X's Day: Personality-Driven Virtual Human Behavior Generation

Haoyang Li[1], Zan Wang[1], Wei Liang[1,2†], and Yizhuo Wang[1†]

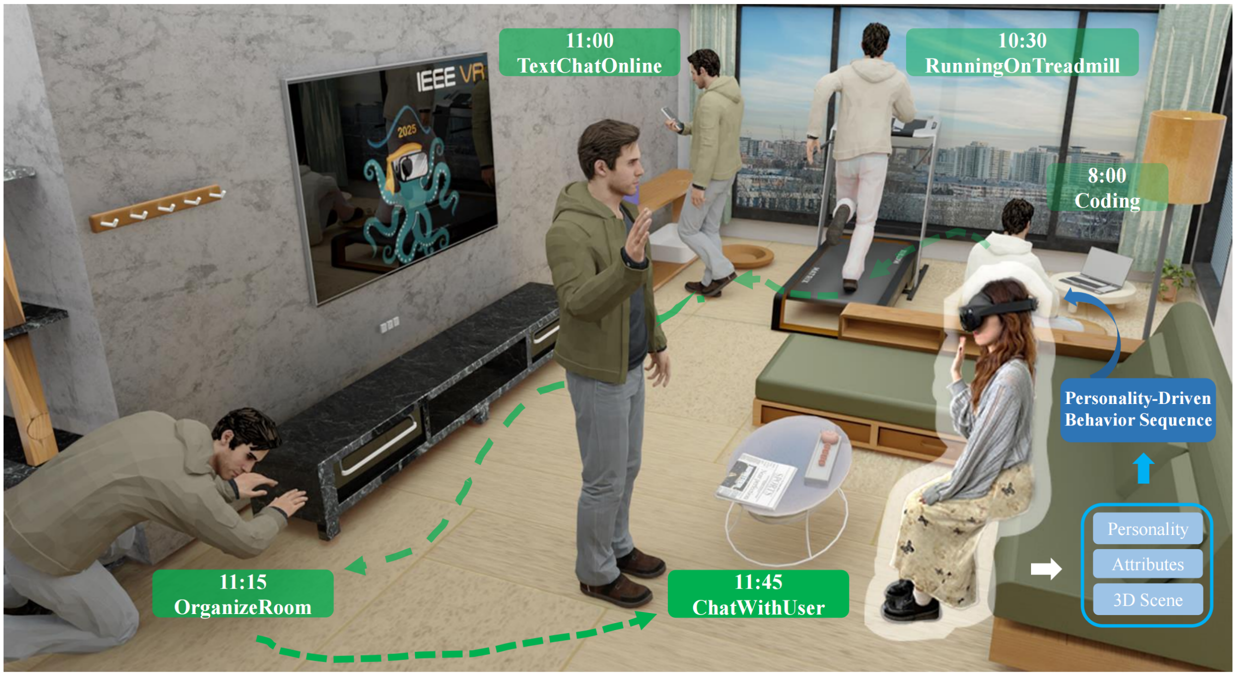


Fig. 1: **X's day**. Our approach generates a one-day activity sequence for the virtual human X by integrating personality traits, attributes, and 3D scene information. The colors representing X range from light to dark, indicating the progression of behaviors over time. Each green box corresponds to a completed activity. The "ChatWithUser" activity is currently active, during which X is greeting the user, who can observe and interact with X through the VR headset. Our approach enhances user engagement by providing scene-aware interactions that reflect distinct personality traits, making virtual humans more relatable and dynamic, with potential applications in training simulations, educational environments, and social VR experiences.

Abstract—Developing convincing and realistic virtual human behavior is essential for enhancing user experiences in virtual reality (VR) and augmented reality (AR) settings. This paper introduces a novel task focused on generating long-term behaviors for virtual agents, guided by specific personality traits and contextual elements within 3D environments. We present a comprehensive framework capable of autonomously producing daily activities autoregressively. By modeling the intricate connections between personality characteristics and observable activities, we establish a hierarchical structure of Needs, Task, and Activity levels. Integrating a Behavior Planner and a World State module allows for the dynamic sampling of behaviors using large language models (LLMs), ensuring that generated activities remain relevant and responsive to environmental changes. Extensive experiments validate the effectiveness and adaptability of our approach across diverse scenarios. This research makes a significant contribution to the field by establishing a new paradigm for personalized and context-aware interactions with virtual humans, ultimately enhancing user engagement in immersive applications. Our project website is at: https://behavior.agent-x.cn/.

Index Terms—Personality-driven Behavior, Behavior Generation, Contextual Scene

1 Introduction

Compelling and lifelike virtual human behavior is essential for creating immersive experiences in virtual reality (VR) and augmented reality (AR) [28, 54]. Furthermore, endowing these virtual humans with distinct personality traits and the ability to reflect various dynamic influencing factors within 3D scenarios—rather than relying on rigid behavior patterns can deepen user engagement, satisfaction, and trust by fostering more relatable and emotional interactions [13, 62]. This is particularly significant in applications such as virtual training [32, 73], educational simulations [57], and social VR [28], where personalized and engaging interactions can positively impact learning outcomes, skill acquisition, and user satisfaction.

Automatically generating behaviors and actions for virtual agents is long a research hotspot, attracting significant attention from researchers. Some researchers explore the generation of high-precision motions under various conditions, such as actions [36, 60], text [1, 15], audio [30, 34], and scene [39, 66]. Recently, some studies such as [7]

- [1] *Beijing Institute of Technology, Beijing, China.*
- [2] *Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing, China.*
- [†] *indicates corresponding authors.*
  E-mail: {lihaoyang, wangzan, liangwei, frankwyz}@bit.edu.cn

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

consider psychological personality traits but primarily focus on short-term interactions, realizing virtual human behaviors through predefined scenarios and behavior matching. Despite these efforts, some methods for generating virtual human behaviors and actions still have several limitations. For instance, they often focus on generating isolated actions or movements without considering the underlying personality traits, preferences, and habits of virtual humans [60, 74]. Moreover, they frequently rely on predefined scenarios [7] and scripts [26], or rules [9], limiting virtual humans' flexibility and adaptability in diverse and dynamic environments.

The recent adoption of LLMs generates significant interest in character personalization, particularly in 2D sandbox environments. The virtual sandbox environment provides a visualized and extensible platform for agent society, bridging the gap between simulation and reality [70]. 2D sandbox environments typically adopt an overhead perspective to simulate agents' interactions. Examples like AgentSims [38] and Generative Agents [44] achieve this by integrating comprehensive maps, agent avatars, and symbolic representations (e.g., emojis) to convey real-time positions, actions, and states. In these settings, static traits like background and psychological characteristics are typically conveyed through natural language and psychological indicators [25, 33, 44, 68]. While LLM-based frameworks in 2D sandboxes show promise, they are insufficient for VR/AR applications. These frameworks fail to address the complexities of 3D virtual environments. In 3D scenes, factors such as a virtual human's position, orientation, and the scene's layout can influence behavior [4, 45, 48], directly impacting user experiences. These critical aspects are overlooked in 2D frameworks, limiting their realism and interactivity in virtual reality applications.

To enrich user experiences in VR/AR applications by endowing virtual agents with distinct and vivid personality traits, we propose automatically generating virtual humans' behaviors based on desired personality attributes and contextual scenes. Achieving this goal involves addressing two significant challenges. First, modeling the relationship between personality and daily actions is complex, as personality is a multifaceted construct influenced by various factors. Similarly, translating these abstract traits into observable, context-dependent behaviors presents considerable difficulty. Second, virtual human behaviors must adapt to dynamic environments, where interactions between the virtual human and the scene can alter the behavior context.

To tackle these challenges, we introduce a novel framework to dynamically generate long-term behavior sequences for virtual humans in 3D environments, effectively reflecting their desired personality traits. We represent the behavior space as a hierarchical structure comprising three levels: Needs, Task, and Activity. This explicit representation facilitates the modeling of complex relationships between activities and personality traits, providing a clear framework for decision-making by moving from broader needs to specific activities. Our framework includes a Behavior Planner module and a World State module that iteratively generates virtual human behaviors with personality. Initially, we leverage the Large Language Model (LLM) to generate a sampling result of the behavior space over a specific personality trait according to the given personality traits and the 3D scenes; we employ GPT-4-turbo as the backbone of our LLM. Then, the Behavior Planner utilizes a Condition Reasoning method and a Behavior Sampling method to select activities hierarchically from top to bottom, while the World State module monitors the virtual environment's status, including historical activities, time, and scene information. As activities are generated, the probability distribution over the behavior space is updated in response to changes in the World State, ensuring that behaviors remain contextually relevant and adaptive to local circumstances, as well as showing the desired personality trait. We show a generated example in Fig. 1.

