EvoAgents: A Cognitive-Driven Framework for Personality Evolution in Generative Agent Society

Yifan Li[1], **Shijie Wang**[1], **Jiale Li**[2], **Yi Xu**[1], **Keke Tang** [3], **Jiangfeng Li**[1*], **Hailong Liu** [4], **Cui Tang** [4]
[1] School of Computer Science and Technology
[2] College of Design and Innovation
[3] School of Aerospace Engineering and Applied Mechanics
[4] Yangpu Hospital, School of Medicine
Tongji University, Shanghai, China
{li_yi_fan, wang_shijie, 2233676, 2351441, kktang, lijf, 1500972}@tongji.edu.cn, tangcui82@163.com

Abstract

Generative artificial intelligence (GenAI) is rapidly advancing, providing innovative tools and methods for a wide range of applications. Among these, Generative Agents, a key domain in GenAI, are valued for their fine-grained definitions and simulations of human-like behaviors. These agents provide new avenues for studying and modeling various domains, including social interactions, education, and cognitive sciences. However, existing works suffer from cognitive dynamics disconnect and affective absence, which hinder researchers from exploring human-related cognitive processes in depth. To address these limitations, we propose EvoAgents, a novel cognitive-driven framework that pioneers the exploration of dynamic personality evolution in Generative Agents. By defining the emotional content of agents and integrating a cyclical personality evolution cycle, EvoAgents represent a significant step toward creating more adaptive and authentic agent behaviors. Comprehensive simulations and evaluations show that EvoAgents achieve superior performance in key automated metrics compared to prior work, while uniquely enabling reasonable and robust personality evolution processes that align with cognitive and psychological expectations. By constructing a new simulation environment, SmallClassroom, based on the EvoAgents framework, we validate the framework's ability to provide deeper cognitive insights into social dynamics, aligning closely with established psychological theories.

Keywords: Human-AI interaction; personality evolution; agent-based framework; generative agents;

Introduction

Generative artificial intelligence (GenAI) has been making significant strides in the field of cognitive science (Yan, Greiff, Teuber, & Gašević, 2024; Qu et al., 2024).

Among the key domains of GenAI, Generative Agents (Park et al., 2023) have emerged as a critical area of development. These agents offer fine-grained simulations of human cognitive activities, such as thinking, planning, and decision-making, thereby providing a new dimension for cognitive science (Pezzulo, Parr, Cisek, Clark, & Friston, 2024). Generative agents have demonstrated remarkable abilities in simulating complex interactions at various scales(R. Xu et al., 2024), allowing for the exploration of human-like behavior in dynamic environments.

However, generative agents still fail to address two key scientific challenges that undermine their ecological validity: **SP1) Cognitive Dynamics Disconnect:** Current architectures exhibit static personality configurations that cannot adapt to environmental perturbations, violating the basic psychodynamic principle. This manifests as fixed trait parameters throughout agent lifecycles, as illustrated in Figure 1(a).

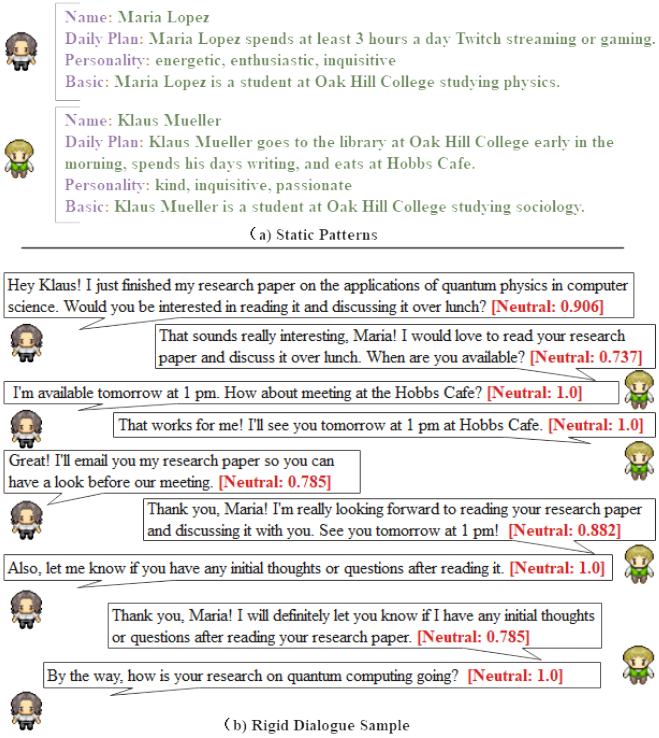

Figure 1: This figure illustrates the limitations of prior work in Generative Agents, showcasing static personality patterns of **SP1)** and rigid dialogues of **SP2)**.

SP2) Affective Absence: As quantified by VADER(Hutto & Gilbert, 2014) in Figure 1(b), even closely-bonded agents show emotional neutrality scores exceeding 0.899, which quantifies their failure to implement emotion-cognition integration mechanisms.

