applied sciences

MDPI

Article

Extroversion–Introversion Rescheduler in Generative Agent via Few-Shot Prompting

Sungwon Cho [1], Youngmin Ji [2] and Yunsick Sung [2,*]

[1] Computer Science and Engineering, Dongguk University-Seoul, Seoul 04620, Republic of Korea; 2020112037@dgu.ac.kr
[2] Department of Computer Science and Artificial Intelligence, Dongguk University-Seoul, Seoul 04620, Republic of Korea; ji4642@dgu.ac.kr
* Correspondence: sung@dongguk.edu

Abstract

Generative Agent (GA) has emerged as a promising framework for simulating human-like behaviors. However, it is required for GA to generate a schedule that consistently reflects the agent's E-I trait particularly in the extroversion–introversion (E-I) category to improve the realism of GA. We propose an E-I evaluation and rescheduling method that adjusts the agent's schedule. Specifically, our method takes as input a one-hour schedule segmented into five-minute tasks and a corresponding E-I trait classified into seven degrees ranging from extremely high extroversion to extremely high introversion. Using the *Evaluator* powered by GPT-4o mini, each task is assessed for the alignment with the E-I traits. Each task that fails to meet a threshold is regenerated using few-shot prompting based on a collected successful schedule. This process is repeated until all tasks are aligned with the corresponding traits. Finally, the evaluator accesses the overall E-I consistency of the schedule that contains the tasks. Therefore, it is possible for the proposed method to enable E-I-consistent schedule generation in GA without retraining any models. In experiments, the proposed framework improved E-I alignment from an average of 14.7% to that of 78.4% with only 1.38 iterations on average, demonstrating both practical effectiveness and computational efficiency.

Keywords: generative agent; large language models; few-shot prompting; personality types

1. Introduction

Recent advancements in artificial intelligence have enabled the development of an intelligent agent capable of simulating human-like thinking, emotion, and behavior. In particular, the emergence of Large Language Models (LLMs) [1–3] has significantly enhanced the capabilities of natural language-based reasoning, dialogue generation, and action planning. LLMs now function as central components in systems that aim to replicate human-like cognition, moving beyond the role of simple text generators [4,5].

One representative framework that leverages LLMs is the Generative Agent (GA) system [6–8], which simulates daily human life by integrating memory, planning, and action execution through natural language. GA retrieves past experiences, perceives current contexts, and generates future-oriented behaviors in text, enabling coherent interactions that are socially, temporally, and contextually aligned [9,10].

The GA system has been applied in various domains such as education, gaming, affective computing, and digital humans [11–13]. However, they primarily treat *extroversion–introversion (E-I)* as metadata [14,15] or embed it implicitly via memory recall. In most

Check for updates

Academic Editors: Marina Anca Cidota and Clara Maathuis

Received: 22 December 2025
Revised: 12 January 2026
Accepted: 13 January 2026
Published: 15 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

cases, assigned personalities are not explicitly reflected in generated behaviors or schedules. For example, an agent configured as highly introverted often performs actions that are inherently extroverted, such as initiating social conversations or leading public events [16,17]. Such inconsistencies diminish coherence and realism in user experiences.

To enable E-I consistent simulation, the generated tasks should be aligned with the agent's personality in both semantics and expression. This necessitates an external intervention framework [7,18,19] that evaluates E-I alignment in pre-generated behaviors and revises those that are misaligned. Ideally, this should not require modifications to the internal GA system, supporting modularity and scalability [8,20].

Previous studies have explored prompt-based approaches to control E-I expression in LLM outputs [16,21,22], embedding E-I related descriptions into generation instructions. However, these approaches focus on single-sentence control [23,24] and are not applicable to sequential or schedule-based behavior generation. Moreover, many methods require fine-tuning or internal modification of the LLM or GA pipeline [25,26], leading to increased complexity and reduced portability.

This paper proposes an E-I-based reschedule framework that operates independently of the GA system. The framework retrieves pre-generated tasks in a schedule, evaluates the alignment of tasks, and revises misaligned tasks using a few-shot prompting technique based on GPT-4o [27]. During this rescheduling, the core intents of the behaviors of each task are preserved, while their expressions are tailored to reflect a target E-I trait. The rescheduling loop continues until all tasks in a schedule meet the corresponding alignment criteria. The proposed method focuses on the E-I dimension of personality [16,28,29] suited for reflecting tasks. Rather than binary classification, the E-I dimension is divided into seven traits, from Extremely High Introversion to Extremely High Extroversion [12], enabling fine-grained evaluation. These E-I traits can be more effective results than dividing E and I in a dichotomous way [30].

This paper makes the following contributions:

- We design and implement an external intervention framework that modifies pre-generated tasks to match the agent's personality without changing the internal GA system [5].
- We introduce a subdivision of the E-I dimension into seven traits, enabling finer personality alignment beyond the binary [16,28].
- We develop an evaluator and rescheduling loop that repeatedly assesses E-I alignment and refines misaligned tasks via few-shot prompting [22,31], while using majority voting and median rules to stabilize LLM variability [23,26].
- We validate our framework in the GA system measuring the similarity between the traits of pre-generated tasks and target E-I traits [15,21].
- We present a generalizable E-I alignment framework that can be extended to other personality dimensions such as Thinking–Feeling or Sensing–Intuition [29,32].

This paper is set out as follows. Section 2 introduces the related research, Section 3 explains the framework that we propose, Section 4 discusses the experimental results, and Section 5 discusses the conclusions.

2. Related Work

2.1. Generative Agent and Personality Representation

Generative Agent (GA) simulates human-like behavior by combining memory retrieval, planning, and action execution via natural language [6,7]. Park et al. (2023) introduce a GA framework enabling the agent to form intentions from structured memories and generate daily schedules using LLM prompting [6]. These systems are capable of coherent temporal and contextual behavior simulation, as demonstrated in environments

like Smallville [9,10], and have been applied in complex domain simulations across gaming and affective computing [11,13].

Despite impressive results, existing GA frameworks lack mechanisms for embedding E-I traits into behavior generation. Personality is often added only as metadata or implicitly via memory reflection [14,28], without affecting behavior scheduling. Consequently, an agent configured as an introvert often performs extroverted actions, indicating a disconnect between label and action [16,17]. GA systems also do not provide processes for evaluating personality alignment or revising misaligned tasks [8,19]. This gap necessitates a post-generation intervention capable of adjusting behaviors according to personality while preserving agent intent and narrative structure [4,20]. Although the GA system improves temporal and contextual coherence through memory retrieval, planning, and action execution, personality is typically treated as metadata or reflected only implicitly; explicit mechanisms for evaluating personality alignment and revising misaligned task entries are generally absent, leaving trait–action mismatches at the task traits.

2.2. Prompt-Based Personality Control via In-Context Learning

In-context learning is a powerful LLM mechanism that enables models to follow behavioral patterns based on examples in prompts—without parameter updates [3,33,34]. Techniques such as zero-shot and few-shot prompting have been used to guide personality-consistent outputs from LLMs, especially using personality descriptors [16,21,22]. PersonaLLM (2023) showed that GPT-3.5/4 can output Big-Five-aligned text when prompted with personality examples [21]. Also Pan et al. (2023) demonstrated evaluation and induction of E-I traits via prompting [22]. Other work has used few-shot personality example prompting to produce text indistinguishable from a target individual's style [23,27].

However, these prompting-based approaches typically apply at the sentence level [23,24,35] and are not designed for multi-step behavioral sequences like schedules [12,17]. They also lack evaluation or correction loops to ensure that generated behaviors align properly with target personality traits [22,31]. Moreover, in-context prompting alone cannot guarantee deterministic outputs, due to randomness and token-sampling variations in LLMs [25,26,35]. Thus, reliable personality-consistent behavior generation requires additional evaluator, filter, or correction mechanisms [20,22,33]. Work on persona prompting and personality evaluation/induction shows that in-context examples (zero-/few-shot) can steer outputs toward target traits at the sentence level, partially addressing controllability; nevertheless, these methods remain sentence-scoped and typically omit explicit evaluation–correction loops and whole-schedule verification, leaving task traits alignment and stability unaddressed.

3. Behavior Reschedule Using LLM

This section introduces a method that intervenes in the task rescheduling process, which occurs during the execution of the Generative Agent (GA). The primary objective is to adjust the schedule in such a way that it better aligns with a specific E-I trait pre-assigned to each agent. By incorporating this intervention, we aim to ensure that the reschedule is consistent with the pre-assigned E-I traits of an agent.

Our framework differs by (i) optimizing toward a pre-assigned target E-I trait, (ii) treating E-I evaluation and rescheduling as separable roles, and (iii) enforcing task-fidelity constraints so that E-I alignment does not override functional behavior.

3.1. Overview

During rescheduling, it is essential to preserve behaviors from tasks in a schedule. Based on this premise, we propose a method to generate a reschedule that reflects the

agent's E-I traits within a GA. The overall structure of the proposed method, which consists of *Data Process*, *E-I Task Process*, and *E-I Schedule Process* is shown in Figure 1. Data Process obtains a schedule including tasks from Generative Agent and delivers those to E-I Task Process. E-I Task Process modifies each task to reflect a target E-I trait. The E-I Schedule Process evaluates the extent to which the modified content reflects the target E-I trait. Finally Data Process updates revised tasks. The architecture of the frameworks that reschedule through this process is as follows.

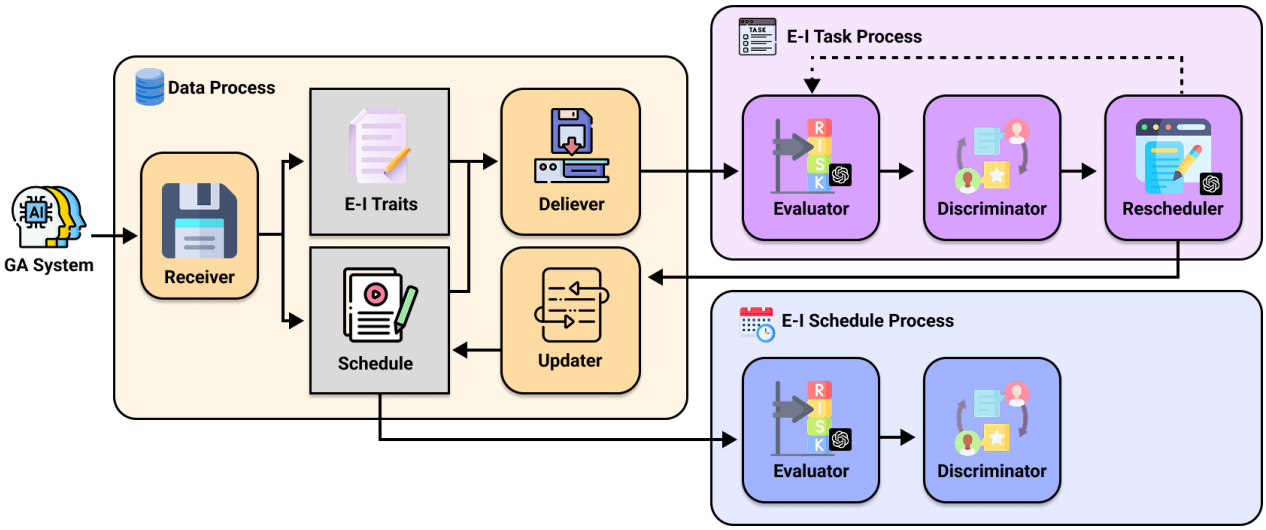


Figure 1. Reschedule framework architecture for GA system.

3.2. Data Process

Data Process refers to *Receiver*, *Deliver*, and *Updater*. First, Receiver determines a target E-I trait of a GA and retrieves schedules. Each schedule contains multiple tasks where each task is defined per five to ten minutes and contains execution duration. Each task is performed by multiple behaviors. Next, Deliver transfers the E-I trait and schedule. Finally Updater receives a revised schedule from E-I Task Process and updates the schedule received previously with the revised schedule. Updater transfers the revised schedule to E-I Schedule Process. If the result by E-I Schedule Process fails, this is reflected in the evaluation of the experimental results.

This flow aims to transform one of the input datasets, S, by reflecting the E-I trait of T. The process to generate this output proceeds as follows. First, the input data is loaded in the form of a text file using Receiver. The variable R is initialized at 0 and is incremented by 1 each time a detailed task is successfully modified to match the E-I trait (T) and serves as the termination condition for the while loop. S.size represents the number of tasks entries in S. For each entry, an Evaluator is applied. If the evaluation result matches the E-I trait as determined by the Discriminator, R is incremented; otherwise, the entry in S is modified using Rescheduler. This process is repeated until all entries reflect the designated E-I trait. However, the total number of iterations is limited to three to address the possible infinite loop risk. Once complete, the modified contents of S itself are then returned as the final output. The overall rescheduling procedure is summarized in Algorithm 1.

Algorithm 1: Data Process and E-I Task Process.

Input: E-I trait T^{Target}, Schedule S
Output: Schedule S
/* Data Process */
$T^{Target}, S \leftarrow$ `Deliver(Receiver());`
/* E-I Task Process */
$LC \leftarrow 1$;
while $R \neq S.size$ & $LC \leq 3$ **do**
  $R \leftarrow 0$;
  $LC \leftarrow LC + 1$;
  for $i \leftarrow 1$ **to** $S.size$ **do**
    $E \leftarrow$ `Discriminator(Evaluator(`$T^{Target}, S[i]$`));`
    if E *is*"*Success*" **then**
      $R \leftarrow R + 1$;
    else
      S$[i] \leftarrow$ `Rescheduler(`$S[i]$`);`
return S;

3.3. E-I Task Process

E-I Task Process consists of three main sub-processes. First, *Evaluator* evaluates the E-I trait of each task. Rather than performing a single evaluation, it conducts the evaluation three times. If all three evaluation results are same, the evaluation is utilized. If all three results differ, the median value among them is utilized. In all other cases, the result that appears twice is selected as the final output. The prompt used in this Evaluator is shown in Figure 2.

The prompt can be divided into four distinct sections, each serving a specific purpose. Section 1 provides a brief overview of the information contained in the entire prompt, guiding the LLM on what type of information it should extract or focus on. Section 2 offers the basis for Evaluator to classify a given sentence into one of the seven traits. It includes a list of 100 words for each E-I trait, where each word is semantically associated with specific behavioral traits. Similarly, Section 3 also provides the basis for classification, but instead of individual words, it presents 100 behavior-related topics for each E-I trait. Lastly, Section 4 defines how Evaluator should respond to the input sentence, detailing the format and method by which the answer should be derived.

Next, *Discriminator* compares the E-I trait with the target E-I trait and determines whether the evaluation of each task is a Success or a Fail. If Evaluator result matches the target E-I traits, the task is classified as Success; otherwise, it is classified as Fail. Success indicates that the tasks reflect the target E-I traits, while Fail indicates that they does not.

Finally *Rescheduler* alters the failed tasks classified as Fails to better reflect the target E-I trait and reprocesses them from Evaluator. This iterative process continues until all tasks are classified as Successes by Discriminator. Rescheduler operates based on the Discriminator result of each detailed tasks. If the task is classified as Success, the Rescheduler is not executed. However, if it is classified as Fail, the Rescheduler reconstructs the detailed tasks to better reflect the target E-I traits. The prompt used in this Rescheduler is shown in Figure 3.

Rescheduler requires examples in order to apply the few-shot learning technique. To support this, the prompt explains to the LLM how to interpret the rescheduler prompt template and how to derive a desired output. When there are successfully modified tasks

preceding or following the current tasks, they are added for few-shot as examples. Since these examples are contextually connected to the target E-I trait, they guide the LLM to generate more coherent and natural modifications. In contrast, if no successfully modified tasks are available, the prompt instructs the LLM to transform the input sentence directly according to the target E-I traits.

Section 1 (Lines 1–7)

1: The belows are the information that contains the words and the topics of the
2: behaviors of extroversion/introversion in personality as a guideline for evaluating
3: sentences by the degrees of extroversion and introversion in personality. Those are
4: expressed by 7 levels: extremely high extroversion, high extroversion, somewh
5: at extroverted, balanced, somewhat introverted, high introversion, and extre
6: mely high introversion.Here is a list of behavior-related words categorized by
7: the degree of extroversion/introversion in personality.

Section 2 (Lines 8–14)

8: Words_of_Extremely_High_Extroversion = [talkative, loud, and etc.]
9: Words_of_High_Extroversion = [articulate, outgoing, sociable, and etc.]
10: Words_of_Somewhat_Extroverted = [friendly, open-minded, and etc.]
11: Words_Balanced = [flexible, adaptable, composed, and etc.]
12: Words_Somewhat_Introverted = [reserved, thoughtful, and etc.]
13: Words_High_Introversion = [quiet, introspective, and etc.]
14: Words_Extremely_High_Introversion = [silent, reclusive, and etc.]

Section 3 (Lines 15–23)

15: The following are behavior-related topics categorized by the degree of extrov
16: ersion/introversion in personality.

17: Topics_Extremely_High_Extroversion = [....]
18: Topics_High_Extroversion = [....]
19: Topics_Somewhat_Extroverted = [....]
20: Topics_Balanced = [....]
21: Topics_Somewhat_Introverted = [....]
22: Topics_High_Introversion = [....]
23: Topics_Extremely_High_Introversion = [....]

Section 4 (Lines 24–32)

24: Based on the personality extroversion and introversion guidelines that includes wo
25: rds and topics related to Extroversion and Introversion that I provided, please
26: evaluate following sentence using one of the seven levels mentioned above (e
27: xtremely high extroversion, high extroversion, somewhat extroverted, balanc
28: ed, somewhat introverted, high introversion, and extremely high introversion).
29: The output should be exactly one of the following values: extremely high extrov
30: ersion, high extroversion, somewhat extroverted, balanced, somewhat introve
31: rted, high introversion, or extremely high introversion. Do not include any ad
32: ditional text or formatting. If the evaluation is not possible, output only IDK.

Figure 2. Evaluator prompt template.

3.4. E-I Schedule Process

E-I Schedule Process consists of two sub-process. First, *Evaluator* assesses the E-I traits of both the modified schedule by E-I Task Process and the original schedule by Receiver. Evaluator in E-I Schedule Process works in the same way as Evaluator in E-I Task Process, evaluating the E-I trait of a given input. The key difference lies in the input format: while Evaluator in E-I Task Process receives tasks, Evaluator in E-I Schedule Process takes the entire schedule structured in hourly units as input, and evaluates the E-I traits based on this full schedule.

(Prompt)
The target personality has one of the seven levels classified according to the degree of extroversion/introversion in the personality. The names of each level are as follows, with higher levels indicating a tendency closer to E and lower levels indicating a tendency closer to I, using "balanced" as the midpoint:

Extremely High Extroversion
High Extroversion
Somewhat Extroverted
Balanced
Somewhat Introverted
High Introversion
Extremely High Introversion

Following the Example below, transform the Input sentence into one that fits the Target personality. If the Result sentence is blank, generate a sentence appropriate for the Target personality. If the Result sentence provided in the example is not blank, consider that the new sentence occurs as a continuation of the Result sentence. However, if it is difficult to make a natural connection, you may disregard the continuity. You must maintain the structure and core content of the Input sentence, and the sentence must be written as a single sentence. Also, do not include the phrase "Result sentence:" in your output.

(Prompt Template)
Example
Target personality: {personality}
Input sentence: {sentence}
Result sentence: {result}

Figure 3. Rescheduler prompt template.

Next, *Discriminator* compares the E-I trait results of the two sets of schedules against the target E-I trait, and quantitatively evaluates the extent to which the modified tasks better reflect the target E-I traits compared to the original one. Discriminator quantifies how well the entire schedule reflects the target E-I traits. It first converts the E-I traits of both the original tasks and the modified tasks, as well as the target E-I traits, into corresponding numerical representations based on a predefined mapping. Then, using a custom-designed formula, it calculates how closely each task aligns with the target E-I traits and expresses the result as a percentage, indicating the degree of alignment.

4. Experiments

The purpose of this experiment is to modify a schedule suitable for the target E-I trait. However, there can be no issues in applying the reschedule to the GA system by maintaining the behavior of the agent.

4.1. Experimental Setup

A pre-generated schedule and target E-I trait are passed to Receiver, and the reschedule is returned as a result. The following experiments were conducted in the environment using RTX 5070Ti, Ubuntu-22.04, and python 3.10.

4.2. Data Composition

In the GA system, each agent's schedule is generated by the hour. Figure 4 is part of the hourly schedule generated in an actual GA system.

When a one-hour schedule in Figure 4 is created, the schedule is subdivided into five-minute units. Figure 5 is the tasks in which the 06:00 AM schedule in Figure 4 is

subdivided into five-minute tasks in the GA system. Figure 6 shows the agent persona in the GA system. We add the E-I trait to the Agent Persona. This E-I trait becomes a target E-I trait for evaluating and modifying tasks in our reschedule framework.

[ID:yvTaux] Monday February 13 – 06:00 AM Activity: Isabella is waking up and completing her morning routine

[ID:nB94XF] Monday February 13 – 07:00 AM Activity: Isabella is preparing breakfast and getting ready for the day

[ID:SHUBrR] Monday February 13 – 08:00 AM Activity: Isabella is opening Hobbs Café and getting ready for the day's customers

[ID:AbhCuS] Monday February 13 – 09:00 AM Activity: Isabella is serving customers at Hobbs Café

Figure 4. Hourly schedule.

Traits are arranged from extroversion to introversion as follows: Extremely High Extroversion, High Extroversion, Somewhat Extroverted, Balanced, Somewhat Introverted, High Introversion, and Extremely High Introversion. A corresponding behavior-based prompt is defined for each trait and is used in the E-I Task Process. Data in this format was provided to Receiver. The following is Table 1, which summarizes the E-I traits used in the experiment. For ease of reference in presenting the experimental results, each of these trait subtypes was abbreviated using the initial letter of its descriptor.

Table 1. Personality to Symbol.

Personality	Initial
Extremely High Introversion	EHI
High Introversion	HI
Somewhat Introverted	SI
Balanced	BA
Somewhat Extroverted	SE
High Extroversion	HE
Extremely High Extroversion	EHE

['waking up and completing her morning routine (waking up and turning off her alarm)', 5]

['waking up and completing her morning routine (stretching in bed to wake up her body)', 5]

['waking up and completing her morning routine (getting out of bed and making her bed)', 5]

['waking up and completing her morning routine (heading to the bathroom to wash her face)', 5]

['waking up and completing her morning routine (brushing her teeth)', 5]

['waking up and completing her morning routine (taking a shower)', 10]

['waking up and completing her morning routine (drying off and getting dressed for the day)', 10]

['waking up and completing her morning routine (doing her hair and applying makeup)', 10]

['waking up and completing her morning routine (checking her phone for messages and notifications)', 5]

Figure 5. Schedule.

daily_plan_req : Isabella Rodriguez opens Hobbs Cafe at 8am everyday, and works at the counter until 8pm, at which point she closes the cafe.

name : Isabella Rodriguez

first_name : Isabella

last_name : Rodriguez

age : 34

innate : friendly, outgoing, hospitable

learned : Isabella Rodriguez is a cafe owner of Hobbs Cafe who loves to make people feel welcome. She is always looking for ways to make the cafe a place where people can come to relax and enjoy themselves.

EI : extremely high extroversion

Figure 6. Agent Persona.

4.3. Experimental Results

The experiment followed the procedures outlined below: Original task sentences were composed of neutral or slightly introverted everyday expressions. Target E-I trait was set to various traits of extroversion one by one. Each task was evaluated against the current E-I trait. As a result, more than 65% of initial tasks already followed the same trait of the target E-I traits after the first iteration. The remaining tasks reached the target trait within 1–2 iterations in most cases. The average number of iterations required for alignment was approximately 1.38, indicating that a high degree of alignment could be achieved with minimal repetition. In particular, the final modified schedule, when evaluated as a whole, showed an average similarity of over 78% with the target E-I trait. Additionally, as the experiments were repeated, the variance in alignment decreased, indicating stable and consistent performance in alignment.

Table 2 presents the results of Evaluator in the E-I Task Process for original tasks in which the agent, whose E-I trait was categorized as Extremely High Introversion, was working at a café.

Table 2. Evaluating a schedule description by Extremely High Introversion in the E-I task process.

Schedule Description	Results
taking customer orders	SE
preparing drinks	SI
preparing food items	SI
serving food and drinks to customers	SI
checking on customers to ensure satisfaction	BA
cleaning tables and the counter area	SI
restocking supplies (napkins, utensils, etc.)	SI
taking a quick break to hydrate	SI
preparing for the next wave of customers	BA

Despite the agent being assigned the E-I trait of Extremely High Introversion, it can be observed that the original tasks contain less activities that align with this trait. After undergoing the E-I Task Process, the resulting modified tasks and the inferred E-I traits determined by Evaluator in the E-I Schedule Process are shown in Table 3.

Table 3. Evaluating the modified schedule description by Extremely High Introversion in the E-I Task Process.

Modified Schedule Description	Results
avoids eye contact by staring fixedly at the monitor while mechanically reciting lines rehearsed internally a dozen times to get through the interaction as fast as possible.	EHI
focuses intensely on the precise measurements and mechanics of the beverage making to create a protective bubble of silence, consciously ignoring the chatter on the other side of the counter.	EHI
meticulously assembles the ingredients with head bowed low, using the intense focus on the task as a shield to avoid any chance of eye contact with the waiting customers.	EHI
gently slides the tray forward with a barely audible mumble of thanks, immediately retreating to the safety behind the counter to minimize the duration of the social interaction.	EHI
discreetly scans the table from across the room to verify they are eating happily, choosing to rely on visual cues rather than risking an intrusive verbal interaction.	EHI
wipes down the surfaces with rapid efficiency and an averted gaze, using the physical intensity of the task as a shield to discourage anyone from striking up a conversation.	EHI
treats the task as a stealth mission, moving with calculated silence and speed to replenish items without drawing a single glance from the customers nearby.	EHI
retreats to the most secluded corner of the back room, sipping water while glued to a phone screen to signal total unavailability for casual small talk.	EHI
uses the brief moment of silence to obsessively organize the workspace and take deep stabilizing breaths, mentally steeling against the impending social energy drain.	EHI

The modified tasks in Table 3 were updated with more specific descriptions in order to preserve the behaviors of the original tasks. Evaluator also confirmed that it aligns with the agent's assigned E-I traits by Extremely High Introversion. Table 4 presents the evaluation results for both the original and modified tasks during the E-I Schedule Process.

Table 4. Evaluting the modified schedule's E-I trait.

Original	Schedule	Modified Schedule
EHI	BA	EHI

Equations (1) and (2) were used to numerically confirm the experimental results. Equation (1) quantifies the difference between the evaluation result for the original schedule and the evaluation result for the target E-I trait. Equation (2) quantifies the difference between the evaluation result for the modified schedule and the evaluation result for the target E-I trait.

$$\text{Similarity Score} = \left(1 - \frac{|\text{Map}(T^{Task}) - \text{Map}(T^{Target})|}{\text{MaxDiff}}\right) \times 100 \quad (1)$$

$$\text{Modified Similarity Score} = \left(1 - \frac{|\text{Map}(T^{Modified}) - \text{Map}(T^{Target})|}{\text{MaxDiff}}\right) \times 100 \quad (2)$$

The following two equations quantify how well the task has been rescheduled. Equation (1) uses the E-I trait of the original task and the original E-I traits to compute the score, where T^{Task} refers to the E-I traits derived from the original task, and T^{Target} refers to the original E-I traits. The part labeled Map denotes the numerical representation of each E-I trait. In this mapping, Extremely High Introversion (EHI) is assigned the value 1, and Extremely High Extroversion (EHE) is assigned the value 7. The E-I traits between

these two extremes are assigned values from 2 to 6 in sequential order. MaxDiff refers to the maximum possible difference in Map values; in this study, it is set to 6 since the E-I trait is divided into 7 traits. However, this value can be adjusted if the E-I trait is further subdivided. Finally, the entire expression is multiplied by 100 to convert the result into a percentage. Based on the experimental results from Tables 1–4, if Map(T^{Task}) is 4, Map(T^{Target}) is 1, and MaxDiff is 6, the resulting Similarity Score is 50.0. Equation (2) is similar to Equation (1), except that T^{Target} is replaced with $T^{Modified}$, which uses the E-I trait derived from the modified task instead of the original one. All other values remain the same, and in this case, Map($T^{Modified}$) is 1, resulting in a Similarity Score of 100.0. T^{Task} presents the Similarity Scores of the original and modified tasks for each E-I trait. The overall experimental results are summarized in Table 5.

Table 5. Total result.

Personality	Schedule (Zero-Shot)	Modified Schedule (Few-Shot)
EHI	10.73	86.91
HI	12.51	71.92
SI	8.89	70.66
BA	52.82	84.23
SE	4.75	72.19
HE	2.57	80.21
EHE	10.63	85.77

The results indicate that the original tasks tend to be classified primarily as Balanced, with a general bias toward introversion rather than extroversion. Using the proposed framework, it was confirmed that E-I traits can be reflected in tasks with a minimum accuracy of 70%.

Several key insights were drawn from the experimental results:

When the initial schedule is composed of overly neutral expressions or vocabulary, GPT-based evaluations tended to converge to labels such as Balanced or Somewhat Introverted. This is attributed to the nature of Schedule, which primarily consisted of quiet routines (e.g., making the bed, washing up), offering limited Extroverted behavioral indicators. Nevertheless, GPT was able to consistently modify each task into one of the seven traits based on predefined evaluation criteria.

During the E-I Task Process using few-shot prompts and example templates, the behavior of the original task was preserved, while its contextual or expressive style was altered. This resulted in an E-I Task Process that was more likely to receive labels such as High Extroversion or Extremely High Extroversion from GPT, thereby validating the effectiveness of the prompt design.

To mitigate the non-deterministic nature of GPT responses, the parameter was set relatively low. Each task was evaluated three times, and the final classification was determined based on majority voting or the median score. This strategy helped reduce erratic predictions and contributed to maintaining the system's overall performance above average.

4.4. Comparison Experiment

The proposed framework shows the similar performance in other models. Llama3.1-405B instruction was used, and Qwen2-72B Base was used. The experimental method was performed by changing the model used to Llama3.1-405B and Qwen2-72B, although the method of the proposed framework is the same. The following table was the comparative experimental result.

Table 6 summarizes the results of applying the Score equation using the proposed framework, Llama3.1-405B, and Qwen2-72B, respectively. In the proposed framework, Llama3.1-405B showed similar results overall, and was more accurate than the framework proposed by the BA trait. In contrast, Qwen2-72B had lower inference performance than the proposed framework, showing an overall weaker appearance in E-I trait inference.

Table 6. Comparative experimental results.

Model	Accuracy (%)
Ours	78.84
Llama3.1-405B	79.66
Qwen2-72B	67.03

This confirms that our proposed framework was dependent on the generative and inference performance of the model and achieves similar results on models with similar performance to the proposed framework.

5. Conclusions

This paper successfully designed and experimentally validated a GPT-based reschedule system that modifies schedule to align with a target E-I trait, with the focus on the Extroversion/Introversion dimension subdivided into seven traits. A traditional prompt-based E-I trait system typically performs single-pass generation and provides limited compliance guarantees under stochastic decoding. In contrast, we proposed a reschedule framework for a Generative Agent that treats E-I trait alignment as a specification compliance problem. Given pre-generated tasks and a target E-I trait, our system enforces alignment through evaluating and modifying tasks, without fine-tuning the underlying LLM. The proposed system achieved high alignment performance with most tasks converging to the target E-I trait, and the final outputs consistently achieving over 78% similarity with the target E-I trait. Furthermore, this paper demonstrated the feasibility of adjusting E-I traits in text while preserving behavior meanings through the use of few-shot learning. The application of majority voting and median-based stabilization strategies in the evaluation sub-process effectively controlled GPT's non-determinism and ensured consistency in results. These findings indicate that the system goes beyond simple classification and can be applied to personalized text generation, E-I traits design, and behavior recommendation systems based on affective traits. Future research may expand this approach to other categories such as Thinking–Feeling, Sensing–Intuition, and Judging–Perceiving, leading to the development of more precise and multidimensional personality-based text adjustment systems.

Author Contributions: Conceptualization, S.C., Y.J. and Y.S.; Methodology, S.C., Y.J. and Y.S.; Software, S.C.; Investigation, S.C.; Writing—original draft preparation, S.C. and Y.J.; writing—review and editing, S.C., Y.J. and Y.S.; Visualization, S.C.; Supervision, Y.S.; Project administration, Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2026-RS-2023-00254592) grant funded by the Korea government(MSIT). This research was supported by the "Regional Innovation System & Education (RISE)" through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government. (2025-RISE-01-007-04).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and materials used in this study consist of behavior schedule generated by a generative agent system and prompts constructed for personality alignment using personality-based labels. All core components of the proposed method, including the style evaluator, prompt examples, and behavior revision logic, are fully described within the article. Since the model operates through OpenAI's GPT-4o-series APIs (with separate Rescheduler and judge roles), no additional model training or proprietary datasets were used. Any further inquiries about implementation details can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used OpenAI GPT-4o series models for language polishing and for generating example prompt templates. The authors reviewed and edited the content and take full responsibility for the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Park, J.S.; O'Brien, J.C.; Cai, C.J.; Morris, M.R.; Liang, P.; Bernstein, M.S. Generative Agents: Interactive Simulacra of Human Behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), Francisco CA, USA, 29 October–1 November 2023. [CrossRef]
2. Mairesse, F.; Walker, M.A. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Comput. Linguist.* **2011**, *37*, 455–488. [CrossRef]
3. Pennebaker, J.W.; King, L.A. Linguistic styles: Language use as an individual difference. *J. Personal. Soc. Psychol.* **1999**, *77*, 1296–1302. [CrossRef]
4. Ziems, C.; Held, W.; Shaikh, O.; Chen, J.; Zhang, Z.; Yang, D. Can large language models transform computational social science? *arXiv* **2023**, arXiv:2305.03514. [CrossRef]
5. Costa, P.T.; McCrae, R.R. The Revised NEO Personality Inventory (NEO-PI-R). In *The SAGE Handbook of Personality Theory and Assessment*; Boyle, G.J., Matthews, G., Saklofske, D.H., Eds.; SAGE Publications: Thousand Oaks, CA, USA, 2008; Volume 2, pp. 179–198.
6. Madaan, A.; Tandon, N.; Gupta, P.; Hallinan, S.; Gao, L.; Wiegreffe, S.; Alon, U.; Dziri, N.; Prabhumoye, S.; Yang, Y.; et al. Self-Refine: Iterative Refinement with Self-Feedback. *arXiv* **2023**, arXiv:2303.17651. [CrossRef]
7. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS 33), Virtual, 6–12 December 2020; pp. 1877–1901.
8. Aher, G.V.; Arriaga, R.I.; Kalai, A.T. Using large language models to simulate multiple humans and replicate human subject studies. In Proceedings of the 40th International Conference on Machine Learning (ICML). PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 337–371.
9. Andreas, J. Language models as agent models. *arXiv* **2022**, arXiv:2212.01681. [CrossRef]
10. Ganesan, A.V.; Lal, Y.K.; Nilsson, A.H.; Schwartz, H.A. Systematic evaluation of GPT-3 for zero-shot personality estimation. *arXiv* **2023**, arXiv:2306.01183. [CrossRef]
11. Jiang, G.; Xu, M.; Zhu, S.C.; Han, W.; Zhang, C.; Zhu, Y. Evaluating and inducing personality in pre-trained language models. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), Abu Dhabi, United Arab Emirates, 7–11 December 2022.
12. Jiang, G.; Xu, M.; Zhu, S.C.; Han, W.; Zhang, C.; Zhu, Y. MPI: Evaluating and inducing personality in pre-trained language models. *arXiv* **2022**, arXiv:2206.07550.
13. Li, T.; Zheng, X.; Huang, X. Tailoring personality traits in large language models via unsupervisedly-built personalized lexicons. *arXiv* **2023**, arXiv:2310.16582.
14. Mao, S.; Zhang, N.; Wang, X.; Wang, M.; Yao, Y.; Jiang, Y.; Xie, P.; Huang, F.; Chen, H. Editing personality for LLMs. *arXiv* **2023**, arXiv:2310.02168.
15. Shao, Y.; Li, L.; Dai, J.; Qiu, X. Character-LLM: A trainable agent for role-playing. *arXiv* **2023**, arXiv:2310.10158. [CrossRef]
16. Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; Weston, J. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv* **2018**, arXiv:1801.07243. [CrossRef]
17. Yang, T.; Shi, T.; Wan, F.; Quan, X.; Wang, Q.; Wu, B.; Wu, J. PsyCoT: Psychological questionnaire as powerful chain-of-thought for personality detection. *arXiv* **2023**, arXiv:2310.20256.
18. Goldberg, L.R. The structure of phenotypic personality traits. *Am. Psychol.* **1993**, *48*, 26–34.
19. Karra, S.R.; Nguyen, S.T.; Tulabandhula, T. Estimating the personality of white-box language models. *arXiv* **2022**, arXiv:2204.12000.

20. Myers, I.B. *The Myers-Briggs Type Indicator: Manual*; Educational Testing Service: Princeton, NJ, USA, 1962.
21. Jiang, H.; Zhang, X.; Cao, X.; Breazeal, C.; Roy, D.; Kabbara, J. PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024, Mexico City, Mexico, 16–21 June 2024; pp. 3605–3627.
22. Pan, K.; Zeng, Y. Do LLMs possess a personality? Making the MBTI test an amazing evaluation for large language models. *arXiv* **2023**, arXiv:2307.16180. [CrossRef]
23. Safdari, M.; Serapio-García, G.; Crépy, C.; Fitz, S.; Romero, P.; Sun, L.; Abdulhai, M.; Faust, A.; Matarić, M. Personality traits in large language models. *arXiv* **2023**, arXiv:2307.00184. [CrossRef]
24. Mairesse, F.; Walker, M.A.; Mehl, M.R.; Moore, R.K. Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Intell. Res.* **2007**, *30*, 457–500. [CrossRef]
25. Wang, Z.M.; Peng, Z.; Que, H.; Liu, J.; Zhou, W.; Wu, Y.; Guo, H.; Gan, R.; Ni, Z.; Zhang, M.; et al. RoleLLM: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. *arXiv* **2023**, arXiv:2310.00746.
26. Kennedy, R.B.; Kennedy, D.A. Using the Myers-Briggs Type Indicator® in career counseling. *J. Employ. Couns.* **2004**, *41*, 38–43.
27. Liu, Y.; Iter, D.; Xu, Y.; Wang, S.; Xu, R.; Zhu, C. G-Eval: NLG evaluation using GPT-4 with better human alignment. *arXiv* **2023**, arXiv:2303.16634. [CrossRef]
28. Shinn, N.; Cassano, F.; Berman, E.; Gopinath, A.; Narasimhan, K.; Yao, S. Reflexion: Language Agents with Verbal Reinforcement Learning. *arXiv* **2023**, arXiv:2303.11366. [CrossRef]
29. Hirsh, J.B.; Peterson, J.B. Personality and language use in self-narratives. *J. Res. Personal.* **2009**, *43*, 524–527. [CrossRef]
30. Choong, E.J.; Varathan, K.D. Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum. *PeerJ* **2021**, *9*, e11382. [CrossRef] [PubMed]
31. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *arXiv* **2022**, arXiv:2203.02155. [CrossRef]
32. Wang, X.; Fei, Y.; Leng, Z.; Li, C. Does role-playing chatbots capture the character personalities? Assessing personality traits for role-playing chatbots. *arXiv* **2023**, arXiv:2310.17976.
33. OpenAI. GPT-4 Technical Report. Technical report, OpenAI. *arXiv* **2023**, arXiv:2303.08774.
34. Cao, X.; Kosinski, M. ChatGPT can accurately predict public figures' perceived personalities without any training. *arXiv* **2023**, arXiv:2308.01920.
35. Rao, H.; Leung, C.; Miao, C. Can ChatGPT assess human personalities? A general evaluation framework. *arXiv* **2023**, arXiv:2303.01248. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.