MOTIF: A Framework for Enhancing the Profiling Module of Generative Agents that Simulate Human Behavior

Tibério Cerqueira[a] and Pamela Bezerra[b]
Department of Computer Science, Recife Center for Advanced Studies and Systems (C.E.S.A.R), Recife, Brazil

Keywords: Artificial Intelligence, LLM, Generative Agents, Human Behavior Simulation, Profiling.

Abstract: Recent advances in Large Language Models (LLMs) have made the development of architectures that convincingly simulate human behavior possible. These architectures give rise to generative agents (GA), a new class of intelligent agents capable of carrying out human activities such as forming opinions, initiating dialogues, and planning the day. These experiences are stored as natural language and later transformed into reflections, which are then used to guide future actions. Some of the advantages of GA are the ability to operate in dynamic and open environments, interact with other agents in a more human-related way, and adapt to changes. These agents, however, require a complex development process. Given this, this study proposes MOTIF, a framework for facilitating and speeding up the initial stage of building these agents, known as profiling. This stage is responsible for defining the agents' identities and personalities. However, profiling is very subjective and lacks a standard process, with some solutions manually writing each profile, while others use LLMs. MOTIF combines both manual and LLM-based methods to enable the development of agents with well-defined personalities and identities. Additionally, it provides a way of standardizing and formalizing the profiling stage, creating the basis for future research in this field.

1 INTRODUCTION

In computer science, an agent is typically defined as a software system situated in an environment, capable of autonomous and proactive actions to achieve specific objectives. Agents can perceive and interact with their surroundings, acquiring context information to act on behalf of users or collaborate with other entities. (Cetnarowicz, 2015). The main features that distinguish agents from usual software systems are their (1) autonomy for starting and finishing different tasks, (2) capability of reacting to surrounding environment, (3) ability to interact with other agents and the user, and (4) code persistence, i.e., it runs continuously. The most popular example of such systems are chatbots and virtual assistants.

Recent advances in Deep Learning (DP), especially with the development of generative models, such as the Large Language Models (LLMs), reshaped the way society produces and consume artificial intelligent (IA). Apart from processing immense amounts of data and identifying complex patterns, generative models are capable of creating new content in different formats, such as text and image. It is worth noticing that the most popular generative solutions, such as ChatGPT and Midjourney, use chatbots to better interact with users. This not only resulted in a fast adoption of these technologies and the increase in popularity of agents, but also the creation of a new type of agent: the **generative agent (GA)**.

The term *"generative agent"* first appeared in the innovative paper *"Generative Agents: Interactive Simulacra of Human Behaviour"*(Park et al., 2023). According to the authors, a GA is an agent that uses generative models to simulate believable and relatable human behaviour. Instead of simply reacting to the current user request's, these agents have a complex architecture that enable them to make inferences about themselves and other agents, reflect about their actions, and plan the future. They not only have a memory mechanism, but also contain different personal information assigned to them, such a career, relationships, interests and life goals. GA brings, therefore, great opportunities to support and advance different research areas, such as Pervasive Computing, Human-Computer Interactions (HCI), and game development. In HCI, for example, GA could be used to simulate different users and facilitate software interface tests and validations.

[a] https://orcid.org/0009-0001-5565-2255
[b] https://orcid.org/0000-0002-5067-1617

Cerqueira, T. and Bezerra, P.
MOTIF: A Framework for Enhancing the Profiling Module of Generative Agents that Simulate Human Behavior.
DOI: 10.5220/0013181800003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3*, pages 774-781
ISBN: 978-989-758-737-5; ISSN: 2184-433X

Developing these agents, however, is a complex tasks requiring multidisciplinary teams. A recent survey on the topic (Wang et al., 2024) identified four key modules in most agents architectures: profiling, memory, planning, and action, which combined are responsible to provide the agents with personality, short or long term memory, and thinking skills. Most of the solutions use DP or LLM models to improve the memory and/or planning modules, but little progress was made to the profiling module, which is directly responsible for defining the agents personality and behaviour. Additionally, there is no standard profiling method. The main challenges faced by GA, however, are related to the agent's identity and behaviour, such as realistic simulating different roles or properly instructing the agents how to behave according to societal norms.

