Artificial Intelligence and Social Computing, Vol. 122, 2024, 10–20
https://doi.org/10.54941/ahfe1004637

Application of Large Language Models in Stochastic Sampling Algorithms for Predictive Modeling of Population Behavior

Yongjian Xu[1], Akash Nandi[2], and Evangelos Markopoulos[3]

[1]Akasha-AI, London, N1C 4PF, United Kingdom
[2]Harvard University, Department of Economics, Cambridge, MA, 02138, United States of America
[3]University of Turku, Department of Mechanical Engineering, Turku, FI-20014, Finland

ABSTRACT

Agent-based modeling of human behavior is often challenging due to restrictions associated with parametric models. Large language models (LLM) play a pivotal role in modeling human-based systems because of their capability to simulate a multitude of human behavior in contextualized environments; this makes them effective as a mappable natural language representation of human behavior. This paper proposes a Monte Carlo type stochastic simulation algorithm that leverages large language model agents in a population survey simulation (Monte-Carlo based LLM agent population simulation, MCLAPS). The proposed architecture is composed of a LLM-based demographic profile data generation model and an agent simulation model which theoretically enables complex modelling of a range of different complex social scenarios. An experiment is conducted with the algorithm in modeling quantitative pricing data, where 9 synthetic Van Westendorp Price Sensitivity Meter datasets are simulated across groups corresponding to pairings of 3 different demographics and 3 different product types. The 9 sub-experiments show the effectiveness of the architecture in capturing key expected behavior within a simulation scenario, while reflecting expected pricing values.

Keywords: Large language models, Agent-based modelling, Turing experiments, Stochastic simulations, Monte-carlo simulations, Population behaviour modelling, Van Westendorp PSM, Pricing

INTRODUCTION

Agent-based modelling (ABM) has been widely used by economists and social science researchers to model a variety of complex and dynamic social systems; some examples include (Aher et al. 2023, Park et al. 2023, Axtel and Farmer 2022, Argyle et al. 2022, Loyall 1997, Macal 2016). The general methodology used in agent-based modelling requires the definition of an interactable state agent, the agent's environment, and the agent interaction mechanism that dictates how an individual agent behaves under certain conditions. Creating agent simulation models requires forming pre-existing logical hypotheses which guide the construction of the agent simulation conditions that are based on economic and behavioral theories; this causes the model's prediction of the dynamic process to be only as valid as the underlying assumptions. ABM systems are fundamentally chaotic and thus sensitive to changes in initial conditions; thus, marginally distinct underlying assumptions may lead to significantly different model predictions.

Traditional ABM methods are also limited by the complexity scope of the underlying process being modelled. As a result, creating effective ABMs requires the underlying logical assumptions to capture key process determinants.

As a demonstration of concept in ABM's ability to model group dynamics, a simple experiment is created with rudimentary state agent models to demonstrate emergent wealth distribution patterns within a group environment:

- **State Agent**: Each iteration, the agent can either *Replicate* or *Die* depending on the amount of resources consumed by the agent in the current turn. The outcome of each iteration is decided by a random sampling process that is a function of the current resources of the state agent:
  - If the agent's available resources $r < 1$, the agent consumes all available resources in an attempt to stay alive, where $P_{\text{death}} = 1 - r_{\text{consumed}}$
  - If the agent's available resources $r \geq 1$, the agent randomly consumes a quantity between $1 \leq r_{\text{consumed}} \leq r$ in an attempt to replicate (as survival is no longer a concern); $P_{\text{replication}} = r_{\text{consumed}} - 1$, so replication is guaranteed when 2 units of resources are consumed
  - If the agent consumes 0 to 10 times the guaranteed survival resource amount, the agent dies with probability $P_{\text{death}} = \frac{1}{9}(r_{\text{consumed}} - 1)$
- **Environment**: The environment hyperparameters are the number of state agents in the system, the total available resources per iteration among agents, and the resource distribution among agents:
- **Interaction Mechanism**: The resource is randomly uniformly distributed among all alive agents in the iteration. Adjusting the **distribution** variation to be less than one ensures a more equally random outcome for agents; a distribution with zero variance ensures complete equality of resource distribution.

