic, accessed March 2025) drafted a Day-in-the-Life scenario from each set of fragments, and the authors checked every draft against the source transcript. The result is a set of visual storyboards of personas grounded in real testimony (Fig. 2).

Three personas anchor this platform, each embodying one of the dimensions measured by the system: a school teacher whose route depends on bus frequency (mobility), a local business owner watching her block transform (housing), and a longtime resident negotiating access to shops and services (amenities). The personas are synthesized from the interview transcripts described above, so their trajectories are not generic: they carry the specific frictions residents actually reported.

Each persona also became an agent in an agent visualization of the corridor (Fig. 3). The model was built from the corridor's street network, land uses, and transit stops, and assigned each agent a daily routine drawn from its micro-stories: where it lives, works, shops, and how it moves between them. The model therefore reproduces, at street level, the routines the interviews described. In parallel, we read the transcripts a second time for what changes the residents requested, and grouped those desires into a small set of interventions (for example, more ground-floor commerce, higher bus frequency). Each intervention was translated into a change in the model and into a revised storyboard for the affected personas, so that the same testimony that produced the present also produced a future scenario.

We brought both states together in an installation designed for group deliberation. The installation is large-scale and built for groups: four screens arranged in two pairs. The left pair shows the corridor as it is, with ABM pedestrian flows below, Day-in-the-Life storyboards above, both grounded in what residents described in the interviews. The right pair shows the corridor as those same residents said they wished it to be: the desires recorded in the transcripts, more local commerce, better transit, safer streets, translated into interventions, re-simulated in the ABM flows below, and rendered as the personas'

transformed days above (Fig. 4). Each pair thus shows the same corridor in two registers, the aggregate in the flows and the individual in the stories, following an idea the Aspen Movie Map explored decades earlier: a place is better understood when several ways of seeing it are held together. The present and possible future face each other, and both are built from the neighborhood's own words.

This approach draws heavily on the thought experiment developed by John Bordley Rawls (1921–2002), an influential 20th century American political philosopher. Renowned for his seminal work *A Theory of Justice* (1971), Rawls famously introduced the concept of the "veil of ignorance" to conceptualize a society structured around absolute fairness rather than self-interest [46]. The point of this arrangement is collective understanding: seeing what a policy does to people who are not you. Here, residents do not forget who they are, but they are asked to step into the shoes of someone else on the same corridor.

3. Resident Platform

Whereas the Neighborhood Platform presented thus far asked residents to evaluate the city through personas built from other people's testimony, the second part of the project, the Resident Platform, inverts this: residents supply their own testimony, and the system simulates it. The end user becomes the protagonist, represented by an agent grounded in their self-report.

3.1 From Interview to Resident Agent

The session begins with an interview conducted by the platform itself. The resident is asked about their commute, their preferences, and their daily routine. From those answers, the system constructs a resident agent that walks the streets the resident actually walks. Testimony and simulation, two separate components of the Neighborhood Platform, here collapse into a single session: the resident first describes their daily life, and by the end of the session the agent is simulating it on the screen. This design was practical in nature. We wanted to reach the San Francisco residents that other outreach programs had not, so the platform had to run on simple hardware in order to travel to wherever people were.

Under the hood, the system runs on LLM-augmented agent-based modeling: classical agent-based simulation, but with a language model as the engine of each agent. Traditional ABM agents follow fixed behavioral rules, a limitation that has constrained urban simulation for decades. To address this limitation, our agent instead reasons in natural language about each scene along its route, and writes that reasoning down where the resident can read it, and express their agreement or disagreement.

The system takes three inputs (the interview transcript, Street View imagery along the resident's route, and census data for the surrounding blocks) and processes them through two modules (the agent and an image editor), producing three outputs: modified street imagery, agent trajectories, and well-being scores, using StreetScore—a computer vision algorithm designed to estimate how humans visually perceive urban attributes such as safety, beauty, liveliness, and wealth [47](Fig. 5).

3.2 Proposing and Simulating Interventions

Once the agent exists, the loop is simple. The resident can pause the agent's walk at any point. For example, they might stop at a corner that has always bothered them, and say what they would change about it in plain language: "remove cars, add greenery," or "clean debris, add an outdoor cafe." The system takes the actual Street View image of that point on the resident's route and edits it with an instruction-following image model: the cars are gone, the greenery is in, the cafe has tables on the sidewalk, but the buildings, the light, and the geometry of the street remain their street (Fig. 6). The resident is not looking at an artist's rendering of a generic improvement. They are looking at a photograph of their own corner, changed only in the way they asked.

The resident then sees their entire route twice, side by side: the sequence of street images as it is today, and the same sequence with their changes applied. Below, their agent's response unfolds along both. If the modified street changes how the agent walks, a block it now prefers, a detour it no longer takes, the new path appears on the map (Fig. 7). And along the full length of the route, the well-being indicators are plotted as continuous lines.

This sequence of street views, traversed by an agent on the resident's behalf, is a form of what the Aspen Movie Map called surrogate travel. That system let a viewer drive through a town they had never visited, and its designers found that with repeated exposure people built an internal model of the place: they recognized corners, anticipated turns, and could later find their way in the real Aspen. Our platform runs this logic in reverse. The resident already holds that internal model; it is the agent that must match it. We therefore worked with long-time residents, people who have walked their routes for many years, because the comparison between what the agent does and what the resident knows is only informative when the resident's knowledge is rich. Every time the agent takes a turn the resident would not, or lingers somewhere she would hurry past, the gap is a measurement of the system, not of her.

Data Export

Part of the argument for this work is that existing tools of planning are esoteric and opaque, whereas these platforms are accessible and transparent. However, due to administrative inertia, it is likely that it will take a long time for municipal governance to catch up. So in the meantime, providing a pipeline from these platforms to existing planning systems is a necessary interim step, ensuring that community proposals can be exported directly into the standard formats that planners already use. To achieve this, the system captures block-level citizen proposals; exporting geolocations, intervention types, and before/after metrics as GeoJSON, while strictly excluding personal profile data.

3.3 Field Evaluation

We initially deployed the Resident Platform in San Francisco's Inner Richmond, hosting a workshop at the Internet Archive with a representative from the SF Planning Department [48] (Fig. 10). To reach a robust

sample size of thirty long-time residents and validate the approach across a fundamentally different urban fabric, we subsequently deployed the platform in Boston. We recruited people who knew their daily routes well. Ages ranged from 22 to 57. The questionnaire was used only in Boston; 20 of the 22 completed it (n = 20; 14 men, 6 women; ages 23–40, M = 27.9, SD = 5.3).

MIT COUHES determined the study exempt. All participants gave written informed consent at the start of the session, including anonymized data use. Separate consent covered identifiable images in figures; other faces were blurred.

Sessions were individual, with a researcher present (~ 25 min): (1) consent; (2) intake through the platform interface; mobility patterns, demographics, and open-ended preferences, recorded in Google Sheets; (3) the system built the resident agent from these inputs; (4) the participant watched the agent walk their route, proposed changes in plain language, and saw its reasoning, new path, and scores (the platform ran on a research server that stored model outputs to edit images and recompute routes); spoken comments were recorded as field notes; (5) AFP questionnaire (Supplementary Appendix A).

