Generative AI Agents for Travel Behaviour: Applications in Surveys and Modelling

Tareq Alsaleh*[1] and Bilal Farooq[2]

[1]PhD Candidate, Laboratory of Innovation in Transportation (LiTrans), Toronto Metropolitan University, Canada
[2]Associate Professor, Laboratory of Innovation in Transportation (LiTrans), Toronto Metropolitan University, Canada

Short summary

We explore the potential of Generative Artificial Intelligence (AI) agents created using open-access and locally hosted Large Language Models (LLMs) in replicating human survey behaviour and mode choice preferences in scenario-based travel surveys. The aim is to establish performance and validation benchmarks for utilizing AI agents in travel behaviour analysis, agent-based simulations, and other case uses. Accordingly, we developed a systematic scientific approach to assess the performance of seven open-access foundational LLMs, with parameters ranging from one to seventy billion, which can be generalized for creating and validating the performance of Generative AI agents in various applications. The AI agents were developed using a zero-shot learning approach, incorporating both unrestricted sociodemographic and static prompting, as well as a dynamic restricted sociodemographic prompting strategy. The performance of these agents was validated against the human benchmark dataset, evaluating their effectiveness and reliability in capturing and replicating nuanced travel behaviour.
Keywords: Generative AI, AI Agents, Large Language Models (LLMs), Travel Behavioural Modelling, Travel Surveys

1 Introduction

Travel surveys are the backbone of effective transportation planning and operations. Over time, travel diaries and scenario-based surveys have significantly evolved, transitioning from paper-based methods to smartphone applications Patterson et al. (2019), and even to virtual reality scenario-based solutions Ansar et al. (2023). These advancements aim to address longstanding operational and technical challenges—whether related to hypothetical bias, cost, or low participation rates Stopher & Greaves (2007). As generative AI is advancing in all fields, many researchers are investigating its potential application in transportation operations and planning and its possible contribution to solving some of the longstanding issues of human mobility, traffic operation and planning.

Mahmud et al. (2025) discussed the potential of foundational LLMs such as BERT, and GPT in broader transportation applications, especially in intelligent transportation systems (ITS) such as in traffic management, autonomous driving, and infrastructure optimization. Yan & Li (2023) on the other hand, discussed the potential of generative AI and LLMs in traffic perception, traffic prediction, simulation, and decision-making, as well as in human mobility prediction. Hence, attempting to address the traditional operational challenges from the regular recording of human mobility patterns. In a recent study by Wang et al. (2023), the authors presented the LLM-Mob framework using GPT-3.5-turbo via the OpenAI API, which aims to predict the next location a person will visit based on their past mobility patterns. The research used public human mobility datasets to demonstrate that LLM-Mob outperforms deep learning models in predictive performance and offers superior interpretability. GPT-4 was also used in a recent study on travel diaries generation for urban mobility, where the LLM was used to generate realistic and personalized travel diaries incorporating individual profiles and contextual reasoning. In this study, the authors highlighted the need to enhance data diversity and address privacy concerns when applying such AI models in mobility studies Li et al. (2024).

Despite the growing interest, research on LLMs in transportation remains limited. Most studies have focused on specific use cases or a single LLM model or framework, restricting the generalizability of findings across diverse contexts. High computational costs often hinder research efforts, while sensitive or proprietary datasets remain inaccessible to many researchers. Data privacy concerns are particularly significant when dealing with personal mobility data, which highlights the need for a comprehensive testing framework. Such a framework should evaluate open-access, computationally feasible LLMs, detailing their potential and limitations across various applications. In this study, we address these challenges by evaluating seven open-access models with different architectures against human-based mobility responses, utilizing locally hosted, computationally feasible solutions.

2 Methodology

We develop a methodology for the complete AI agent generation process and the verification of the agent's responses. At first we provide details regarding the LLMs setup and the local API server, along with the AI agents and their persona generation approach. The details on the case study of the Stated Preference (SP) data used to verify the AI agent's responses are then provided.

AI Agents Generation

As detailed in Figure 1, we used seven open-access LLMs for the generation of the AI agents using zeroshot learning approach. Two methods were followed as part of the zero-shot learning. First, we generated individual agents with static, context and scenario-based prompting where agent sociodemographics were unrestricted and determined by the LLMs for each run along with the mode choice for the described travel scenario. The second zeroshot AI agents were developed using dynamic prompting and restricted sociodemographic. In the second approach, the AI agent persona is fed to the LLM dynamically for each loop, creating an agent with matching socioeconomic characteristics of the corresponding human respondent and returning the agent selected travel mode according to the described scenario.