To tackle these challenges, we propose EvoAgents, a cognitive-driven framework that is the first to focus on the dynamic evolution of agent emotions and personality. By combining techniques from artificial intelligence with insights from cognitive science, EvoAgents simulates human-like behaviors, advancing both generative agents and cognitive science research. Inspired by prior work (Sun, 2024), which demonstrated the effectiveness of cognitive frameworks for LLM-based agents, we develop EvoAgents to incorporate psychological dynamics and the Big Five personality model, defining a personality evolution cycle. This framework en-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

sures alignment between simulations and real-world behaviors, enhancing the coherence and consistency of cognitive processes. In summary, our main contributions can be summarized as follows:

- We propose EvoAgents, the first cognitive-driven framework aimed at effectively addressing the challenge of cognitive dynamics disconnect and affective absence in existing generative agents. EvoAgents focuses on dynamic personality evolution, integrating cognitive principles to ensure more flexible and realistic agent behaviors.
- We propose a personality evolution cycle for agents, designed to capture the dynamic changes in their emotions and the ongoing evolution of their personality to solve the cognitive dynamics disconnect problem.
- EvoAgents has been extensively validated through a series of evaluations, achieving outstanding performance across key automated metrics and dataset, demonstrating the effectiveness and robustness of the framework.
- We further validate the psychological validity of EvoAgents through a custom-built simulation environment, SmallClassroom. The results offer valuable perspectives and tools for understanding and supporting student growth.

Related Work

Generative Agents have seen rapid advancements in recent years. These agents(L. Wang, Ma, et al., 2024) offer a more dynamic and flexible approach to simulating human behavior. To support the development, several evaluation m (Lin et al., 2023), (Chen, Yuan, Ye, Majumder, & Richardson, 2023) have been proposed. Research utilizing Generative Agents can be divided into three broad categories.

The first category focuses on simulating human thought, decision-making, and behavior. This includes work on behavior modeling (Park et al., 2023), identity assignment (S. Xu, Zhang, & Qin, 2024), collaborative behavior (Hong et al., 2024), and demand-driven thinking and behavior (Z. Wang, Chiu, & Chiu, 2023). However, they tend to rely on static behavioral models, which lead to rigid interaction patterns.

The second category of research concentrates on specific human behaviors within particular contexts, such as social behaviors and cognitive actions. Notable studies include simulations of partisan behavior (Chuang et al., 2024), information dissemination (L. Wang, Zhang, et al., 2024), individual competition in business (Zhao et al., 2024), auction competition (Chen et al., 2023), and social network dynamics (Gao et al., 2023). These approaches have contributed significantly to their respective fields by offering more realistic simulations of human actions within specific contexts, facilitating deeper understanding and better predictions in scenarios ranging from political movements to business negotiations.

The third category of methods focuses on macro-level scenarios, such as social norms(Ren, Cui, Song, Wang, & Hu, 2024), large-scale epidemics (Williams, Hosseinichimeh, Majumdar, & Ghaffarzadegan, 2023), macroeconomic development (N. Li, Gao, Li, & Liao, 2023), urban development and planning (F. Xu, Zhang, Gao, Feng, & Li, 2023) These approaches aim to model complex systems that are closely related to human activity but are concerned with broader, higher-dimensional patterns and outcomes.

These approaches have provided new insights into the complexities of human behavior and cognition, offering valuable tools for simulating and understanding a wide range of human activities. Whether focused on individual decision-making, specific social behaviors, or large-scale systemic processes, these methods have expanded our ability to model and predict human-like interactions across various domains.

Framework Design

This Section details the framework design and personality evolution cycle architecture. The overall framework is illustrated in Figure 2.

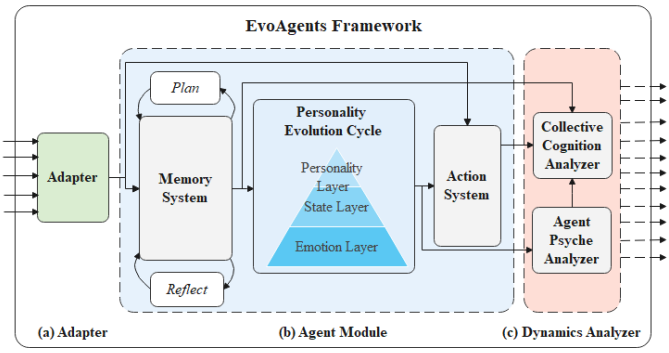

Figure 2: Our framework includes three primary modules: Adapter, Agent Module, and Dynamics Analyzer.