Each session produced a timestamped archive: (i) the agent's record at each decision point; image pair, reasoning, nine tag scores, route choice, in CSV/JSON [50]; (ii) the participant's own image choices and two tags per intersection; (iii) the original and edited route image sequences. Questionnaire answers were stored separately.

The agent is an LLM digital twin built with the two-step prompt described in ref. 50 (full prompts in Supplementary, adapted from ref. 50). Step 1 sets the identity: an embodied persona template, injected demographics (gender, age, income, city familiarity), the participant's preferences verbatim, and a memory module distinguishing city-familiar from unfamiliar agents. Step 2 is the task: the agent compares Street View image pairs on six Place Pulse dimensions (safety, beauty, liveliness, wealth, boringness, depressingness), writes 1–2 sentences of first-person reasoning, scores nine tags (1–7) keeping the top two as stated reasons, and picks a route. Model: Agent reasoning used GPT-4o mini (gpt-4o-mini, OpenAI) via the OpenAI API, with default sampling parameters (temperature = 1.0, not explicitly set), accessed between September 2025 and March 2026.

Interventions were plain-language edit instructions ("remove the cars," "add green infrastructure") applied to one point or the whole route with an instruction-following image model [VERIFY: which], which changed only what was asked and preserved geometry and lighting. Invalid image pairs (identical views, rendering errors, misalignment) were removed [ref. 50]; overlapping front/rear views remain a known limitation. After each edit, the agent re-walked the route with the same profile. Both original and edited sequences were scored on the six perceptual dimensions with a computer-vision model trained on Place Pulse data [22], producing the well-being curves shown to participants.

AFP questionnaire (Supplementary A), four subscales rated 1–7: Behavioral Mirroring (5 items, one reverse-coded), Priority Alignment (2), Recognition & Ownership (3), Trust & Delegation (3). General trust items were adapted from Jian et al. (2000); the trust checklist was shortened from eight to four items

after the first five sessions, and analyses used the four common items. Open questions collected reflections.

Our evaluation focuses on the human audit of a digital surrogate. We place the agent in the simulated world and ask the resident to review and correct its choices, using those very corrections to uncover the latent constructs driving the human's own spatial reasoning.

Two primary insights emerged from this evaluation. First, participants bypassed abstract exploration to test the model against their lived realities. One resident immediately navigated to her own block, replaced a surface parking lot with trees, and watched her agent reroute. Second, the evaluation clarified how generative agents process hypothetical urban design. Because these agents resolve spatial choices semantically rather than geometrically, relying on conditional inference and text-based analogies rather than historical GPS logs, they can generate a plausible reaction to a street that does not yet exist. The agent reroutes through the trees based on the generalized statistical association that "greenery is preferable to asphalt," providing a computable baseline that the human resident can then accept, contextualize, or correct.

How well does the agent correspond to the person it stands for? The results reflect the underlying mechanics of conditional inference. As shown in Fig. 8, the Agent Fidelity Perception rating (ranging from 1–7) reported across four subscales show that participants accepted it as a reasonable proxy without fully embracing it. The system captured what they cared about (Priority Alignment, M = 5.40) better than how they would actually behave (Behavioral Mirroring, M = 4.8) or whether they would trust it to act for them (Trust & Delegation, M = 4.83). This discrepancy is expected: while the agent successfully reflects explicitly stated priorities, it must improvise unobserved behaviors by falling back on generalized correlations learned from the internet. This reliance on statistical clichés rather than personal idiosyncrasies explains why Recognition & Ownership averaged lowest (4.92) with an unusually wide confidence interval [2.2, 6.0]—a polarized reaction to the engine's interpolation that we will return to in the Discussion.

The platform is open source and requires no institutional deployment. Given a daily route (Fig. 9b), home to work or home to school, the system retrieves the streets, amenities, and conditions along that trajectory and grounds the simulation in the places the resident actually moves through.

4. Discussion

The first lesson we learned is that seeing a proposal through someone else's day changes how people deliberate. When participants evaluated an intervention from inside another person’s routine, the discussion became more concrete and, at the same time, broader in its implications. Conventional participation confines residents to hyperlocal negotiations, such as the placement of a tree, the aesthetics of street lighting, or what Alfasi [1] calls "local, specific, and one-time" events. This time horizon ignores the impacts that will matter most: the consequences that surface not next year but next generation. Our sessions often began there too, with suggestions for better lighting or a bench, but they

rarely stayed there. The questions started with proximate concerns, such as how to calm traffic on a specific street, but quickly became more abstract and hypothetical: what happens to a neighborhood if cars disappear altogether? In San Francisco, where an actual upzoning proposal was on the table, this shift was not merely academic [49]. Residents held two fears at once—losing what made the neighborhood special, and missing out on benefits that had never been offered before. By seeing the trajectory of a neighborhood represented concretely it became much easier to choose which future they preferred.

The second lesson concerns the agent itself, and it took us some time to see it. Participants rated its grasp of what they cared about highest (Priority Alignment, M = 5.40), with its mirroring of how they would actually behave (Behavioral Mirroring, M = 4.87) and their willingness to let it act for them (Trust & Delegation, M = 4.83) somewhat lower, though overlapping confidence intervals make this ordering descriptive. Our first instinct was to treat this as a shortcoming to be engineered away. We now think that perspective misunderstands what the agent is for. The agent sits between two parties who cannot otherwise read each other. To the resident, it is a rehearsal of their own memory of the neighborhood: their corner, their detour, their unease after dark. To the planner who receives it through the export pipeline, the same agent is a structured proposal: a geolocation, an intervention type, a set of well-being metrics ready for a GIS layer. It is a boundary object: stable enough to travel between the two worlds, flexible enough to be read differently in each.

Seen this way, the agent getting it wrong is not a failure—it is often where the useful part happens. A companion study also found that agents frequently chose the same route as their resident, but for entirely different reasons [50]. And when a resident watches their agent take a street they would never take, they do not shrug. They object: "I would never walk there, that corner is badly lit, and something happened to a neighbor once." That sentence is knowledge no survey would have captured, and it only came out because the agent was wrong in a way the resident could see and correct. But this only works if the resident is there to argue. An agent that gets someone wrong inside a planning pipeline, with nobody watching, is a misrepresentation. So fidelity still matters, not to make the agent right, but to keep it close enough to be worth arguing with.

This leads to a third lesson: Conventional participatory planning is notoriously extractive [51]: it mines communities for feedback, treats residents as data points, and strands them in tokenistic consultations where they can only passively complain [1]. Our aim is to shift this paradigm from extractive consultation to generative design. By letting users immediately modify the environment and debate the outcomes, the interface shifts civic engagement from passive complaining to active design.

Ethical and societal risks of generative agents in interactive systems

The literature has raised three objections to agents as “synthetic participants” in planning: what William Agnew calls “the illusion of artificial inclusion” to refer to representation without presence that simulates community participation without granting a real voice to those affected [32]; the “flattening” warned of by

Wang, Morgenstern, and Dickerson, where the generic agent erases individual differences and harmfully portrays identity groups [31]; and the risk pointed out by Burton et al. that language models propagate an “illusion of consensus,” by converging toward synthetic responses that mask the true diversity of opinions and manufacture artificial agreement on conflicts that still require negotiation [33]. We understand it as a living and malleable database of testimonies, routines, and daily frictions, not a substitute for lived experience. By presenting the agent as a continuous draft rather than a final verdict, the simulation does not replace community deliberation, but catalyzes it: the resident assumes the role of editor of their own “reality” and the technology returns to its legitimate function as a tool to facilitate, not silence, urban dialogue.

