Towards the use of Foundation Models and Embeddings for Nuanced Perception and Decision-Making in Agents

Nick Malleson and Ed Manley

School of Geography, University of Leeds, Leeds, UK,
N.S.Malleson@leeds.ac.uk,
E.J.Manley@leeds.ac.uk
www.nickmalleson.co.uk

Abstract. Following the emergence of powerful generative large language models (LLMs), there has been a flurry of interest in the use of LLMs to control agents in agent-based models. Proponents argue that using the information about humans and human behaviour contained within an LLM could lead to the creation of agents who exhibit more complex and nuanced behaviour than those whose actions are driven by traditional behavioural frameworks. This paper begins to explore the use of a specific concept that underpins LLMs; that of embeddings. An embedding is a vector-based numerical representation of a piece of text that captures aspects of its meaning and context. We hypothesise that conceptualising agents' characteristics through embeddings, rather than with discrete state variables, may offer a more nuanced and expressive foundation for representing agent characteristics and behaviours. We demonstrate the potential of this approach by recreating the Schelling residential segregation model using rich text descriptions of household agents and converting these to embeddings as a means of defining agents. The results show how agents can self-organise into more diverse and emergent clusters than is possible when they are defined with a small number of discrete attributes. This offers a path toward more realistic, high-dimensional representations of agent heterogeneity.

Keywords: agent-based modelling, foundation model, large-language model, embeddings, schelling

Short paper presented to *The Social Simulation Conference* (SSC), Delft, 25–29th August, 2025

1 Introduction

Modelling human behaviour in agent-based models (ABMs) remains one of the key challenges facing the discipline (An et al, 2021; Heppenstall et al, 2021). Traditionally, behaviour in ABMs has been implemented either through explicit rule-based logic or by using more elaborate behavioural frameworks such as the popular Beliefs, Desires and Intentions (BDI). More recently, techniques like neural networks and genetic algorithms (DeAngelis and Diaz, 2019) are showing promise as well as, in part due to the recent availability of huge GPU clusters, reinforcement learning (e.g. see Ale Ebrahim Dehkordi et al, 2023). Building on this, the emergence of generative large-language models (LLMs) such as ChatGPT, has stimulated a huge amount of interest interest among researchers who are exploring how these models might enable more sophisticated and realistic simulations of human behaviour and decision-making (Gao et al, 2023; Xi et al, 2023; Cheng et al, 2024; Gürcan, 2024; Wang et al, 2024). Initiatives such Concordia (Vezhnevets et al, 2023) and MetaGPT (Hong et al, 2024) are investigating whether LLMs can equip agents with advanced abilities like natural language understanding, reasoning, and planning (see Ma et al, 2024).

Two key innovations that underpin the success of generative LLMs are *embeddings* and *self-attention*. After tokenisation – the process of splitting up text into discrete units (words, symbols, sub-words, etc.) – an additional vector is attached to each token to capture the token's deeper meaning. These vectors are called 'embeddings'. Early embeddings were static, which limited their ability to account for the different contexts that can apply to words (for example the different meaning of the word 'bank' in 'river bank' and 'savings bank'). The Transformer architecture (Vaswani et al, 2017) addressed this limitation through a mechanism called self-attention, which produces dynamic embeddings for each word that depend on the context in which they are used. This innovation ultimately seeded the powerful LLMs that are emerging today. Figure 1 provides a hypothetical example of what an embedding vector may look like for some words.

	Embedding dimension				
	plural	politics	animal		wealth
Word	0	1	2	...	383
rich	0.1	0.2	0		0.9
Conservative	0	0.9	0		0.3
kittens	1	0	1		0
orange	0.5	0	0.2		0
deprived	0	0.3	0		-0.8

Fig. 1: Hypothetical word embeddings showing how different words could be represented as numeric vectors across abstract dimensions.

Although there is a very new and growing literature on use of LLMs to drive the behaviour of agents, much less attention has been paid to the benefits that

may be realised through the use of the underlying technologies that underpin LLMs. To begin to fill this gap, this paper explores the use of embeddings as a means of capturing richer agent characteristics than is currently possible when agents are defined using discrete variables (e.g. for age, affluence, behavioural preferences, etc.). These embeddings – in theory – encode latent characteristic features of the population, that may be otherwise difficult to extract from conventional data and small-scale surveys (which themselves suffer from biases). The paper implements a simple residential segregation model that is based on Schelling's well known model of dynamic segregation (Schelling, 1971). By describing household agents using text, we demonstrate that clusters of homogeneity still emerge, as with Schelling's original work, but the clusters themselves are much richer and more diverse. A criticism of Schelling's original model, and of some ABMs more generally, is that by reducing human behaviour to a few simple characteristics, we disregard "dominant explanations like structural racism, white flight, and red-lining" which "eradicates the possibility of collective solutions to address the problem" (Larooij and Törnberg, 2025). Embeddings, and the related technologies that underpin LLMs, may offer a means of increasing the diversity and accuracy of ABMs. Whilst this is early work and requires extensive further exploration before any firm conclusions can be drawn, it suggests opportunities for new ways of capturing agent heterogeneity and, as future work, using similar technologies to potentially model more nuanced behaviour.

2 Method

The code created to run these experiments in their entirety is relatively straightforward. It is written in Python and available in full on GitHub.

2.1 Creating Household Description Embeddings

To explore how embeddings can be used to represent complex agent heterogeneity, we developed a parsimonious ('toy') spatial agent-based model inspired by Schelling's model of residential segregation (Schelling, 1971). For this preliminary work, we describe household agents using three dimensions: household structure, income and political beliefs. Rather than representing agents using simple discrete attributes, one for each dimension, we create hypothetical, rich, text-based household descriptions. For illustrative purposes these are simply produced with the use of an open-source generative LLM, *Llama-4-Maverick-17B-128E-Instruct-FP8*[1], executed using the API provided by the together.ai service. We use this paid service because it offers access to much larger LLMs than could be made available locally. Of course these household descriptions are entirely fabricated and will undoubtedly be biased towards those types of households who are best represented in the LLM training data – bias is an ongoing problem for all LLM-related work (Navigli et al, 2023; Park et al, 2023;

[1] https://huggingface.co/meta-llama/Llama-4-Maverick-17B-128E-Instruct-FP8

Vezhnevets et al, 2023; Wang et al, 2025) – but they are simply used here to demonstrate the value in converting the descriptions to embeddings. The prompt used to generate the descriptions is as follows, where N is the requested number of households.

> *Produce N one-sentence, anonymous, detailed descriptions of stereotypical UK households, describing their household structure, income and political beliefs. Output in CSV format with one line per household description and nothing else.*

As an example, one of the household descriptions produced by the LLM, chosen arbitrarily, is:

> *A retired couple living alone in a semi-detached house in a suburban area, relying on state pensions and modest savings, strongly supporting the Conservative party*

Each household description was then converted to numerical embedding vector using a pre-trained sentence transformer model, *MiniLM-L6-H384-uncased*[2] (Wang et al, 2020). This model is small enough to be executed locally using the HuggingFace Transformers python library. First, each text description is tokenised into subword units and passed through the transformer, which uses self-attention to generate contextualised vector representations for each token. These token-level vectors are then aggregated into a single fixed-length sentence embedding using mean pooling, whereby each token's embedding is averaged. The result is a vector of 384 dimensions, where vectors representing semantically similar household descriptions are located closer together, and where households can be defined by embeddings that reflect nuanced socio-political characteristics.

2.2 The Agent-Based Segregation Model

The simulation, written in Python, has been designed to replicate Schelling's original model of residential segregation (Schelling, 1971). The environment consists of a 2D grid of fixed size (default 20×20), initially populated with a user-defined number of agents (default 300). On initialisation, each agent is assigned a free position on the grid and is assigned a randomly-chosen description and the description's associated embedding. We create 350 separate descriptions (see Section 2.1) so most agents will be unique, but it is likely that some will be assigned to the same description.

During each iteration of the simulation, all agents assess the similarity between their own embedding and those of their immediate 8 neighbours (the Moore neighbourhood). Similarity is computed using cosine similarity, which is a widely used metric for comparing text embeddings (Reimers and Gurevych, 2019). Agents are considered "happy" if the mean similarity to their neighbours exceeds a configurable threshold. Here the threshold was chosen to manually to

[2] https://huggingface.co/nreimers/MiniLM-L6-H384-uncased

be the largest value that still permitted the model to reach an equilibrium where nearly all agents were happy. Unhappy agents relocate to a randomly selected vacant cell, and the process is repeated for a fixed number of iterations (currently 200). A potential drawback with the embedding approach used here is that, unlike with the original Schelling model, the similarity threshold itself is largely meaningless. Previously a threshold of, say, 50% meant that an agent would be happy if at least half of its neighbours were of the same type. Similarity loses its clear meaning here, although this could be advantageous as the assumption of binary 'sameness' is highly unrealistic anyway. The use of embeddings here allows a much more nuanced estimate of household similarity.

In the original Schelling model, visualising the results is trivial because each agent is represented by a binary state. Here, however, each agent is represented by a 384-item vector. To attempt to capture some of the spatial distribution of the different agents, the high-dimensional embeddings were reduced to a three-item vector using Principal Components Analysis. These three dimensions can then be mapped to RGB colour values where agents who are similar in the embedding space should be represented with similar colours.

3 Results and Discussion

3.1 Analysis of the Household Description Embeddings

Before running the model in earnest, creating agents from a set of 350 household descriptions, we first experiment with an example with only 5 arbitrarily-chosen household descriptions. This is for the purposes of experimenting with the embeddings themselves. Table 1 illustrates the chosen descriptions. Given the prompt to produce household descriptions for the UK, there is notably strong reference to UK policy and political parties. Note that the first and second descriptions have been made deliberately similar to ensure that their associated embeddings are also similar while the remaining three are designed to be diverse.

Figure 2 illustrates the similarity between the five example embeddings. As expected, the first two embeddings have very high similarity, with the remaining five showing much lower similarity.

The example model is then executed and reaches equilibrium after approximately 50 iterations. Figure 3 illustrates the spatial locations of the agents at the beginning and end of the simulation. As there are only five agent types it is easy to distinguish them using their colours. Note the similarity in the colours used to distinguish household types 0 and 1; these represent the two similar households and clearly cluster together as they share many similarities in their descriptions and, hence, in their embeddings also.

3.2 Agent-Based Model Results

After running the example simulation with only 5 distinct household descriptions, the model is then executed in earnest with 200 agents and a total of 350 descriptions that are drawn randomly upon agent initialisation.

0	A retired couple living alone in a semi-detached house in a suburban area, relying on state pensions and modest savings, strongly supporting the Conservative party.
1	An elderly couple residing in a suburban, semi-detached house, drawing income from their savings and their state pensions, voting for the Conservative party consistently.
2	A young, single professional renting a studio flat in a city centre, earning a salary around £35,000 from a career in marketing, voting for the Liberal Democrats and actively campaigning for environmental causes.
3	A large, multi-generational family residing in a terraced house, with the patriarch working as a manual labourer on a zero-hours contract, the matriarch a part-time carer, and several children, identifying as Labour supporters and strongly union-backed.
4	A single parent with three children, living in a council flat, surviving on a tight budget that includes Universal Credit and Child Tax Credits, and staunchly supporting the Labour party, particularly its more left-wing elements.

Table 1: Five example household descriptions.

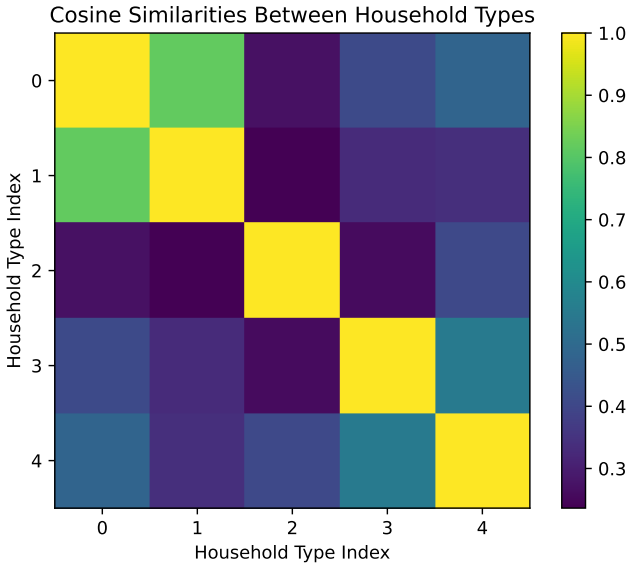
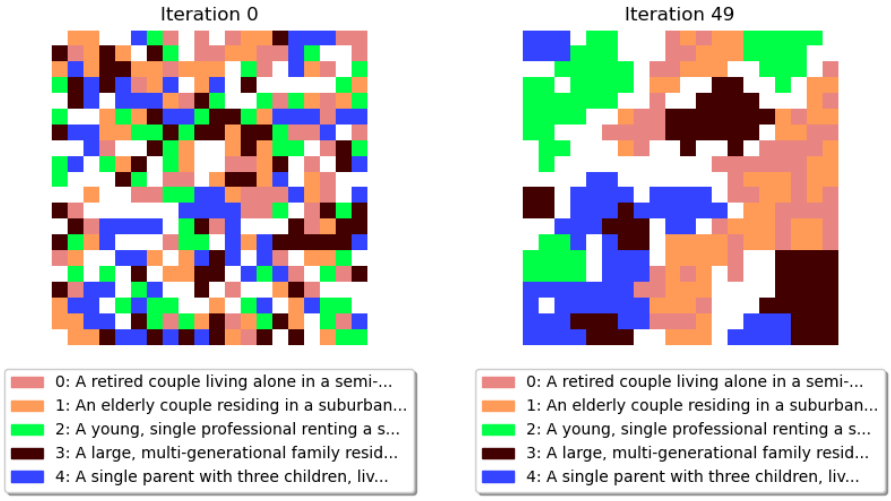


Fig. 2: Similarity of the example embeddings



(a) Initial agent locations. (b) Final agent locations

Fig. 3: The locations of the agents at the beginning and end of the example simulation (with only five distinct agent types).

Figure 4 illustrates the number of happy agents over the course of the simulation. By iteration 200 the simulation has reached equilibrium; if the simulation were run for a larger number of iterations there would be no noticeable change in the number of happy agents.

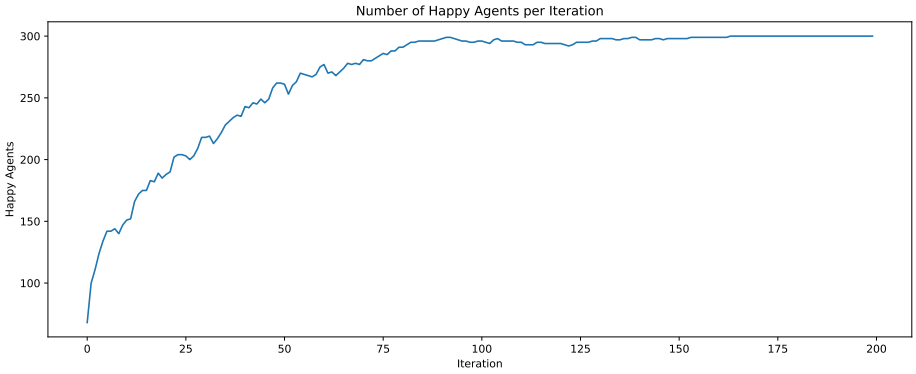
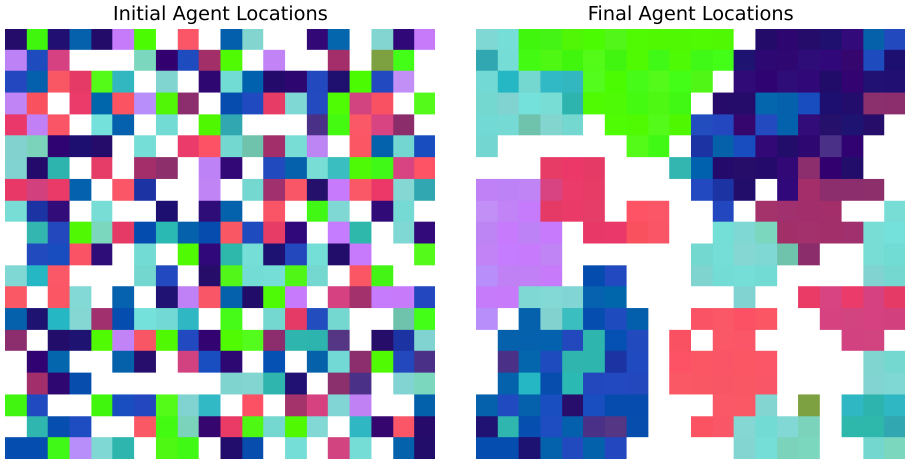


Fig. 4: The number of happy agents over time.

Figure 5 illustrates the positions of the household agents at the beginning and end of the simulation. Households cluster according to the similarity in their embeddings, which is entirely expected. This is consistent with the original Schelling model. The most interesting observation is that because the descriptions of the

agents are so heterogeneous, the clusters are also heterogeneous. Looking closely at the figure it becomes apparent that what appear to be contiguous areas of colour are actually made up of different, but similar, household types. This potentially represents a much more realistic population of households that no longer need be distinguished according to a few coarse variables that necessarily simplify diverse, heterogeneous socio-economic-demographic characteristics.



(a) Initial agent locations. (b) Final agent locations

Fig. 5: The locations of the agents at the beginning and end of the full simulation.

4 Challenges and Future Directions

This paper has demonstrated that one of the core methods that underpins the success of modern large language models (LLMs), that of text *embeddings*, can be used to create rich, heterogeneous agents. Although the model used here is necessarily parsimonious and no firm conclusions should yet be drawn from such preliminary work, the paper shows that the use of embeddings, and potentially additional related LLM technologies, could be very valuable for creating more nuanced, detailed ABMs.

Part of the motivation for this work is the realisation that although LLM-backed agents could be extremely powerful – the number of recent reviews is a testament to this (Gao et al, 2023; Xi et al, 2023; Cheng et al, 2024; Gürcan, 2024; Wang et al, 2024) – the need to interface with an agent using purely text is extremely limiting. Describing a rich environment with language, and requiring an agent to explain their actions with text, will necessarily required considerable abstractions. Recent innovations in foundation models (LLMs are a type

of foundation model) have shown that embeddings for multi-modal data (text, image, spatial, audio, etc.) can be used simultaneously (Balsebre et al, 2024; Huang et al, 2024; Mai et al, 2024) to describe environments in a way that is much richer and more nuanced. Hence the use of embeddings to describe household agents in this paper could ultimately form part of a more comprehensive approach to agent-based modelling that uses foundation models, large-language models and the knowledge embedded within them. In that vein, immediate future work will explore whether an LLM or foundation model could be utilised to model the *behaviour* of the agents, moving beyond a simple similarity threshold. Although very preliminary, the recent cited work that uses LLMs to control agents is intriguing – see Park et al (2023) in particular.

There are, of course, a number of challenges that need to be considered with future work. To begin with, bias remains a significant problem for models that have been trained on data from the Internet (Li et al, 2023; Park et al, 2023; Vezhnevets et al, 2023; Wang et al, 2024, 2025). The bias in our household descriptions is inconsequential, as these are simply used to demonstrate the method and will, in future, be based on real descriptions. However, it is likely that the *embeddings* will suffer from bias, as these are also generated using a transformer that will most likely have been trained on Internet data. It may be that the transformer will be better at capturing nuanced aspects in household descriptions from certain socio-economic groups than others. For example, the may be cultural or social norms that are present in the descriptions but not well represented in the ultimate embeddings. Fortunately, methodological innovations parameter-efficient transfer learning (Houlsby et al, 2019) – such as Low-Rank Adoption (LoRA) (Hu et al, 2021) – may provide an opportunity to use fine-tuning to reduce biases.

Secondly, although there are large numbers of pre-trained open-source LLMs available, running a pre-trained model can be extremely expensive. If these models are then used to control the behaviour of large numbers of agents the computational requirements will quickly become unmanageable.

Finally, work is required to develop methods to validate ABMs that are backed by LLMs or other foundation models. New issues include inconsistencies due to the stochastic nature of LLM responses (Chopra et al, 2024), their sensitivity to specific prompts (Vezhnevets et al, 2023) and the ongoing problem of hallucinations (Chen et al, 2024). Further clarity is needed around which LLMs are most appropriate for the production of agent descriptions, and the contexts in which each LLM is most appropriate. In addition, the use of the Schelling model here is perhaps overly simplistic; immediate future work will look to utilise a more representative ABM.

As this area of research rapidly evolves, embedding-based representations offer a promising route toward more expressive and human-like agents. By building on these foundations, future agent-based models could more effectively capture the richness, complexity, and subtlety of real-world social systems.

Bibliography

An L, Grimm V, Sullivan A, Turner II B, Malleson N, Heppenstall A, Vincenot C, Robinson D, Ye X, Liu J, Lindkvist E, Tang W (2021) Challenges, tasks, and opportunities in modeling agent-based complex systems. Ecological Modelling 457:109,685, DOI 10.1016/j.ecolmodel.2021.109685

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, DOI 10.1111/gean.12267

DeAngelis DL, Diaz SG (2019) Decision-Making in Agent-Based Modeling: A Current Review and Future Prospectus. Frontiers in Ecology and Evolution 6:237, DOI 10.3389/fevo.2018.00237

Ale Ebrahim Dehkordi M, Lechner J, Ghorbani A, Nikolic I, Chappin É, Herder P (2023) Using Machine Learning for Agent Specifications in Agent-Based Models and Simulations: A Critical Review and Guidelines. Journal of Artificial Societies and Social Simulation 26(1):9, DOI 10.18564/jasss.5016

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. DOI 10.48550/ARXIV.2312.11970

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. DOI 10.48550/ARXIV.2309.07864

Cheng Y, Zhang C, Zhang Z, Meng X, Hong S, Li W, Wang Z, Wang Z, Yin F, Zhao J, He X (2024) Exploring Large Language Model based Intelligent Agents: Definitions, Methods, and Prospects. DOI 10.48550/ARXIV.2401.03428

Gürcan Ö (2024) LLM-Augmented Agent-Based Modelling for Social Simulations: Challenges and Opportunities. In: Lorig F, Tucker J, Dahlgren Lindström A, Dignum F, Murukannaiah P, Theodorou A, Yolum P (eds) Frontiers in Artificial Intelligence and Applications, IOS Press, DOI 10.3233/FAIA240190

Wang L, Ma C, Feng X, Zhang Z, Yang H, Zhang J, Chen Z, Tang J, Chen X, Lin Y, Zhao WX, Wei Z, Wen JR (2024) A Survey on Large Language Model based Autonomous Agents. Frontiers of Computer Science 18(186345), DOI 10.1007/s11704-024-40231-1

Vezhnevets AS, Agapiou JP, Aharon A, Ziv R, Matyas J, Duéñez-Guzmán EA, Cunningham WA, Osindero S, Karmon D, Leibo JZ (2023) Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia. 2312.03664

Hong S, Zhuge M, Chen J, Zheng X, Cheng Y, Zhang C, Wang J, Wang Z, Yau SKS, Lin Z, Zhou L, Ran C, Xiao L, Wu C, Schmidhuber J (2024)

MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. DOI 10.48550/arXiv.2308.00352, 2308.00352

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. DOI 10.48550/ARXIV.2402.00262

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in Neural Information Processing Systems, Curran Associates, Inc., vol 30

Schelling TC (1971) Dynamic models of segregation. The Journal of Mathematical Sociology 1(2):143–186, DOI 10.1080/0022250X.1971.9989794

Larooij M, Törnberg P (2025) Do Large Language Models Solve the Problems of Agent-Based Modeling? A Critical Review of Generative Social Simulations. DOI 10.48550/arXiv.2504.03274, 2504.03274

Navigli R, Conia S, Ross B (2023) Biases in Large Language Models: Origins, Inventory, and Discussion. Journal of Data and Information Quality 15(2):1–21, DOI 10.1145/3597307

Park JS, O'Brien J, Cai CJ, Morris MR, Liang P, Bernstein MS (2023) Generative Agents: Interactive Simulacra of Human Behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, ACM, San Francisco CA USA, pp 1–22, DOI 10.1145/3586183.3606763

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? DOI 10.48550/arXiv.2501.08579, 2501.08579

Wang W, Wei F, Dong L, Bao H, Yang N, Zhou M (2020) MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. DOI 10.48550/arXiv.2002.10957, 2002.10957

Reimers N, Gurevych I (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, pp 3980–3990, DOI 10.18653/v1/D19-1410

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, ACM, Boise ID USA, pp 87–97, DOI 10.1145/3627673.3679662

Huang W, Wang J, Cong G (2024) Zero-shot urban function inference with street view images through prompting a pretrained vision-language model. International Journal of Geographical Information Science 38(7):1414–1442, DOI 10.1080/13658816.2024.2347322

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, DOI 10.1145/3653070

Li Y, Zhang Y, Sun L (2023) MetaAgents: Simulating Interactions of Human Behaviors for LLM-based Task-oriented Coordination via Collaborative Generative Agents. DOI 10.48550/ARXIV.2310.06500

Houlsby N, Giurgiu A, Jastrzebski S, Morrone B, Laroussilhe QD, Gesmundo A, Attariyan M, Gelly S (2019) Parameter-efficient transfer learning for NLP. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, PMLR, Proceedings of Machine Learning Research, vol 97, pp 2790–2799

Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, Wang L, Chen W (2021) LoRA: Low-Rank Adaptation of Large Language Models. DOI 10.48550/ARXIV.2106.09685

Chopra A, Kumar S, Giray-Kuru N, Raskar R, Quera-Bofarull A (2024) On the limits of agency in agent-based models. DOI 10.48550/arXiv.2409.10568, 2409.10568

Chen C, Yao B, Ye Y, Wang D, Li TJJ (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