In summary, the contributions of this paper are three-fold:

- We introduce a novel task of generating long-term behaviors for virtual humans in 3D environments that effectively convey specific personality traits.
- We develop a framework that can automatically generate human daily activities in an autoregressive manner. This allows the generated behaviors to simultaneously reflect the desired personality traits and adapt to the dynamic 3D environment.
- We conduct extensive experiments to evaluate the efficacy of the proposed framework thoroughly. Additionally, we demonstrate the versatility of our approach by applying it across various new application scenarios.

2 Related Work

2.1 Representation of Human Daily Activities

Previous research proposes various representations to more accurately and effectively describe human daily activities. These representations include vectors, tables, trees, Markov models, and graphs, each capable of reflecting the diversity and complexity of human behaviors to varying degrees [11].

Vectors, which provide a simple sequential list, typically use verb phrases to describe linear and time-series data [55]. Tables, which employ a schedule template to represent daily behaviors, effectively manage large datasets of behavior events [22]. Tree structures, which are another commonly used representation, are particularly suited for representing hierarchical behavior relationships. Examples include behavior trees [14, 49] and decision trees [56]. Silverman et al. [53] develop realistic human behavior models in virtual environments using tree structures. Markov models [10], which are widely used statistical models for behavior modeling, capturing temporal dependencies in behavior sequences through state transitions [63]. Graph structures, which represent the complex relationships and dependencies between behaviors through nodes and edges, make them particularly suitable for representing complex parallel and interactive tasks. Graph structures that infer human daily activities generally including finite state machines [31], process mining techniques using graph structures [42], abstract activity sketches [36], and attribute graphs [2].

By integrating existing representation methods, hierarchical features effectively capture the complexity and structure of daily human activities. Specifically, across different hierarchical levels, abstract human behaviors with inclusive or progressive relationships can be represented, while within the same level, similar categories are depicted.

2.2 LLM-Based Generative Agent

LLM-based agents designed for simulating human-like behavior are digital entities capable of replicating human-like interactions and personalities. In recent years, LLM-based agent systems are primarily categorized into two main types: **Multi-Agent Systems** and **Single-Agent Systems**.

Multi-Agent Systems enhance LLM functionalities by specializing them into multiple agents with distinct capabilities. These agents interact with each other to simulate more complex real-world environments, thus providing advanced functionalities [42]. However, our research focuses on the Personality-Driven behavior generation and dynamic interaction of a single virtual human rather than the simulation of human groups within Multi-Agent Systems.

In Single-Agent Systems, the agent's capabilities are typically divided into three categories: Decision-making Thought, Tool-use, and Memory [19]. Lin et al. [37] propose a novel agent framework inspired by the dual-process theory of human cognition in LLM capability enhancement. In individual task implementations, Jin et al. [24] develop a human-like generative driving agent to simulate complex driving behaviors. Regan et al. [46] study the emotional state evolution of generative LLM agents when they perceive new events. Shao et al. [51] construct generative agents capable of assuming specific roles. There are also studies of generating virtual student profiles and learning behaviors [41, 71].

Existing studies demonstrate the powerful capabilities of LLMs with less emphasis on Personality-Driven behavior generation within 3D environments. Building on this foundation, we design a novel approach that fully leverages the capabilities of LLMs while incorporating intrinsic and extrinsic factors and dynamic changes to achieve complex, real-time behavior with personality decisions in 3D environments.

2.3 Personalization of Virtual Human

Virtual humans' personalization typically manifests in visual appearance and character attributes. Most current research focuses on the

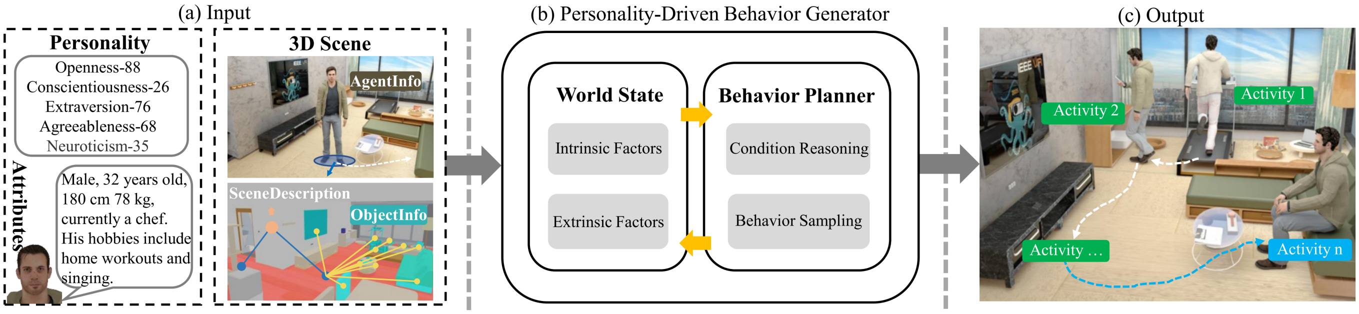


Fig. 2: **Overview of framework**. (a) The input comprises the virtual human's personality, attributes, and 3D scene information. We utilize the Big Five personality traits to represent the personality, while the attributes described by natural language encompass physiological traits, social characteristics, hobbies, and preferences. The 3D scene information includes scene layout (represented by orange and blue dots), object relations (indicated by yellow dots), and the spatial positioning of the virtual human and objects. (b) In the generator, the World State continuously monitors the environmental status, while the Behavior Planner generates activities autoregressively. (c) The output is an activity sequence, with the activity labeled in blue indicating the current activity, while the others represent completed activities.

personalization of virtual human appearance, covering various aspects such as facial features [65], stylization [6], animation [3, 40], and full-body models [8, 47]. Our research primarily focuses on behavioral personalization, therefore not involving appearance personalization.

Regarding character personalization, current studies typically focus on defining the "Profile" of the virtual agents, including their personality traits and relevant attributes information, to facilitate the evolution and development of interactions within social simulations. Initially, virtual characters' personalities are often defined using simple textual descriptions. For example, several studies [44, 68] use natural language to initialize each agent's basic information, attributes, and simple adjectives to describe their personality traits (e.g., friendly, kind). Subsequently, some research begins representing virtual character personalities based on psychological theories, utilizing human personality assessment tools such as the Big Five model [23, 25, 33] and the MBTI model [52]. These studies are primarily conducted in 2D sandbox environments, where multiple characters are defined randomly or simply to simulate the evolution and development of different personalities within a social group.

In related research within 3D environments, some research achieves personalization by body language [5, 12, 58, 61]. In addition, for the whole sequence of individual behavior, Cai et al. [7] use the Big Five personality indicators to introduce guiding instructions for personality modeling from a small sample of psychological tests. However, their research primarily focuses on short-term interactions and relies on predefined scenarios and settings. In contrast, our research enables the generation of real-time dynamic and long-term Personality-Driven behavior decisions within 3D environments, allowing virtual characters to better adapt to users' personality traits and immediate needs.

3 Overview

The framework of our approach is outlined in Fig. 2. It incorporates user-defined personality traits, attributes, and 3D Scene as inputs (Fig. 2 (a)) to develop a Personality-Driven Behavior Generator (Fig. 2 (b)), which automatically generates sequences of daily activities for virtual humans (Fig. 2 (c)). Following similar methodologies [25] and [33], we utilize the Big Five personality traits [16] to represent virtual human personalities, which encompasses five key dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. These dimensions delineate individuals' fundamental characteristics and predispositions across emotional, behavioral, and cognitive domains. The attributes of the virtual human are also essential, as these characteristics influence decision-making and behavioral motivations, enabling virtual humans to exhibit actions that align with users' expectations. Consistent with the definition of agent information in other studies [33, 44, 68], we allow users to use natural language to describe the virtual human's physiological attributes (such as gender, age, height, and weight), social attributes (such as occupation and educational background), hobbies, and preferences. These attributes can enhance the believability of the virtual human, making their actions more appropriate and realistic. The 3D Scene includes structural information such as the relationships between objects and rooms and the spatial data of the virtual human and objects.

The Personality-Driven Behavior Generator consists of two key modules: **World State** and **Behavior Planner**. The generator operates in an autoregressive manner, producing activities iteratively. Each activity is informed by the one generated previously, ensuring a coherent and contextually relevant progression of behavior over time.

The **World State** is designed to continuously monitor the status of the virtual environment, providing critical input to the Behavior Planner. It comprises intrinsic factors that remain static, such as personality traits and attributes, and extrinsic factors that change dynamically, including historical activities, time, and scene information. At each iteration, the World State updates its status based on the previous iteration, ensuring the behavior sequence remains responsive to external changes.

