Enhancing Spatial Reasoning and Behaviour in Urban ABMs with Large-Language Models and Geospatial Foundation Models

Nick Malleson*[1], Andrew Crooks[2], Alison Heppenstall[3] and Ed Manley[1]

[1]School of Geography, University of Leeds, UK
[2]Department of Geography, University at Buffalo, US
[3]School of Social and Political Sciences, University of Glasgow, UK

February 14, 2025

Summary

Modelling human behaviour continues to be a significant challenge for the field of agent-based modelling, and one that prohibits the development of comprehensive empirical ABMs for urban applications, such as Urban Digital Twins. However, two recent methodological advances offer the potential to transform empirical agent-based models. Early evidence suggests that large-language models (LLMs) can be used to represent a wide range of human behaviours, with models responding in realistic ways to given prompts. Indeed there is already a flurry of activity that focusses on implementing LLM-backed agents – i.e. agents who are controlled by LLMs. At the same time, the concept of the foundation model is also being applied in domains beyond text analysis. Of particular interest are geospatial foundation models that automatically encode spatial data in such a way as to associate different spatial objects in numerous and nuanced ways that have otherwise alluded manual classification schemes. Taken together, these two technologies offer considerable potential for a new generation of agent-based models that contain agents who can behave in response to spatial and social prompts in a way that is realistic and has so far proven impossible to replicate using manually-programmed behavioural rules.

This paper presents a discussion of the state of the art in both LLMs and geospatial foundation models in the context of their potential role in agent-based modelling. It discusses the transformational potential of these technologies and outlines the critical questions that need to be addressed before they can be used to create robust, reliable and trustworthy models for empirical policy applications that support decision-making.

KEYWORDS: Agent-based Modelling; Large language model; Geospatial foundation model; Urban Modelling.

*Corresponding author n.s.malleson@leeds.ac.uk. ORCID: 0000-0002-6977-0615

1 Introduction

Agent-based modelling (ABM) is a form of computer simulation in which the individual components that drive the system (the 'agents') are modelled directly. This contrasts with aggregate approaches where higher-level equations are used to attempt to represent the underlying components. ABM has shown promise as a useful method for simulating systems where the behaviours and interactions of the individual components are essential in understanding and predicting the behaviour of the wider system (Crooks et al. 2019). However, simulating the behaviour of *people* remains a key challenge (Heppenstall et al. 2021) that prevents ABMs from being more widely used in policy applications (Badham et al. 2018). This is especially true in the realm of Urban Digital Twins, where solutions largely fail to embed humans (Malleson et al. 2024), leaving them "sterile" (Fotheringham 2023).

Behaviour in agent-based models (ABMs) has typically been implemented directly through logical rules or indirectly through the use of more comprehensive behavioural frameworks or even neural networks and genetic algorithms (DeAngelis & Diaz 2019). However, all of the previous approaches have serious limitations – see DeAngelis & Diaz (2019) for details. The rise of two technologies: large language models (LLMs) and geospatial foundation models (GFMs) may offer a breakthrough. Early evidence suggests that LLMs are able to respond in a 'human-like' way to given prompts, generating "believable" (Park et al. 2023) behaviour. At the same time, geospatial foundation models are emerging as tools that automatically encode spatial data, capturing complex associations between spatial objects in nuanced ways that often elude manual classification schemes (Lu et al. 2024). Taken together, these two technologies offer considerable potential for a new generation of agent-based models that contain agents who can behave in response to spatial and social prompts in a way that is realistic and has so far proven impossible to replicate using manually-programmed behavioural rules.

2 Agents as Large-Language Models

The rapid rise in popularity of LLMs, such as ChatGPT, has lead to a flurry of activity as researchers experiment with the potential opportunities offered by LLMs to produce more realistic and nuanced simulations of human behaviour and decision-making (Gao et al. 2023). Projects such as AutoGPT, BabyAGI, Generative Agents, MetaGP (for details see Ma et al. 2024) are exploring the extent to which LLMs can give agents advanced capabilities such as natural language understanding, reasoning and planning. Some have already produced toy agent-based models with agent behaviour controlled by distinct LLMs (for example see Park et al. 2023, Li et al. 2023).

However, the integration of an LLM into an ABM is not straightforward for a number of reasons. Firstly, unlike with traditional ABMs where the agent is programmed in the context of their environment and can interact with their environment directly, LLMs interact with the outside world via text prompts. Hence it is necessary to *describe* the agent's environment using language as the LLM input, and translate the LLM output language into actions. To this end, bespoke frameworks to allow LLMs to represent agents are emerging (e.g. Li et al. 2023, Vezhnevets et al. 2023, Williams et al. 2023, Li et al. 2024) although these are still in early stages.

A second challenge relates to the computational complexity of an LLM. Even trained LLMs have considerable compute and memory requirements that are exacerbated by attempting to run multiple LLMs simultaneously (i.e. one per agent). Instead, Chopra et al. (2024) suggest the use of "archetypes" whereby individual agents are grouped and assigned to a small number of distinct LLMs, such that the LLM only needs to be queried once to simulate the behaviour of a whole group of agents. This leads to fewer LLM computations but will reduce the behavioural heterogeneity of the agents.

Thirdly, the use of LLMs will clearly bring issues of bias and representation to the fore. Although "LLMs have been trained on massive amounts of human culture" (Vezhnevets et al. 2023) the "long-tail effect" (Xi et al. 2023) means that they may not be representative of minorities, marginalised groups, or those who do not engage digitally to the same extent as others (i.e. the digital divide). In addition, as training data are predominantly sourced from English speakers, they will exhibit cultural biases as well as others such as gender, occupational, socio-economic, etc. (Wang et al. 2025).

Finally, there will be issues with model validation that need to be overcome. Traditional validation approaches – such as comparing model outputs to real-world data – will be useful, but there are questions about LLM consistency (i.e. stochasticity) (Chopra et al. 2024) and robustness (i.e. sensitivity to prompts) (Vezhnevets et al. 2023), problems with halluncinations where LLMs invent fiction (Chen et al. 2024), train-test contamination (Vezhnevets et al. 2023), and behavioural subjectivity (Ma et al. 2024) that need to be explored. The ongoing development of bespoke evaluation metrics and standard benchmarks (see Chen et al. 2024) is encouraging.

3 Multimodal and Geospatial Foundation Models

Foundation models (FMs) (Bommasani et al. 2021) are a generalisation of LLMs that are not limited to text inputs and outputs. Vision-based FMs such as SAM and DALL-E have manifested the capabilities of analysing and generating images. These approaches extend to *multimodal* FMs, which encode and combine data from different sources, such as text, images, video, and sound. The multimodal FMs learn from large-scale cross-modality mutual enhancement, and therefore they can generate highly effective and transferable representations, making them versatile for a wide range of real-world applications.

In this context, *geospatial* foundation models are emerging as method for encoding novel spatial data representations, capturing complex associations between spatial objects in nuanced ways that often elude manual classification schemes. GFMs enable the assimilation of diverse spatial, temporal, and other sources – such as street view images, geotagged social media data, video, GPS trajectories, points-of-interest – incorporating them in construction of rich, multidimensional spatial representations (known as 'embeddings') of geographic features (Mai et al. 2024). One early implementation of GFMs, the CityFM model (Balsebre et al. 2024), shows a strong capability for predicting traffic speeds based on urban morphology, land use, and road network data.

But GFMs offer even greater opportunities where considering human perception of the environment. GFMs can potentially encode spatial representations that capture latent measures of the environment that may otherwise be difficult to collect at scale in many contexts (e.g. building function, noise, smell). The potential implications for geographic agent-based modelling are appealing – geographic features are remembered, recalled and used by agents for a variety of reasons. By integrating massive and diverse data, representations derived from GFMs create new views of the space that extend on the limitations of conventional geographic features. As such, geospatial foundation models provide a new opportunity to generate rich representations of geographic features, filling a gap in conventional geographic analysis.

Yet, GFMs will not be the 'silver bullet' in meeting all challenges relating to spatial representation and agent decision-making. Early evidence suggests that LLMs are poor predictors of navigation performance (Momennejad et al. 2024, Zeng et al. 2024), and reproductions of city images are inconsistent and indicate bias in underlying data (Jang et al. 2024). Understanding these limits and biases inherent to these models remains an important research task.

4 Towards Multi-Modal Foundation Models for ABMs

The integration of Large Language Models (LLMs) with ABMs is already proving to be an exciting development. It has the potential to offer more realistic, nuanced behaviours that are potentially more realistic and dynamic in their implementation. However, a significant hurdle is that, ultimately, the interface to and from the agent's 'brain' (the LLM) must be through text. The environment must be converted into text and an agent's behaviours and actions must be expressed textually. A potential solution to this problem could be a specialised geospatial foundation model that was trained not only on spatial data but also on large corpora of text. Such a model could capture both the 'understanding' about human behaviour that LLMs exhibit, as well as a a detailed and rich picture of 'place' (Tuan 1977) that the agents can use to interpret their virtual surroundings.

Of course, there are many hurdles that must be overcome. These include access to data (although spatial data are abundant through sources such as Open Street Map, it is not clear whether there is sufficient text data available, especially for minority groups); computational access and costs (the training and executions of these models may be beyond academic capabilities or budgets); skills barriers (social scientists will be key to ensuring behavioural accuracy, but do they have the expertise to engage with advanced foundation models?). If these difficulties can be overcome, the technologies might lay the foundations for the integration of ABMs to create richer, more comprehensive urban digital twins.

References

Badham, J., Chattoe-Brown, E., Gilbert, N., Chalabi, Z., Kee, F. & Hunter, R. F. (2018), 'Developing agent-based models of complex health behaviour', *Health & Place* **54**, 170–177.

Balsebre, P., Huang, W., Cong, G. & Li, Y. (2024), City foundation models for learning general purpose representations from openstreetmap, *in* 'Proceedings of the 33rd ACM International Conference on Information and Knowledge Management', pp. 87–97.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R. E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S. M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K. & Liang, P. (2021), 'On the Opportunities and Risks of Foundation Models'.

Chen, C., Yao, B., Ye, Y., Wang, D. & Li, T. J.-J. (2024), Evaluating the LLM Agents for Simulating Humanoid Behavior, *in* '1st HEAL Workshop at CHI Conference on Human Factors in Computing System', Honolulu, HI, USA.

Chopra, A., Kumar, S., Giray-Kuru, N., Raskar, R. & Quera-Bofarull, A. (2024), 'On the limits of agency in agent-based models'.

Crooks, A., Malleson, N., Manley, E. & Heppenstall, A. (2019), *Agent-Based Modelling and Geographical Information Systems: A Practical Primer*, Sage, London, UK.

DeAngelis, D. L. & Diaz, S. G. (2019), 'Decision-Making in Agent-Based Modeling: A Current Review and Future Prospectus', *Frontiers in Ecology and Evolution* **6**, 237.

Fotheringham, A. S. (2023), 'Digital twins: The current "Krays" of urban analytics?', *Environment and Planning B: Urban Analytics and City Science* **50**(4), 1020–1022.

Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F. & Li, Y. (2023), 'Large Language Models Empowered Agent-based Modeling and Simulation: A Survey and Perspectives'.

Heppenstall, A., Crooks, A., Malleson, N., Manley, E., Ge, J. & Batty, M. (2021), 'Future developments in geographical agent-based models: Challenges and opportunities', *Geographical Analysis* **53**(1), 76–91.

Jang, K. M., Chen, J., Kang, Y., Kim, J., Lee, J., Duarte, F. & Ratti, C. (2024), 'Place identity: a generative ai's perspective', *Humanities and Social Sciences Communications* **11**(1), 1–16.

Li, N., Gao, C., Li, M., Li, Y. & Liao, Q. (2024), 'EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities'.

Li, Y., Zhang, Y. & Sun, L. (2023), 'MetaAgents: Simulating Interactions of Human Behaviors for LLM-based Task-oriented Coordination via Collaborative Generative Agents'.

Lu, S., Guo, J., Zimmer-Dauphinee, J. R., Nieusma, J. M., Wang, X., VanValkenburgh, P., Wernke, S. A. & Huo, Y. (2024), 'AI Foundation Models in Remote Sensing: A Survey'.

Ma, Q., Xue, X., Zhou, D., Yu, X., Liu, D., Zhang, X., Zhao, Z., Shen, Y., Ji, P., Li, J., Wang, G. & Ma, W. (2024), 'Computational Experiments Meet Large Language Model Based Agents: A Survey and Perspective'.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R. & Lao, N. (2024), 'On the Opportunities and Challenges of Foundation Models for GeoAI (Vision Paper)', *ACM Transactions on Spatial Algorithms and Systems* **10**(2), 1–46.

Malleson, N., Franklin, R., Arribas-Bel, D., Cheng, T. & Birkin, M. (2024), 'Digital twins on trial: Can they actually solve wicked societal problems and change the world for better?', *Environment and Planning B: Urban Analytics and City Science* p. 23998083241262893.

Momennejad, I., Hasanbeig, H., Vieira Frujeri, F., Sharma, H., Jojic, N., Palangi, H., Ness, R. & Larson, J. (2024), 'Evaluating cognitive maps and planning in large language models with cogeval', *Advances in Neural Information Processing Systems* **36**.

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'.

Tuan, Y.-F. (1977), *Space and Place: The Perspective of Experience*, University of Minnesota Press, Minneapolis, MN.

Vezhnevets, A. S., Agapiou, J. P., Aharon, A., Ziv, R., Matyas, J., Duéñez-Guzmán, E. A., Cunningham, W. A., Osindero, S., Karmon, D. & Leibo, J. Z. (2023), 'Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia'.

Wang, Q., Wu, J., Tang, Z., Luo, B., Chen, N., Chen, W. & He, B. (2025), 'What Limits LLM-based Human Simulation: LLMs or Our Design?'.

Williams, R., Hosseinichimeh, N., Majumdar, A. & Ghaffarzadegan, N. (2023), 'Epidemic Modeling with Generative Agents'.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X. & Gui, T. (2023), 'The Rise and Potential of Large Language Model Based Agents: A Survey'.

Zeng, Q., Yang, Q., Dong, S., Du, H., Zheng, L., Xu, F. & Li, Y. (2024), 'Perceive, reflect, and plan: Designing llm agent for goal-directed city navigation without instructions', *arXiv preprint arXiv:2408.04168* .