Adaptive Agent Networks for Representative Synthetic Population Surveys

1st Francesco Cosimo Mazzitelli
dept. of Engineering
University of Sannio
Benevento, Italy
f.mazzitelli@studenti.unisannio.it

2nd Angelo Furno
EMob-Lab
University Gustave Eiffel / ENTPE
Bron, France
angelo.furno@univ-eiffel.fr

3rd Eugenio Zimeo
dept. of Engineering
University of Sannio / CINI
Benevento, Italy
zimeo@unisannio.it

***Abstract*—Synthetic survey data enables scalable, reproducible experimentation in social science and urban behavior modeling. This work presents a modular AI framework that generates context-aware, statistically realistic survey responses by instantiating synthetic citizens as autonomous, Large Language Model (LLM)-driven agents. By combining demographic conditioning with dynamic memory mechanisms, the proposed system bypasses hard-coded branching logic to simulate citizen perceptions. We outline a four-stage pipeline: preprocessing, agent construction, survey execution, and post-processing, highlighting its capability to serve as a behavioral engine for cognitive urban digital twins and smart city analytics.**

***Index Terms*—Smart Cities, Cognitive Digital Twins, Large Language Models, Generative Agents, Synthetic Populations.**

I. Introduction

Modern smart city development relies on understanding how citizens react to new urban policies, transit systems, or sustainability initiatives. Traditional human behavioral research relies on physical surveys, which are severely constrained by high collection costs, delays, and privacy regulations.

To overcome these barriers, synthetic population generation and cognitive digital twins have emerged as powerful alternatives. Recent advancements demonstrate that Role-Playing Agents (RPAs) can mimic human reasoning and replicate aggregate statistical profiles. However, concerns remain about validation, demographic fidelity, reproducibility, and the propagation of biases from the training data [1], [2].

We propose an agentic AI framework (SYNAPSE) tailored for smart city applications that transforms raw demographic profiles and survey metadata into an ecosystem of autonomous generative agents. Each agent acts as a virtual resident, evaluating multi-modal survey questions while leveraging an adaptive internal memory window to maintain temporal and contextual logic. This framework eliminates manual rule-crafting, providing urban planners with a realistic simulation engine to test hypothetical scenarios before physical deployment.

II. Related Works

The shift from static data visualizations to cognitive urban digital twins is accelerated by open communication layers like the Model Context Protocol (MCP), which enables seamless interoperability between generative AI and analytical services [3]. In urban analytics, generative frameworks are increasingly coupled with scientific toolkits to simulate complex systems. This paradigm aligns directly with modular platforms optimized for real-time urban monitoring, traffic flow tracking, and sustainable waste collection infrastructure [4].

Simultaneously, LLMs have transitioned into active behavioral simulators. Recent work evaluates conversational agents that conduct automated phone surveys or answer validated psychometric questionnaires [5]. These studies highlight that while LLM-generated personas match human response means, they introduce specific styling biases [6].

Consequently, embedding strict architectural constraints, such as the deterministic pipeline and JSON formatting schemas introduced here, is essential to build robust behavioral simulators capable of augmenting empirical mobility datasets and safeguarding population statistical distributions [2].

III. SYNAPSE Architecture

The framework is engineered as a four-stage processing pipeline (Fig. 1) that isolates intermediate artifacts to ensure modularity and scalability.

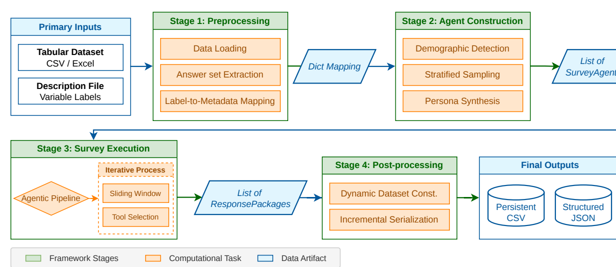

Fig. 1. SYNAPSE's four-stage pipeline

Stage 1: Preprocessing. This module processes tabular data files and companion metadata. The companion metadata file contains: question id, question, description, expected answers, type of answer (eg. TEXT, NUMERIC). If the type is not defined it will be inferred by a keyword-based fallback mechanism.

Stage 2: Agent Construction. To represent a human collectivity without excessive computational overhead, we apply a multi-tier stratified sampling approach:

1) Demographic tags (e.g., age, income) are mapped from description variables using an LLM-assisted classification block.
2) Relative frequencies are derived across demographic vectors to construct unnormalized weights. Let C represent demographic columns, and $f_j(x_{ij})$ denote the frequency of option x within column j. The raw weight for sample index i is:

$$\tilde{w}_i = \prod_{j \in C} \frac{1}{f_j(x_{ij})} \quad (1)$$

3) Normalization bounds individual weights ($w_i = \tilde{w}_i / \sum \tilde{w}_k$), driving a weighted sampling method without replacement to generate a representative synthetic population [2].