That said, the sorting of residents in the room into “recognizers” and “non-recognizers” of their own agents reveals clear limitations in the system’s ability to accurately represent all participants. This is likely a function of flattening, and one that would only increase if the process were truly and broadly inclusive. As of now, there are sharp demographic boundaries determining who gets into the room at all. Opening the system to the community does not guarantee that the community that shows up is a representative sample. In fact, this is one of the primary reasons why the use of agents was proposed for urban participation in the first place: to fill what Fulay, Dimitrakopoulou, and Roy describe as the "empty chair", a methodological and symbolic tool designed to guarantee equity by introducing absent voices into decision-making tables [53].

Relying on artificial intelligence to fill this void is a double-edged sword. While simulating missing voices can spark new discussions, these authors caution that AI struggles to accurately characterize underrepresented groups and must only complement, not replace, genuine participation. As Wang, Brignac, Mao, and Fang (2026) warn, language models exhibit strong stereotype biases, distorting real-world demographics by consistently linking specific traits to certain groups [52].

Consequently, the system reflects the inputs it draws from: in neighborhoods with less documentation or no institutional presence, design decisions get harder and the cost of getting them wrong gets higher. Ultimately this could be facilitated by the aggregation of increasingly granular data records, but this runs the risk of model lock-in, where a prior world state is over-weighted compared to changing conditions in the real world. Continual open-ended community participation is intended to reduce this risk, but this relies on optimistic assumptions about the extent to which community members would be willing and able to engage with the platforms in order to iteratively refine their performance. So, genuine participation today often falls short: the classic 'Thursday at 4 PM' workshop remains a structural barrier to access.

Future Work and Limitations

Two present limitations of this work stem from shortcomings in the evaluation. First, more work is required to compare the two platforms directly, positioning a scenario built from other residents' stories against the one built from their own behavior patterns, to test whether the personalized version produces

more engagement and deeper reflection. The second missing piece is a look at agent ownership: whether the split we observed between residents who recognized their agent and those who did not follows some sort of demographic line such as age, income, language, or digital literacy. Further, we don't know which part of the session (the interview, the visible reasoning, the ability to correct) contributes most to the ability to recognize one's own agent. One candidate variable is interview length. Our interviews are short compared with the two-hour protocol of Park et al., so a direct test of this would help determine if a longer interview yields a more recognizable proxy for the resident.

The small scale and narrow scope of pilot data collection is a further limitation, one that greatly limits any generalizable claims of implications for actual community planning processes. To expand participation, the solution lies not only in simplifying the technology, or relying on more people to directly seed it with data, but perhaps in making participation more indirect, through calibration with actual behaviors in the context in which they occur.. This vision traces back to Patrick Geddes, an influential Scottish biologist and urban planner who combined civics education with town planning, and argues in his book Cities in Evolution (1915) that town planning must not begin merely with mechanical communications or technical blueprints, but must be rooted in an active, continuous understanding of the city's historic essence and collective life.

To make this collective life visible to citizens, through his civic exhibitions, Geddes argued that urban planning should be a process of constant public education and reflection, spatially integrated into the city itself, aiming to reunite the material needs of a place with the civic consciousness of its people [37]. If we deploy these interactive platforms not as behind-closed-doors workshop tools, but as accessible, permanent urban infrastructure, participation might cease to be an isolated event. The system becomes a continuous civic interface where residents can interact with, debate, and correct the simulations at their own pace, within the spaces they inhabit daily.

Reaching communities with less access need not mean abandoning the method, however, because the method does not start with technology; it starts with testimony. The present user interface can be further augmented with asynchronous online systems, such as dedicated resident platforms, designed to capture the qualitative reasoning behind community feedback. Through these extensions, the system can aim to capture underlying reasons *why* a specific urban change would affect an individual in a particular way. These platforms could explore a variety of mechanisms to map the lived impact of urban interventions, ensuring the planning process remains centered on complex human reasoning rather than binary preferences.

These tools are a work in progress, but we believe they gesture toward an honest version of what community participation could be. Behind them is a simple conviction: working with communities has to be generative, not extractive. This is why we go to the places where people live and develop tools to provoke reflection on their daily lives in those places. And what we hear, frequently, are possibilities no planner would have thought to model, because no planner lives those realities.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author Contribution

L.I. and J.N. wrote the main manuscript text, L.I. prepared the figures. All authors reviewed the manuscript

Acknowledgement

I would like to acknowledge everyone who made this project possible. For the Neighborhood Platform, my gratitude goes to Yasushi Sakai for guiding development of the first interface; Miguel and Naroa Coretti for developing the first version of the agent-based model; Thomas Sanchez Lengeling for his work on the agent visualization; and Gabriela Bila for contributing to the graphics of the interface's ‘upzoning’ screen. I also thank Luis Alberto Alonso, and Markus Elkatsha for their work on the metrics component of the platform, which, although not explicitly detailed in this paper, is clearly visible in the system images. For the Resident Platform and its community deployment, I am grateful to Peter Hirshberg for organizing the exhibition event at the Internet Archive in San Francisco; Rachel Tanner and the San Francisco Planning Department for their invaluable support, feedback, and vital liaison with the local community; and Miguel Pliego for his help in conducting the user interviews. Additionally, special thanks go to Adrián Mora for his technical support in deploying the server code, making the platform universally accessible. Finally, I would like to thank Pattie Maes and the entire MIT City Science extended team for their support throughout this project.

Data Availability

The Resident Platform is open source at cs-futurecities.media.mit.edu (last accessed in June 2026); the study version is archived at Github [LeticiaIG/synthetic-stories]. The repository contains code, and anonymised survey responses.

References

1. Alfasi, N. (2021). Why public participation isn’t a tool for democratizing planning. A comment. *Planning Theory*, *20*(2). https://doi.org/10.1177/1473095221991487

2. Hall, P. (2014). *Cities of Tomorrow: An Intellectual History of Urban Planning and Design Since 1880* (4.ª ed.). Wiley-Blackwell.
3. UK Parliament. (s.f.). *1842 Report on the Sanitary Condition of the Labouring Population of Great Britain*. Living Heritage: A Nation's Health. UK Parliament
4. Snow, J. (1855). On the Mode of Communication of Cholera (2.ª ed.). John Churchill.
5. Lowry, I. S. (1964). *A Model of Metropolis* (Memorandum RM-4035-RC). RAND Corporation. RAND Corporation
6. Forrester, J. W. (1969). *Urban Dynamics*. M.I.T. Press.
7. Von Neumann, J. (1966). *Theory of Self-Reproducing Automata* (A. W. Burks, Ed.). University of Illinois Press.
8. Epstein, J. M., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press. MIT Press
9. Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press. Princeton University Press
10. Freidrichs, C. (dir.) (2011). The Pruitt-Igoe Myth (Documentary)
11. Jacobs, J. (1961). The Death and Life of Great American Cities
12. Alexander, C. et al. (1977). A Pattern Language: Towns, Buildings, Construction
13. Habraken, N. John. *Supports: An Alternative to Mass Housing*. Translated by B. Valkenburg, Architectural Press, 1972. (Original work published 1961).
14. Gibson, T. (1979). *People Power: Community and Character*. Penguin Books.
15. Certeau, M. de. (1984). *The practice of everyday life* (S. F. Rendall, Trad.). University of California Press.
16. Holgersson, H. (2016). Challenging the hegemonic gaze on foot: Walk-alongs as a useful method. En T. Shortell & E. Brown (Eds.), *Walking in the European city: Quotidian mobility and urban ethnography* (pp. 207–225). Routledge. ResearchGate
17. Urban Ethnography Lab. (2025). *Go alongs* [blog]. Urban Ethnography Lab. Urban Ethnography Lab
18. Intille, S. S., Larson, K., Munguia Tapia, E., Beaudin, J. S., Kaushik, P., Kukla, J., & McLeish, R. (2006). The PlaceLab: A live-in laboratory for pervasive computing research. En Pervasive Computing (pp. 14–34). Springer.
19. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things Journal, 1(1), 22–32.
20. Shoval, N., Schvartzman, B., Rogerson, R. J., & Isaacson, M. (2018). Tracking the emotional response of tourists in time and space using wearable biosensors. Tourism Management, 69, 141–153.
21. Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. PLOS ONE, 8(7), e68400.
22. Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. A. (2014). Streetscore: Predicting the perceived safety of urban environments. En IEEE CVPR Workshops (pp. 779–785).

