Enhancing spatially-disaggregated simulations with large language models

ilya zaslavsky*
San Diego Supercomputer Center, University of California San Diego
izaslavsky@ucsd.edu

Jiaxi Lei
San Diego Supercomputer Center, University of California San Diego
jil1119@ucsd.edu

Rishi Graham
School of Medicine, University of California San Diego
regraham@health.ucsd.edu

Mark S. Handcock
Department of Statistics, University of California Los Angeles
handcock@ucla.edu

Eliah Aronoff-Spencer
School of Medicine, University of California San Diego
earonoffspencer@health.ucsd.edu

ABSTRACT

We present our experience integrating large language models (LLMs) and simulation engines to enhance spatially-disaggregated simulation, taking advantage of the spatial knowledge and spatial reasoning capabilities of LLMs. The examples illustrate LLM integration with different variations of compartmental epidemiological models, including agent-based models (ABM) in the context of modeling COVID-19 infection spread in a school setting, and LLM integration with a system dynamics model which supports a serious game focused on strategies for responding to disease outbreaks at the county level. We present the architecture of the integrated LLM-simulation system, demonstrate the initial results, and discuss the challenges of the current approach, related to LLM's understanding of spatial information and spatial relationships, their reasoning capabilities, and model performance and scalability.

CCS CONCEPTS

• **Modeling and simulation**; • **Natural language processing**; • **Knowledge representation and reasoning**;

KEYWORDS

System Dynamics, Large Language Models, Agent-Based Models, Spatially-Disaggregated Models, Epidemiological Modeling

ACM Reference Format:
ilya zaslavsky*, Jiaxi Lei, Rishi Graham, Mark S. Handcock, and Eliah Aronoff-Spencer. 2024. Enhancing spatially-disaggregated simulations with large language models. In *2024 7th International Conference on Mathematics and Statistics (ICoMS 2024), June 23–25, 2024, Amarante, Portugal*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3686592.3686595

1 INTRODUCTION

The goal of this study is to improve strategies and assess human responses to public health interventions during emergencies, such as the COVID-19 pandemic. These responses are influenced by a variety of local factors and individual characteristics, making them difficult to accurately represent in formal models. In epidemiology, the spread of disease is commonly viewed as a spatial process, as the transmission of infectious diseases largely relies on the spatial arrangement and interactions among individuals or populations. This includes modeling both direct (person-to-person contact) and indirect (e.g., via a shared environment with a high viral load) spatial interactions, tracing and forecasting infection pathways, and devising health interventions, which often necessitates data on geographic locations, the distances between individuals and populations, and their movement patterns. To map disease transmission, agent-based models (ABM) and metapopulation models (e.g., [1], [2], [3]) explicitly incorporate spatial interactions at the level of individuals or subpopulations in different locations, respectively. The SEIR model, a prevalent compartmental epidemiological model, categorizes the population into Susceptible (S), Exposed (E), Infectious (I), and Recovered (R) compartments and describes the flow of individuals through these compartments over time, influenced by transmission rates, incubation periods, and recovery rates [4]. In this study, we examine models that merge SEIR methodologies with both ABM and metapopulation models, resulting in spatially-aware compartmental ABMs and spatially-disaggregated system dynamics (SD) models for tracking disease spread.

Implementing such models and simulating effective public health strategies requires extensive local data and an understanding of spatial configurations and movements. Trained on vast text corpora, large language models (LLMs) can potentially offer the necessary spatial intelligence and refine these simulations by adapting to local conditions already recognized by the models. This paper outlines our initial efforts to integrate LLMs with ABM and SD SEIR models, aiming to create more realistic and location-sensitive simulations without the need for explicitly programmed spatial information. In addition, we aim to investigate the types and volumes of additional local demographic and socio-economic data, along with patterns of spatial relationships, which could further enhance LLMs for spatial decision-making. We also seek to develop a modular modeling framework that combines LLMs with traditional simulation engines, capitalizing on the strengths of each to allow for their independent update and calibration while ensuring their interoperability through an API. The next section describes prior work and our methodology for augmenting simulations with LLMs. It is followed by a summary

ICoMS 2024, June 23–25, 2024, Amarante, Portugal

ACM ISBN 979-8-4007-0722-3/24/06
https://doi.org/10.1145/3686592.3686595

of findings with respect to LLM capabilities to enhance simulations and concludes with future directions.

