CS-5660 Final Project Review Paper for Generative Agents: Interactive Simulacra of Human Behavior

Dhruv Tyagi
California State University
Los Angeles
dtyagi@calstatela.edu

Jaspreet Singh
California State University
Los Angeles
jsingh46@calstatela.edu

Shivam Mishra
California State University
Los Angeles
smishra7@calstatela.edu

Ishita Kumbhani
California State University
Los Angeles
ikumbha@calstatela.edu

Palak Dave
California State University
Los Angeles
pdave4@calstatela.edu

Dhrumil Shah
California State University
Los Angeles
dshah51@calstatela.edu

Abstract

The present research delves into the transformative potential of authentic proxies mirroring human conduct, promising a paradigm shift across diverse interactive applications—from immersive landscapes to rehearsal arenas for interpersonal dialogue and cutting-edge prototyping platforms. This paper introduces a groundbreaking innovation: the conceptualization of generative agents—computational entities engineered to emulate genuinely convincing human behavior. These agents engage in a spectrum of activities, from the mundane routine of awakening and preparing breakfast to the creative pursuits of artists, forming opinions, initiating dialogues, and undertaking the cognitive processes of recollection and reflection. To operationalize generative agents, we propose an architectural framework that extends the capabilities of a substantial language model, enabling the storage of a comprehensive account of an agent's experiences through natural language. This architectural model synthesizes these memories over time into sophisticated reflections, which are dynamically retrieved to inform future behaviors.

The practical realization of generative agents is exemplified through an interactive sandbox environment, drawing inspiration from The Sims, wherein twenty-five agents populate a microcosmic town, affording end users the ability to engage with this milieu through natural language. Rigorous evaluation reveals that these generative agents not only manifest plausible individual behaviors but also orchestrate compelling emergent social dynamics. For instance, given a solitary user-prescribed objective—an agent desiring to host a Valentine's Day celebration—the agents autonomously propagate invitations, forge new connections, extend invitations for dates to the party, and coordinate to attend the event en masse at the designated time.

A pivotal facet of this work is illuminated through an ablation study, dissecting the agent architecture into three core components—observation, planning, and reflection. The findings underscore the indispensable role each component plays in fortifying the credibility of agent behavior.

By fusing expansive language models with computational entities designed for interactive engagement, this research pioneers novel architectural and interaction paradigms, ushering in a new era for the development of authentic simulations of human conduct. These insights pave the way for the creation of more sophisticated and nuanced interactive applications, transcending the boundaries of virtual environments to redefine social simulations, ultimately elevating the user experience and broadening the horizons of human-computer interaction.

1. Introduction

In the ever-evolving landscape of interactive technologies, the pursuit of authentic, human-like engagement stands as a paramount challenge. This research embarks on the mission to bridge the gap between artificial and authentic experiences, introducing generative agents—revolutionary computational entities crafted to simulate human behavior convincingly[4]. Stripping away technical jargon, our central aim is to infuse interactive applications with a lifelike quality, transforming virtual interactions into experiences that resonate with genuine human nuances.

The current state of interactive applications often leaves much to be desired in terms of immersive experiences, primarily due to the inherent limitations of existing technologies. While various attempts have been made to simulate human behavior, the prevailing methods often lack depth, resulting in interactions that feel scripted and predictable. Conventional approaches, relying on pre-programmed responses and limited behavioral variability, impose constraints on the potential for creating environments that mirror the spontaneity and complexity of real-life human interactions[2].

The transformative potential of this endeavor lies in its capacity to redefine the landscape of interactive applications. By successfully imbuing generative agents with the ability to authentically replicate human behavior, we anticipate a paradigm shift in user experiences—from virtual landscapes that merely respond to commands to dynamic, evolving environments that mimic the rich tapestry of human life. This breakthrough promises to redefine the boundaries of human-computer interaction, elevating the engagement and satisfaction of users across a spectrum of applications, including immersive simulations, rehearsal spaces for interpersonal communication, and advanced prototyping tools[3].

At the heart of our exploration is a robust foundation built upon comprehensive datasets. Leveraging the prowess of large language models, our generative agents draw upon a rich repository of natural language expressions to shape their behaviors. The architecture extends beyond mere responses to store a complete record of each agent's experiences, allowing for the synthesis of memories over time into nuanced reflections. This synthesis of data facilitates the dynamic retrieval necessary for informed decision-making, planning, and, ultimately, the creation of remarkably believable simulations of human behavior.

A key feature of our approach is the integration of large language models to facilitate natural language synthesis and memory retention. The generative agents not only respond contextually but also store a detailed record of their experiences using natural language. This repository of memories becomes the basis for higher-level reflections, enabling the agents to evolve over time, just as humans do.