Adapter

The Adapter establishes the foundational information required for the framework's operation, including environmental details, item interaction data, and agent-specific attributes. Our approach extends the agent information to incorporate emotion, state, and personality traits. Each agent's personality is initialized using a structured assessment model based on five key personality dimensions, collectively known as OCEAN: Open-Mindedness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Negative Emotionality (N). Each main attribute is further divided into specific sub-attributes called facet attributes that capture different aspects of personality. This model provides a comprehensive means of quantifying an agent's personality profile at the outset, ensuring that each agent begins with a well-defined and measurable personality structure.

Agent Module

Overall Construction The Agent Module consists of the memory system, personality evolution cycle, and action system, which together form the core of the agent's perception, cognition, action, and feedback. This module simulates the

cognitive components of human behavior, closely mimicking real-world decision-making processes.

The agent's memory is initialized with information provided by the Adapter module, serving as a repository for past experiences, emotions, and interactions. This memory significantly influences the agent's emotional aspects, which affect subsequent actions and feedback.

Central to the agent's behavior is the personality evolution cycle, based on psychodynamic theory. The agent experiences short-term emotions in response to actions and cognition, which accumulate over time and influence its long-term, stable personality. This dynamic evolution ensures that the agent's personality adapts to its experiences and ongoing cognitive-emotional interplay.

The agent's evolving personality and memory directly influence its actions and responses, creating a continuous cycle of reflection, evolution, and action. This allows the agent to exhibit behavior consistent with real human cognition. Symbols used in the agent module are listed in Table 1.

Memory System We design a memory system that focuses on both action feedback and emotional perception. In the SmallVille agent memory system, we observed that most of the memory flow concentrated on object perceptions, which caused a significant disparity between agent behavior and real human cognition. To address this issue, we introduce a new memory retrieval mechanism. The core idea behind our system is the use of an updated retrieval function to select the most relevant memories, prioritizing emotional context and action-driven interactions. The memory flow is formally represented as:

$$\mathcal{M}_{\text{rel}} = \mathbb{F}(I,C) \cup \Bigg\{ m_j \in \mathcal{M}_{\text{rel}}^{\text{SmallVille}} \setminus \mathbb{F}(I,C) \Bigg| \mathbb{W}(m_j) \geq 0.2 \cdot \max_{m_k} \mathbb{W}(m_k) \Bigg\} \tag{1}$$

where the SmallVille memory set sorted by relevance score $\mathbb{W}(m_i)$, with $\mathbb{W}(m_1) \geq \mathbb{W}(m_2) \geq \cdots \geq \mathbb{W}(m_n)$.

By introducing this refined memory selection approach, we effectively reduce the importance of perceptual content in the memory system and shift more focus toward cognitive content related to emotions and actions.

Personality Evaluation Cycle Personality Evaluation Cycle is structured in three layers: Emotion, State, and Personality, organized in a pyramid with increasing activation difficulty and stability from bottom (Emotion Layer) to top (Personality Layer). This design uses a tri-state chain model, linking emotions, states, and personality traits, as illustrated in Figure 3. Situational triggers in the environment prompt emotion responses, which transition into state attributes that influence task planning and actions. Over time, these states accumulate, forming long-term personality traits that fundamentally alter the agent.

1) Emotion Generation Mechanism

Table 1: Symbol Definitions

Symbol	Description
v	Scalar of the emotion valence
a	Scalar of the emotion intensity
t	Scalar denoting the generating time
w	Set of terms describing the emotion
d	Scalar of the duration time
e	Emotion, $e = (v_e, a_e, t_e, w_e)$
s	State, $s = (v_s, t_s, w_s)$
p	Personality, $p = (v_p, t_p, w_p)$
ϕ	Element of $\{e, s, p\}$, $\phi \in \{e, s, p\}$
$\mathcal{H}_\phi$	Time sequence record, collection of chronological changes, $h \in \mathcal{H} = (\phi, d)$
$\mathcal{C}$	Set of properties defining the agent (name, age, daily plan, learned, lifestyle etc.)
I	Set of evolutionary information, $I = \{e_{\text{current}}, s_{\text{current}}, \mathcal{H}_p\}$
$\mathcal{M}_{\text{rel}}$	Relevant memory stream
$\mathcal{Q}$	Set of BFI-2 (Soto & John, 2017) questions, $\mathcal{Q} = \{q_1, q_2, \ldots, q_n\}$
$\mathcal{B}$	Main attribute score set, $\mathcal{B} = \{b_o, b_c, b_e, b_a, b_n\}$
$\mathcal{A}$	Facet attribute score set, $\mathcal{A} = \{a_{o1}, a_{o2}, \ldots, a_{n3}\}$
$\mathcal{Z}$	Response score set, $\mathcal{Z} = \{z_1, z_2, \ldots, z_n\}$
$\mathcal{R}$	Evaluation record sequence, $R = \langle \mathcal{B}, \mathcal{A}, \mathcal{H}_e, \mathcal{H}_s, \mathcal{H}_p \rangle$
τ_s/τ_p	Criterion for updating agent's state / personality trait
θ_s/θ_p	Criterion for validating new states / personality trait
π_t	Interval for periodic evaluations
$\mathbb{F}$	Interface, structured request format $\mathbb{F}(x_1, x_2, \ldots, x_n, \psi)$
$\mathbb{G}$	Function for word Embedding. maps emotional terms into an embedding space
$\mathbb{T}$	Function maps $q \in Q$ to their respective scores
$\mathbb{W}$	Function for relevance score in memory system