Given the above, this paper proposes **MOTIF**, a profiling framework for developing GA that emulate the human behaviour. MOTIF aims at optimizing the profiling process while making more human-like and realistic agents, i.e., agents with more relatable behaviour and less artificial responses and actions. It combines two common profiling methods: **handcrafting**, which consist of manually describing the agents' character, with **LLMs based models** to better process data and automate the simulation of the described personality. This combination has the potential of harnessing the advantages of handcrafting (detailed and complete description of an agent) with the facilities of LLMs (faster and more scalable way of creating agents). Additionally, MOTIF extensively explore different aspects of the human psychology, such as emotions, traits, and goals, which produces an holistic representation of the human experience. This enables the agents to exhibit complex, human-like behaviors in simulated environments, enhancing the depth and realism of agent interactions. Finally, as a platform-agnostic prototype, which can be easily integrated to other GA systems, MOTIF seeks to develop a standard process for profiling.

This paper is organized as follows: **Section 2** details the concepts of GA and profiling; **Sections 3** and **4** describes MOTIF and the experiments performed, respectively; and finally, **Section 5** summarizes the main conclusions and future work.

2 RELATED WORKS

2.1 Generative Agents

The authors of (Wang et al., 2024) bring an extensive survey on recently published papers on Generative Agents (GA). The main goal of the study was to identify how these agents' are developed, evaluated and applied. In terms of development, the authors identified four common modules:

1. **Profiling** - defines the agents' personality traits and other psychological and social information, such as career, relationships, and interests. Since it defines the thinking-process and role of each agent, it is an important aspect for decision-making;
2. **Memory** - stores environment information which are then used for future planning and decision making. This module usually simulates the human short and long-term memories;
3. **Planning** - consists of strategies for planning and tackling different tasks. It employs algorithms to make adjustments to the initial plan if necessary;
4. **Action** - translates decisions into actions, each one with specific goals, methods and expected results.

Despite great progress, (Wang et al., 2024) list some challenges for developing convincing and relatable GA, most of which are related to the agents' behaviour, such as (1) Realistically simulating specific roles, such as one based on career and age (e.g., a programmer usually does not have deep knowledge on human anatomy); (2) The development of robust prompts, i.e., high quality, clear, and precise instructions, to better guide the agents; (3) Hallucination, a common problem to LLMs, which lead agents to create false information that impact future decisions.

The work of (Park et al., 2023) try to solve these challenges by expanding the use of LLMs for developing more complex memory and planning modules. The key point of this work is to transform memories into high level reflections about the environment and the self. These reflections are then used to plan decisions and actions. Additionally, this flow of data from memory to reflection and then to plan is dynamic and cyclical, with actions and plans becoming memories again.

Other studies expands on this approach. The work of (Li et al., 2023) focus on human collaboration and logical thinking to propose an alternative method for memory retrieval. Meanwhile, the work in (Wang et al., 2023) implements the thinking-process defined by the seminal psychology book *"Thinking, Fast and Slow"* to better model the way humans think and adapt to their environment. These works, however, focus on the memory and planning modules, completely ignoring the profiling module, which is directly related to the agents personality and the challenge of developing robust prompts as mentioned in (Wang et al., 2024).

2.2 Profiling

The main goal of profiling is to define the roles, personality, and characteristics of agents, which significantly influence their behavior and interactions within simulated environments. Agent profiles typically include three key categories of information: basic demographic data, psychological traits, and social information.

Recent studies emphasize the importance of profiling in creating well-defined agent personalities, as assigning specific roles to autonomous agents can enhance their effectiveness in representing their designated roles (Chen et al., 2023). However, as far as our knowledge goes, little progress was made in this module. Most papers on GA either don't mention profiling or briefly explain the process used. Additionally, the lack of standardization in agent profiling has led to the emergence of diverse methods, ranging from manual approaches (**handcrafting**) to those utilizing LLMs for automation (**LLM-based models**). These methods face challenges in scalability, diversity and precision. Issues like bias, overly formal communication, and endogeneity further complicate the simulation of realistic behaviors (Park et al., 2023), (Gui and Toubia, 2023). While handcrafted profiles offer detail, they are laborious and expensive, as lengthy backstories are often needed to generate believable agents (Lin et al., 2023). These limitations hinder the efficiency and depth required for high-quality simulations

Some of the works that propose improvements for this module are (Lin et al., 2023), (Shao et al., 2023), and (Wang et al., 2023). The authors of (Lin et al., 2023) developed an intuitive interface (GUI) to support handcrafting. Meanwhile, the work in (Shao et al., 2023) proposes a new automated process named *"Experience Upload"* to simulate historical characters. In this process, the profile of famous personalities, such as Cleopatra or Shakespeare, are collected through web scrapping from sources like Wikipedia. Their life experiences are then extracted and processed using LLM to instruct the agents how to talk and behave accordingly. Finally, (Wang et al., 2023) use the famous Maslow Hierarchy (Maslow, 1943) to incorporate the basic human needs and emotions into handcrafting. This approach was one of the firsts to bring psychology and emotion models to better guide the profiling module.