23. Pink, S., Berg, M., Lupton, D., & Ruckenstein, M. (Eds.). (2022). *Everyday automation: Experiencing and anticipating emerging technologies*. Routledge.
24. Pink, S., Duque, M., & Melis, A. (2022). Anthropological filmmaking for automated futures: Speculative ethnography in the smart home. *Visual Anthropology Review*, *38*(1), 45–68.
25. Bleecker, J., Foster, N., Girardin, F., & Nova, N. (2022). *The manual of design fiction*. Near Future Laboratory.
26. Izquierdo, L., Nawyn, J., & Larson, K. (2026). Comparative assessment of women's safety and mobility research methods in informal neighborhoods. *npj Urban Sustainability*, *6*(1), Article 103. doi.org.
27. Izquierdo, L., Palomo, G., Grignard, A., Alonso, L., Siller, M., & Larson, K. (2021). An agent-based model to evaluate the perception of safety in informal settlements. In *Proceedings of the 2020 Conference of the Computational Social Science Society of the Americas*. Springer Proceedings in Complexity.
28. Park, J. S., Field, S. H., & Bernstein, M. S. (2022). Social Simulacra: Creating populated social spaces for training AI systems. En *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–18). Association for Computing Machinery. doi.org
29. Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. En *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–22). Association for Computing Machinery. doi.org
30. Park, J. S., Schoelkopf, L., Mehta, S., Wang, J., Jagadish, S., West, R., … Bernstein,M. S.(2024). *Generative agent simulations of 1,000 people*(Preprint arXiv:2411.10109). arXiv
31. Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. arXiv preprint.
32. Agnew, W., Bergman, A. S., Chien, J., Díaz, M., El-Sayed, S., Pittman, J., Mohamed, S., & McKee, K. R. (2024). The Illusion of Artificial Inclusion. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24).
33. Burton, J. W., Lopez-Lopez, E., Hechtlinger, S., Rahwan, Z., Aeschbach, S., Bakker,M. A., Becker, J. A., Berditchevskaia, A., Berger, J., Brinkmann, L., Flek, L., Herzog,S. M., Huang, S., Kapoor, S., Narayanan, A., Nussberger, A.-M., Yasseri, T., Nickl,P., Almaatouq, A., … Hertwig, R. (2024). How large language models can reshape collective intelligence. Nature Human Behaviour, 8(9), 1643–1655. https://doi.org/10.1038/s41562-024-01959-9
34. Batty, M., & Torrens, P. M. (2001). Modelling complexity: The next generation of geographic information systems. *Environment and Planning B: Planning and Design*, *28*(5), 745–768.
35. Batty, M. (2024). Cap. 10, 12 y 13. En *The computable city: Histories, technologies, stories, predictions*. The MIT Press.
36. Batty, M. (2018). Cap. 1. En *Inventing future cities*. The MIT Press.

37. Geddes, P. (1892). Outlook Tower. En *Cities in evolution: An introduction to the town planning movement and to the study of civics*. Williams & Norgate.
38. Welter, V. M. (2002). The Outlook Tower. En *Biopolis: Patrick Geddes and the city of life* (caps. 3–4). The MIT Press.
39. Negroponte, N. (1970). Architecture Machine Group. En *The architecture machine: Toward a more human environment*. The MIT Press.
40. Negroponte, N (1978). The Media Room. Report for ONR and DARPA. MIT, Architecture Machine Group, Cambridge, MA.
41. Bolt, R. A. (1980). "Put-that-there": Voice and gesture at the graphics interface. En *Proceedings of the 7th annual conference on Computer graphics and interactive techniques* (pp. 262–270). Association for Computing Machinery. doi.org
42. Lippman, A. (1980). Movie-Maps: An Application of the Optical Videodisc to Computer Graphics. En *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. ACM.
43. Cartwright, W. (2006). Using non-immersive multimedia to build surrogate travel tools. *KN - Journal of Cartography and Geographic Information*, *56*(3), 131–136. doi.org
44. Noyman, A., Sakai, Y., & Larson, K. (2018). CityScopeAR: Urban Design and Crowdsourced Engagement Platform. En *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM.
45. Izquierdo, L., Nawyn, J., Xiao, V., Fernandez, F., Noyman, A., & Larson, K. (2026). *The Urban Rehearsal Machine: Immersive Simulation and Generative Co-Creation in Urban Planning*.
46. Rawls, J. (1971). *A theory of justice*. Harvard University Press.
47. Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. A. (2014). Streetscore: Predicting the perceived safety of urban environments. En *IEEE CVPR Workshops* (pp. 779–785). IEEE. doi.org
48. Hirshberg, P. (2025, September 25). *Visualizing Family Zoning, a City Science Fair* [Workshop]. Internet Archive, San Francisco, CA, United States. Luma. Retrieved June 15, 2026, from luma.com/u5of5sf5
49. San Francisco Planning Department. (2025, December 12). *San Francisco Family Zoning Plan*. City and County of San Francisco. Retrieved July 15, 2026, from https://sfplanning.org/sf-family-zoning-plan#maps-policy
50. Izquierdo Garcia, L., Ureña-Pliego, M., Mora-Carrero, A., & Larson, K. (2026). My digital twin walks the city: Decisional symmetry in human—agent urban navigation. In *Companion Publication of the 2026 ACM Designing Interactive Systems Conference* (DIS '26 Companion), 194–204. https://doi.org/10.1145/3802974.3809415
51. Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press. doi.org
52. Wang, D., Brignac, E., Mao, M., & Fang, X. (2026). Measuring stereotype and deviation biases in large language models. *Scientific Reports*, *16*, Article 23661.

53. Fulay, S., Dimitrakopoulou, D., & Roy, D. (2025). The Empty Chair: Using LLMs to Raise Missing Perspectives in Policy Deliberations. *arXiv preprint arXiv:2503.13812*.