The aim is to generate N number of AI agents, representing a survey respondents. As they are generated for each LLM model we are testing, each of them is provided sociodemographic characteristics aligned with a set of predetermined characteristic variables. For the purpose of representing the process we define the following parameters:

- $\mathbf{M}$: the LL model (Open Access locally Hosted LLM).
- Q_i: the combined prompt for agent i, consisting of system and user prompts.
- $f(\mathbf{M}, Q_i)$: a function that queries the model $\mathbf{M}$ with Q_i and returns a raw text response R_i.
- $g(R_i)$: a function that attempts to parse a valid JSON object J_i from the raw text R_i.
- $\mathcal{O} = \{k_1, k_2, \dots, k_c\}$: the set of c predetermined characteristic variables and the choice.
- $h(J_i, \mathcal{O})$: a function that standardizes the variables characteristics in J_i to the set $\mathcal{O}$, producing a dictionary D_i.

In practice, the procedure to generate the N agents with their sociodemographic characteristics and mode choice can be achieved with the following steps:
For each agent i, we have:

$$Q_i \;=\; (\text{system prompt},\ \text{user prompt}), \tag{1}$$

- Model Generation, Query the model:

$$R_i \;=\; f\big(\mathbf{M}, Q_i\big). \tag{2}$$

- Extract and validate JSON, attempt to parse R_i:

$$J_i \;=\; g\big(R_i\big). \tag{3}$$

Standardize Keys Map the JSON keys to the official schema:

$$D_i \;=\; h\big(J_i, \mathcal{O}\big). \tag{4}$$

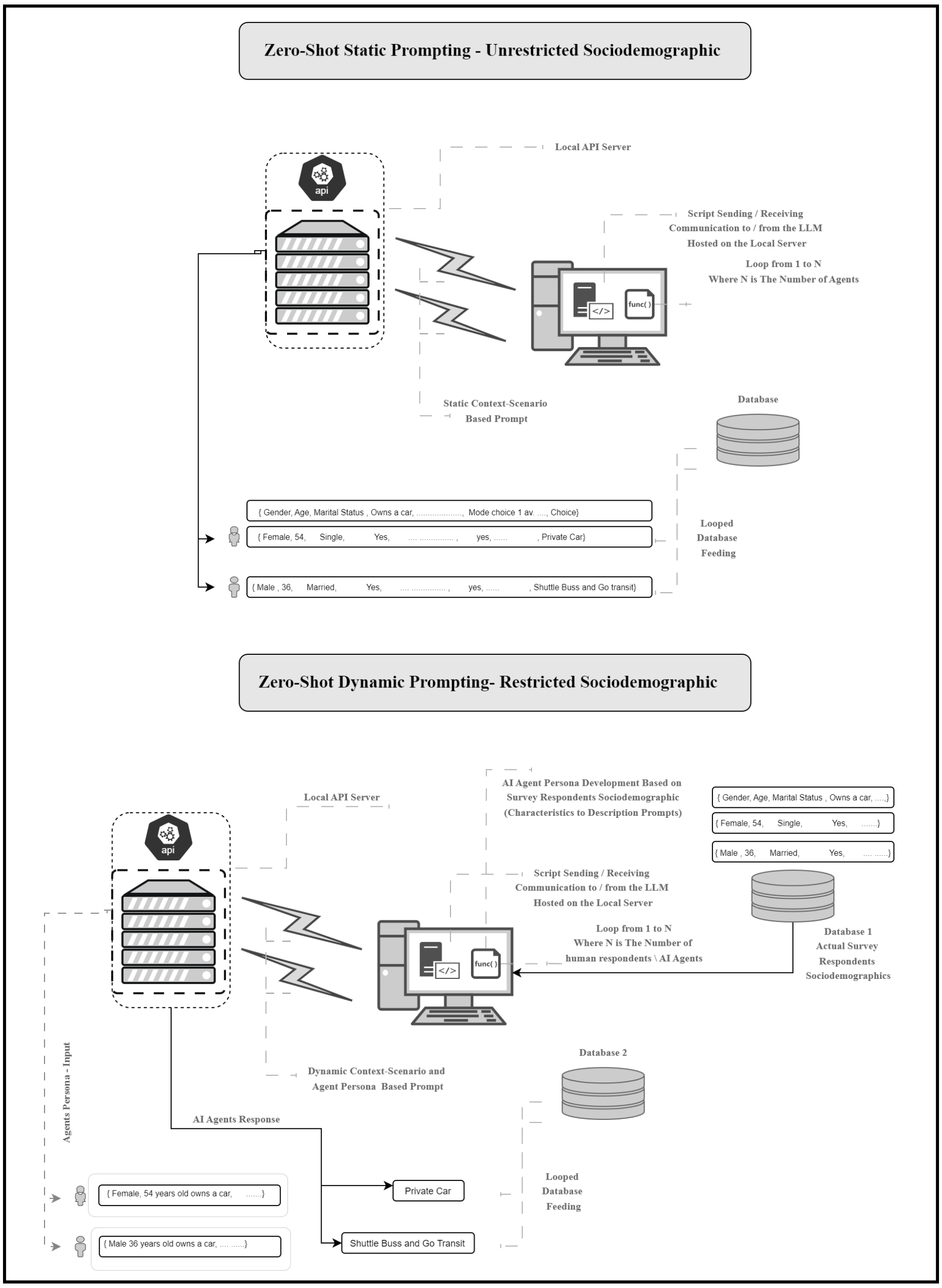

Zero-Shot Static Prompting - Unrestricted Sociodemographic
api
Local API Server
Script Sending / Receiving Communication to / from the LLM Hosted on the Local Server
Loop from 1 to N Where N is The Number of Agents
</>
func( )
Static Context-Scenario Based Prompt
Database
{ Gender, Age, Marital Status , Owns a car, ..., Mode choice 1 av. ...., Choice}
{ Female, 54, Single, Yes, ... , yes, ... , Private Car}
{ Male , 36, Married, Yes, ... , yes, ... , Shuttle Buss and Go transit}
Looped Database Feeding
Zero-Shot Dynamic Prompting- Restricted Sociodemographic
api
Local API Server
AI Agent Persona Development Based on Survey Respondents Sociodemographic (Characteristics to Description Prompts)
{ Gender, Age, Marital Status , Owns a car, ...}
{ Female, 54, Single, Yes, ...}
{ Male , 36, Married, Yes, ...}
Script Sending / Receiving Communication to / from the LLM Hosted on the Local Server
Loop from 1 to N Where N is The Number of human respondents \ AI Agents
</>
func( )
Database 1 Actual Survey Respondents Sociodemographics
Dynamic Context-Scenario and Agent Persona Based Prompt
Database 2
Agents Persona - Input
AI Agents Response
{ Female, 54 years old owns a car, ...}
{ Male 36 years old owns a car, ...}
Private Car
Shuttle Buss and Go Transit
Looped Database Feeding

Figure 1: AI Agents Generation Methodology

Row Construction For each $k \in \mathcal{O}$:

$$r_{i,k} \;=\; \begin{cases} D_i[k], & \text{if } k \in D_i, \\ \text{"N/A"}, & \text{otherwise.} \end{cases} \tag{5}$$

Hence the row for agent i is

$$\mathbf{r}_i \;=\; \bigl(r_{i,k_1},\, r_{i,k_2},\, \dots,\, r_{i,k_c}\bigr). \tag{6}$$

Append to database Insert $\mathbf{r}_i$ into the database. After iterating $i = 1, \dots, N$, we obtain:

$$\text{database} \;=\; \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_N \end{bmatrix}, \tag{7}$$

an $N \times c$ matrix where each row is as in (6).

For the dynamic restricted approach, in a *previous scenario* we defined a more explicit prompt

$$Q_i = \bigl(S,\, U(d_i)\bigr) \tag{8}$$

where S is the system message and $U(d_i)$ injects some dynamic content d_i. Now we adapt this idea to incorporate *real survey respondents' sociodemographics* rather than synthetic data, and to capture a single key outcome: the chosen travel mode.

1. **Redefine $\boldsymbol{d_i}$** as *actual survey data* for respondent i, e.g. age, income, household size, etc. This replaces any previously synthetic d_i.
2. **Dynamic Prompt Construction.** As in Equation (1)

   $$Q_i \;=\; \bigl(S,\, U(d_i)\bigr),$$

   but now $U(d_i)$ embeds *real* attributes. The subsequent calls (6)–(7) remain the same, with R_i, J_i, and D_i unchanged except for the fact they come from real-data prompts.
3. **Extracting One Key Response: Choice.**

   Instead of collecting many fields from D_i, we now want only the single chosen mode. Define a new function $\sigma(\cdot)$ to extract one field from J_i. We introduce:

   $$\text{choice}_i \;=\; \sigma\bigl(J_i\bigr), \tag{8}$$

   which yields the chosen travel mode (e.g. `"Private Car"`). All other data in D_i can be ignored or omitted.
4. **Final Output.** Equation (7) provided a CSV-like structure that stores all 58 official keys. Now, each row need only keep the real input (d_i) and the single chosen mode choice$_i$. Formally,

   $$\text{CSV}_{\text{survey}} \;=\; \bigl[i \quad d_i \quad \text{choice}_i\bigr]_{i=1,\ldots,N}\,. \tag{9}$$
5. **Overall Loop.** The iteration $i = 1, \ldots, N$ in (8)–(2) and (2) remains the same, but we store only choice$_i$ from (8). Symbolically,

   $$\boxed{\text{For } i = 1 \text{ to } N: \quad \bigl(Q_i,\; R_i,\; J_i,\; \text{choice}_i\bigr).} \tag{10}$$

