Personality as Vector Space for Agent-Based Simulation

Danila Chernousov

March 12, 2025

Abstract

This research conceptualizes an agent's personality as a point in a vector space, where each dimension represents a distinct trait or behavioral tendency. This approach allows for selecting an appropriate subspace tailored to specific tasks, while disregarding irrelevant traits. The Big Five personality framework, often considered a basis set for human personality, serves as a reference point for our model. Expanding on this idea, we propose a more generalized method for profiling agents by combining orthogonal traits to generate diverse populations. In addition, incorporating a probability distribution can better mimic real-world personality variations. LLMs have demonstrated the ability to maintain consistent personalities, but it remains uncertain whether this consistency persists under alternative personality frameworks.

Keywords: Artifical Intelligence, AI Agents, Large Language Model, Agent-Based Simulation, Agent Profiling .

1 Introduction

In recent years, large language models (LLMs) have demonstrated remarkable capabilities in simulating human-like behavior, raising interest in their potential for modeling artificial agents with distinct personalities. However, existing approaches often rely on rigid predefined personality types, limiting flexibility and adaptability in various applications, such as social simulations[Park et al., 2024, Park et al., 2023], virtual assistants, and economic simulations[Leng, 2024]. This research aims to develop a more flexible approach to defining agent personalities using a vector space representation. By treating personality as a point in a high-dimensional space, where each axis represents a distinct trait or behavioral tendency, we enable more granular and dynamic personality modeling.

The key object of this research is the representation of an agent's personality using LLMs. Personality, in psychological terms, is typically defined as a set of behavioral traits that influence how individuals interact with their environment [1]. Several frameworks exist for personality modeling, with the Big Five Personality Traits being one of the most widely accepted structures, and it has received significant attention in the field [John and Srivastava, 1999]. Current research suggests that LLMs can exhibit consistent personalities when prompted correctly[Serapio-García et al., 2023, Frisch and Giulianelli, 2024, Klinkert et al., 2024]. One possible explanation for the consistency of LLM-generated personalities in psychometric tests is the extensive body of research on personality traits that is likely present in their training data. However, in cases where behavioral data is scarce or nonexistent, such as niche psychological traits or hypothetical personality constructs, LLMs may struggle to generate consistent behaviors. It remains uncertain whether they maintain this consistency when modeling personality traits that lack well-documented psychological research. This raises concerns about the reliability of LLM-generated agents, particularly in simulations that require nuanced or underexplored personality dimensions.

The primary challenge lies in ensuring consistency and coherence in agent behavior when personalities are defined using a mathematical basis.

Compared to existing methods, our approach offers several advantages. Unlike rule-based personality modeling, which requires extensive manual tuning, a vector-based model provides a systematic and scalable way to define personalities. Additionally, this method allows for smooth interpolation between personality types, enabling the creation of hybrid personalities that are difficult to define in categorical models. The

scalability of this approach also allows for realistic population sampling by drawing from parametric probability distributions tuned to mimic real-world human personality distributions.

This research contributes to the growing field of computational personality modeling by proposing a novel, mathematically grounded approach to agent personality definition. By combining vector-based representations with LLM capabilities, we aim to enhance the flexibility, scalability, and realism of artificial personality modeling in simulations and interactive AI systems.

References

Frisch, I., & Giulianelli, M. (2024). Llm agents in interaction: Measuring personality consistency and linguistic alignment in interacting populations of large language models. https://arxiv.org/abs/2402.02896

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, 102–138.

Klinkert, L. J., Buongiorno, S., & Clark, C. (2024). Driving generative agents with their personality. https://arxiv.org/abs/2402.14879

Leng, Y. (2024). Can LLMs mimic human-like mental accounting and behavioral biases? *Proceedings of the 25th ACM Conference on Economics and Computation*, 581–581.

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. https://arxiv.org/abs/2304.03442

Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. https://arxiv.org/abs/2411.10109

Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., Abdulhai, M., Faust, A., & Matarić, M. (2023). Personality traits in large language models. https://arxiv.org/abs/2307.00184