The **Behavior Planner** generates activities by leveraging information from the World State. To enhance the quality and coherence of these activities, we employ a large language model (LLM) and adopt the Chain of Thought (CoT) framework to enable more structured and reasoned decision-making. The planning process is divided into two stages: **Condition Reasoning** and **Behavior Sampling**. During the Condition Reasoning stage, the LLM assesses the current World State to infer contextual conditions. In the Behavior Sampling stage, activities are hierarchically selected, from high-level goals to more granular activities, based on the conditions derived in the previous step.

For clarity in this technical discussion, we illustrate our framework using behaviors within a home environment. However, the framework can generate diverse, personality-driven activities across various scenarios, as Section 7 demonstrates.

4 Hierarchical Representation of Human Behavior

To facilitate the generation process, we represent human behavior using a hierarchical structure. This is motivated by the inherent complexity of human behavior, which is influenced by various potential factors. By breaking this complexity down into multiple layers, the model provides a clear framework for decision-making.

We categorize behavior into three layers: Needs, Task, and Activity. The specific definitions of these three levels are outlined below and illustrated in Fig. 3.

Needs: The Needs Layer defines fundamental requirements that humans strive to fulfill. Needs, as outlined in various psychological theories such as Maslow's hierarchy [43], drive individuals to pursue specific goals and actions. Previous studies [53, 72] demonstrate the utility of Maslow's needs theory in determining the priority of behavioral motivations. Personality traits significantly influence how these needs are expressed and prioritized. For example, someone high in

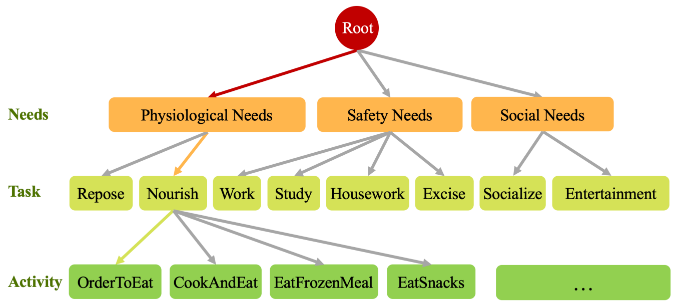


Fig. 3: **Hierarchical behavior representation**. This representation facilitates the generation process, progressing from broad goals to specific activities, as illustrated by the colorful arrow lines.

extraversion may prioritize social needs and seek relationships and community, while someone high in conscientiousness might focus more on achievement-related needs. Thus, we position needs as the top layer.

Due to the abstract nature and long-term impact of esteem and self-actualization needs, we adopt the approach of [72] to simplify home activities into three categories: **Physiological Needs**, **Safety Needs**, and **Social Needs**. These categories effectively capture the motivations behind daily human behaviors [27] and provide a basis for subsequent selection.

Task: The Task Layer further refines the Needs Layer, with each category of needs corresponding to multiple task types that address specific requirements. We consult three expert psychologists from research institutions specializing in Cognitive, Personality, and Behavioral Psychology. Utilizing semi-structured interviews and focus group discussions, we systematically categorize home behavior tasks into eight types: **Repose**, **Nourish**, **Work**, **Study**, **Housework**, **Exercise**, **Entertainment**, and **Socialize**. These task types align with specific needs, as Fig. 3 illustrates.

Activity: The Activity Layer provides the most detailed representation of the Task Layer, defining specific behavioral activities. Each task type encompasses multiple specific activity options, and the design of these activities is informed by collected daily behavior data. The information of the participants and the collection methods for the collection of daily behavior data are as follows:

We conduct a systematic survey to obtain a representative activity set that accurately reflects daily human behavior patterns. The objective is to gather comprehensive information on activities in daily home environments. A total of 50 participants are recruited, consisting of 25 females and 25 males, whose ages range from 12 to 70 years ($M = 36.42, SD = 18.64$). The participants represent a variety of occupations, including students, those in the service industry, workers, professionals, those in corporate management, those in government, and others (retirees, freelancers, and homemakers).

The data collection process includes both questionnaires and interviews, with the latter mainly targeting minors and elderly individuals who can not complete the questionnaires independently, ensuring the completeness and accuracy of the data. We gather basic demographic information and focus specifically on the specific activities participants may engage in under eight major task categories. Participants are guided to recall their typical daily behaviors at home and to detail all possible activities for each task category. After data collection, we organize and analyze the responses, extracting a comprehensive set of activities for each task category. The detailed data of participants' profiles and extracted activity data are presented in supplementary materials.

5 Personality-Driven Behavior Generator

In the Personality-Driven Behavior Generator, there are two critical modules: **World State** and **Behavior Planner**. World State involves the key factors influencing virtual human behavior to exhibit personalized preferences and temporal-spatial rationality. These factors are integrated into Conditions for Behavior Planner in the Condition Reasoning process (Fig. 4) and further used in Behavior Sampling (Fig. 5), enabling long-term dynamic Personality-Driven behavior generation for virtual humans.

5.1 World State

The World State module aims to manage the status of the virtual environment continuously, providing essential information to the Behavior Planner module. There are two types of factors in the World State that influence virtual human behavior: intrinsic factors and extrinsic factors. The detailed format on the usage of these factors is presented in supplementary materials.

Intrinsic factors are specified by users and include the personality and attributes of virtual humans. Recall that we utilize the Big Five personality traits to represent these personalities. In particular, it includes five dimensions: Extraversion, Openness, Conscientiousness, Agreeableness, and Neuroticism. Attributes encompass physiological characteristics (such as gender, age, height, and weight), social attributes (such as occupation and educational background), and hobbies and preferences.

Extrinsic factors, on the other hand, are influenced by both individual characteristics and spatiotemporal features, including context, time, and environment, as well as past activities [4, 18, 50, 64]. Based on these insights, we define time, completed activities, and 3D scene information as extrinsic factors. These elements enhance the timeliness and contextual appropriateness of virtual human behaviors, improving the adaptability of activities with personality across various scenarios.

In particular, time refers to the current timestamp experienced by the virtual human, which assists the language model (LLM) in assessing the urgency and suitability of tasks, influencing the selection of activities. Completed activities represent all the virtual humans completed up to now. These historical events function similarly to the memory module described in [44].

Regarding the 3D Scene, we focus on three main components: SceneDescription, AgentInfo, and ObjectInfo. SceneDescription describes the structural information of the 3D environment where the virtual agent is located. It includes the containment relationships between Rooms, Places, and Objects and the connectivity relationships between different Places. Similar to [17], we adopt Dynamic Scene Graphs (DSG) to represent 3D scenes, parsing DSG into natural language descriptions, thus providing contextual information about the environment. The implementation of DSG we make is presented in supplementary materials. AgentInfo indicates the current location of the virtual human, while ObjectInfo describes the shortest navigable paths and distances between all objects and the virtual human.

Upon completing a generated activity in each iteration, the factors in the World State will be updated accordingly. Specifically, for **time**, the duration of the activity will be added to the current timestamp. For the set of **Completed Activities**, the latest generated activity, along with its configuration, such as the time of the activity and the interacted objects, will be recorded. Regarding the **3D Scene**, the position of the virtual human will be updated in AgentInfo, while the interacted objects will be updated in ObjectInfo. Consequently, SceneDescription will also be modified to reflect these changes.

5.2 Behavior Planner

The Behavior Planner is the core module responsible for generating activities by leveraging information from the World State module. It primarily consists of two stages: **Condition Reasoning** and **Behavior Sampling**. They are used to infer contextual conditions and behavior-selecting processes. To illustrate the generation process, we use the first activity of a virtual person in the morning as an example for one planning iteration in Fig. 4 and Fig. 5.

The consideration of using two stages rather than relying solely on simple function evaluations for behavior selection is that we employ the idea of the Chain of Thought (CoT) [69] to enable more reasonable decision-making. In the Condition Reasoning phase, the LLM evaluates the current state of the world to deduce the relevant contextual

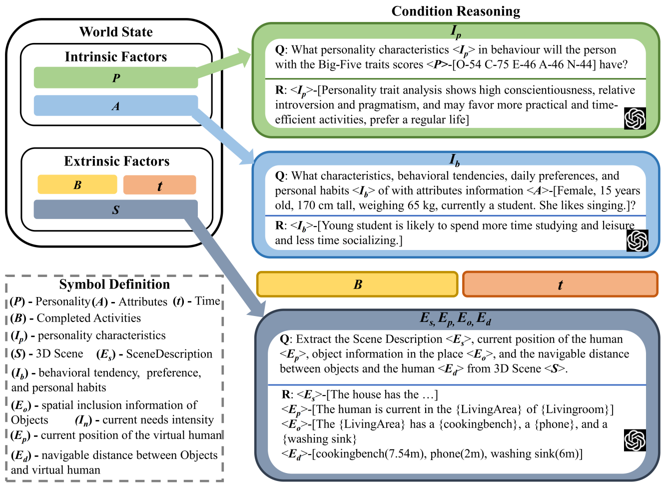


Fig. 4: **Example of Condition Reasoning process**. **Q** denotes the input provided to the large language model (LLM) at the current step, while **R** represents the response generated by the LLM. The dashed box labeled "Symbol Definition" explains the symbols' meanings.