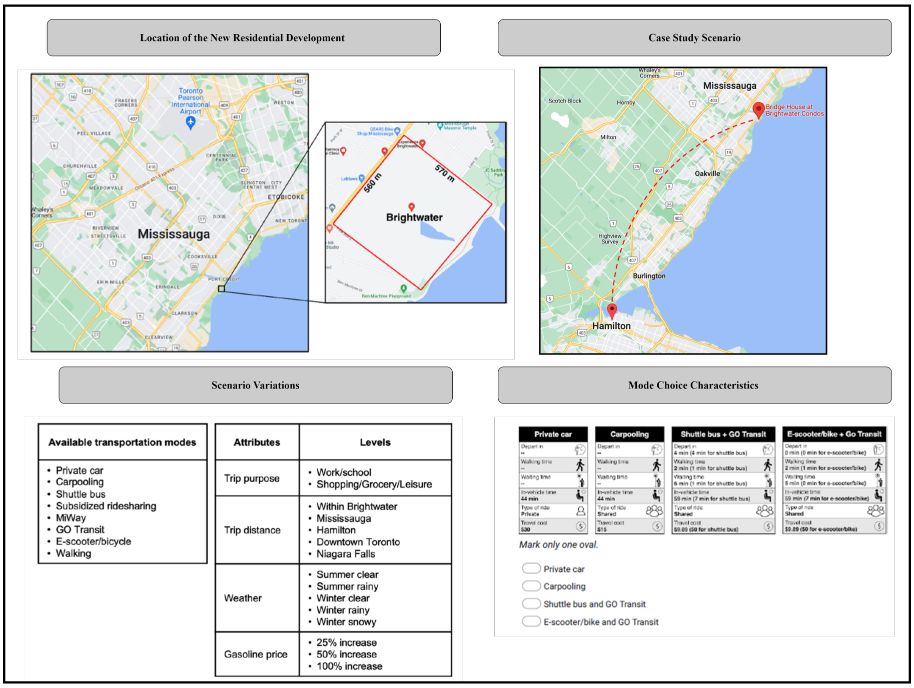
   In short, real data d_i is used to build the prompt, the model returns J_i, and we record choice$_i$.

   All other steps (reading data, forming the loop over $i = 1, \ldots, N$, querying $\mathbf{M}$, parsing R_i, and writing output) remain identical to the original procedure, except that we store only choice$_i$ from each agent's JSON response.

   The equations define the complete looped process for generating N synthetic agents, extracting and standardizing their responses, and storing them in a tabular CSV format.

Case Study

To verify the AI agent's ability to replicate human behaviour effectively, its behavioural performance must be tested against actual human benchmarks in identical contextual and travel scenarios. While testing for the agent's behaviour in different spatial and temporal frames is underway, in this submission, we present the metrics of different LLM agents' performance with reference to a recently stated preference (SP) conducted in the Brightwater community. Brightwater community is a master-planned, mixed-use community on 72 acres of waterfront land in Port Credit, Mississauga, including over 2,900 residential units and 300,000 sq. ft. of commercial space, to be completed in 2029 (estimated), with first residents occupation in 2023. The data collection campaign ran from October 31, 2022, to November 20, 2022. A total of 159 future residents completed the survey which included key sociodemographics information as well as perception related questions and scenario based travels from Brighwater location to key neighbouring cities and areas in different seasons and weather conditions. Figure 2 shows the case study location and the key scenario variations in addition to the travel scenario used for the context development and LLM agent's response verification.


Location of the New Residential Development
Case Study Scenario
Mississauga
Brightwater
560 m
570 m
Toronto Pearson International Airport
Bridge House at Brightwater Condos
Oakville
Burlington
Hamilton
Scenario Variations
Mode Choice Characteristics

Available transportation modes
Private car Carpooling Shuttle bus Subsidized ridesharing MiWay GO Transit E-scooter/bicycle Walking

Attributes	Levels
Trip purpose	Work/school Shopping/Grocery/Leisure
Trip distance	Within Brightwater Mississauga Hamilton Downtown Toronto Niagara Falls
Weather	Summer clear Summer rainy Winter clear Winter rainy Winter snowy
Gasoline price	25% increase 50% increase 100% increase

Private car	Carpooling	Shuttle bus + GO Transit	E-scooter/bike + Go Transit
Depart in --	Depart in --	Depart in 4 min (4 min for shuttle bus)	Depart in 0 min (0 min for e-scooter/bike)
Walking time --	Walking time --	Walking time 2 min (1 min for shuttle bus)	Walking time 2 min (1 min for e-scooter/bike)
Waiting time --	Waiting time --	Waiting time 6 min (1 min for shuttle bus)	Waiting time 5 min (0 min for e-scooter/bike)
In-vehicle time 44 min	In-vehicle time 44 min	In-vehicle time 59 min (7 min for shuttle bus)	In-vehicle time 59 min (7 min for e-scooter/bike)
Type of ride Private	Type of ride Shared	Type of ride Shared	Type of ride Shared
Travel cost $30	Travel cost $15	Travel cost $9.89 ($0 for shuttle bus)	Travel cost $9.89 ($0 for e-scooter/bike)

Mark only one oval.