Figures

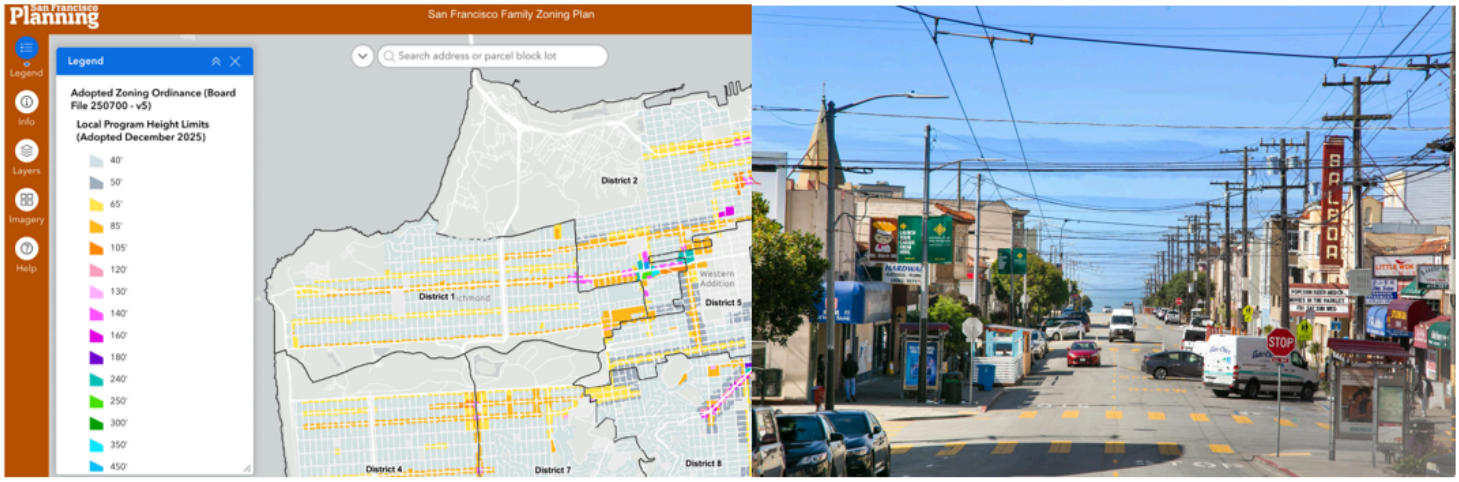

Figure 1

(Left) San Francisco Planning Department interactive map showing the local program height limits adopted in December 2025 under the Family Zoning Plan (Board File 250700). The legend details the permitted residential height increases in feet (ranging from 40' to 450') across city transit corridors, including District 1 (Richmond). (Right) Image showing a street view of Richmond.

Figure 2

Two Inner Richmond residents interviewed about the Family Zoning proposal, and how they experience daily life in the neighborhood. Each row follows one person from left to right: the interview, her situation today, and the future she hopes for. Top row, a teacher: (a) her interview; (b) in the mornings she has to

let several full buses pass before she can board; (c) the future she would like: more public space and housing near her school, so she can live close to her students and understand their needs. Bottom row, a local business owner: (d) her interview; (e) the shops on her street are closing because there is not enough foot traffic; (f) the future she would like: more people and better transit around her shop. Panels (b), (c), (e), and (f) are illustrations produced with a generative AI tool (Midjourney) from the interview content; panels (a) and (d) are video stills reproduced with the participants' permission.

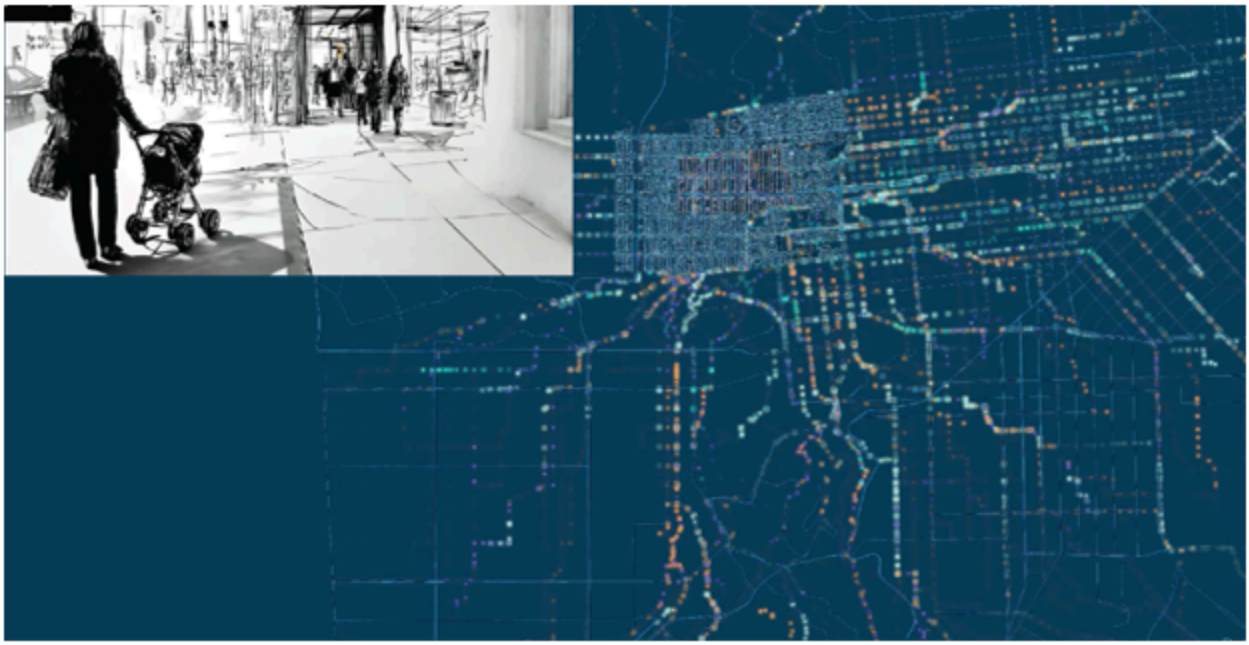

Figure 3

The corridor as an agent-based model. Each colored dot is an agent moving through the corridor's street network on a daily routine derived from the interviews. The inset shows the storyboard panel for the persona whose routine is being simulated at that moment, so that the aggregate flows on the map and the individual story is always read together. [Photo credits: Thomas Sanchez and Kent Larson]