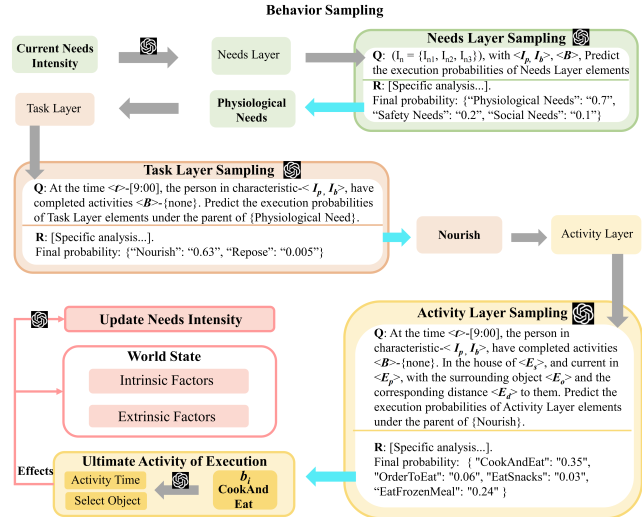


Fig. 5: **Behavior Sampling process**. We use the first activity of a virtual person in the morning as an example for one iteration. **Q** denotes the input provided to the LLM at the current step, while **R** represents the response generated by the LLM. The result after sampling in each layer is pointed out by the blue arrow. The symbols' meanings are in the dashed box labeled "Symbol Definition" in Fig. 4

conditions. Following this, in the Behavior Sampling phase, activities are selected hierarchically, progressing from broad goals to more specific activities, all guided by the conditions identified in the earlier stage. This design enhances the framework's ability to effectively adapt behavior to the context. For prompt engineering, our designed prompts are based on JSON templates and, in line with CoPB [50], adopt the CoT approach to enhance reasoning capabilities.

Condition Reasoning: To derive the necessary conditions from the information in the World State, the Behavior Planner employs the LLM to reason through several vital considerations. First, based on the personality traits, the LLM is queried to identify the personality characteristics of the virtual human according to specific scores across the five personality metrics. Regarding attribute information, the LLM generates insights into the behavioral tendencies, daily preferences, and personal habits typical of individuals with these attributes in a home environment. Time information and completed activities are incorporated as extended instructions. For the 3D Scene, the LLM extracts the scene description, the current position of the virtual human, all object information within the scene, and the shortest navigable distances between the objects and the virtual human. Fig. 4 illustrates an example of the Condition Reasoning process.

Behavior Sampling: Guided by the detailed conditions, we further request the LLM to predict selection probabilities to facilitate specific node selection in each layer. The element with the highest probability in each layer will be chosen. Fig. 5 demonstrates the detailed prompt and conditions. The processes for each layer are as follows:

For **Needs**, we adopt the first three levels of Maslow's hierarchy of needs, with each need's intensity dynamically changing over time. The research by Yuan et al. [72] captures the dynamic nature of needs, primarily manifest as two dynamic processes: "Spontaneous Flow" and "Instantaneous Jump." In our work, we incorporate dynamic processes and use LLM to predict needs changes and the subsequent need levels to be fulfilled with information. The LLM assesses the urgency of the three needs based on the current need intensity, categorizing them into selection probability.

For **Task** under corresponding needs, we use all conditions except for the 3D Scene to predict the execution probabilities of each task element.

For **Activity** under the corresponding task type, we use all conditions to predict the execution probabilities of Activity Layer elements. After selecting the activity with the highest probability, these predictive variables are also as combined inputs to predict the object most likely to be selected to complete this activity, along with an estimate of the required time.

6 Result and Experiment

Our work aims to enable virtual humans to dynamically generate long-term behaviors that align with their personality traits and attribute information within a given 3D scene based on their current state. To validate the effectiveness of this framework, we analyze the results generated by virtual humans with opposite personality traits in different scenes and experiment with a user study.

6.1 Behavioral Results with Notable Individual Differences in Different Scenarios

In this experiment, we concentrate on two key personality traits, "Openness" and "Conscientiousness," for the analysis. Our method is applied to two distinct 3D home environments, defining two contrasting personality types for each trait (e.g., high vs. low openness, high vs. low conscientiousness, high score is 100, low score is 0, and the other factors in World State are all the same), resulting in two groups of personality combinations. Subsequently, behavior sequences of virtual humans exhibiting these four distinct personalities are predicted in two scenes over an entire day.

As shown in Fig. 6, the experiment generates eight behavior sequences corresponding to different personality and scene combinations. We define the same Attributes of the four types of virtual human personalities as: "A computer science student with hobbies that include home exercise, music, and singing". From Fig. 6, which depicts the overall behavioral trajectories of Scene 1 and Scene 2, we observe that as the complexity of the scene increases and the number of interactive objects grows, the personality traits of virtual humans become more pronounced in their behaviors.

Virtual humans with High-Openness traits tend to explore a diverse range of activities without adhering to a fixed pattern in object selection. For example, in Scene 1, they engage in activities 3, 5, and 6 on different computers, demonstrating that the proximity of objects does not influence their choice of activities. In contrast, virtual humans with Low-Openness traits clearly prefer familiar activities and objects, repeatedly engaging in specific actions across both scenes. In Scene 1, for instance, they play PC games on the same computer during activities 2 and 4, while in Scene 2, they repeatedly use the same mobile device to listen to music for rest during activities 2, 4, and 7.

Virtual humans with High-Conscientiousness traits exhibit more structured and consistent behavior, often choosing activities that re-

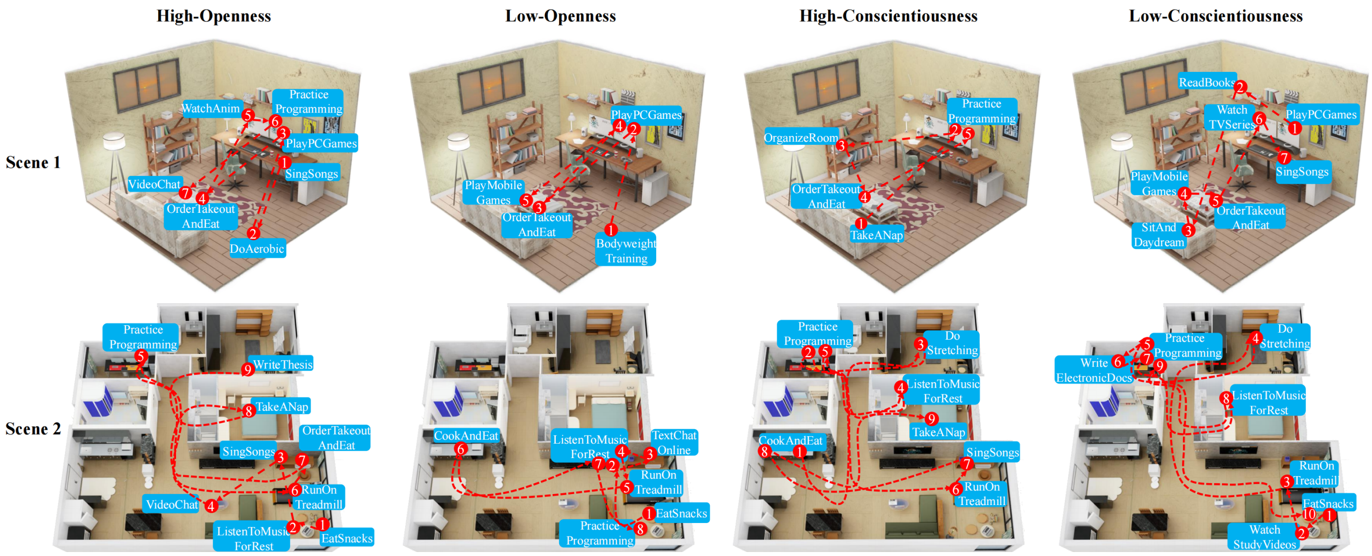


Fig. 6: **Generated behavior sequences**. Examples of generated behavior sequences illustrate four distinct personalities across two scenes. The activity sequences in Scene 1 represent afternoon and evening activities, while those in Scene 2 depict morning activities. The virtual human performs various activities (shown in blue boxes) sequentially (indicated by the numbers), interacting with different objects (represented by red circles). The dashed line with an arrow indicates the path to the following activity location. An activity may occur multiple times, resulting in multiple red circles with different numbers within a single blue box.

quire planning and effort, such as studying or exercising. The activity path information in the figure reveals that these individuals prefer objects closer to proximity when selecting tools for their tasks. In contrast, virtual humans with Low-Conscientiousness traits display less structured activity patterns, favoring rest and leisure activities. They are often easily distracted during tasks, as demonstrated in Scene 2, where the "Practice Programming" in activities 5, 7, and 9 is interrupted by other activities, such as activities 6 and 8.