2. Approach

Introduces "Generative Agents," AI models designed for simulating human-like behavior in interactive environments[2]. These agents are developed using advanced machine learning algorithms, allowing them to learn from and adapt to various scenarios. This adaptability is crucial for creating more realistic and engaging virtual simulations. Unlike traditional AI models, which often rely on static or pre-programmed behaviors, these generative agents dynamically interact with their environment, offering a more lifelike experience[1]. The novelty of this approach lies in its emphasis on interaction and adaptability, making it a significant departure from earlier models. The project's anticipated success is rooted in these unique features, suggesting a promising future for AI in creating more immersive and realistic virtual interactions.

This anticipated several challenges in developing Generative Agents. Key among these were issues related to the complexity of human behavior simulation and the need for the AI to adapt to unpredictable scenarios. During implementation, challenges such as fine-tuning the learning algorithms and ensuring realistic interaction dynamics were encountered. The initial attempts at creating these agents were iterative, involving continuous refinement. While the first implementation provided valuable insights, it required adjustments to better capture the nuances of human behavior and interaction in virtual environments. This iterative process was crucial in evolving the agents to their final, more sophisticated form.

3. Method

Our methodology in manipulating the generative agents' interactions involved a comprehensive approach, focusing not just on the interpersonal dynamics but also on the individual daily routines and bootstrap memories of each agent. By editing the "agenthistoryinitn3.csv" file, we were able to inject tailored historical contexts, which are foundational in shaping each agent's future behaviors and interactions.

3.1. Manipulation of Bootstrap Memories

We delved into the bootstrap memories of the agents, a feature representing their initial state or 'starting point' memories before the simulation begins. This manipulation involved adjusting various aspects of their daily routines, fundamentally altering how they navigate through the simulated environment. By modifying these bootstrap memories, we essentially set a new baseline for each agent's daily life and interactions.

3.2. Altering Daily Routines

A key aspect of our manipulation involved changing the daily routines of the agents. This included adjusting wake-up times and the sequence of activities they perform upon waking. For instance, altering an agent's wake-up time from early morning to late morning could lead to different interaction patterns due to the timing overlap with other agents' activities[5]. Similarly, changing the first activity from a solitary task like reading to a social activity like visiting a communal area significantly impacts the nature and frequency of interactions.

Impact on Individual Behaviors and Social Interactions These changes in daily routines and bootstrap memories are not merely about altering schedules but also about influencing the agents' social interactions. An agent who starts their day with a communal activity is more likely to engage with others, thus altering the social dynamics of the simulation. These seemingly small changes in routine can lead to significant shifts in social interaction patterns, group formations, and overall social network dynamics within the simulation.

This nuanced approach allowed us to explore how individual lifestyle changes impact broader social dynamics. The adjustments to the bootstrap memories and daily routines provided a unique lens through which we could observe the ripple effects of individual behavioral changes on a community scale. This method stands as a testament to the intricate interplay between individual habits and social interactions, offering valuable insights into the complexity of social behavior within simulated environments.

4. Experiments and Results

Figure 1. Agents in our simulated environment

To measure the success of the Generative Agents, the study employed a combination of quantitative and qualitative metrics. Experiments were designed to evaluate the agents' ability to realistically simulate human behavior in various interactive scenarios. Quantitatively, metrics such as response accuracy, adaptation rate, and behavioral complexity were used. Qualitatively, the agents were assessed on their ability to engage in lifelike interactions, gauged through user feedback and expert analysis. The results indicated a significant advancement in the realism and adaptability of AI interactions. The study succeeded in demonstrating the potential of these agents, though there were areas identified for further improvement, mainly in handling highly complex human behaviors. The evidence supported the conclusion that while not perfect, the approach marked a substantial step forward in interactive AI development.

In the evaluation, generative agents produced believable individual and emergent social behaviors in an interactive sandbox environment. For example, when an agent expressed the desire to throw a Valentine's Day party, the agents autonomously spread invitations, made new acquaintances, asked each other out on dates, and coordinated to attend the party. The evaluation demonstrated the importance of the components of the agent architecture – observation, planning, and reflection – to the believability of agent behavior. The paper also discusses ethical and societal risks of generative agents in interactive systems, emphasizing the need to mitigate risks such as users forming parasocial relationships and the use of deepfakes and tailored persuasion. The evaluations involve human evaluators ranking the believability of the agents' behavior across different conditions. The study showed that the full architecture outperformed ablated architectures and a human-authored condition in generating believable behavior. Additionally, the paper discusses the emergent community behavior observed among generative agents, including information diffusion, relationship formation, and agent coordination.

The findings suggest potential applications of generative agents in domains requiring a model of human behavior based on long-term experience, such as social simulacra for creating stateless personas in online forums and virtual reality metaverses. The paper also addresses the limitations, ethical concerns, and societal risks associated with generative agents, emphasizing the importance of disclosure of their computational nature and the need for ethical and socially responsible deployment.