○ Private car
○ Carpooling
○ Shuttle bus and GO Transit
○ E-scooter/bike and GO Transit

Figure 2: Case Study Location and Travel Scenario Details

3 Results and discussion

For the unconstrained static promoting, a total of four LLMs were examined, namely the LLama 3.2 3B, Qwen 2.5 7B, Stealth 1.2 7B and Mistral 7B Q4. These models were selected to cover the most prominent open access foundational models, and so that it is of a medium to small size and can be hosted and operated on top of a line user machine, but not at an enterprise scale. On the other hand, the models used for the restricted AI agent generation covered the same four models in addition to Llama 3.2 1B, Llama 3.3 70B, and Qwen 2.5 32B. The three additional models were possible due to the difference in token intensity between the restricted and unrestricted agent generation, especially on the output side, which allowed the testing of larger models.

For the analysis of the unrestricted LLM agents, the distribution of their sociodemographic characteristics was compared to that of the actual respondents. The comparison was conducted using the normalized Root Mean Square Error (nRMSE) to assess the frequency distribution of categorical sociodemographic variables, as well as the Chi-square test and the P-value for the goodness-of-fit test. The null hypothesis states that there is no significant difference between the categorical distribution of the AI agents' sociodemographic variables and that of the actual respondents.

Table 1 presents the results of this exercise for the four models. The nRMSE varied across different variables, with an average ranging from 0.27 for Stealth 1.2 7B to 0.32 for Mistral 7B Q4, considering all sociodemographic characteristics. This indicates that, on average, the predicted distributions of the sociodemographic characteristics deviate by 27% to 32% from the actual distribution of respondents' sociodemographics. While the AI-generated agents capture broad patterns of residents' sociodemographics, the observed error suggests clear discrepancies in capturing finer details of respondents' characteristics based solely on context. This is further supported by the Chi-square test and P values, which led to the rejection of the null hypothesis for all variables across all models, except for scooter ownership in the Llama 3.2 3B model. The rejection of the null hypothesis indicates that the predicted distributions significantly differ from the actual data. While it is important to acknowledge that the actual respondent's sociodemographic variables distribution is limited in size and by the scope of the study, the models had little heterogeneity in agents' sociodemographics despite being generated independently. Further investigation is underway on different datasets with larger population sizes. One key consideration to improve this limitation is to establish a guided sociodemographic generation, where aggregate level statistics per variable category are provided for the model to develop granular level agents sociodemographic personas.

Table 1: Sociodemographic Analysis of Unrestricted LLM Agents Compared to Survey Respondents

Variable	Llama 3.2 3B			Qwen 2.5 7B			Stealth v1.2 7B			Mistral 7B Q4		
	nRMSE	Chi-Square	P-Value	nRMSE	Chi-Square	P-Value	nRMSE	Chi-Square	P-Value	nRMSE	Chi-Square	P-Value
Young	0.8931	1401.26	0.0	0.1006	17.79	0.000025	0.1006	17.79	0.000025	0.8805	1362.06	0.0
Adult	0.5220	173.86	0.0	0.4717	141.96	0.0	0.4717	141.96	0.0	0.5094	165.58	0.0
Middle Age	0.3019	68.75	0.0	0.3019	68.75	0.0	0.3019	68.75	0.0	0.3019	68.75	0.0
Old	0.0692	11.81	0.000587	0.0692	11.81	0.000587	0.0692	11.00	0.000911	0.0692	11.81	0.000587
Gender	0.3182	64.66	0.0	0.4679	139.84	0.0	0.2478	39.22	0.0	0.4679	139.84	0.0
Marital Status	0.7987	642.16	0.0	0.1949	38.50	0.0	0.1949	38.50	0.0	0.1194	14.46	0.000143
Student	0.0314	5.12	0.0235	0.0314	5.16	0.0230	0.0314	5.16	0.0230	0.0314	5.16	0.0230
Employment	0.1949	38.50	0.0	0.1949	38.50	0.0	0.1949	38.50	0.0	0.1949	38.50	0.0
Driving License	0.0253	4.12	0.0421	0.0253	4.12	0.0421	0.0253	4.12	0.0421	0.0253	4.12	0.0421
Vehicle Ownership	0.0440	7.32	0.0068	0.0440	7.32	0.0068	0.0377	5.37	0.0203	0.0440	7.32	0.0068
Transit Pass	0.0859	4.69	0.0302	0.3441	81.45	0.0	0.3630	90.63	0.0	0.3630	90.63	0.0
Bicycle Ownership	0.3924	105.04	0.0	0.6603	309.16	0.0	0.6415	291.75	0.0	0.6603	309.16	0.0
Scooter Ownership	0.0182	1.94	0.1628	0.0251	4.10	0.0428	0.0251	4.10	0.0428	0.0251	4.10	0.0428
Residential Safety Perception	0.1132	20.29	0.000007	0.1132	20.29	0.000007	0.1132	20.29	0.000007	0.1132	20.29	0.000007
Residential Walking Friendly Perception	0.2138	43.24	0.0	0.2138	43.24	0.0	0.2138	43.24	0.0	0.2138	43.24	0.0
Residential Cycling Friendly Perception	0.2785	61.36	0.0	0.2435	46.93	0.0	0.2785	61.36	0.0	0.2785	61.36	0.0
Residential Scooter Friendly Perception	0.3145	63.80	0.0	0.5094	167.44	0.0	0.2138	29.50	0.0	0.5597	200.88	0.0
Public Transport Accessibility	0.4873	151.14	0.0	0.3804	92.10	0.0	0.4496	128.64	0.0	0.4810	147.27	0.0
Housing Type	0.6918	593.79	0.0	0.8490	894.37	0.0	0.8364	868.07	0.0	0.8490	894.37	0.0
Low Income	0.0828	13.45	0.000245	0.0828	14.35	0.000151	0.0828	14.35	0.000151	0.0828	14.35	0.000151
Medium Income	0.3790	86.92	0.0	0.5605	202.78	0.0	0.5605	202.78	0.0	0.5605	202.78	0.0
High Income	0.4777	133.53	0.0	0.4777	145.42	0.0	0.4777	145.42	0.0	0.4777	145.42	0.0
Toddler Child	0.1257	21.72	0.000003	0.1257	22.87	0.000002	0.1257	22.87	0.000002	0.1257	22.87	0.000002
Young Child	0.0377	5.92	0.0149	0.0377	6.23	0.0125	0.0377	6.23	0.0125	0.0377	6.23	0.0125
Household Size	0.5911	250.07	0.0	0.6666	318.00	0.0	0.6666	318.00	0.0	0.6666	318.00	0.0
Number of Cars	0.5283	210.56	0.0	0.6981	367.68	0.0	0.2955	65.92	0.0	0.3018	68.75	0.0