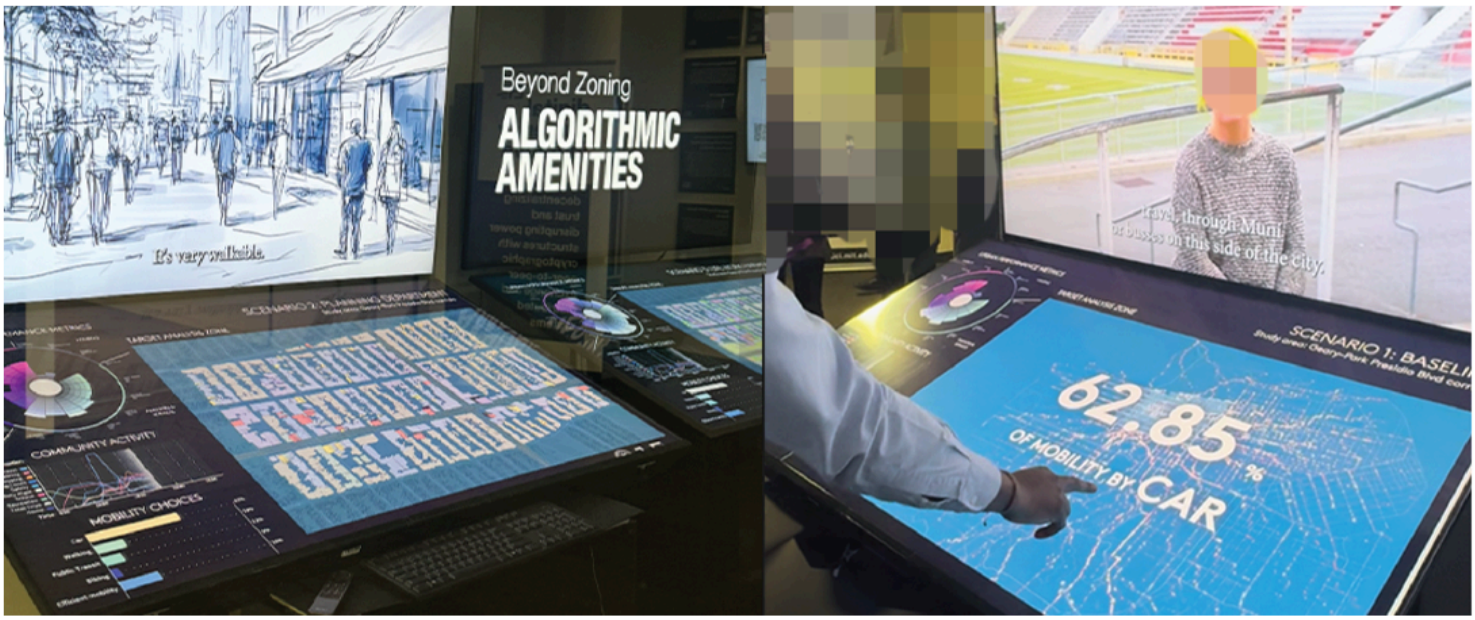

Figure 4

The interactive installation in use, featuring a four-screen comparative layout. The left pair displays the baseline state of the corridor, pairing current simulated pedestrian flows (bottom) with 'Day-in-the-Life' storyboards for specific personas (top). The right pair visualizes the corridor post-intervention, integrating the San Francisco Planning Department's 2025 'family zoning' or 'beyond zoning' initiatives with resident-driven design changes. These updated screens demonstrate the spatial impact through re-simulated pedestrian dynamics (bottom) and reflect the transformed daily experiences within the personas' storyboards (top).

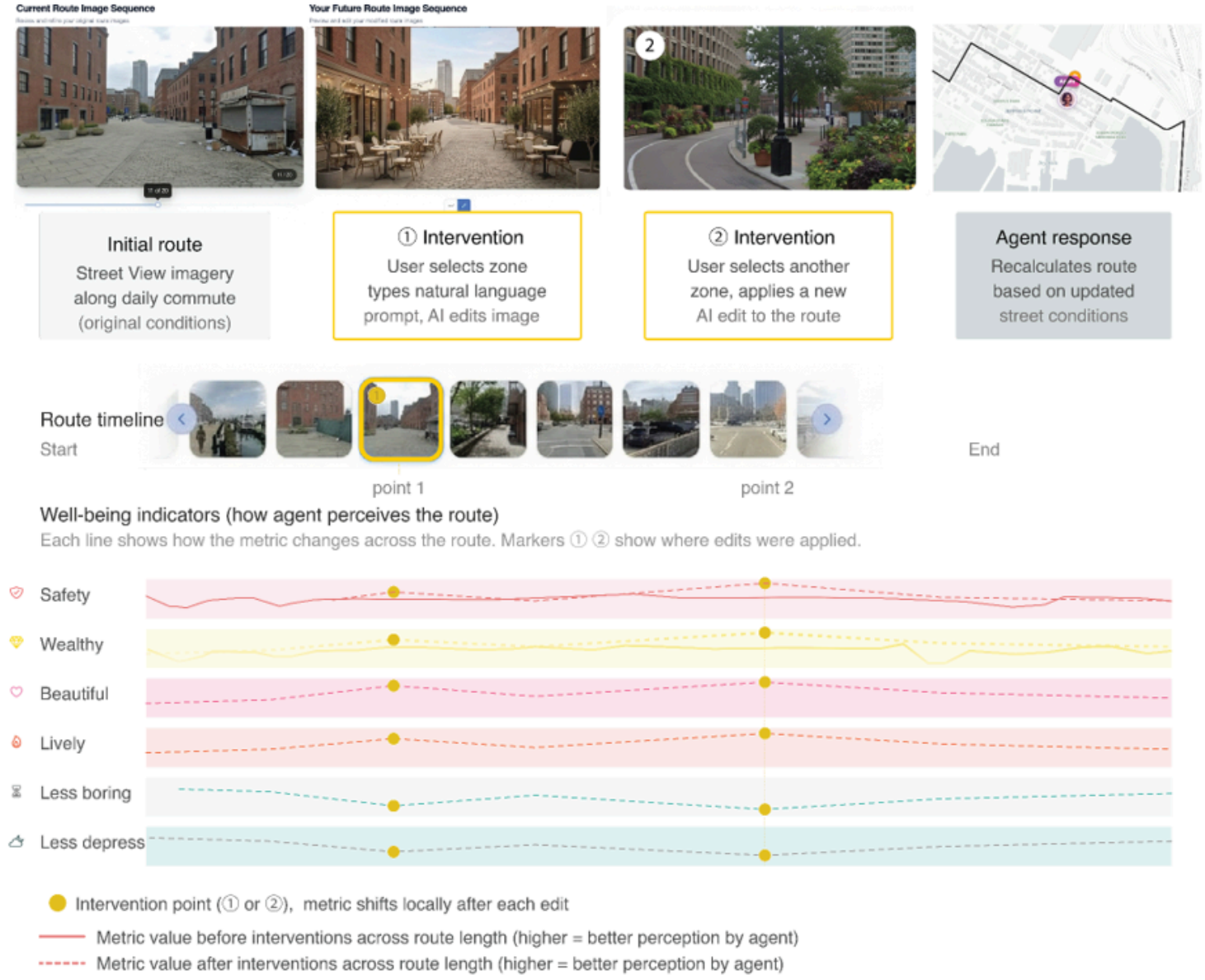

Figure 5

Overview of the human-agent computational pipeline. The system utilizes foundational data (interview transcripts, Street View panoramas, and census demographics) to drive a dual-module framework comprising an image editing engine and a spatial resident agent. The resulting workflow, depicted in Figure 4, enables users to apply targeted interventions to baseline routes. The system sequentially outputs the modified street-level imagery, the agent's adaptive routing trajectories, and longitudinal streetscores that quantify changes in perceived safety, liveliness, and aesthetics across the journey.