2 METHODOLOGY

The rapid advancement of Large Language Models (LLMs) has significantly expanded their capabilities, yet their spatial analysis and reasoning abilities are not fully understood. Recent research examined a range of pre-trained LLMs, including testing geospatial skills of ChatGPT in terms of basic concepts of spatial analysis, mapping and spatial data management [5] and qualitative spatial reasoning [6], and comparing performance of several models on prompts involving understanding of geospatial predicates [7]. The Llama-2 LLM learns basic linear representations of space and time, as shown in [8]; while [9], [10], [11] demonstrated that LLMs can capture spatial structure and spatial relationships given various types of geospatial location embeddings, and [12] described LLM's potential in assisting with the generation of SQL code for querying spatial databases. Despite these advancements, each study underscored the need for further improvement in LLMs' spatial reasoning capabilities and raised additional research questions for future work.

Building on the pioneering work that introduced generative agents [13], subsequent studies explored the concept of adding information about agent personas to refine agent-based models, resulting in what are known as Generative ABMs (GABMs) [14], [15], [16], [17]. This approach has been applied to various fields, including epidemiology [18]. The studies also documented challenges in performance and scalability, as the models were limited to managing only a small number of generative agents.

In the realm of System Dynamics (SD) models, while LLMs have been employed to aid in the construction and evaluation of simulation models [19], their use has not extended to enhancing simulations through interactive engagement between a simulation engine and an LLM at regular time steps, to our knowledge.

In our work, we have integrated LLMs with both spatially-aware ABM and SD models. The two examples presented below extend our earlier and ongoing modeling work on disease outbreak prediction and prevention, primarily using COVID-19 infection data for the County of San Diego. In each case, the LLM serves to augment an existing model by examining simulation outputs at selected steps and adjusting the weights of different spatial preferences based on key infection characteristics (as computed by the simulation engine) and local knowledge (as retrieved from LLM).

2.1 The ABM model of infection transmission in schools

This first case is an ABM developed to forecast the impacts of pharmaceutical and non-pharmaceutical interventions on the spread of COVID-19 in an elementary school setting [20], [21]. The model describes the dynamics of infection among students and school staff, who were categorized into several states: susceptible, exposed, infected symptomatic, infected asymptomatic, and removed from school/isolated. Importantly, the model included spatial information about classrooms and other school areas (including school buses), characterized by their viral loads, ventilation quality, and

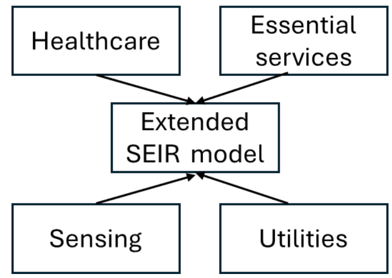

Figure 1: Main modules in the SD model underlying the series game focused on community resilience to disease outbreaks.

spatial layouts. It simulated regular school operations over a two-week period, mirroring common school activities and using school floor plans from the San Diego Unified School District. These activities, including studying at desks, classroom group projects, lunch, and recess in the school yard, were associated with specific movement patterns, which influence the likelihood of close contacts with an infected person, based on spatial proximity and duration. Also, an agent's change to an exposed or infected status could be triggered by agent's location within a given space (room) with elevated viral load. The model proved its effectiveness in assessing the efficacy of various intervention strategies, such as mask mandates, staggered attendance schedules, designated seating arrangements on buses, adjustments to HVAC settings, relocating lunch spaces, and increasing testing frequency — on reducing infection rates under different viral strains and local conditions.