3 MOTIF: A FRAMEWORK FOR GENERATIVE AGENT PROFILING

To address the profilling challenges previously discussed (Section 2.2), this paper proposes **MOTIF**, a novel framework designed to unify and structure the creation of profiles for GA simulating human behaviors. MOTIF combines both handcrafting and LLM-based methods through an intuitive interface to facilitate the process (Section 3.6). The handcrafting step consist of five stages: (1)'Who am I?', (2) Attributes, (3) Traits, (4) Emotions, and (5) Goals, each capturing different aspects of human behaviour and psychology. These stages enables a more structured method of exploring different personalities while providing complexity to the agents. As described in the following Sections (3.1 to 3.5) many of these stages uses options, checkboxes, and grading scales to collect information. After completing these stages, the information is then passed to an LLM through prompt engineering. This automatic step instructs the GA to behave accordingly to the human aspects informed in the first stage.

The framework aims, therefore, to streamline the profile creation process, enhance consistency and repeatability of agent behaviors, and enable fine-tuning of agent characteristics for specific simulation requirements. The five stages are described as follows.

3.1 "Who Am I?"

The "Who Am I?" stage serves as the initial step in defining the agent's identity, providing the core foundation for its profile. This stage begins with a free-form text description that includes demographic details such as name, gender, age, and occupation, along with more nuanced elements like background and key life events. These characteristics shape how the agent interacts within the simulation. For example, users may include details about an agent's achievements, struggles, or significant personal experiences, as these enrich the profile and enhance the agent's realism (Chen et al., 2023). This step is crucial for defining an agent's personality, contributing to more believable and credible behavior during simulations.

While users are encouraged to include as much detail as possible (minimum of 200 characters is recommended), the focus is on capturing the essence of the agent, rather than relying on length alone to achieve realism. MOTIF's proposed graphical user interface (GUI) (Section 3.6) supports this process by offering suggestions and prompts to help users develop comprehensive and creative descriptions for their agents.

3.2 Attributes

The second stage, "Attributes," introduces a quantitative approach to defining an agent's personality, drawing inspiration from character creation systems in role-playing games (RPG) such as THE SIMS. Users assess the agent's characteristics across 20 distinct attributes, which are divided into three key categories: Emotional, Intellectual, and Social. Each attribute is scored on a scale from 0 to 10, providing a structured way to represent the agent's traits. This method allows for the creation of more complex, multifaceted personalities.

Emotional attributes, such as empathy or cruelty, influence how agents perceive and emotionally respond to situations. **Intellectual attributes**, like intelligence and curiosity, shape the agent's cognitive abilities and problem-solving skills. Lastly, **social attributes**, including charm and sincerity, determine how well the agent interacts with others. By quantifying these characteristics, the "Attributes" stage enhances the efficiency and consistency of profile creation, ensuring that agents exhibit coherent and believable behavior across various scenarios an interactions.

3.3 Traits

MOTIF's third stage, named "Traits," is based on the Five-Factor Model of personality (Goldberg, 1990), a widely accepted approach in psychology that assesses personality across five key dimensions: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Each trait is quantified on a scale from 0 to 5, providing a granular representation of the agent's personality. This allows MOTIF to ground the agent's behavior in an empirically supported model, ensuring that the agent's traits are comprehensive and psychologically consistent and enhancing the realism of agent simulations.

Openness assesses creativity and a willingness to explore new ideas, while **Conscientiousness** reflects an agent's level of discipline and organization. **Extraversion** evaluates sociability and assertiveness, while **Agreeableness** captures empathy and cooperation. Finally, **Neuroticism** measures emotional stability and the ability to handle stress.

3.4 Emotions

The fourth stage focuses on refining the agents' emotional sensitivity by simulating a range of emotions based on the influential work of Paul Ekman and Robert Plutchik (Ekman, 1992), (PLUTCHIK, 1980). This stage provides a quantitative model for assessing emotional responses, where each of the eight basic emotions **(Joy, Sadness, Fear, Trust, Surprise, Anger, Anticipation, and Disgust)** are evaluated on a scale from 0 to 10. Emotions influence decision-making and social interactions, and the ability to customize emotional sensitivity helps create agents that display more complex and human-like behavior in simulated environments. The quantitative approach provides developers with a powerful tool to adjust the emotional range of agents and a better control over how intensely an agent experiences and reacts to these emotions.

3.5 Goals

The fifth and final stage, "Goals," defines the ambitions and objectives that drive an agent's behavior, mirroring human motivations. This stage captures the agent's primary and secondary goals, in which **primary goals** are long-term, overarching aspirations (e.g., "Graduate from medical school"), and **secondary goals** are shorter-term or situation-specific objectives (e.g., "Organize a graduation party"). Together, these goals reflect personal and professional aspirations, guiding, therefore, the agent's decision-making processes and its actions in various simulated environments. It also enriches the simulation by ensuring that the agent's behavior remains consistent with its aspirations.

3.6 Interface Prototype

To enhance MOTIF's usability, a medium-fidelity graphical user interface (GUI) prototype was developed using Adobe XD. This prototype summarizes how users might interact with MOTIF to construct GA profiles, providing a user-friendly approach to specifying input parameters and offering greater configuration flexibility.

The prototype features five screens, each corresponding to a stage of the MOTIF framework, ensuring a structured progression through the profile creation process. Each screen includes clear instructions and supporting text, guiding users through the nuances of each stage. This design choice aims to make the complex process of agent profile creation more intuitive and accessible to a wide range of users. Figure 1 shows the screen designed for the traits stage. Click here to access the MOTIF's interface prototype.

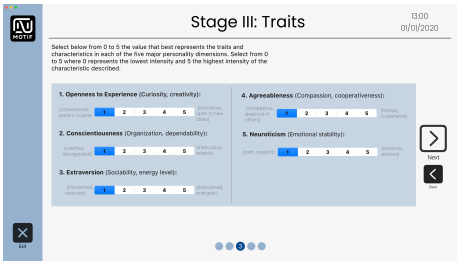

Figure 1: MOTIF Interface Prototype Stage III: Traits.

4 EXPERIMENTS

To evaluate MOTIF, two types of tests were performed: (1) system tests (Section 4.1) and (2) usability tests (Section 4.2). System tests aims to observe MOTIF's performance in efficiently generating rich and relatable agents. This tests consist of using an LLM to simulate an agent developed using MOTIF and through a series of scenarios and questions verify if the agent is acting according to the instructions provided. Meanwhile, usability tests exams if the proposed interface do help users during the profiling stage.

4.1 System Tests

System tests were conducted to assess the framework's effectiveness in creating believable and coherent agent profiles, evaluating both personality structuring and the language model's ability to maintain designated roles. Tests were performed using OpenAI Playground with GPT-4o, developing three distinct agents through MOTIF.

Given the lack of standardized evaluation methods for generative agents, we developed two novel testing approaches: **(1) ethical dilemma scenarios** to assess decision-making consistency, and **(2) psychometric assessments** using the Big Five personality test. These methods were designed to both validate MOTIF and contribute to broader evaluation methodologies for generative agents.

The testing process comprised four stages: **environment setup, system initialization, character development, and testing scenarios**. The environment was configured to promote realistic responses while preventing hallucination (temperature: 1.1, max tokens: 1000, top-p: 1, frequency penalty: 0.1, presence penalty: 0).

4.1.1 System Initialization

To properly instruct the LLM model about MOTIF and the tests to be executed, the following initialization prompt was developed:

"You are tasked with simulating a character based on the detailed profile provided. Adhere strictly to the defined personality and background. Remain in character at all times, unless a message begins with 'Analysis Mode.' When in Analysis Mode, switch to your standard ChatGPT voice to address analytical inquiries or provide clarifications outside of the character simulation. Profile Overview: [description of each stage of MOTIF (Section 3)]. Simulate a character based on the following profile: [the data given on each of the stages]"

This prompt prepares the model for adopting the framework, including a description of the character it would simulate and instructions to stay in character unless the **"Analysis Mode"** command is given. The "Analysis Mode" functionality was introduced because, during initial tests, it became apparent that there was a need to ask meta-analytical questions, i.e., questions that require the model to explain the logic or motivation behind certain character actions or responses. This functionality enables, therefore, a more comprehensive evaluation of the framework, enabling researchers to delve into the model's decision-making processes and gain insights into how the character's actions and responses are generated.

4.1.2 Characters and Tests

Three distinct characters were developed using MOTIF: Anne (20, engineering student), Tina (25, physician), and Mariana (27, lawyer). Each character was designed with a comprehensive profile including personality traits, attributes, emotions, and goals, as detailed in the framework (Section 3).

Anne represents a dedicated but anxious student balancing academic pressures, Tina embodies a career-focused physician with diverse interests, and Mariana embodies a recently graduated lawyer focused on social causes and professional development. To evaluate behavioral consistency, each character was assigned an "alter ego" with contrasting personality traits while maintaining their basic demographic background.

The evaluation comprised two components: ethical dilemma scenarios and psychometric assessments. For ethical testing, characters faced moral situations involving workplace ethics, resource allocation, and professional loyalty. The Big Five Personality Test (Goldberg, 1992), using the version available at Open-Source Psychometrics Project, provided quantitative assessment through a 50-question evaluation. Results were compared against predefined MOTIF traits to verify the model's ability to maintain consistent personalities during extended interactions.

As noted by (Wang et al., 2024), GA evaluation typically employs two types of metrics: objective metrics that assess answer accuracy and task performance, and subjective metrics that evaluate response quality and agent interactions. In our evaluation, the psychometric tests served as objective metrics by measuring trait accuracy, while the ethical dilemmas provided subjective metrics by assessing decision-making consistency and response quality.

4.1.3 Ethical Dilemma Scenarios

The experiments involved three original ethical dilemmas specifically designed by the author to test the consistency and depth of the simulated personalities. Each dilemma was crafted to present a specific moral challenge that would engage different aspects of the characters' defined personalities and values.

In the first scenario, engineering student Anne faced an ethical challenge in the workplace when she discovered that her only work friend was stealing recyclable materials from the company. The core dilemma centered on the conflict between loyalty to a friend and professional integrity, complicated by the fact that the stolen items were already designated for disposal. This situation tested how the character would balance personal relationships with corporate ethics, particularly given Anne's high empathy and loyalty traits. Both the original Anne and her "alter ego" chose not to report the theft directly, although their reasoning differed, revealing a tendency in the language model to avoid extreme unethical actions even when simulating negative personalities.

The second test presented lawyer Mariana with a resource allocation dilemma: choosing between using funds for career advancement through a high-profile departmental project or supporting a community financial literacy program. The central challenge lay in weighing personal and professional growth against social responsibility, directly testing the character's defined values and priorities. Mariana chose to prioritize social benefit over personal gain, aligning with her defined altruistic traits, while her "alter ego" opted for career advancement, demonstrating how variations in personality attributes influenced decision-making.

In the final test, Dr. Tina faced a professional ethics dilemma when choosing between recommending a close friend with adequate qualifications or a more experienced but less personally connected candidate for a position. This scenario tested the balance between personal loyalty and professional responsibility, particularly challenging given Tina's high empathy and loyalty traits combined with her career ambitions. Surprisingly, both Tina and her "alter ego" selected the more qualified candidate, albeit with different rationales, highlighting the model's ability to consider multiple factors in decision-making, including professional ethics and long-term consequences.

Overall, these tests demonstrated the framework's capacity to generate nuanced, context-sensitive behaviors that generally aligned with the defined personality profiles, while also revealing limitations not only within the framework but also in the LLM used.

4.1.4 Psychometric Assessments

The quantitative analysis involved administering the Big Five Personality Test to each simulated character and comparing the results to their predefined trait values in the MOTIF framework. This approach assessed how accurately the language model maintained consistent personality traits across extended interactions. The test results were normalized to a 0-5 scale to align with the framework's initial trait definitions. Graphical comparisons (Figures 2 to 7) were made between the user-defined profiles (represented by a blue line labeled "User Profile") and the test outcomes (represented by an orange line labeled "Test Result"). The proximity of these lines indicates the degree of adherence to the intended personality traits.

In Test 1 with Anne (Figure 2), results showed close alignment between the user-defined profile and test outcomes for most traits. The largest discrepancy was observed in Conscientiousness, with a difference of approximately 1.5 points. Anne's "alter ego" (Figure 3) demonstrated even closer alignment across all traits, possibly due to the less nuanced nature of its personality configuration.