On the other hand, the analysis for the mode choice distribution from the different LLM Agents responses is shown in Figure 4. The Agents were given the option to select among six mode choices, as well as to determine the choice availability for the described travel scenario. Stealth 1.2 7B showed superior performance with nRMSE value of 0.13 compared to 0.4 for Mistral 7B Q4 and around 0.3 for both Llama and Qwen models. The results suggest that Stealth 1.2 7B provides a closer approximation to the actual mode choice distribution, with lower deviations in predicting respondents' travel preferences. The relatively lower nRMSE of 0.13 indicates that this model shows a better understanding of mode choice behaviour, capturing the underlying patterns of the different sociodemographic influences on the mode choice preference. Stealth was developed by Jan and is part of a new experimental family designed to enhance Mathematical and Logical abilities. It is important to highlight two takeaways here; first, it is important to note that the differences in performance across models are indicators of the varying capabilities of LLMs in processing and interpreting contextual information related to travel behaviour. Therefore, we are investigating the impact of the temporal and spatial context variation on the model performance. Second, It was noted during the analysis that all models were able to select only valid mode choices for travel scenarios despite being presented with six mode choices, out of which two do not operate on the selected route. This highlights the potential use of AI Agents for choice availability analysis, an area of great impact, especially in the context of Revealed Preference (RV), where the non-selected choices availability is difficult to obtain.

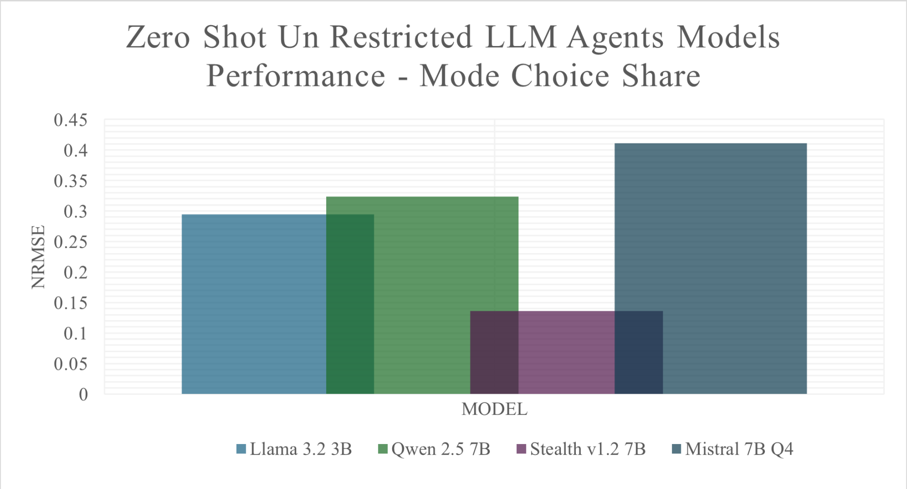


Figure 3: nRMSE of the Mode Choice Distribution for Unrestricted LLM Agents Compared to the Actual Respondents