Stage 3: Survey Execution. Rather than treating survey sections as disjoint tasks, agents traverse questions sequentially. To prevent logical contradictions, a sliding-window memory tracks the five most recent context pairs and presents them within the instruction block as an active psychological anchor. Every query response is bound to a valid JSON payload specifying one of four predefined operation schemas, passed to agents as *tools*: *single selection*, *multiple selection*, *numeric bounds*, or *unstructured text*.

Stage 4: Post-Processing. The final layer captures runtime JSON payloads, parses tool identifiers, aligns choices with the original schema strings, and formats the output. Data is flushed incrementally to dual-format targets: a flat, tabular CSV matrix for statistical analysis, and a nested JSON schema that retains complete agent thought logs and metadata for auditability.

IV. Evaluation

Two distributional metrics are computed for categorical variables to compare synthetic responses with human ground truth.

Normalized Shannon Entropy (H). For each categorical variable, the entropy is normalized by $\log K$ (where K is the number of distinct categories), yielding $H \in [0, 1]$. A value of $H = 1$ indicates a uniform distribution; $H = 0$ denotes a degenerate mass. The median entropy across variables is reported as the aggregate scalar following Salvador et al. [7].

Kullback–Leibler Divergence (D_{KL}). This metric quantifies the information lost when the agent distribution Q approximates the oracle distribution P. A smoothing constant $\varepsilon = 10^{-8}$ is added to every bin to avoid undefined terms. Lower values indicate closer alignment.

Non-response filtering. Non-response codes $(-1, -7, -8, -9)$ are mapped to a sentinel value `<NR>` and *excluded* from distributional computations to ensure metrics reflect substantive answer quality rather than abstention patterns.

V. Results and Discussion

As shown in Table I, SYNAPSE achieves consistent improvements in entropy alignment across all evaluated models. `gemma2:27b` shifts from $H = 0.527$ to 0.685, `gemma2:9b` from 0.447 to 0.655, and `llama3.3:70b` from 0.319 to 0.738, all converging toward or surrounding the oracle value of 0.529. This uniform upward shift suggests that the sliding-window memory and JSON-constrained tool schema successfully reduce the *mode-collapse* tendency documented in [7], where agents over-concentrate responses on dominant categories.

TABLE I
Framework Comparison on the My Daily Travel Survey.

		Baseline [7]		**SYNAPSE**	
Model	**B**	**H**	**KL**	**H**	**KL**
Oracle	—	0.566	—	0.529	—
`gemma2:27b`	27	0.527	2.863	0.685	2.337
`gemma2:9b`	9	0.447	3.900	0.655	2.654
`llama3.3:70b`	70	0.319	3.250	0.738	3.475

KL divergence results further support this trend, showing reductions for `gemma2:27b` and `gemma2:9b`. `llama3.3:70b` shows a marginal increase (3.475), which is attributable to the broader evaluation scope: SYNAPSE covers all 178 categorical columns, whereas the baseline evaluated a curated subset. No monotonic relationship between model size and distributional fidelity emerges, confirming that architectural choices and instruction-tuning dynamics outweigh simple scale parameters.

VI. Conclusions

This paper presented SYNAPSE, a modular agentic framework for generating statistically realistic synthetic survey responses. By replacing hard-coded branching graphs with a sliding-window memory and a structured JSON tool schema, SYNAPSE produces synthetic populations whose distributional properties closely match empirical ground truth. Future work will expand the evaluation to include additional datasets and introduce cross-tabulation consistency checks.

References

[1] M. Larooij and P. Törnberg, “Validation is the central challenge for generative social simulation: a critical review of llms in agent-based modeling,” *Artificial Intelligence Review*, vol. 59, no. 1, p. 15, 2025.

[2] J. Bisbee *et al.*, “Synthetic replacements for human survey data? the perils of large language models,” *Political Analysis*, vol. 32, no. 4, pp. 401–416, 2024.

[3] X. Hou *et al.*, “Model context protocol (mcp): Landscape, security threats, and future research directions,” *ACM Transactions on Software Engineering and Methodology*, 2025.

[4] L. Goglia, F. C. Mazzitelli, and E. Zimeo, “Sustainable planning of pick-up and delivery for waste collection,” in *2025 IEEE International Smart Cities Conference (ISC2)*. IEEE, 2025, pp. 1–6.

[5] L. V. Vogelsmeier *et al.*, “Delving into the psychology of machines: Exploring the structure of self-regulated learning via llm-generated survey responses,” *Computers in Human Behavior*, p. 108769, 2025.

[6] J. C. de Winter, T. Driessen, and D. Dodou, “The use of chatgpt for personality research: Administering questionnaires using generated personas,” *Personality and Individual Differences*, vol. 228, p. 112729, 2024.

[7] I. Salvador, A. Furno, and S. Derrible, “Large language model-enhanced general transportation agent framework for human mobility forecasting and synthetic travel survey data generation,” *Transportation Research Record*, 2026.