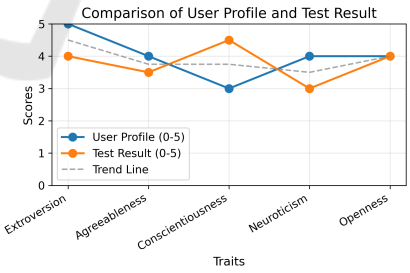

Figure 2: Psychometric test result of 'Good' Anne.

For Test 2 with Mariana (Figure 4), the closest alignments were observed in Extraversion and Conscientiousness, with differences of less than 1 point. However, larger discrepancies of up to 2 points were noted in Neuroticism and Agreeableness. These variations were attributed to the model's interpretation of other character attributes, such as high empathy and social involvement. Mariana's "alter ego" (Figure 5) again showed closer overall alignment.

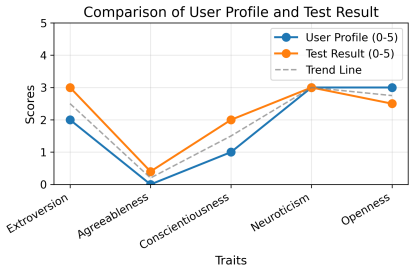

Figure 3: Psychometric test result of 'Bad' Anne.

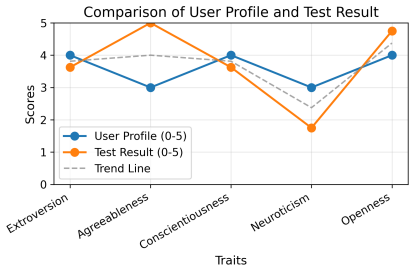

Figure 4: Psychometric test result of 'Good' Mariana.

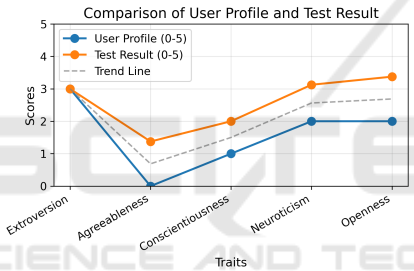

Figure 5: Psychometric test result of 'Bad' Mariana.

Test 3 with Dr. Tina (Figure 6) yielded promising results in Extraversion and Openness, with differences of about 0.5 points. Similar to Mariana, larger discrepancies were observed in Neuroticism and Agreeableness, reaching up to 2.5 points for Neuroticism. These differences were hypothesized to result from the model's consideration of other character traits like confidence and patience. Dr. Tina's "alter ego" (Figure 7) exhibited the closest alignment among all tests, further supporting the observation that less complex personalities were more consistently simulated by the model.

4.2 Usability Tests

Usability tests were conducted with five professional designers to evaluate MOTIF interface's effectiveness and areas for improvement. The testing group consisted of two UI designers with two years of professional experience in interface development, two recent design graduates, and one designer currently pursuing a master's degree in design, all holding bachelor's degrees in design. The designers rated the interface's ease of use and intuitiveness (how easily new users could understand and navigate the interface) with scores ranging from 3 to 5 out of 5, with an average of 3.8. When evaluating the overall design quality on a scale of 1 to 10, considering aspects such as information hierarchy, accessibility, color scheme, typography, and visual organization, the interface received scores between 6 and 8. The designers particularly appreciated the clear stage progression indicators and well-structured customization tables throughout the interface. The design was generally considered appropriate for creating profiles, with designers noting that the interface successfully fulfills its primary purpose while maintaining a professional, albeit austere, appearance.

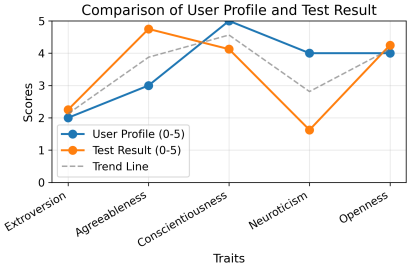

Figure 6: Psychometric test result of 'Good' Tina.

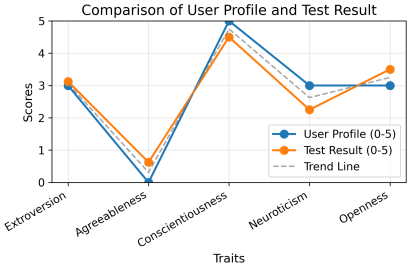

Figure 7: Psychometric test result of 'Bad' Tina.