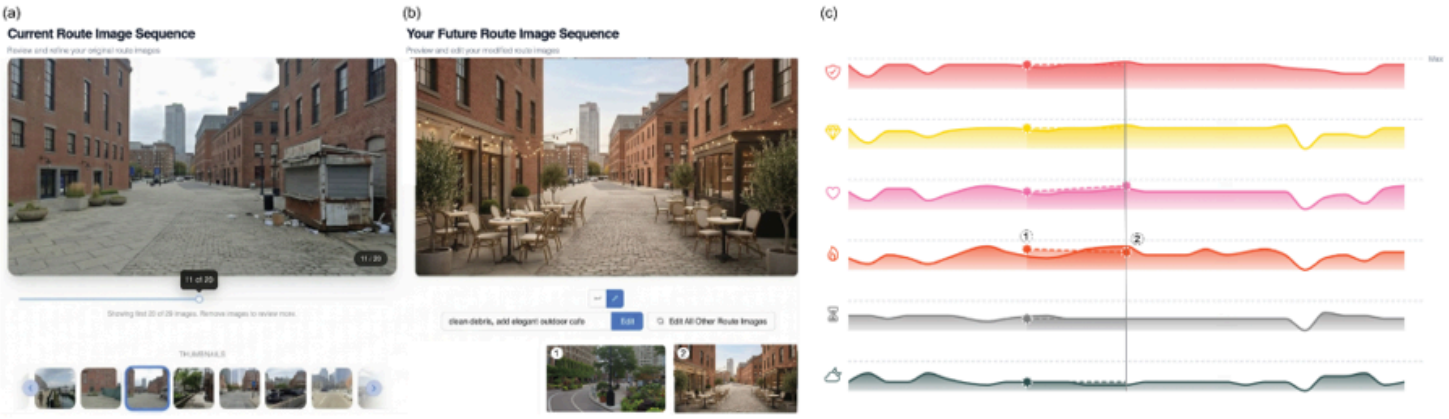

Figure 6

A corner on one resident's route before (left) and after (right) the change she asked for in plain language. The image model edits only what was requested; the buildings, light, and street geometry are preserved. [clean debris, add elegant outdoor cafe]

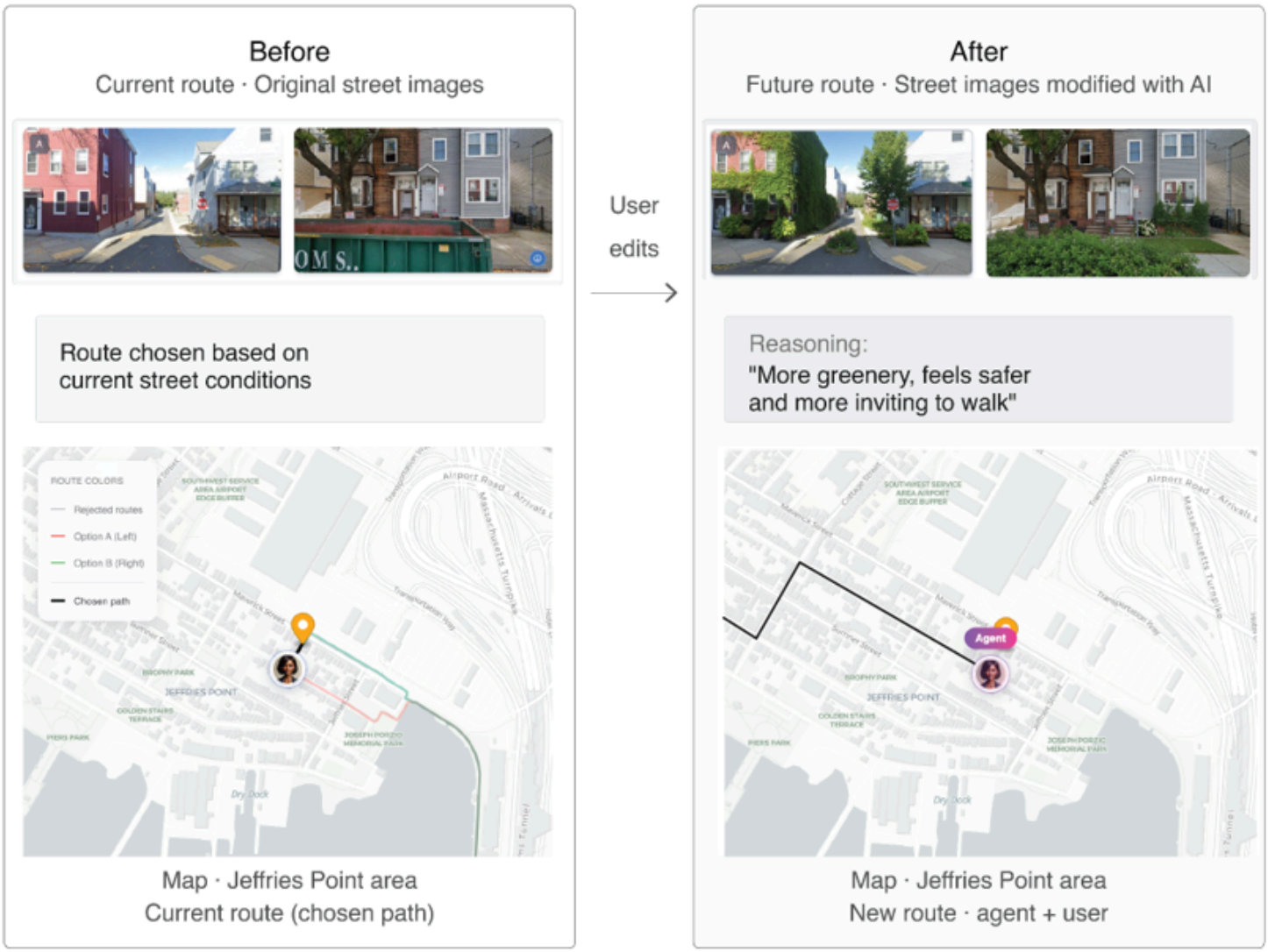

Figure 7

Screenshot of the Residents Platform App. Left (Before): the agent's route based on original street conditions. Right (After): street images modified with AI reflect the user's proposed changes, and the agent recalculates its route, explaining its preference for the modified street in terms of greenery, safety, and walkability.