The results reveal significant differences in behavior choices, activity sequences, and scene interactions with objects chosen among virtual humans with different personalities. These results indicate that our proposed framework can generate distinct long-term behaviors for virtual humans based on different personalities and scenarios. Moreover, the choice and order of behaviors, objects chosen, and types of activities vary significantly across personality traits, highlighting that personality characteristics profoundly influence the daily behavior patterns of virtual humans.

6.2 User Study

We conduct a user study through experiments to validate the effectiveness of our framework from the perspectives of general users. This study aims to explore the impact of factors in World State and the Needs that serve as the primary driving force for activity selection within our framework on the behavior generation process, as well as to evaluate the rationality and personality with preference alignment of the generated behaviors. Detailed visualization and data collection results are presented in supplementary materials.

According to the experimental requirements of our institute, we conduct relevant data collection and user study. Furthermore, all data are collected with prior informed consent, and all results are anonymized and presented with the explicit consent of the participants who contribute to this section and the collection of behavior data in Section 4.

6.2.1 Experiment Settings

Participants: For the general user group, we recruit 14 participants, including seven females and seven males, aged between 14 and 55 ($M = 30.50, SD = 10.53$), covering various occupational backgrounds and diverse educational and social backgrounds. All participants have no impairments in reading text or watching videos. Detailed information regarding the attributes of the participants is provided in the supplementary materials.

Compared methods: To evaluate the influence of each condition on virtual human behavior generation, we design the experiment with seven combinations of comparative methods. By systematically removing or retaining specific conditions, we analyze the critical role of these factors in behavior generation. The comparison approaches include: (1) **Ours**, without removal. (2) **Sandbox**, adopting the method similar to other existing agent simulation of 2D sandbox [33, 44] with Memory, Planning, and Reflection. Note that in the Sandbox Method, similar to these sandbox simulations, an initial daily plan is generated at the beginning, followed by subsequent decomposition and real-time planning updates as activities are completed. In Ours and the following methods (3)-(7), there is no initial full-day activity planning; instead, a new activity is generated in each generation round. (3) **W/O Time**, removal of time. (4) **W/O Event**, removal of completed activities. (5) **W/O Need**, removal of the needs of virtual human. (6) **W/O PA**, removal of Personality and Attributes. (7) **W/O 3D**, removal of 3D Scene.

Data: Initially, we gather Personality and Attributes data from 14 participants. Personality traits are quantified using the enhanced BFI-2 scale [59], which assesses the Big Five personality traits through five scores and requires approximately five minutes to complete. Attributes data encompass physiological characteristics (e.g., gender, age, height, weight), social attributes (e.g., occupation, educational background), and personal interests. Leveraging each participant's Personality and Attributes information, we employ seven comparative methods to generate seven distinct behavior sequences for virtual humans, simulating a full day of activity for each participant. This process results in 14 groups and 98 behavioral sequence outputs.

Procedure: Each participant is asked to evaluate seven behavior sequence results generated using the comparative methods based on their Personality and Attributes. The results are presented through videos, and the evaluation takes approximately 15 minutes. The order of the seven different methods is randomized within each group. After observing each result, participants complete a five-point Likert scale (1 = "strongly disagree," 5 = "strongly agree") based on the following two criteria:

1) **Behavior Rationality**: Does the virtual human's behavior exhibit temporal and spatial coherence and conform to typical human

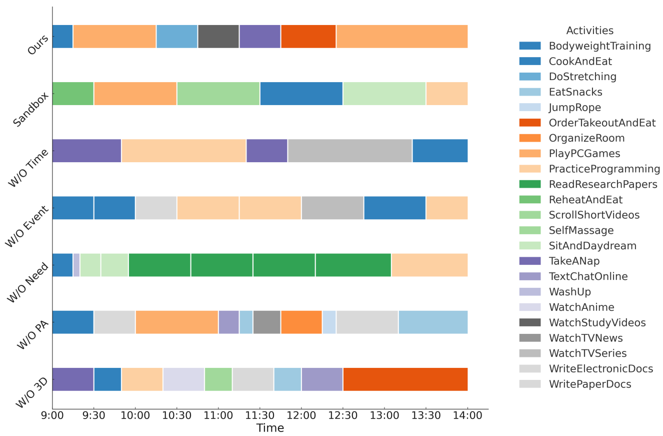


Fig. 7: **Visualization of behavior sequence**. Visualization of a participant's generated behavior sequence under seven conditions. The Big-Five personality scores are {O-69, C-38, E-75, A-58, N-17} and Attributes are "Male, 24 years old, 170 cm tall, weighing 90 kg, currently a student with a bachelor's degree. His hobbies include playing video games, and his unique habit at home is lying down.". Each sequence consists of a period of time, and each activity is shown by one color.

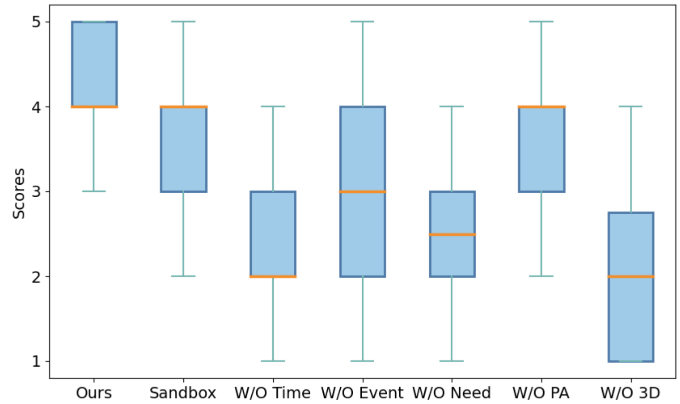


Fig. 8: **Box plots of Behavior Rationality**. The box plots of the participants' ratings on the behavior sequences' Behavior Rationality in 7 different conditions. The bottom and top edges of the box represent the 25th and 75th percentiles, respectively. The horizontal line, depicted in orange, represents the median rating. The whiskers extend to the most extreme data points. Hollow circles indicate outlier ratings.

behavioral patterns?

2) **Personality Alignment**: Based on your judgment, does the virtual human's behavior preference resemble yours?

Participants are not explicitly informed of the methods used to generate each result during the experiment. To manage the high volume of daily behavior instances, which ranges from approximately 30 to 50 entries per day, and to enhance readability, facilitate intuitive judgment, and reduce the time required for user scoring, each day is divided into three distinct periods: 9:00–14:00, 14:00–19:00, and post–19:00. The 14 participants are evenly distributed into three groups based on age and gender, with 5 participants assigned to period 1, 5 to period 2, and 4 to period 3. Each group maintains a balanced gender ratio of approximately 1:1 and includes at least one participant from each age category, ensuring an even distribution of age demographics across the three time periods. Each group evaluates the behavior sequences of participants within its designated time segment for a given day. Fig. 7 illustrates an example of a generated behavior sequence for a participant. Please refer to the supplementary materials for all other results.

6.2.2 Statistical Analysis

To evaluate the impact of different methods on the scores, we employ the Friedman test to analyze the statistical significance of user ratings. Additionally, to further clarify the differences between methods, we use the Wilcoxon Signed-Rank Test with Bonferroni correction for pairwise comparisons of the ratings.

User Rating Analysis for Behavior Rationality: Fig. 8 shows the visualization of the participants' ratings using box plots. The results of the Friedman test indicate significant differences in user ratings across seven compared methods of **Behavior Rationality** ($\chi^2 = 51.09, p < 0.05, \mathrm{d}f = 6$) at the $\alpha = 0.05$ significance level. A post-hoc test using Wilcoxon Signed-Ranks Test with Bonferroni correction (at the correlated significance level of $\alpha = 0.007$) reveals a significant difference that the mean rating of Ours (without any removal, $M = 4.36, SD = 0.63$) is statistically higher than that of the W/O Time (time removal, $M = 2.21, SD = 0.89$) ($W = 0, p < 0.007, r = 0$), W/O Event (completed activities removal, $M = 3.00, SD = 1.24$) ($W = 7.00, p < 0.007, r = 1.87$), W/O Need (needs-driven removal, $M = 2.36, SD = 0.92$) ($W = 2.00, p < 0.007, r = 0.53$), and W/O 3D (3D Scene removal, $M = 2.00, SD = 0.96$) ($W = 0, p < 0.007, r = 0$). Moreover, the post-hoc test does not find any significant difference between Ours' mean rating and Sandbox's and W/O PA's.