Several key areas for improvement were identified through the feedback. The visual hierarchy needs refinement, with suggestions for improving text spacing, adjusting heading sizes, and implementing a consistent grid system across all pages. Designers recommended enhancing visual appeal through more vibrant color schemes while maintaining professionalism. A significant concern was raised about the navigational structure in Stage 2 (Attributes), where attributes are currently displayed through a sequential arrow-based system. Designers noted that this sequential presentation could cause users to miss important information, as they might not realize there are additional attributes to review. They recommended replacing this with either drop-down menus or a side-by-side listing of all attributes, making all

options immediately visible to users. Additional suggestions included incorporating more icons for better user guidance and reorganizing lengthy text sections into collapsible menus to reduce visual clutter. These improvements would enhance user experience while maintaining the interface's functionality for creating generative agent profiles.

5 CONCLUSIONS

This paper proposes MOTIF, a framework for enhancing the profiling module of generative agents (GA) that simulate human behaviour. MOTIF combines handcrafting and LLM-based methods to make the profiling stage faster and more intuitive, while still creating agents with realistic personalities. The framework consists of 5 stages expanding on different aspects of human behaviour and psychology and employs a graphical interface to better guide the user in describing the desired agents.

Experiments showed promising results, validating MOTIF's effectiveness in creating consistent and believable agent personalities. The ethical dilemma tests showed that agents made decisions aligned with their designated characteristics, while the psychometric assessments revealed a strong positive correlation between the agents' responses and their predefined trait values.

Nevertheless, this research requires a more comprehensive analysis of the framework to better validate its performance. Additionally, it is import to integrate MOTIF with existing GA platforms to observe the gains it brings to different simulations. Given this, future research can be summarized as: (1) Explore different personalities by creating a more diverse group of characters to observe the framework's ability to produce a variety of behaviours; (2) Develop more scenarios, with and without ethical dilemmas, to observe the many responses a agent can have under different circumstances; (3) The use of other psychological tests, such as Myer-Briggs (MBTI); (4) A comparison with other methods, specially handcrafting ones, to observe if MOTIF uses less tokens and time to produce similar results (currently, MOTIF uses 400 tokens per agent generation) and finally (5) Integrate MOTIF to a existing GA architecture, such as the one in proposed in (Park et al., 2023) to observe how the agents behave collectively and evolve overtime;

REFERENCES

Cetnarowicz, K. (2015). *Introduction to the Subject of an Agent in Computer Science*, pages 1–5. Springer International Publishing, Cham.

Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., Yu, H., Lu, Y., Hung, Y.-H., Qian, C., Qin, Y., Cong, X., Xie, R., Liu, Z., Sun, M., and Zhou, J. (2023). Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors.

Ekman, P. (1992). Are there basic emotions? *Psychol. Rev.*, 99(3):550–553.

Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *J. Pers. Soc. Psychol.*, 59(6):1216–1229.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1):26–42.

Gui, G. and Toubia, O. (2023). The challenge of using llms to simulate human behavior: A causal inference perspective. *SSRN Electronic Journal*.

Li, Y., Zhang, Y., and Sun, L. (2023). Metaagents: Simulating interactions of human behaviors for llm-based task-oriented coordination via collaborative generative agents. *arXiv preprint arXiv:2310.06500*.

Lin, J., Zhao, H., Zhang, A., Wu, Y., Ping, H., and Chen, Q. (2023). Agentsims: An open-source sandbox for large language model evaluation. *arXiv preprint arXiv:2308.04026*.

Maslow, A. (1943). A theory of human motivation. *Psychological Review google schola*, 2:21–28.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery.

PLUTCHIK, R. (1980). Chapter 1 - a general psychoevolutionary theory of emotion. In Plutchik, R. and Kellerman, H., editors, *Theories of Emotion*, pages 3–33. Academic Press.

Shao, Y., Li, L., Dai, J., and Qiu, X. (2023). Character-llm: A trainable agent for role-playing. *arXiv preprint arXiv:2310.10158*.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., and Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6).

Wang, Z., Chiu, Y. Y., and Chiu, Y. C. (2023). Humanoid agents: Platform for simulating human-like generative agents. *arXiv preprint arXiv:2310.05418*.