Implemented using the MESA and MESA-GEO Python libraries [22], the model integrates ABM simulation with Geographic Information System (GIS) capabilities. To enrich the simulation with advanced decision-making and social interaction dynamics, we interfaced the model with pre-trained LLMs like OpenAI's ChatGPT 3.5 and Anthropic's Claude 2, as well as with a local LLM developed using the Flan-Alpaca framework [23]. The refined model incorporates detailed agent personas, detailing their mobility and social preferences. For instance, it characterizes students by their likelihood to play outside or spend time with friends. These personas maintain unique narratives, which the LLM utilizes to simulate agent actions. The model considers several characteristics including the agent's location, identity and health status of surrounding agents, and the physical and environmental attributes of the room itself. The duration and intensity of social interactions for each agent's behavior are then used to simulate the health status of each agent at subsequent time steps. Decoupling LLM from the ABM allowed us to explore how different models enable dynamic and context-sensitive generation of agent actions based on the unique personas and situational variables.

2.2 The system dynamics model underlying serious games on responding to disease outbreaks.

The second model is a System Dynamics (SD) model used within an experimental serious game framework,

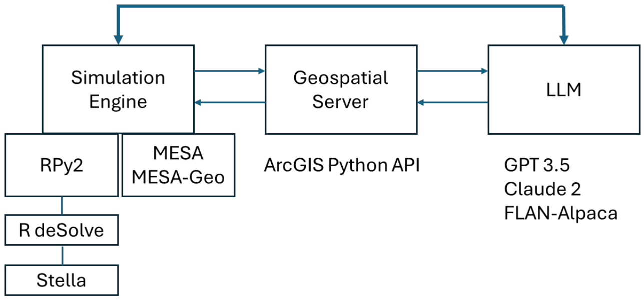

Figure 2: The overall architecture of LLM integration with simulation systems

designed to aimed at enhancing response strategies to infectious disease outbreaks in regions such as San Diego County [24]. The serious game was designed to foster community resilience to disease outbreaks by enhancing individual capacities in governance and security, essential services, healthcare, sensing, surveillance, vaccine development, and combating misinformation. These capacities, including financial, human and social capacities, represent the main stocks in the model (Figure 1). Participants in the game could select from a range of intervention strategies based on recommendations from authoritative bodies such as [25], or to devise their own strategies. The strategies included investments in detection systems (public health surveillance, contact tracing, laboratory testing), community mitigation measures (public health messaging, social media, and law enforcement measures), medical interventions (working with hospitals, vaccine manufacturers, and pharmaceutical companies), improvements in healthcare system preparedness (such as managing hospital surge capacity), and investments into scientific infrastructure. To evaluate the effectiveness of the suggested strategies and their combinations, game judges translated them to scenario parameters and ran the simulations using a system dynamics model implemented using the Stella software platform. This model allowed players to explore trajectories and long-term outcomes of their strategic decisions, providing insights into the potential impacts of various interventions on community resilience and outbreak management.

The diversity within a population, variations in community characteristics, differing healthcare capacities, and other variables make it difficult to quantify the model, accurately assess resilience, and compare the effectiveness of various response strategies across a region as large and varied as San Diego County. To accommodate the spatial heterogeneity inherent in such a large area, the SD model was enhanced to allow for the simulation of selected strategies at the zip code level. The choice of zip code-specific strategies was facilitated by a LLM, leveraging its extensive knowledge base on the characteristics and demographics of different locations. Meanwhile, the computation of regional capacities and SEIR stock values is carried out within the Stella platform. The integration of an LLM into the strategy selection process enabled allocation of resources tailored to the unique needs and characteristics of each area.