Further statistical analysis of the results reveals that, for the Behavior Rationality of the virtual human's behaviors as represented by the activity sequences, participants rate the activities' reasonableness highly when all components and variables are included in Ours.

Compared to the method analogous to the 2D sandbox approach, no significant difference is observed between the Sandbox method and our approach regarding perceived reasonableness. This is attributable to the Sandbox method's initial comprehensive planning of the entire day's activities, which establishes a robust temporal structure. Subsequent activities are then allocated by segmenting this initial plan into discrete time intervals, with dynamic adjustments to the overarching schedule as needed. Conversely, apart from the Sandbox method, Ours and other methods do not rely on exhaustive pre-planning. Instead, they predict activities in real-time before each action's commencement, showcasing our approach's high adaptability and dynamic nature. Despite the absence of a predefined daily plan, our framework maintains a high degree of behavioral rationality.

Excluding Personality and Attributes in W/O PA does not result in significant differences compared to our approach, suggesting that the generated sequences of human daily behavior retain a certain level of rationality even without these factors. This indicates that psychological personality traits, along with individual physiological, social, and interest-related attributes, have a limited impact on the rationality of the generated behavior.

In contrast, significant differences in perceived reasonableness emerge when temporal information (W/O Time), completion status (W/O Event), human needs (W/O Need), and 3D scene context (W/O 3D) are omitted compared to our entire framework. This finding suggests that these four factors are essential for ensuring the rationality of activities. For example, the appropriateness of an activity at a given time, the avoidance of repeating certain activities after completion, and the contextual relevance of objects used in specific scenarios are all significantly influenced by the inclusion or exclusion of these conditions.

User Rating Analysis for Personality Alignment: Fig. 9 shows the visualization of the participants' ratings using box plots. The results of the Friedman test indicate significant differences in user ratings across seven compared methods of **Personality Alignment** ($\chi^2 = 23.97, p < 0.05, \mathrm{d}f = 6$) at the $\alpha = 0.05$ significance level. A post-hoc test using Wilcoxon Signed-Ranks Test with Bonferroni correction (at the correlated significance level of $\alpha = 0.007$) reveals a significant difference that the mean rating of Ours (without any removal, $M = 3.64, SD = 0.93$) is statistically higher than that of the W/O Time (time removal, $M = 1.93, SD = 1.14$) ($W = 0, p < 0.007, r = 0$) and W/O PA (Personality and Attributes removal, $M = 2.14, SD = 0.94$) ($W = 0, p < 0.007, r = 0$). Moreover, the post-hoc test finds no significant difference between the other methods.

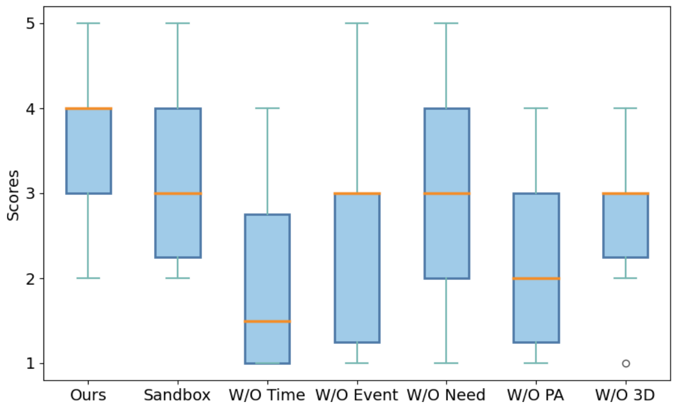


Fig. 9: **Box plots of Personality Alignment**. The box plots of the participants' ratings on the behavior sequences' Personality Alignment in 7 different conditions. The bottom and top edges of the box represent the 25th and 75th percentiles, respectively. The horizontal line, depicted in orange, represents the median rating. The whiskers extend to the most extreme data points. Hollow circles indicate outlier ratings.

Based on the aforementioned statistical results, significant differences are observed only when Time, Personality, and Attributes are excluded, compared to our complete framework. At the same time, other conditions do not exhibit substantial variations. This suggests that the factors of Personality, Attributes, and Time have a pronounced personalized impact on overall behavioral sequences. In contrast, other variables exert relatively minimal influence on the comprehensive behavioral pattern.

We meticulously label all object names to further investigate personalization aspects between our complete framework **Ours** and the sandbox-like method **Sandbox**. Under both methods, we provide detailed path information of the virtual agents' activity sequences within each time segment. Additionally, we highlight the available object options generated at the onset of each new activity and the final choices made by the virtual agents.

Our analysis reveals that virtual agents consistently select the same objects under the Sandbox approach when performing identical activities. In contrast, under our complete framework, object selection varies following the virtual agents' personality traits, as influenced by the 3D Scene context. For instance, highly conscientious and neurotic agents tend to choose the nearest and least effortful objects. In contrast, those with high openness are inclined towards selecting objects that are more distant but perceived as more attractive or novel.

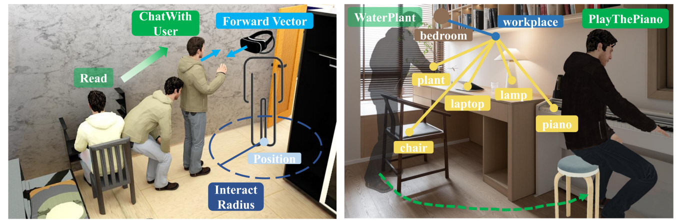
Participants are then asked to evaluate which object preferences better align with their characteristics across the two conditions. The results indicate that 11 participants (78.6%) perceive the preferences generated by Ours as more closely matching their inclinations. This finding suggests that our approach captures general preference patterns and demonstrates personalized object selection, thereby enhancing personalization in virtual agent behavior.

6.2.3 User Feedback

In this study, we collect and analyze user feedback on the generated virtual human behavior sequences, tailored according to their personality and attributes, yielding several key insights:

For groups with higher rationality scores, participants generally perceive the activities as well-aligned with typical daily routines, exhibiting appropriate timing, location, and diversity. Conversely, lower rationality scores are attributed to misaligned activity timing, inappropriate venue or object selection, and sequences that deviate from normative human behavior. A lack of activity diversity or repetitive execution of essential activities like eating also contributes to lower ratings.

For groups with higher personality alignment scores, participants report that the activity choices and sequences closely mirror their behaviors, with completion times reflecting their habits, thereby enhancing alignment with individual preferences. In contrast, lower personality alignment scores are often linked to unique personal habits or unconventional preferences that are not strictly personality-driven. Examples include skipping dinner due to dieting, avoiding TV dramas, or having irregular meal times.



(a) Chat with User in VR (b) Noon Routine of Real Scene in AR

Fig. 10: **Applicaitons in VR (a) and AR (b) scenario.**

7 Applications

To demonstrate the practical applicability and scalability of our Personality-Driven Behavior Generation Framework for virtual humans, we expand its implementation to VR and AR environments and apply it across diverse application domains, including healthcare, education, and games.

7.1 Application in VR and AR Scenarios

Our approach effectively extends to VR and AR environments, enabling virtual humans to perform in realistic interactive scenarios. Virtual humans can engage with objects in the virtual environment and directly interact with users.

As depicted in Fig. 10 (a), the virtual human's current activity shifts from "Read" to "ChatWithUser." The user's position and the forward vector are obtained in real-time via VR devices and integrated into the 3D Scene. When the virtual human enters the user's interaction radius, two potential actions are generated based on its personality traits: interacting with the user or continuing its current activity. We define three specific activities—"Chat with the user," "Accompany the user," and "Provide information and displays," organized under the "**Socialize**" Task Layer. During behavior execution, the virtual human dynamically adjusts interactions based on real-time user data, ensuring responsive and context-aware engagement.

Fig. 10 (b) illustrates virtual human applications in AR environments, where real-world scenes are scanned and reconstructed into 3D scenes represented by the DSG structure. Here, the virtual human, guided by real-world context, can exhibit personalized behaviors, as represented by the three colored circles denoting Room, Place, and Object Nodes in the scene.