As for the dynamic prompting and restricted sociodemographic agents that align with the actual respondent's characteristics. Figure 4 shows the drop in the nRSME values with the introductions of Agents personas. Mistral and Qwen had the most significant reductions with 75% and 74% reduction rates compared to their unrestricted versions. Stealth maintained its robust performance recording one of the lowest nRMSEs of 0.1. Llama 3.2 3B model, on the other hand, didn't show significant improvement, with only a 14% reduction in its overall nRMSE value for the mode choice distribution. The reduction in the nRMSE mainly highlights the limitations of the LLMs to create diverse agent personas on repeated static contextual prompts. The heterogeneity that was introduced by the dynamic prompting for agents' persona creation improved the choice distribution for all models, although to varying degrees. The nRMSEs for all seven models are shown in Figure5. Stealth, Mistral and Qwen 2.5 7B showed the best performance compared to the rest of the models.

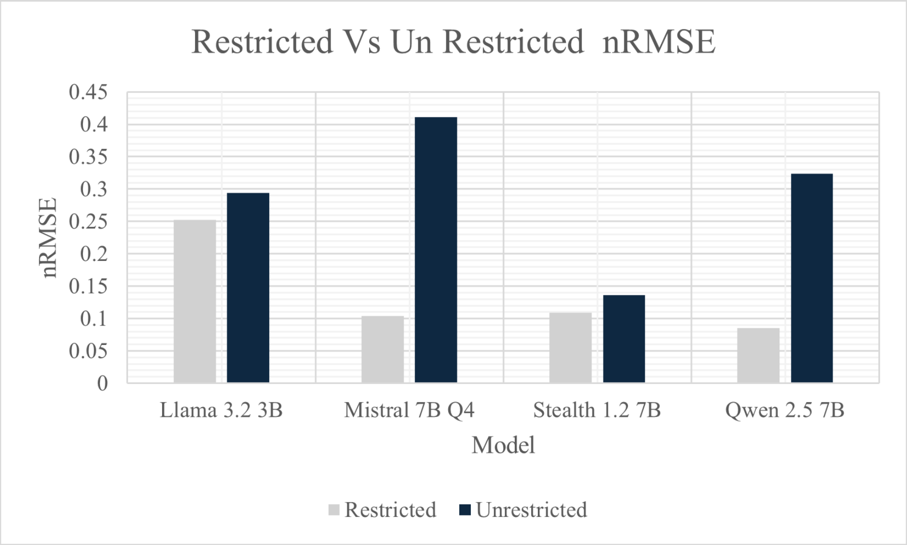


Figure 4: nRMSE of the Mode Choice Distribution for Restricted LLM Agents Vs Unrestricted LLM Agents Compared to the Actual Respondents

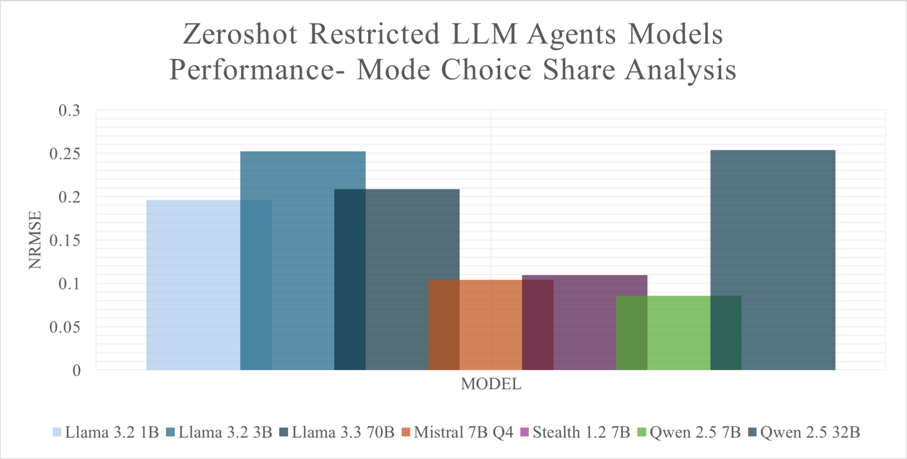
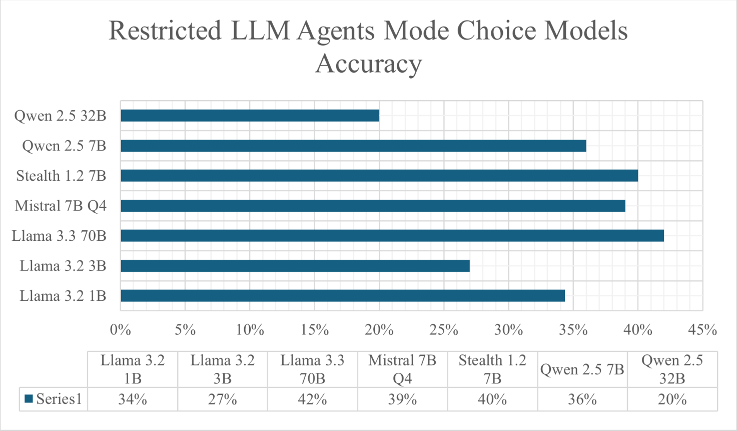