The integration of the LLM with the Stella model is realized through a Python module that facilitates communication between the LLM and Stella at simulation steps. The Stella model itself is encapsulated within an R wrapper, which uses the model's XMILE representation to enable dynamic parameter adjustments, model loading, and execution within the Stella environment via R's deSolve package. This hybrid design approach, where the LLM's execution is decoupled from the simulation engine, offers greater flexibility in model calibration and sensitivity analysis, as adjustments to the simulation parameters or to the LLM prompts can be made independently, and the components can be independently validated and tested. At the same time, this decoupling increases computational overhead and slows the simulation process, especially for large and complex models. We expect that the simulation stack will continue to evolve towards achieving an optimal balance between execution efficiency and the analytical depth and flexibility provided by this hybrid approach.

The overall framework in Figure 2 integrates a geospatial server, which hosts multiple layers of local spatial data accessible via ArcGIS Python API. This setup is a key part of our Retrieval-Augmented Generation (RAG) [26] system, designed to enhance LLM prompts with external information from document repositories and databases. At present, due to token limitations, the geospatial server does not feed information directly to the LLM during simulations. Instead, it contributes to the initial setup of the model, ensuring that spatial data informs the simulation's parameters and environment from the outset.

3 RESULTS

3.1 The ABM – LLM integration

Our initial observations from the integration of Large Language Models (LLMs) with the ABM described above, have yielded several insights regarding agent behavior, consistency challenges, and model performance and scalability limitations:

- The integration successfully captured complex behaviors exhibited by agents, which are influenced by individual mobility preferences. This detailed representation offers a more authentic and nuanced depiction of how various factors, including social interactions and movement patterns, impact the spread of infectious diseases.
- One of the challenges we encountered involves maintaining consistency in agent interactions. In some instances, for example, Agent A attempted to interact with Agent B while Agent B was simultaneously involved in another activity. Such inconsistencies highlight the need for coordination mechanisms within the model to ensure seamless agent behaviors across simulations.
- In the current implementation, LLM is queried multiple times per time step for each agent, which presents serious performance and scalability challenges. As the complexity and number of agents and their personas increase, the model becomes increasingly slow. To address this limitation, we are refining the model's performance and consistency, e.g., using heuristics based on insights from previous model interactions.

3.2 The SD – LLM integration

From our first experiments in integrating LLM with the Stella system for spatially-disaggregated modeling of disease outbreak responses, we make the following observations regarding the interplay between LLM geospatial reasoning capabilities, spatial data integration, and computational efficiency.

- The LLMs have generated meaningful insights into spatially-disaggregated scenarios, proposing tailored dynamic resource allocation strategies for specific areas, such as prioritizing research efforts and vaccine distribution among students in zip codes housing universities —despite the absence of explicit programming of such localized details into the model. However, the performance varied across different areas, showing a tendency to offer more detailed insights on well-documented areas, such as those with universities or downtown events, which are likely to have a higher density of contact and textual descriptions in the training corpus. For the same reason, focusing on spatial disaggregation by named neighborhoods rather than zip codes appears a more promising approach. This method aligns better with the inherent strengths of LLMs in recognizing and processing information based on named locations.
- Token limits of pre-trained LLMs restrict the inclusion of detailed spatial data in prompts. Training LLMs with various types of spatial data and developing a hybrid system where an LLM's partial spatial understanding is supplemented by a specialized spatial reasoner appears a promising, though resource-intensive and complex, strategy.
- A robust software framework is essential for systematically assessing how LLMs integrate spatially-disaggregated SEIR data from models into decision-making about resource allocations. Such a framework would address the inefficiencies of our current implementation, with delays in processing within the Stella wrapper and during LLM queries.
- Given the model's substantial computational requirements, identifying methods to group similar zip codes promises to reduce the frequency of LLM queries.
- Resources relevant to outbreak response are typically managed by various county agencies rather than a single regional authority. This complexity mirrors the organization of the serious game, where players often represent different organizational domains. In the next iteration of the model, we will include several county agencies, characterized to an LLM by their mission statements and available strategies, similar to agent personas in the LLM-ABM integration.