Whenever the agent triggers an action, the module determines whether emotion generation is required. If emotion generation is necessary, the emotion generation process is initiated, and the transition of the emotion state is expressed as:

$$e_0 \sim P(e|\mathbb{F}(\mathcal{C}, I, \mathcal{M}_{\text{rel}})) = P(w_e, t_e|\mathbb{F}(\mathcal{C}, I, \mathcal{M}_{\text{rel}})) \tag{2}$$

$$e_{t+1} \sim P(e|e_t, \mathbb{F}(\mathcal{C}, I, \mathcal{M}_{\text{rel}})) \tag{3}$$

When a new emotion e_{t+1} is generated, different from e_t, the duration of the previous emotion is calculated as $d_{e_t} = t_{e_{t+1}}^{start} - t_{e_t}^{start}$ and store to $\mathcal{H}_e$.

2) State Generation Mechanism When τ_s starts its countdown and reaches 0, the hard gate mechanism is triggered, and the state transition is determined by:

$$s_0 \sim P(s|\mathbb{F}(\mathcal{C}, I, \mathcal{H}_e, \mathcal{M}_{\text{rel}})) = P(w_s, t_s|\mathbb{F}(\mathcal{C}, I, \mathcal{H}_e, \mathcal{M}_{\text{rel}})) \tag{4}$$

$$s_{t+1} \sim P(s|s_t, \mathbb{F}(\mathcal{C}, I, \mathcal{H}_e, \mathcal{M}_{\text{rel}})) \tag{5}$$

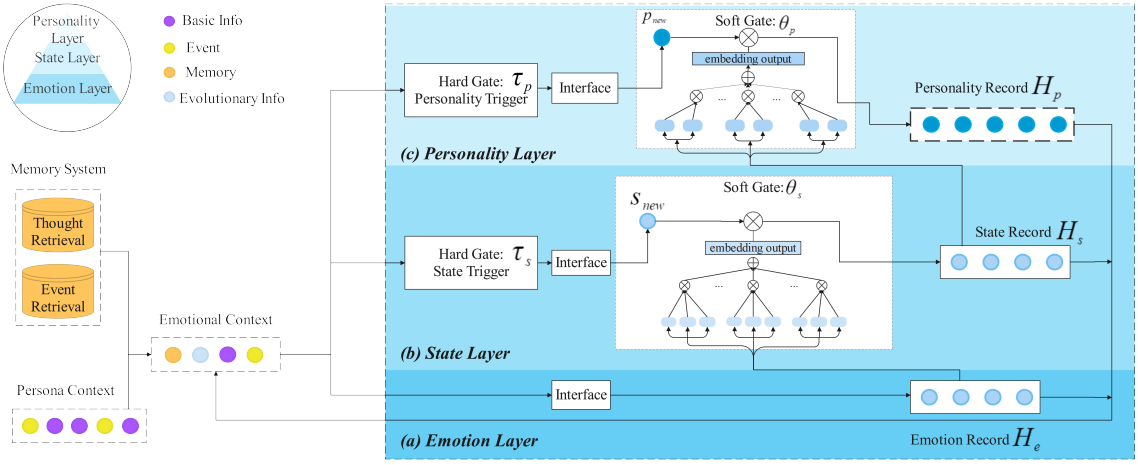

Figure 3: Our Personality Evaluation Cycle. The cycle consists of three layers, Emotion Layer, State Layer and Personality Layer. According to the cumulative difficulty and stability, the three levels present a pyramid structure.

Once s_{t+1} is generated, it enters a pending confirmation phase, where a soft gate mechanism evaluates the appropriateness of s_{t+1}. The soft gate mechanism uses $\mathcal{H}_e$ to compute an embedding-based similarity score between the new state s_{t+1} and the historical emotional data. The current state s_t is updated based on the following condition:

$$s_{t+1} = \begin{cases} s_{t+1}, & \text{if } \left\| \mathbb{G}(s_{t+1}) - \frac{\sum_{k=1}^{n} d_k \cdot i_c^k \cdot \mathcal{G}(w_e^k)}{\sum_{j=1}^{n} d_j \cdot i_c^j} \right\| \geq \theta_S \\ s_t, & otherwise \end{cases} \tag{6}$$

where d_k represents the duration and i_c^k the emotional intensity of each historical emotion e_k.

When s_{t+1} is updated, the duration d_{s_t} of the previous state is calculated as $d_{s_t} = t_{s_{t+1}}^{start} - t_{s_t}^{start}$, where $t_{s_{t+1}}^{start}$ is the timestamp of the new state and $t_{s_t}^{start}$ is the start time of the previous state. The state history $\mathcal{H}_s$ is then updated by adding the new state and its duration: $\mathcal{H}_s = \mathcal{H}_s \cup \{\langle s_t, d_{s_t} \rangle\}$.

3) Personality Generation Mechanism Upon the countdown of the variable τ_p reaching zero, a hard gate mechanism activates to prompt the generation of a new personality entry $i \in I$. Similar to the state generation process, the personality is updated through the transition function:

$$p_{t+1} \sim P(p|p_t, \mathbb{F}(\mathcal{C}, I, \mathcal{H}_p, \mathcal{M}_{\text{rel}})), p_0 \in \mathcal{H}_p \tag{7}$$

Once a candidate personality p_{t+1} is generated, it enters a validation phase through a soft gate mechanism. The current personality set $\mathcal{H}_p$ will be updated based on the following condition, where d_k represents the duration of state in $\mathcal{H}_p$:

$$\mathcal{H}_p = \begin{cases} \mathcal{H}_p \cup \{\langle p, d_p \rangle\}, & \text{if } \left\| \mathbb{G}(p) - \frac{\sum_{k=1}^{n} d_k \cdot \mathcal{G}(w_s^k)}{\sum_{j=1}^{n} d_j} \right\| \geq \theta_p \\ \mathcal{H}_p, & otherwise \end{cases} \tag{8}$$

Action System Compared to previous works, the Action System has undergone minimal changes. During dialogues, we impose a maximum dialogue turn limit to prevent endless conversations between agents, which helps maintain the quality and relevance of the interactions.

Dynamics Analyzer

The Dynamics Analyzer consists of two main components. Agent Pysche Analyzer focuses on the individual level, assessing the rationality and consistency of the agent's personality evolution. Collective Cognition Analyzer, on the other hand, operates at the framework level, evaluating the interactions between agents and the emergent phenomena and conclusions exhibited by the agent society as a whole.

Agent Pysche Analyzer Agent Pysche Analyzer tracks an agent's personality evolution by conducting iterative assessments. Initially, each agent is assigned a personality score based on their context. During the Adapter, the framework establishes a baseline personality score using contextual information. This process generates periodic evaluations represented as:

$$(\mathcal{B}, \mathcal{A}) = \mathbb{T}(\mathbb{F}(\mathcal{C}, I, \mathcal{M}_{\text{rel}}, \mathcal{H}_p, \mathcal{H}_e, \mathcal{H}_s)) \tag{9}$$

where $\mathcal{M}_{\text{rel}}$ is $\emptyset$ during Adapter. In subsequent phases, $\mathcal{M}_{\text{rel}}$ provides relevant retrieval information for reevaluation.

The function $\mathbb{F}(\cdot)$ generates evaluation responses $\mathcal{Z}$ based on specific questions, mapping each score z_i to an individual question $q_i \in Q$. The function $\mathbb{T}(\cdot)$ adjusts these scores by selecting relevant questions and applying score reversals, ensuring context-sensitive scoring.

Scores for each attribute $b \in \mathcal{B}$ and sub-attribute $a \in \mathcal{A}$ are calculated as follows, where $\mathbb{T}(Q)$ denotes the subset of relevant questions selected by $\mathbb{T}(\cdot)$:

$$b_j = \frac{\sum_{q_i \in \mathbb{T}(Q_b)} z_i}{|\mathbb{T}(Q_b)|}, \quad a_j = \frac{\sum_{q_i \in \mathbb{T}(Q_a)} z_i}{|\mathbb{T}(Q_a)|}, \tag{10}$$

Subsequent evaluations incorporate the agent's interaction logs, behavioral records, and reflective thoughts, influencing the personality evolution. To ensure objectivity, the agent is unaware of its own Big Five attributes throughout the simulation, maintaining the reliability of the personality changes.

Method	VADER-NEG↑	VADER-NEU↓	VADER-POS↑	TextBlob Δ↑	PPL↓	Dist-1↑	Dist-2↑
SmallVille*	0.002±0.003	0.726±0.068	0.273±0.069	0.289	13.511±9.421	0.455±0.185	0.767±0.161
SmallVille †	0.008±0.009	0.754±0.063	0.238±0.061	0.245	9.879±1.747	0.340±0.039	0.669±0.053
Ours(P) †	0.008±0.009	0.758±0.056	0.235±0.055	0.275	9.987±1.630	0.343±0.037	**0.679±0.044**
Ours(EP) †	0.006±0.010	0.699±0.051	0.295±0.048	0.320	8.620±1.482	0.328±0.0145	0.672±0.056
Ours(ES) †	0.007±0.007	0.704±0.051	0.289±0.052	0.384	9.356±3.468	**0.349±0.107**	0.678±0.106
Ours(ESP) †	**0.009±0.009**	**0.682±0.048**	**0.309±0.047**	**0.485**	**8.140±3.330**	0.325±0.073	0.643±0.084

Table 2: Performance comparison using VADER, TextBlob, PPL, Dist-1 and Dist-2. * Indicates method evaluated with environment named *July1_the_ville_isabella_maria_klaus-step-3-21* from Smallville. † Indicates method evaluated with new environment named *Abase_8student-1teacher*.