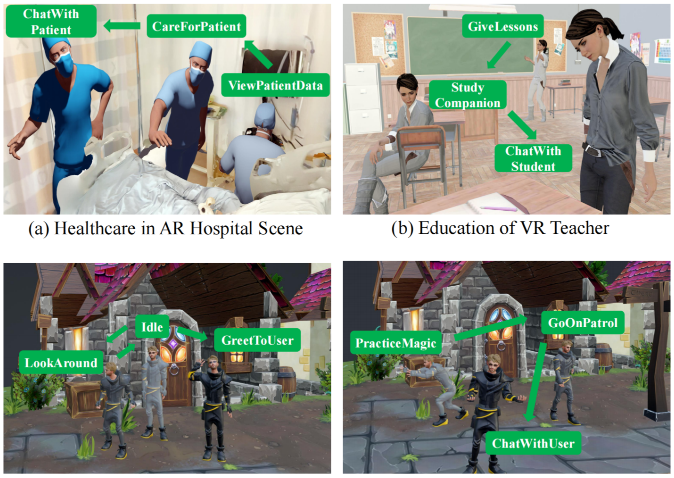
7.2 Applications of Healthcare, Education, and Games

To further illustrate the diversity of our approach, we extend the virtual human behavior generation framework to public settings such as healthcare, education, and games by setting predefined personality traits and attributes for different roles.

In **Healthcare** scenarios, virtual humans can function as healthcare providers, generating activities showing traits such as caregiving or consultation based on role-specific requirements. Fig. 11 (a) displays sequences of virtual humans acting as a doctor who "enjoys conversing with patients and shows significant concern for their conditions" in the natural hospital setting in an AR environment, demonstrating the ability to adapt flexibly to dynamic changes.

In **Education** scenarios, virtual humans can take on the roles of teachers, dynamically selecting behaviors such as teaching, discussing, or guiding based on their personalities and contextual settings. Fig. 11 (b) shows demonstrations of virtual human acting as "A teacher who enjoys discussing with students and has an outgoing personality", adapting interaction styles according to the classroom atmosphere and the assigned personality traits.

Our approach also applies in **Gaming** environments, where it generates Personality-Driven behaviors for non-player characters (NPCs).



(a) Healthcare in AR Hospital Scene (b) Education of VR Teacher

(c) Games for NPCs with different Personality and Attributes

Fig. 11: **Applications of Healthcare (a), Education (b), and Games (c).** The characters' colors, from light to deep, indicate the order of behaviors.

Fig. 11 (c) depicts a virtual character functioning as an NPC in games. As shown in Fig. 11 (c), NPCs with different personalities can generate behavior consistent with their respective traits (combining the Personality and Attributes features, the character in the bottom-left image is defined as a "silent and cautious guard," while the character in the bottom-right image is described as a "lively and communicative mage."). This implementation infuses NPCs with greater personalization and interactivity, transforming them from simple pre-set roles into intelligent entities with dynamic behaviors.

These extended applications demonstrate our proposed Personality-Driven behavior generation method's broad applicability for virtual humans across multiple domains, including VR and AR applications in education, healthcare, and gaming. Virtual characters based on our method can adapt to diverse environments and generate contextually appropriate behaviors consistent with predefined Personalities and Attributes, providing innovative solutions for future virtual interactions, personalized services, and immersive experiences.

8 Discussions

We discuss the limitations of our Personality-Driven behavior generation method and propose directions for future enhancements. By critically examining the current constraints and identifying opportunities for improvement, we aim to advance the realism, adaptability, and user engagement of virtual human interactions in immersive environments.

8.1 Limitations

While our framework demonstrates promising results in generating long-term, personality-driven behaviors for virtual humans, several limitations warrant further exploration.

Firstly, the framework lacks advanced low-level motion generation techniques, constraining its applicability to highly interactive environments. The current approach primarily focuses on behavior-level generation, producing action sequences based on predefined labels rather than ensuring realistic motions or strict adherence to spatial constraints within the scene. We rely on corresponding animations to represent activities to facilitate clear and intuitive visualization of behavior sequences. However, users may perceive the avatar's motions as lacking fluidity or naturalness, potentially reducing immersion. The reliance on predefined gestures further limits adaptability in scenarios requiring varied and complex gestures, thereby affecting the believability and expressiveness of virtual humans in dynamic settings. Enhancing the framework with sophisticated motion generation techniques is critical to addressing this limitation and improving the naturalism of virtual human behaviors.

Secondly, although our framework supports continuous long-term generation through iterative updates, it faces challenges stemming from the lack of persistent memory mechanisms in LLMs [20]. Over extended periods, the accumulation of generated data may lead to progressive homogenization, undermining the diversity and spontaneity of virtual human behaviors. This poses a challenge to maintaining authentic and engaging interactions over time.

Additionally, some activities generated by the framework may appear peculiar or inconsistent with user expectations. This issue arises from the black-box nature of LLMs, which limits precise control over inference outputs. Consequently, behaviors occasionally deviate from expected norms, potentially disrupting user immersion and the believability of virtual agents. Future work should incorporate additional constraints or validation mechanisms to reduce such anomalies.

Finally, the framework depends on predefined personality attributes and fixed mappings of psychological traits, which may not fully capture the nuanced variations in human behavior across diverse contexts. This rigidity will limit the personalization and contextual relevance of generated behaviors. Integrating more flexible and data-driven methodologies is essential to enhance the framework's ability to model complex, context-specific personality-driven behaviors accurately.

8.2 Future Work

Several promising avenues remain to enhance the proposed method's realism, adaptability, and applicability across diverse contexts.

Enhancing the realism of motions and their integration within 3D scenes is a critical direction. Future research can explore advanced low-level motion generation techniques and tighter 3D environment integration, such as combining approaches like [21, 29, 67]. These advancements will enable more fluid and lifelike activities, thereby enhancing the immersive experience in VR/AR settings.

Developing adaptive, data-driven personality models presents another key opportunity. Future studies can collect data from broader populations and varied sources to ensure that behavior generation is diverse and representative. Incorporating approaches that infer personality traits from diverse datasets will allow the framework to capture a broader range of behavioral nuances, resulting in more personalized and contextually relevant interactions. Exploring memory mechanisms or diversity-promoting algorithms is recommended to sustain behavioral diversity during continuous generation.

Lastly, conducting real-world deployments and improving environmental perception such as [35] is vital. Integrating the framework into existing VR/AR applications and performing empirical validations through user research will provide valuable feedback for refinement. Additionally, incorporating advanced sensing technologies and spatial mapping will enable virtual agents to interpret and respond to environmental changes more effectively, resulting in more contextually appropriate behaviors.

9 Conclusion

This paper introduces a comprehensive framework for generating Personality-Driven behaviors in virtual humans, utilizing advanced long-term and dynamic decision-making techniques within 3D environments. By incorporating a hierarchical behavior structure, a Behavior Planner, and a World State module, our approach proficiently models long-term, context-aware behaviors that align with individual personality profiles while maintaining temporal and spatial coherence. The demonstration across various VR/AR applications in diverse 3D home environments and extended use cases such as healthcare, education, and gaming demonstrates the robustness and adaptability of our method. These results underscore the potential of personality-driven virtual agents to facilitate authentic and engaging interactions, delivering significant value across applications ranging from virtual companions to dynamic NPCs.

Acknowledgments

This work is supported by the Beijing Natural Science Foundation (L242094).

References

[1] H. Ahn, T. Ha, Y. Choi, H. Yoo, and S. Oh. Text2action: Generative adversarial synthesis from language to action. In *International Conference on Robotics and Automation (ICRA)*, 2018. 1

[2] P. Augustyniak and G. Ślusarczyk. Graph-based representation of behavior in detection and prediction of daily living activities. *Computers in Biology and Medicine*, 95:261–270, 2018. 2

[3] Z. Bai, P. Chen, X. Peng, L. Liu, N. Yao, and H. Chen. Bring your own character: A holistic solution for automatic facial animation generation of customized characters. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2024. 3

[4] E. D. Beck and J. J. Jackson. Personalized prediction of behaviors and experiences: An idiographic person–situation test. *Psychological Science*, 33(10):1767–1782, 2022. 2, 4

[5] U. Bhattacharya, N. Rewkowski, A. Banerjee, P. Guhan, A. Bera, and D. Manocha. Text2gestures: A transformer-based network for generating emotive body gestures for virtual agents. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2021. 3

[6] H. Bülthoff. Stylization of virtual humans. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2018. 3

[7] Z. Cai, J. Jiang, Z. Qing, X. Guo, M. Zhang, Z. Lin, H. Mei, C. Wei, R. Wang, W. Yin, et al. Digital life project: Autonomous 3d characters with social intelligence. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1, 2, 3

[8] H. Cha, B. Kim, and H. Joo. Pegasus: Personalized generative 3d avatars with composable attributes. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[9] C.-C. Chiu and S. Marsella. A style controller for generating virtual human behaviors. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011. 2

[10] K. L. Chung. Markov chains. *Springer-Verlag, New York*, 1967. 2

[11] R. Dunne, O. Matthews, J. Vega, S. Harper, and T. Morris. Computational methods for predicting human behaviour in smart environments. *Journal of Ambient Intelligence and Smart Environments*, 15(2):179–205, 2023. 2