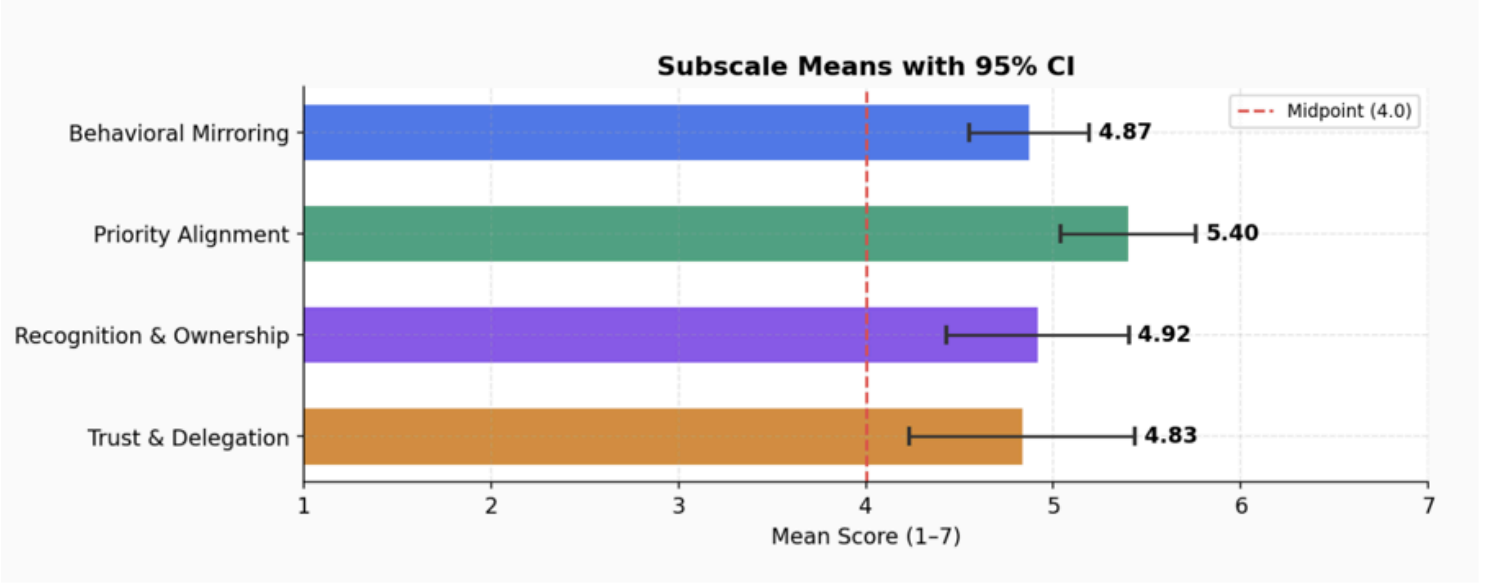

Figure 8

Subscale means with 95% confidence intervals. Horizontal bars show the mean score for each of the four Agent Fidelity Perception (AFP) subscales on a 1–7 scale (n = respondents). Error bars represent 95% confidence intervals (1.96 × SE). The red dashed line marks the scale midpoint (4.0), i.e., a neutral perception. All four subscales scored above the midpoint, and no confidence interval crossed 4.0, indicating significantly positive perceptions across all dimensions. Priority Alignment received the highest rating (M = 5.40), while Recognition & Ownership (M = 4.92), Behavioral Mirroring (M = 4.87) and Trust & Delegation (M = 4.83) clustered closely together at a lower level. The wider intervals for Trust & Delegation and Recognition & Ownership reflect greater variability in responses. Because the confidence interval for Priority Alignment partially overlaps with those of the other subscales, the figure alone does not establish a statistically significant difference between them.

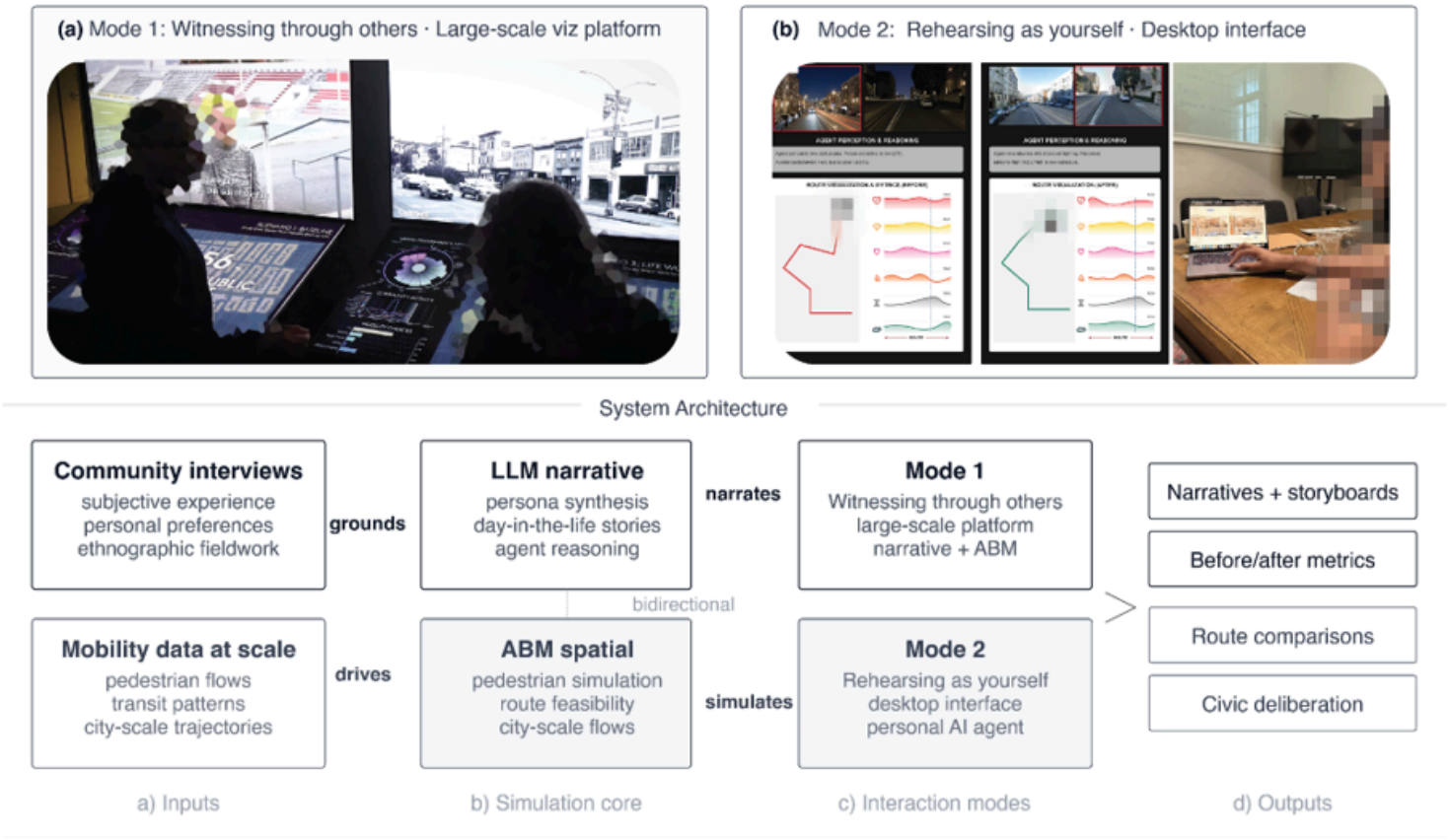

Figure 9

Overview of the deployed systems. Top: the two platforms. (a) Neighborhood Platform: a large-scale installation where groups see the corridor through personas built from community interviews. (b) Resident Platform: a desktop interface where a resident is interviewed, becomes an agent, and tests changes to their own route. Bottom: system architecture. (i) Inputs: community interviews ground the narrative layer; mobility data drive the spatial layer. (ii) Simulation core: an LLM produces personas, stories, and agent reasoning; an agent-based model simulates pedestrian flows and route feasibility; the two exchange information in both directions. (iii) The two interaction modes draw on both layers. (iv) Outputs: narratives and storyboards, before/after metrics, route comparisons, and material for civic deliberation. Participants in (a) and (b) appear with their permission; faces are blurred.

Figure 10

Community engagement workshop at the Internet Archive, San Francisco. (a) Plenary session with the SF Planning Department introducing the platform to all participants, framing urban zoning fundamentals and positioning the interaction as a process to make community concerns and potential policy impacts tangible. (b) Panel discussion with researchers and community members reflecting on the urban interventions proposed during the session. (c) Breakout group: participants collaborating with researchers at the desktop interface to iteratively refine urban interventions and synthesize community priorities. (d) Co-design session: a resident and researcher working side by side using the platform, illustrating the hands-on, dialogue-driven nature of the engagement process.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx