**LLM-based psychological digital twins in social research: opportunities and dangers**

Paweł Sobkowicz,1,2,

1Sano Centre for Computational Medicine, Czarnowiejska 36, 30-054 Kraków, Poland  
2 Narodowe Centrum Badań Jądrowych, , ul. A Sołtana 7, 04-500 Otwock-Świerk, Poland  
*p.sobkowicz @sanoscience.org, pawelsobko@gmail.com***Keywords**: Agent Based Models, ABM, LLM, digital twins, DT

1. Introduction

We discuss here the potential benefits of using LLM engines in social and psychological studies, in particular as a component of coordinated experimental/simulation projects. The advances in LLMs – particularly those trained on general corpora of human generated texts taken from social life – offer a novel opportunity, creating agents that can be defined and tested using normal language.

1. Description of the problem

Agent Based Models( ABMs) have been used in studies of social phenomena for over 55 years. They range from extremely simplistic (where humans are described through a single binary variable (for example in opinion dynamics studies using tools and analogies with the Ising model of magnetism) to more complex approaches, where an agent is characterized by multiple variables, has complex internal rules determining its actions and reactions (goals, preferences, memory), and acts in complex social environment. Despite progress in computing power, these complex models are quite rare. There are several reasons for this. First, human psychology is very complex, and there is no dominant “theory of foundations human behavior”, on which modelers could build their simulation frameworks. Second, the experimental data, necessary to derive and validate the rules determining agent behaviors is fragmented (most of social and psychological studies are not designed with the goal of providing data to facilitate creation of ABMs). Third, such rule-based model is only as good as the completeness of variables and rules sets are, and as the starting conditions correspond to actual situations.

1. A different approach to the problem

In recent years, a new approach has appeared: use of LLMs to create agents (or even agent societies), corresponding to desired psycho-social characteristics (“generative agents models”). The agents are defined through dedicated prompts, and may be “asked” to perform actions, react to situations in a way that corresponds to the predefined profile. This allows simulations to correspond closely to real world situations. The advances in general purpose LLMs since 2024 have resulted in a veritable explosion of the published papers, conference presentations and preprints devoted to uses of such generative agents (for reviews, see e.g [1-3]). At the same time, the nascent field has definite weaknesses and problems [4-7]. We shall discuss potential benefits and limitations of the above approach using simple, illustrative examples (based on imaginary personal descriptions). They show, that while the potential promise of the generative ABMs is great, the current state of the art requires careful validation.

1. Challenges, conclusions and future work

There are two types of challenges facing generative agents’ approach. The first is **technica**l: how to ensure the “independence” and stability of personas created using LLMs. Today’s LLMs still leave a lot to improve. Also, while it is relatively easy to program individual agents or very small groups, simulations of large communities requires new tools and untested scalability.

But far more important are **ethical challenges**. By their very definition, successful psychological digital twins can “reproduce” behavioral characteristics of specific people. This opens way to myriad of possible abuses, some relatively mild (using a DT for marketing/advertising choice purposes), to very dangerous, like political manipulation. Thus, special care should be present in planning research devoted to the topic and in development of the associated tools.

At the same time, the ease with which we can map real world situations into simulations and obtain understandable results is tempting. However, to fulfill the promise of such DT/generative agents, they must be verifies and validates – by comparing them with experiments and with the rule-based ABMs (where the mechanisms are known by definition). Optimally, the LLM-based approach might serve as an improvement engine for traditional models, by discovering the rules of behavior hidden in the LLM structures.

References

1. Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: a survey and perspectives. Humanities and Social Sciences Communications, 11(1), September 2024. ISSN 2662-9992.
2. Xinyi Mou, Xuanwen Ding, Qi He, Liang Wang, Jingcong Liang, Xinnong Zhang, Libo Sun, Jiayu Lin, Jie Zhou, Xuanjing Huang, and Zhongyu Wei. From individual to society: A survey on social simulation driven by large language model-based agents. ARXIV.2412.03563, 2024.
3. D. Helbing and J. Argota Sanchez-Vaquerizo, “Digital Twins: Potentials, Ethical Issues, and Limitations,” SSRN Electronic Journal, 2022, doi: 10.2139/ssrn.4167963.
4. Austin C Kozlowski and James Evans. Simulating subjects: The promise and peril of AI stand-ins for social agents and interactions. September 2024. doi: 10.31235/osf.io/vp3j2.
5. Xuhui Zhou, Zhe Su, Tiwalayo Eisape, Hyunwoo Kim, and Maarten Sap. Is this the real life? Is this just fantasy? The misleading success of simulating social interactions with LLMs. ARXIV.2403.05020, 2024.
6. G. Kovač, R. Portelas, M. Sawayama, P. F. Dominey, and P.-Y. Oudeyer, “Stick to your role! Stability of personal values expressed in large language models,” PLOS ONE, vol. 19, no. 8, p. e0309114, Aug. 2024, doi: 10.1371/journal.pone.0309114.
7. M. Larooij and P. Törnberg, “Do Large Language Models Solve the Problems of Agent-Based Modeling? A Critical Review of Generative Social Simulations.” arXiv, 2025. doi: 10.48550/ARXIV.2504.03274.