4 CONCLUSIONS AND FUTURE WORK

Our experiments integrating LLMs into simulation systems, including spatially-aware ABMs and SD epidemiological models, have yielded findings that align with earlier observations of LLMs' geospatial knowledge and reasoning abilities mentioned in section 2. Furthermore, these experiments have provided added insights and highlighted challenges pertinent to spatially-disaggregated simulation modeling, as LLMs were tasked with allocating resources or fine-tuning behaviors based on agent personas or area characteristics without being directly prompted about these properties.

The main challenges we encountered included understanding the extent of LLM's spatial reasoning abilities and adapting simulations to better align with these capabilities. Capitalizing on LLM's inherent knowledge, which, for instance, shows a stronger familiarity with neighborhood names over zip codes, would lead to more accurate models. Ensuring robustness, refining LLM prompts, conducting sensitivity analysis, and improving the interpretability of results were additional issues we addressed, albeit to a limited extent. One significant observation is the LLMs' underdeveloped ability to use information about spatial proximity between zip codes or other neighborhoods, whether defined through adjacency, k-nearest neighborhoods, or distances. Future work will aim to improve LLMs' spatial reasoning through experimentation with embedding of varying levels of neighborhood data including socio-demographic and economic profiles and contiguity and distance matrices. Given the impracticality of encoding all spatial relationships explicitly, a balance must be found between what spatial knowledge is directly included in simulations and what nuances are left for LLMs to infer. The balance between explicit model details and LLM inferential capabilities is crucial for advancing our understanding of LLM applications in simulation modeling. Addressing the allocation of different types of information into LLMs—whether through training, Retrieval-Augmented Generation (RAG), or prompts—is another critical area of exploration.

Technical implementation challenges also emerged, especially concerning the computational load of querying LLMs, the performance of the Stella wrapper, the management of token limits, and memory retention across iterations. Exploring alternative modeling strategies, such as a hybrid Agent-Based Model-Partial Differential Equation (ABM-PDE) approach, could offer scalable solutions to these issues.

ACKNOWLEDGMENTS

NSF support (awards 2139740, 2200197) is gratefully acknowledged)

REFERENCES

[1] N. Gilbert, Agent-Based Models. SAGE Publications, 2019.
[2] S. Riley, "Large-Scale Spatial-Transmission Models of Infectious Disease," Science, vol. 316, no. 5829, pp. 1298–1301, Jun. 2007, doi: 10.1126/science.1134695.
[3] A. L. Lloyd and R. M. May, "Spatial Heterogeneity in Epidemic Models," J. Theor. Biol., vol. 179, no. 1, pp. 1–11, Mar. 1996, doi: 10.1006/jtbi.1996.0042.
[4] H. W. Hethcote, "The Mathematics of Infectious Diseases," SIAM Rev., vol. 42, no. 4, pp. 599–653, Jan. 2000, doi: 10.1137/S0036144500371907.
[5] P. Mooney, W. Cui, B. Guan, and L. Juhász, "Towards Understanding the Geospatial Skills of ChatGPT: Taking a Geographic Information Systems (GIS) Exam," in Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, in GeoAI '23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 85–94. doi: 10.1145/3615886.3627745.
[6] A. G. Cohn, "An Evaluation of ChatGPT-4's Qualitative Spatial Reasoning Capabilities in RCC-8." arXiv, Sep. 27, 2023. doi: 10.48550/arXiv.2309.15577.
[7] P. Bhandari, A. Anastasopoulos, and D. Pfoser, "Are Large Language Models Geospatially Knowledgeable?," in Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems, in SIGSPATIAL '23. New York, NY, USA: Association for Computing Machinery, Dec. 2023, pp. 1–4. doi: 10.1145/3589132.3625625.
[8] W. Gurnee and M. Tegmark, "Language Models Represent Space and Time." arXiv, Mar. 04, 2024. doi: 10.48550/arXiv.2310.02207.
[9] Y. Yamada, Y. Bao, A. K. Lampinen, J. Kasai, and I. Yildirim, "Evaluating Spatial Understanding of Large Language Models." arXiv, Mar. 05, 2024. doi: 10.48550/arXiv.2310.14540.
[10] Y. Ji and S. Gao, "Evaluating the Effectiveness of Large Language Models in Representing Textual Descriptions of Geometry and Spatial Relations." arXiv, Jul. 04, 2023. doi: 10.48550/arXiv.2307.03678.