In the model above, the described conditions reflect an oversimplified model of the fundamental dynamics of survival in an enclosed group environment with a fixed amount of resources naturally generated per time interval. Logically, there are 6 different possible cases of wealth distribution outcomes, some of which include (More cases are discussed in Online Appendix):

1. If there are sufficiently few resources relative to population size, the population dies off.
2. If the amount of resources relative to population exceeds a minimum threshold and does not exceed a maximum threshold, the population survives with a wealth distribution that varies within a certain range.
3. If the resource amount to population ratio exceeds a maximum threshold, the population dies off.

While simple ABM models are effective in capturing low-complexity social processes, more complex phenomena that depends on a large range of non-linearly behaving parameters and interaction mechanisms are inherently unfeasible for an agent model; this results from the lack of a state framework that can create accurate representations of the multitude of human characteristics and interaction dynamics, as well as the computational feasibility of such an algorithm in application. An example of such a complex process is product pricing; there are a multitude of factors that drive an individual's purchasing decision, which makes the state agent's representation of the act of purchasing highly unpredictable:

- The agent would need to have parameters modeling the purchasing behavior of different individual types, which need to include some representations with regard to price and likelihood of purchase. Fundamentally, this involves creating a utility model for the state agent.
- The agent would need an absolute concept of "money" in its state representation that is relative to its measure of utility.
- There needs to be a mapping of the "product" onto the state representation of the agent, analogous to an interaction mechanism that dictates the utility function of the agent.
- Assuming partial knowledge of the market, which in this case corresponds to other state agent's current parameters, an agent-agent interaction mechanism would need to be defined which affects the state of the observed agent.

The conditions above are in fact analogous to the following underlying assumptions:

1. An agent's willingness to pay is the amount of money an individual is willing to give up obtaining a good/service
2. All products and services are mappable as a function of
3. An individual's willingness to pay for a product is affected by that of other individuals in the environment

Thus, creating parametric models of the conditions above is unrealistic. Large language models, however, can serve as a means to create nonparametric models.

NATURAL LANGUAGE MODELLING OF BEHAVIOR

While natural language is not a unique representation of human behavior, it can be used as a viable representation, assuming that language is a self-referential system which can be used to describe all non-language-based representations. The framework for natural language-based behavioral modelling can be approached from a set theory perspective, whereby defining natural language and human behavior as sets creates a theoretical framework for understanding the theoretical feasibility of such a model; in particular, consider the following postulates and their associated implications:

- **Postulates:**
  - **Set of Tokens:** A finite first order set constructed of the smallest uniquely identifiable possible components of natural language defined by an arbitrary vector embedding model.
  - **Set of Natural Language:** The infinite uncountable set constructed from all possible combinations of the first order set of tokens, under the construction constraint that an element constructed in the set of natural language must possess semantic meaning.
    - **Associated Corollaries:**
      - **Corollary 1:** There exist elements of infinite length within the set of natural language.
      - **Corollary 2:** There exist elements in the set of natural language whose embedding vector representations are invariant relative to each other under a transformation of a high dimensional vector space spanned by the unique vector embeddings of the set of tokens.
    - **Set of Behaviors:** Assuming all possibilities of fundamental first order human behavior, which is defined as distinct forms of actions an individual can execute, can be uniquely and discretely identified, then it is possible to construct a countably infinite set containing all unique behaviors within an infinite amount of time.
- **The Behavior Theorem:** There exists a subset of the set of natural language in which all elements are finite in length and whose vector representation corresponds to a unique element in the set of behaviors.

Assuming the validity of the behavior theorem, then it is theoretically possible to construct the first order algorithm which maps natural language to a unique first order human behavior. Complex behavior can then be constructed from the first order algorithm.

LLM AGENT MODEL

Large language models are generative text-to-text models trained on human produced text, which based on the postulates proposed in the prior section would logically dictate that given a sufficiently large amount of training data the model would converge on a complete mapping of language and behaviour. Hence, a complete model with an infinite amount of text data would be able to encapsulate the infinite set of behaviours. LLMs are trained on human text, so they are theoretically aligned with text generation patterns of humans and their associated behavioral implications.

The key capability provided by LLMs is contextualized text input understanding in the input context window, which enables the model to generate a stimuli response from the input. However, LLMs alone are insufficient to act as a human behavioral simulacra:

- LLMs are static systems, so they cannot make inferencing actions unless data is provided (they lack self-persistency)
- LLM interactions are state-independent, meaning that new actions are not influenced by the state created by previous actions (they lack memory)

To address the limitations, a contextualized state memory and inferencing engine would be necessary to support the model in creating stateful inputs for the LLM to respond to. By defining a self-feeding input-output loop, the LLM can be integrated into a system where they can operate autonomously under some initialization condition and terminate once the halting condition has been met (see Figure 1).

Within the context of an agent-based model, the LLM agents forms the interactable microstates the system, where agent initial conditions are programmed in natural language based on the system environment state.

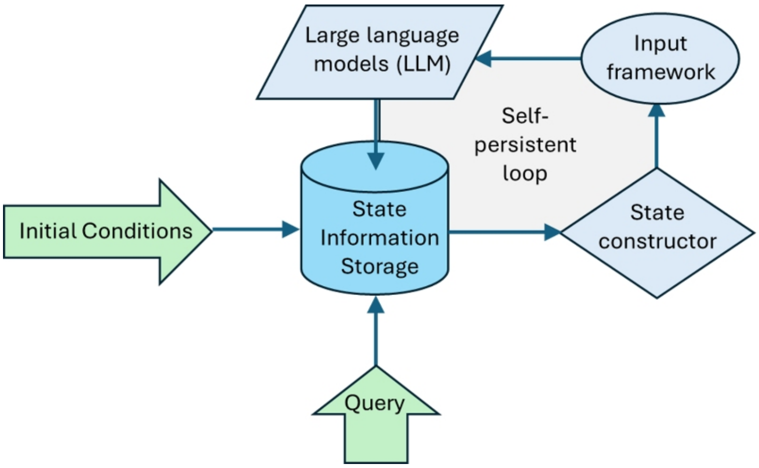

Figure 1: Conceptual diagram of agent state.

Upon definition of some initial agent state $a(t = 0)$, the agent would then evolve from the initial state based on a defined self-persistency rule $\theta(t,\ a(t))$:

$$\frac{da(t)}{dt} = \theta(t,\ a(t))$$

The necessary and sufficient condition for self-persistency is that state evolution must be non-zero, which makes the self-persistency stateful:

$$\frac{\partial\theta}{\partial a \partial t} \neq 0$$

Hence, there would be 3 stateful cases of $\theta(t,\ a(t))$ that determine the general evolution of the agent state:

1. Persistency is only timestep dependent:

$$\frac{\partial\theta}{\partial t} = f(t)$$

2. Persistency is dependent on agent state:

$$\frac{\partial\theta}{\partial t} = f\,(a(t)) \quad \frac{\partial\theta}{\partial a} = f(a),$$

3. Persistency is timestep and agent state dependent:

$$\frac{\partial\theta}{\partial a\partial t} = f(t,\ a(t))$$

LLM AGENT BASED SIMULATION MODELS

Integrating the LLM agent model architecture proposed, it is then possible to construct an ABM simulation comprised of LLM agents as state agents to model interaction processes. In an ABM system, the environment would possess a "world state" which is timestep dependent and evolves with the simulation. Defining the world state Π as a function of timestep and system microstates, there are 4 different cases of Π which can be considered in an ABM simulation:

1. Time-independent: the world state is constant

$$\frac{d\Pi(t)}{dt} = 0$$

2. Time-dependent: the world state evolves based on some predefined evolution rule

$$\frac{d\Pi(t)}{dt} = f(t)$$

3. Microstate dependent: the world state evolves as function of the system's microstates

$$\frac{d\Pi(t)}{dt} = f(a_0\,(\Pi_0(t), t)\,,\ a_1\,(\Pi_1(t), t)\,,\ a_2\,(\Pi_2(t), t)\,,\ \ldots,\ a_n\,(\Pi_n(t), t))$$

4. Microstate-time dependent: defined by a self-evolution condition and is affected by the microstates

$$\frac{d\Pi(t)}{dt} = f(t,\ a_0\,(\Pi_0(t), t)\,,\ a_1\,(\Pi_1(t), t)\,,\ a_2\,(\Pi_2(t), t)\,,\ \ldots,\ a_n\,(\Pi_n(t), t))$$

The system's microstates $a_i\,(\Pi_i\,(t)\,, t)$ are defined as the properties of each state agent at some given timestep. Hence, the system macrostate is defined as some computable global property of the system at a given timestep, which can be expressed as a function of microstates under a defined composition rule:

$$\Psi(t) = \ a_0\,(\Pi_0(t), t) * a_1\,(\Pi_1(t), t) * \ldots * a_n\,(\Pi_n(t), t)$$

In the framework proposed, the definition of the microstate's initial state conditions is defined by the simulation requirements, which can be a set of demographic profiles. Theoretically, if the world state model contains sufficient amounts of structured information, and agent states are individually well-defined, it is possible to simulate any complex environment under such model architecture.

PRICING SIMULATION EXPERIMENTS

A realistic pricing model of a product can be difficult to construct due the dependency of price on both the market as well as the unpredictability in purchasing behaviour of different individuals. However, within a simplified world model, an experiment can be constructed to probe the intrinsic value offered by a certain product towards a demographic group.

For the simplest constructable case of such an instance, the world state function is constant, $\frac{d\Pi(t)}{dt} = 0$. The agent's self-persistency behaviour is time dependent $\frac{\partial \theta}{\partial t} = f(t)$, with the predefined evolution function $f(t)$.

Translating to experiment parameters for a pricing simulation, the static world state is defined as a qualitative product description that does not evolve over time, and the evolution function $f(t)$ is defined as a series of automatically executed Van-Westendorp price sensitivity question queries for the state agent to iterate over until all queries are complete:

1. At what price in GBP would you consider the product to be so expensive that you would not consider buying it?
2. At what price in GBP would you feel the product quality couldn't be very good?
3. At what price in GBP would you consider the product starting to get expensive, so that it is not out of the question, but you would have to give some thought to buying it?
4. At what price in GBP would you consider the product to be a great buy for the money?

The agent initial conditions are constructed from a set of realistic synthetic demographic profiles (appendix) following a predefined structure generated using GPT-3.5-Turbo, which is akin to conducting a random sampling of a population within a demographic group constrained by age and income. The macrostate being measured in this case is the key price points of the cumulative % frequency graphs of Van-Westendorp Price Sensitivity Meter (Survey Monkey n.d.):

- Optimal Price Point (OPP): The price point where the probability of purchase is the highest (SurveyMonkey, 2024)
- Point of Marginal Cheapness (PMC): Price for which purchase probability becomes low due to low perceived quality of the product (SurveyMonkey, 2024)
- Point of Marginal Expensiveness (PME): Price of in which purchase probability becomes low due to the relatively low value to cost ratio of the product (SurveyMonkey, 2024)

Data points from the simulation are then extracted from the output of the state agent at each iteration timestep. The experiment is conducted for

3 different demographic groups and 3 different products of the same type. The product type chosen is smartphones, since their perceived value is well understood, which gives an intuitive foreground for interpreting the synthetic dataset:

- Product 1: Mid-ranged budget smartphone (appendix ref)
- Product 2: High-end foldable smartphone (appendix ref)
- Product 3: High-end flagship smartphone (appendix ref)

The 3 different demographic groups are varied mainly by age and income:

- Demographic 1: age 22-29, income £26000 to £45000
- Demographic 2: age 30-44, income £60000 to £85000
- Demographic 3: age 45-60, income £100000 to £135000

The language model-based synthetic demographic profile generator then creates a realistic full demographic profile based on constraint parameters, which are used to define the initial agent state.

For each of the 9 product-demographic combinations, we conduct a synthetic PSM. Based on the experiment setup, criterions for determining the viability of the synthetic dataset in creating an accurate and realistic predictive model of the perceived value of a product for a given demographic is determined through analysis of the system macrostate; more specifically, we should observe the following:

1. Each synthetic PSM should show that the percentage of agents who report that the product is too expensive at a particular price should be increasing in price.
2. Each synthetic PSM should show that the percentage of agents who report that the product is too cheap at a particular price should be decreasing in price.
3. Each synthetic PSM should show that the percentage of agents who report that the product is expensive but purchasable at a particular price should be increasing in price.
4. Each synthetic PSM should show that the percentage of agents who report that the product is cheap but not concerningly cheap at a particular price should be decreasing in price.
5. Each synthetic PSM should show that the percentage of agents who report that the product is too expensive at a particular price should be weakly greater than the percentage of agents who report the product is expensive but purchasable at that same price.
6. Each synthetic PSM should show that the percentage of agents who report that the product is too cheap at a particular price should be weakly less than the percentage of agents who report the product is cheap but not concerning cheap at that same price.

We define the above to be our six expectations for the results of a PSM experiment. In the following section, we examine their validity relative to the results of our synthetic PSM experiments.

EXPERIMENTAL RESULTS AND ANALYSIS

Consider the results of the PSM corresponding to Product 1 and Demographic 1. Figure 2 shows for any given price the percentage of respondents who believe that the product is too expensive, the percentage who believe that it is too cheap, the percentage who believe it is expensive but purchasable, and the percentage who believe it is cheap but not concerningly cheap (Xu & Nandi, 2024).

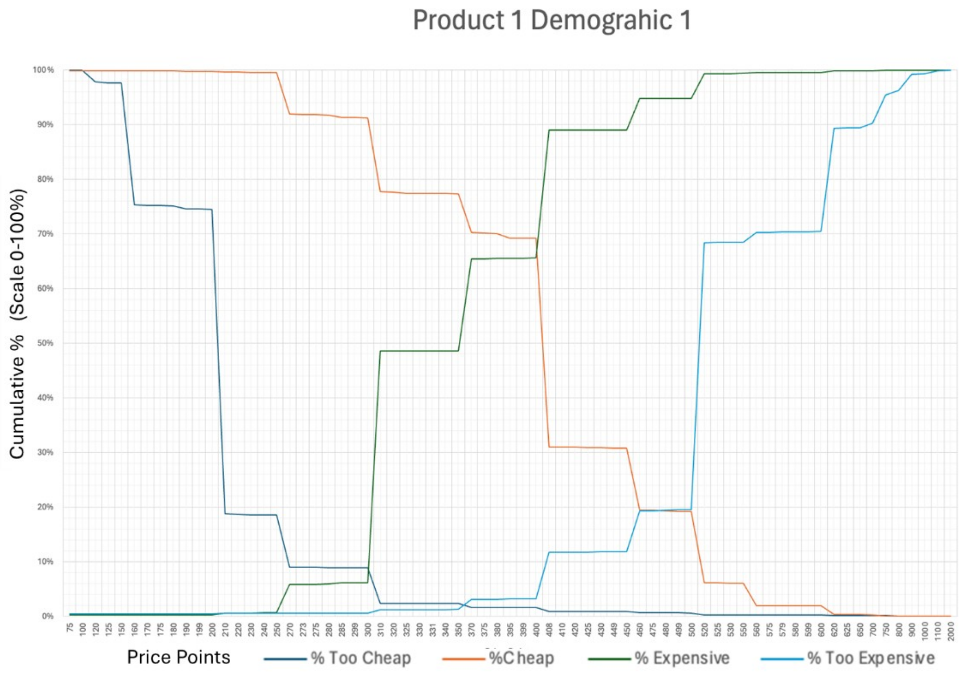

Figure 2: PSM graph for sub-experiment 1-1.

As we can clearly observe, each of the six expectations associated with the results of a natural PSM experiment hold in our synthetic PSM for this product-demographic pair. Furthermore, as shown in the appendix, these six expectations hold for all nine synthetic PSM experiments that we run.

Moreover, as in a natural PSM, we are able to calculate the Point of Marginal Cheapness (PMC), the Optimal Price Point (OPP), and the Point of Marginal Expensiveness (PME). Table 1 lists the observed PMC, OPP, and PME for each of the 9 synthetic PSM experiments.

For each product-demographic group, we are able to calculate the PMC, OPP, and PME by observing the appropriate intersection points of our four trend lines. The fact that our six expectations regarding the results of a PSM experiment hold in all nine synthetic PSM experiments allows us to observe the intersections needed to calculate these values.

Table 1. Key price points found in simulation experiment.

Product-Demographic Group	PMC	OPP	PME
Product 1, Demographic 1	305	355	480
Product 1, Demographic 2	355	470	610
Product 1, Demographic 3	405	505	615
Product 2, Demographic 1	505	620	680
Product 2, Demographic 2	550	750	950
Product 2, Demographic 3	610	810	1065
Product 3, Demographic 1	500	605	830
Product 3, Demographic 2	550	730	1100
Product 3, Demographic 3	625	800	1050

CONCLUSION

Through the experiment, it has been demonstrated for a simple case of a LLM agent-based simulation in PSM pricing that the method is valid in generating expected macroscopic behavioral results from a synthetic population. Further experiments can be conducted with different product types and demographics to further validate the method, as well as construction of more complex interactable world state simulation to simulate time-dependent evolutions of the simulation system.

APPENDIX

Link to online appendix: https://1drv.ms/w/s!Aovld1VbktTPiiVgvxpqJIK1bBnm.

REFERENCES

Aher, G., Arriaga, R. I., & Kalai, A. T. (2023, February 14). Using large language models to simulate multiple humans and replicate human subject studies. arXiv.org. https://arxiv.org/abs/2208.10264

Argyle, L. P., Busby, E., Fulda, N., Gubler, J. R., Rytting, C. M., & Wingate, D. (2022). Out of One, Many: Using Language Models to Simulate Human Samples. Political Analysis, 31, 337–351.

Axtell, R. L., & Farmer, J. D. (2022, June 21). Agent-based modeling in Economics and Finance - inet Oxford. https://www.inet.ox.ac.uk/files/JEL-v2.0.pdf

Loyall, A. B., 1997. Believable Agents: Building Interactive Personalities. Carnegie Mellon University, School of Computer Science, Computer Science Department. https://www.cs.cmu.edu/Groups/oz/papers/CMU-CS-97-123.pdf

Macal, C. M. (2016, May 10). Everything you need to know about agent-based modelling and Simulation - Journal of Simulation. SpringerLink. https://link.springer.com/article/10.1057/jos.2016.7

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, April 7). Generative agents: Interactive simulacra of human behavior. arXiv.org. https://arxiv.org/abs/2304.03442(pdf)

SurveyMonkey. (2024). How to use the van Westendorp Price sensitivity Meter. https://uk.surveymonkey.com/market-research/resources/van-westendorp-price-sensitivity-meter/

Xu, Y., & Nandi, A. (2024). Synthetic Van Westendorp Price Sensitivity Dataset for 3 Different Smartphones in 3 Different Demographics. March 13, 2024. https://1drv.ms/x/s!Aovld1VbktTPiiGQlK_sMCDFRaYG

Xu, Y. (2024). Jerrypanda4563/population_game: A simple jupyter notebook style simulation of the evolution of resource distribution amoung a population over time. GitHub. March 13, 2024. https://github.com/jerrypanda4563/population_game

Xu, Y. (2024a, March 9). Monte Carlo Based Large Language Model Agent Population Simulation. GitHub. March 9, 2024. https://github.com/jerrypanda4563/MCLAPS