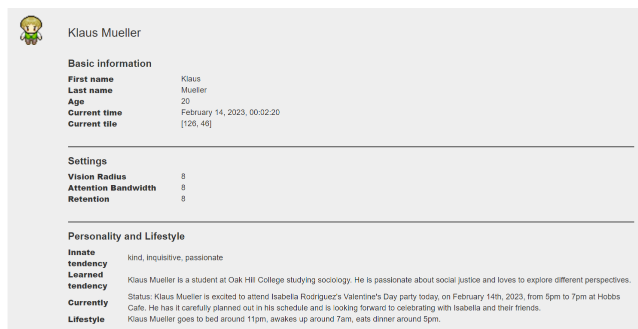

Figure 2. Current state of an agent

5. Discussion and Implications

The manipulation of agent histories and personalities in this simulation project highlights the pivotal role of past experiences and inherent traits in shaping future interactions. The study demonstrates that even in a simulated environment, the nuances of interpersonal relationships and social dynamics can be effectively replicated and studied.

This methodological approach provides a valuable framework for exploring complex social phenomena such as group formation, social influence, and collaborative behavior in a controlled setting. Furthermore, it sheds light on the potential of simulation studies to unravel the intricacies of human social interactions, providing insights that could be transferable to real-world scenarios.

In conclusion, the deliberate alteration of agent histories and personalities not only transformed the interaction patterns among the agents but also offered a deeper understanding of the fundamental drivers of social behavior. This research opens up new avenues for exploring the dynamics of human-like interactions in artificial environments, with potential applications in social science research, AI development, and interactive entertainment.

6. Work Division

The successful completion of our modified project involved a comprehensive and collaborative approach, with each team member playing a vital role in various aspects of the development process. Initially, the team focused on pulling the original code from GitHub, a critical step that laid the groundwork for our subsequent modifications. This phase required careful analysis and understanding of the existing codebase to ensure seamless integration of our custom features.

Once the initial code was secured, team members jointly worked on setting up the simulation environment. This task was crucial as it involved configuring parameters and ensuring that the environment was primed for the new modifications we intended to introduce. Attention to detail was essential here to prevent any compatibility issues with the existing framework.

The core of our project revolved around modifying the agent behaviors and routines. By altering the bootstrap memories and daily routines of the agents in the agenthistoryinitn3.csv file, we fundamentally changed how these agents interacted within the simulation. This modification was not just a technical task but also required a deep understanding of behavioral dynamics to ensure that the changes were realistic and aligned with our project goals.

Parallel to these efforts, a continuous process of testing and debugging was undertaken. This iterative cycle was critical in identifying and fixing bugs, ensuring the stability and reliability of our modifications. Team members actively engaged in this process, contributing to a robust final product.

Documentation was another key aspect of our project, with team members meticulously recording every step of the process, from the initial code pull to the final implementation of modifications. This comprehensive documentation serves not only as a record of our project's journey but also as a valuable resource for future reference and replication of our work.

In essence, our project was a collaborative endeavor that blended technical skills with an understanding of behavioral science. It showcased our team's ability to work cohesively, adapting to challenges and leveraging each member's strengths to achieve a successful outcome. The project not only resulted in a unique and modified simulation environment but also in a rich learning experience for the entire team. 1.

Student Name	Contributed Aspects	Details
Shivam Mishra	Code Integration and Testing	Involved in pulling the initial code from GitHub, integrating modifications, and assisting in testing and debugging.
Jaspreet Singh	Environment Setup and Documentation	Helped set up the simulation environment, and contributed to documenting the setup process and modifications.
Dhruv Tyagi	Agent Behavior and Testing	Worked on modifying agent behaviors and routines, and also participated in testing the modified behaviors.
Dhrumil Shah	Testing, Debugging and Documentation	Engaged in testing and debugging, and collaborated on the documentation process, especially in detailing bugs and fixes.
Ishita Khumbani	Documentation and Environment Setup	Played a key role in documenting the project's progress and also assisted in setting up the simulation environment.
Palak Dave	Collaboration and Code Integration	Facilitated team communication and collaboration, and also contributed to integrating modifications into the main project.

Table 1. Contributions of team members.

References

[1] Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chen Qian, Chi-Min Chan, Yujia Qin, Yaxi Lu, Ruobing Xie, et al. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors in agents. *arXiv preprint arXiv:2308.10848*, 2023. 2

[2] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023. 2

[3] Chen Qian, Xin Cong, Cheng Yang, Weize Chen, Yusheng Su, Juyuan Xu, Zhiyuan Liu, and Maosong Sun. Communicative agents for software development. *arXiv preprint arXiv:2307.07924*, 2023. 2

[4] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008. 1

[5] Yoav Shoham and Kevin Leyton-Brown. Multiagent systems: Algorithmic, game-theoretic, and logical foundations. 2008. 2