Figure 5: nRMSE of the Mode Choice Distribution for Restricted LLM Agents Compared to the Actual Respondents

Model's accuracies, on the other hand, for the one-to-one prediction compared to the actual respondents is presented in Figure 6. Llama 3.3 70B recorded the highest accuracy of 42%, followed by Stealth and Mistral with 40 and 39%, respectively. It is worth highlighting that despite the Llama 3.3 70B outperforming the rest of the models, it is not clearly evident that the higher the parameters, the better the model performance. Stealth and Mistral have way lower number of parameters than Llama 3.3 70B but still scored higher on the nRMSE benchmark and marginally lower on the accuracy percentage. However, Qwen 2.5

7B scored much better in both benchmarks than the version with 32B parameters, same applies for Llama 3.2 1B and 3B parameters, so with the current results, its not evident that the ni=umber of parameters directly impact the accuracy of the model for mode choice prediction. However, this is under investigation with larger and more diverse datasets.



	Llama 3.2 1B	Llama 3.2 3B	Llama 3.3 70B	Mistral 7B Q4	Stealth 1.2 7B	Qwen 2.5 7B	Qwen 2.5 32B
■ Series1	34%	27%	42%	39%	40%	36%	20%

Figure 6: Restricted LLM Agents Mode Choice Prediction Model Accuracy

4 Conclusions

The potential of AI agents to replicate human mode preference in scenario-based surveys was tested for four LL models under unrestricted sociodemographic and static prompting approach. Additionally, this potential was tested using seven LL models under the dynamic restricted agent's creation approach to align with the human sociodemographics. The results from the unrestricted direction do not significantly resemble the actual respondent's sociodemographics and create a mode choice distribution with 0.27 to 0.32 nRMSE values. The results highlight the variation between different LLM models in creating a heterogeneous population, and that correlates with the mode choice preference. A key highlight from this direction is the LLMs agent's potential in mode choice availability studies, in addition to the superior performance of the Stealth 1.2 7B, especially in mode choice distributions compared to the rest of the open access models tested. Under the restricted AI agent creation, a significant improvement was noticed for all models, highlighting the weakness of the tested LL models in population synthesis using looped context-based prompting. The accuracy of the seven tested models maxed at 42% for Llama 3.3 70B parameters, while the accuracy is low, its based on zeroshot prompting. There is a significant potential for increasing the overall models accuracy with few shot learning and parameters hyperparameters tuning, which we are currently investigating. We will also investigate the generalization of the methodology on a range of dataset, including a stated preference experiment for intercity travel by Wong & Farooq (2018).

Acknowledgements

We would like to express our gratitude to Dr. Mohesn Nizami and Dr. Nael Alsaleh, who conducted the original Stated Preference Survey with Brightwater buyers in October 2022 and for making their data available for the validation of the AI agent's responses. This research was funded by a grant from the Canada Research Chair program in Disruptive Transportation Technologies and Services (CRC-2021-00480) and NSERC Discovery (RGPIN-2020-04492) fund.

References

Ansar, M. S., Alsaleh, N., & Farooq, B. (2023). Behavioural modelling of automated to manual control transition in conditionally automated driving. *Transportation research part F: traffic psychology and behaviour*, *94*, 422–435.

Li, X., Huang, F., Lv, J., Xiao, Z., Li, G., & Yue, Y. (2024). Be more real: Travel diary generation using llm agents and individual profiles. *arXiv preprint arXiv:2407.18932*.

Mahmud, D., Hajmohamed, H., Almentheri, S., Alqaydi, S., Aldhaheri, L., Khalil, R. A., & Saeed, N. (2025). Integrating llms with its: Recent advances, potentials, challenges, and future directions. *arXiv preprint arXiv:2501.04437*.

Patterson, Z., Fitzsimmons, K., Jackson, S., & Mukai, T. (2019). Itinerum: The open smartphone travel survey platform. *SoftwareX*, *10*, 100230.

Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, *41*(5), 367–381.

Wang, X., Fang, M., Zeng, Z., & Cheng, T. (2023). Where would i go next? large language models as human mobility predictors. *arXiv preprint arXiv:2308.15197*.

Wong, M., & Farooq, B. (2018). Modelling latent travel behaviour characteristics with generative machine learning. In *2018 21st international conference on intelligent transportation systems (itsc)* (pp. 749–754).

Yan, H., & Li, Y. (2023). A survey of generative ai for intelligent transportation systems. *arXiv preprint arXiv:2312.08248*.