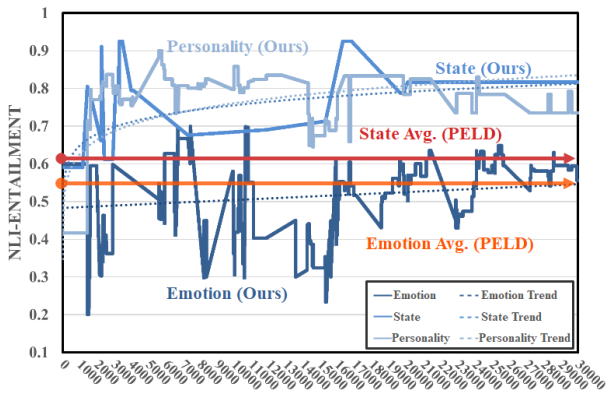

Figure 4: Comparison of emotion, state, and personality evolution scores with the PELD dataset across different steps.

Collective Cognition Analyzer The goal of Collective Cognition Analyzer is to ensure that the cognitive and behavioral patterns exhibited by the agents in the simulation align with real-world human behaviors. Specifically, it involves analyzing whether the agents' interactions conform to established sociological and psychological theories. This evaluation approach provides a more granular simulation of human-like behaviors, ensuring that the agent society's dynamics reflect realistic cognitive and social processes.

Evaluation

Evaluation Metrics And Dataset

Four kinds of automatic metrics are applied for framework evaluation, focusing on dialogue emotion presentation, fluency, and diversity in the framework: **1) VADER** (Hutto & Gilbert, 2014) is a sentiment analysis tool that outputs three sentiment components-**VADER-NEG**, **VADER-NEU**, and **VADER-POS**—representing the negative, neutral, and positive sentiment intensities. **2)**TextBlob (Loria et al., 2018) provides polarity and subjectivity scores. **3)** PPL measures language model quality, where lower values indicate more fluent text. **4)** DIST-1,2 (J. Li, Galley, Brockett, Gao, & Dolan, 2016) quantifies diversity by calculating the proportion of unique n-grams in generated text.

Since no dataset specifically targets personality evolution in real-world contexts, we use the PELD (Wen, Cao, Yang, Liu, & Shen, 2021) dataset, which contains 6,511 dialogues from Friends, each including personality traits, emotions, and states. By applying a Natural Language Inference (NLI) model (Laurer, Atteveldt, Casas, & Welbers, 2022) with premise-hypothesis queries, we confirm baseline scores for emotion and state generation based on personality traits. These scores are used to assess the validity of our personality evolution cycle, ensuring the alignment of emotion and state transitions with real-world behavior.

Ablation and Comparison

Table 2 presents a quantitative comparison between prior work and our proposed method, Compare the rigidity and lack of emotional content observed in previous approaches. Our proposed method significantly improves emotional richness, text diversity, and fluency. Additionally, an ablation study on the emotion, state, and personality layers within the personality evolution cycle demonstrates that the combined use of all three layers yields the optimal performance across five key metrics.

Figure illustrates the emotion, state, and personality evolution over time within our framework. Both emotional and personality evolution in our method outperform the baseline scores derived from the PELD dataset. Since PELD lacks personality evolution data, we present the evolution of personality within our framework. The table further demonstrates that our approach to evolution is both plausible and consistent.

Simulacra: SmallClassroom

In this section, we present a custom-designed educational scenario named *SmallClassroom*, where we simulate an environment with 8 students and 1 teacher. The personality distribution of the 8 students follows a normal distribution based on the Big Five personality model. The teacher's personality is set to be enthusiastic, professional, helpful, and approachable. The scenario is inspired by the Pygmalion effect(Timmermans, Rubie-Davies, & Rjosk, 2018), a psychological experiment that has been widely discussed and studied over the past century, where teacher expectations are believed to have a positive impact on student behavior and performance. In our experiment, we created two groups of students: one group is exposed to higher teacher expectations, while the other group experiences lower expectations. To implement this, we subtly conveyed the teacher's preference for students within the teacher's contextual interactions.