[11] S. Tucker, "A systematic review of geospatial location embedding approaches in large language models: A path to spatial AI systems." arXiv, Jan. 12, 2024. doi: 10.48550/arXiv.2401.10279.
[12] Y. Jiang and C. Yang, "Is ChatGPT a Good Geospatial Data Analyst? Exploring the Integration of Natural Language into Structured Query Language within a Spatial Database," ISPRS Int. J. Geo-Inf., vol. 13, no. 1, Art. no. 1, Jan. 2024, doi: 10.3390/ijgi13010026.
[13] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 2023, pp. 1–22.
[14] L. Wang *et al.*, "A survey on large language model based autonomous agents," ArXiv Prepr. ArXiv230811432, 2023.
[15] A. S. Vezhnevets *et al.*, "Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia," ArXiv Prepr. ArXiv231203664, 2023.
[16] N. Ghaffarzadegan, A. Majumdar, R. Williams, and N. Hosseinichimeh, "Generative agent-based modeling: an introduction and tutorial," Syst. Dyn. Rev., vol. 40, no. 1, p. e1761, Jan. 2024, doi: 10.1002/sdr.1761.
[17] Z. Xi *et al.*, "The rise and potential of large language model based agents: A survey," ArXiv Prepr. ArXiv230907864, 2023.
[18] R. Williams, N. Hosseinichimeh, A. Majumdar, and N. Ghaffarzadegan, "Epidemic modeling with generative agents," ArXiv Prepr. ArXiv230704986, 2023.
[19] A. Akhavan and M. S. Jalali, "Generative AI and Simulation Modeling: How Should You (Not) Use Large Language Models Like ChatGPT," SSRN Electron. J., 2023, doi: 10.2139/ssrn.4675409.
[20] K. Ganapathy, B. Man, J. Lei, and I. Zaslavsky, "Geographically-enabled Agent-based Model for COVID-19 Transmission," presented at the Gateways 2021, 2021. doi: 10.5281/zenodo.5569450.
[21] "GEOACT," GEOACT. Accessed: Mar. 19, 2024. [Online]. Available: https://geoact.org/
[22] B. Wang, V. Hess, and A. Crooks, "Mesa-Geo: A GIS Extension for the Mesa Agent-Based Modeling Framework in Python," in Proceedings of the 5th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation, in GeoSim '22. New York, NY, USA: Association for Computing Machinery, Nov. 2022, pp. 1–10. doi: 10.1145/3557989.3566157.
[23] D. Ghosal, Y. K. Chia, N. Majumder, and S. Poria, "Flacuna: Unleashing the Problem Solving Power of Vicuna using FLAN Fine-Tuning." arXiv, Jul. 05, 2023. doi: 10.48550/arXiv.2307.02053.
[24] Resilient Collective, "A Dynamic Systems Model for Serious Games about Community Resilience to Infectious Disease Outbreaks." 2023.
[25] The White House, "National COVID-19 Preparedness Plan," 2022. [Online]. Available: https://www.whitehouse.gov/wp-content/uploads/2022/03/NAT-COVID-19-PREPAREDNESS-PLAN.pdf
[26] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020, pp. 9459–9474. Accessed: Apr. 04, 2024. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html