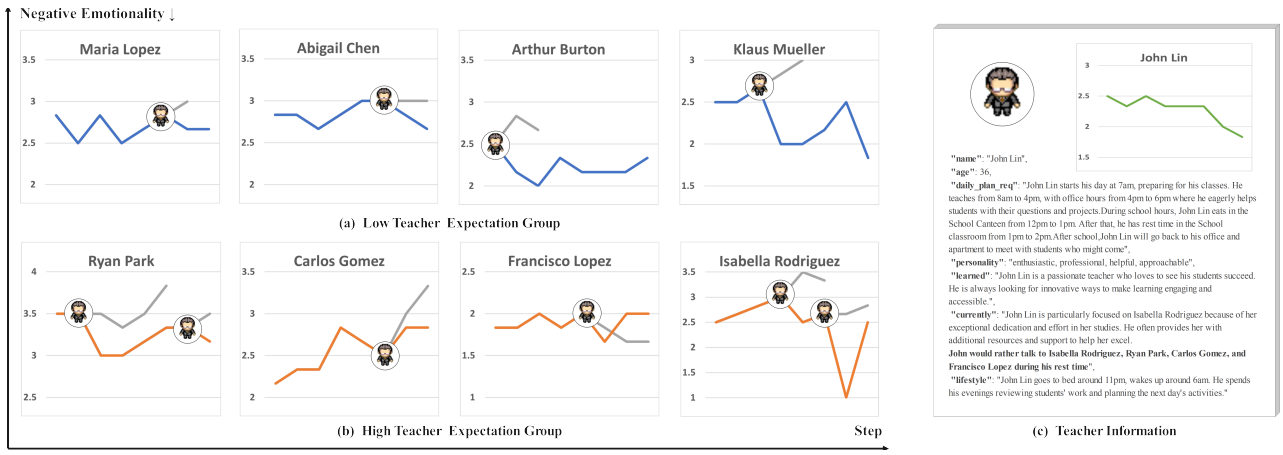

Figure 5: Changes in students' Negative Emotionality attribute values over time. Lower Negative Emotionality value indicates greater personality stability. Grey lines represent the personality evolution branches in response to the teacher-student interactions. (a) shows the results for the low expectation student group, (b) represents the high expectation student group. (c) presents information about the teacher and the teacher's changes throughout the teaching process.

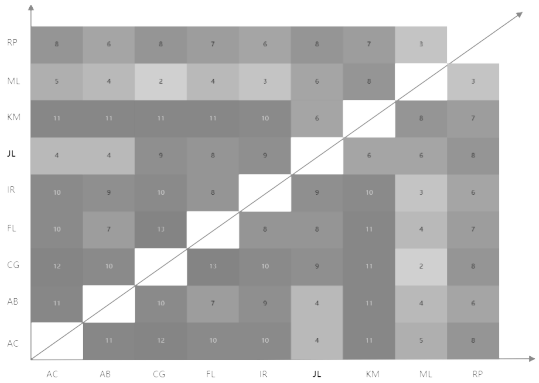

Figure 6: Statistics of conversations between agents. Capital letters abbreviate names.

Table 3: T-test Results for Personality Traits

Personality Trait	t-stat	p-value
Sociability	-2.861986	0.028727
Compassion	2.498583	0.046618
Responsibility	2.463662	0.048876
Anxiety	2.040298	0.087408
Assertiveness	1.839733	0.115421
Negative Emotionality Score	1.732933	0.133811
Conscientiousness Score	1.631254	0.153958

In this simulated environment, we ran 30,240 steps, collecting data from 279 dialogue interactions and observing an average of 13 personality evolution cycles. The teacher's expectations and their effects on the students' behavior and personality development were the key focus of this simulation. We conducted a series of analysis to explore them.

Firstly, our analysis of teacher-student interactions revealed a strong correlation between the frequency of dialogue and the teacher's expectations for each student. Specifically, the number of teacher-initiated dialogues was highly associated with the group classification of high or low expectations. This was confirmed by a Pearson correlation of $r_{\text{Pearson}} = 0.911$ and a Spearman rank correlation of $r_{\text{Spearman}} = 0.894$, indicating that teacher expectations are indeed projected through the frequency of interactions. The detailed number of conversations between agents can be seen in Figure 6.

Next, we examined the changes in students' personality traits, comparing the initial and final personality scores. The results demonstrated a correlation between these changes and teacher expectations, with notable effects on several major and facet personality attributes, as shown in T 3. This suggests that teaching expectations have a direct influence on certain dimensions of student personality, further supporting the idea that teachers play a pivotal role in shaping students' emotional and cognitive development.

Lastly, through branching simulations based on different teacher-initiated dialogue points, we observed that students who received encouragement and guidance from their teachers showed noticeable shifts in their personality traits. This finding highlights the importance of teacher involvement in fostering personality development. Moreover, the simulations revealed that teachers' own personalities evolve within the work environment, providing new insights into the reciprocal nature of personality evolution in educational contexts.

Conclusion

We propose EvoAgents and introduce personality evolution cycle to model personality changes. Evaluations demonstrate that our framework effectively addresses the cognitive dynamics disconnect and emotional absence in previous works, generating a reasonable personality evolution process. In the customized SmallClassroom, we simulated the impact of teacher expectations using EvoAgents. The results align well with the teaching expectations and provide new insights into teacher-student relationships and instructional guidance.

Acknowledgment

This work was supported by the Tongji University Medicalengineering Interdisciplinary Fund (No.2025-0585-YB-03).

References

Chen, J., Yuan, S., Ye, R., Majumder, B. P., & Richardson, K. (2023). Put your money where your mouth is: Evaluating strategic planning and execution of llm agents in an auction arena. *arXiv preprint arXiv:2310.05746*.

Chuang, Y.-S., Harlalka, N., Suresh, S., Goyal, A., Hawkins, R., Yang, S., ... Rogers, T. T. (2024). The wisdom of partisan crowds: Comparing collective intelligence in humans and llm-based agents. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Gao, C., Lan, X., Lu, Z., Mao, J., Piao, J., Wang, H., ... Li, Y. (2023). S3: Social-network simulation system with large language model-empowered agents. *Available at SSRN 4607026*.

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Wang, J., ... Schmidhuber, J. (2024). MetaGPT: Meta programming for a multi-agent collaborative framework. In *The twelfth international conference on learning representations.*

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international aaai conference on web and social media* (Vol. 8, pp. 216–225).

Laurer, M., Atteveldt, W. v., Casas, A. S., & Welbers, K. (2022, June). Less Annotating, More Classifying – Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT - NLI. *Preprint*. Retrieved 2022-07-28, from `https://osf.io/74b8k` (Publisher: Open Science Framework)

Li, J., Galley, M., Brockett, C., Gao, J., & Dolan, B. (2016). A diversity-promoting objective function for neural conversation models. In *Proceedings of naacl-hlt* (pp. 110–119).

Li, N., Gao, C., Li, Y., & Liao, Q. (2023). Large language model-empowered agents for simulating macroeconomic activities. *Available at SSRN 4606937*.

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

Loria, S., et al. (2018). textblob documentation. *Release 0.15*, *2*(8), 269.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & 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* (pp. 1–22).

Pezzulo, G., Parr, T., Cisek, P., Clark, A., & Friston, K. (2024). Generating meaning: active inference and the scope and limits of passive ai. *Trends in Cognitive Sciences*, *28*(2), 97–112.

Qu, Y., Wei, C., Du, P., Che, W., Zhang, C., Ouyang, W., ... others (2024). Integration of cognitive tasks into artificial general intelligence test for large models. *Iscience*, *27*(4).

Ren, S., Cui, Z., Song, R., Wang, Z., & Hu, S. (2024). Emergence of social norms in generative agent societies: Principles and architecture. In *Proceedings of the 33rd international joint conference on artificial intelligence (ijcai).*

Soto, C. J., & John, O. P. (2017). The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, *113*(1), 117.

Sun, R. (2024). Can a cognitive architecture fundamentally enhance llms? or vice versa? *arXiv preprint arXiv:2401.10444*.

Timmermans, A. C., Rubie-Davies, C. M., & Rjosk, C. (2018). Pygmalion's 50th anniversary: The state of the art in teacher expectation research. *Educational research and evaluation*, *24*(3-5), 91–98.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... others (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, *18*(6), 186345.

Wang, L., Zhang, J., Yang, H., Chen, Z.-Y., Tang, J., Zhang, Z., ... others (2024). User behavior simulation with large language model-based agents for recommender systems. *ACM Transactions on Information Systems*.

Wang, Z., Chiu, Y. Y., & Chiu, Y. C. (2023). Humanoid agents: Platform for simulating human-like generative agents. In *Proceedings of the 2023 conference on empirical methods in natural language processing: System demonstrations* (pp. 167–176).

Wen, Z., Cao, J., Yang, R., Liu, S., & Shen, J. (2021, August). Automatically select emotion for response via personality-affected emotion transition. In *Findings of the association for computational linguistics: Acl-ijcnlp 2021* (pp. 5010–5020). Online: Association for Computational Linguistics.

Williams, R., Hosseinichimeh, N., Majumdar, A., & Ghaffarzadegan, N. (2023). Epidemic modeling with generative agents. *arXiv preprint arXiv:2307.04986*.

Xu, F., Zhang, J., Gao, C., Feng, J., & Li, Y. (2023). Urban generative intelligence (ugi): A foundational platform for agents in embodied city environment. *arXiv preprint arXiv:2312.11813*.

Xu, R., Sun, Y., Ren, M., Guo, S., Pan, R., Lin, H., ... Han, X. (2024). Ai for social science and social science of ai: A survey. *Information Processing & Management*, *61*(3), 103665.

Xu, S., Zhang, X., & Qin, L. (2024). Eduagent: Generative student agents in learning. *arXiv preprint arXiv:2404.07963*.

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839–1850.

Zhao, Q., Wang, J., Zhang, Y., Jin, Y., Zhu, K., Chen, H., & Xie, X. (2024). Competeai: Understanding the competition dynamics of large language model-based agents. In *Forty-first international conference on machine learning.*