[12] F. Durupinar, M. Kapadia, S. Deutsch, M. Neff, and N. I. Badler. Perform: Perceptual approach for adding ocean personality to human motion using laban movement analysis. *ACM Transactions on Graphics (TOG)*, 36(1):1–16, 2016. 3

[13] M. L. Fiedler, E. Wolf, N. Döllinger, M. Botsch, M. E. Latoschik, and C. Wienrich. Embodiment and personalization for self-identification with virtual humans. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2023. 1

[14] Y. Francillette, B. Bouchard, K. Bouchard, and S. Gaboury. Modeling, learning, and simulating human activities of daily living with behavior trees. *Knowledge and Information Systems*, 62(10):3881–3910, 2020. 2

[15] A. Ghosh, N. Cheema, C. Oguz, C. Theobalt, and P. Slusallek. Synthesis of compositional animations from textual descriptions. In *International Conference on Computer Vision (ICCV)*, 2021. 1

[16] L. R. Goldberg. The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1):26, 1992. 3

[17] N. Gorlo, L. Schmid, and L. Carlone. Long-term human trajectory prediction using 3d dynamic scene graphs. *IEEE Robotics and Automation Letters (RA-L)*, 9(12):10978–10985, 2024. 4

[18] M. A. Graule and V. Isler. Gg-llm: Geometrically grounding large language models for zero-shot human activity forecasting in human-aware task planning. In *International Conference on Robotics and Automation (ICRA)*, 2024. 4

[19] T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. Chawla, O. Wiest, and X. Zhang. Large language model based multi-agents: A survey of progress and challenges. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2024. 2

[20] K. Hatalis, D. Christou, J. Myers, S. Jones, K. Lambert, A. Amos-Binks, Z. Dannenhauer, and D. Dannenhauer. Memory matters: The need to improve long-term memory in llm-agents. In *Proceedings of the AAAI Symposium Series*, 2023. 9

[21] S. Huang, Z. Wang, P. Li, B. Jia, T. Liu, Y. Zhu, W. Liang, and S.-C. Zhu. Diffusion-based generation, optimization, and planning in 3d scenes. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 9

[22] I. Idrees, S. Singh, K. Xu, and D. F. Glas. A framework for realistic simulation of daily human activity. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2023. 2

[23] H. Jiang, X. Zhang, X. Cao, C. Breazeal, D. Roy, and J. Kabbara. Personallm: Investigating the ability of large language models to express personality traits. In *Findings of the Association for Computational Linguistics: NAACL*, 2024. 3

[24] Y. Jin, R. Yang, Z. Yi, X. Shen, H. Peng, X. Liu, J. Qin, J. Li, J. Xie, P. Gao, et al. Surrealdriver: Designing llm-powered generative driver agent framework based on human drivers' driving-thinking data. In *International Conference on Intelligent Robots and Systems (IROS)*, 2024. 2

[25] S. Jinxin, Z. Jiabao, W. Yilei, W. Xingjiao, L. Jiawen, and H. Liang. Cgmi: Configurable general multi-agent interaction framework. *arXiv preprint arXiv:2308.12503*, 2023. 2, 3

[26] H. Kim, G. Ali, S. Kim, G. J. Kim, and J.-I. Hwang. Auto-generating virtual human behavior by understanding user contexts. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021. 2

[27] J.-S. Kim, H. Jin, H. Kavak, O. C. Rouly, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Location-based social network data generation based on patterns of life. In *IEEE International Conference on Mobile Data Management (MDM)*, 2020. 4

[28] C. Kyrlitsias and D. Michael-Grigoriou. Social interaction with agents and avatars in immersive virtual environments: A survey. *Frontiers in Virtual Reality*, 2:786665, 2022. 1

[29] Y. Lang, W. Liang, and L.-F. Yu. Virtual agent positioning driven by scene semantics in mixed reality. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019. 9

[30] H.-Y. Lee, X. Yang, M.-Y. Liu, T.-C. Wang, Y.-D. Lu, M.-H. Yang, and J. Kautz. Dancing to music. *Advances in Neural Information Processing Systems (NIPS)*, 32, 2019. 1

[31] C. Li, W. K. Cheung, and J. Liu. Elderly mobility and daily routine analysis based on behavior-aware flow graph modeling. In *International Conference on Healthcare Informatics*, 2015. 2

[32] C. Li, W. Liang, C. Quigley, Y. Zhao, and L.-F. Yu. Earthquake safety training through virtual drills. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 23(4):1275–1284, 2017. 1

[33] J. Li, J. Li, J. Chen, Y. Li, S. Wang, H. Zhou, M. Ye, and Y. Su. Evolving agents: Interactive simulation of dynamic and diverse human personalities. *arXiv preprint arXiv:2404.02718*, 2024. 2, 3, 6

[34] R. Li, S. Yang, D. A. Ross, and A. Kanazawa. Ai choreographer: Music conditioned 3d dance generation with aist++. In *International Conference on Computer Vision (ICCV)*, 2021. 1

[35] W. Liang, X. Yu, R. Alghofaili, Y. Lang, and L.-F. Yu. Scene-aware behavior synthesis for virtual pets in mixed reality. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2021. 9

[36] Y.-H. Liao, X. Puig, M. Boben, A. Torralba, and S. Fidler. Synthesizing environment-aware activities via activity sketches. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2

[37] B. Y. Lin, Y. Fu, K. Yang, F. Brahman, S. Huang, C. Bhagavatula, P. Ammanabrolu, Y. Choi, and X. Ren. Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks. *Advances in Neural Information Processing Systems (NIPS)*, 36, 2024. 2

[38] J. Lin, H. Zhao, A. Zhang, Y. Wu, H. Ping, and Q. Chen. Agentsims: An open-source sandbox for large language model evaluation. *arXiv preprint arXiv:2308.04026*, 2023. 2

[39] G. Liu, M. Xu, Z. Pan, and A. E. Rhalibi. Human motion generation with multifactor models. *Computer Animation and Virtual Worlds*, 22(4):351–359, 2011. 1

[40] Y. Liu, Z. Li, Y. Liu, and H. Wang. Texvocab: Texture vocabulary-conditioned human avatars. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[41] X. Lu and X. Wang. Generative students: Using llm-simulated student profiles to support question item evaluation. In *Proceedings of the ACM Conference on Learning@ Scale*, 2024. 2

[42] M. R. Ma'arif. Revealing daily human activity pattern using process mining approach. In *International Conference on Electrical Engineering, Computer Science and Informatics*, 2017. 2

[43] A. Maslow. A theory of human motivation. *Psychological Review Google Schola*, 2:21–28, 1943. 3

[44] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *ACM Symposium on User Interface Software and Technology (UIST)*, 2023. 2, 3, 4, 6

[45] H. Pick. *Spatial orientation: Theory, research, and application*. Springer Science & Business Media, 2012. 2

[46] C. Regan, N. Iwahashi, S. Tanaka, and M. Oka. Can generative agents

predict emotion? *arXiv preprint arXiv:2402.04232*, 2024. 2

[47] A. Salagean, E. Crellin, M. Parsons, D. Cosker, and D. Stanton Fraser. Meeting your virtual twin: Effects of photorealism and personalization on embodiment, self-identification and perception of self-avatars in virtual reality. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2023. 3

[48] T. Sanocki, K. Michelet, E. Sellers, and J. Reynolds. Representations of scene layout can consist of independent, functional pieces. *Perception & Psychophysics*, 68:415–427, 2006. 2

[49] L. Scherf, A. Schmidt, S. Pal, and D. Koert. Interactively learning behavior trees from imperfect human demonstrations. *Frontiers in Robotics and AI*, 10:1152595, 2023. 2

[50] C. Shao, F. Xu, B. Fan, J. Ding, Y. Yuan, M. Wang, and Y. Li. Chain-of-planned-behaviour workflow elicits few-shot mobility generation in llms. *arXiv preprint arXiv:2402.09836*, 2024. 4, 5

[51] Y. Shao, L. Li, J. Dai, and X. Qiu. Character-llm: A trainable agent for role-playing. In *Annual Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023. 2

[52] Z. Shu, X. Sun, and H. Cheng. When llm meets hypergraph: A sociological analysis on personality via online social networks. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, 2024. 3

[53] B. G. Silverman. More realistic human behavior models for agents in virtual worlds: Emotion, stress, and value ontologies. *Philadelphia: U of Penn/ACASA Technical Report*, 2001. 2, 3

[54] E. M. Sims. Reusable, lifelike virtual humans for mentoring and role-playing. *Computers & Education*, 49(1):75–92, 2007. 1

[55] M. Singh, R. Richie, and S. Bhatia. Representing and predicting everyday behavior. *Computational Brain & Behavior*, 5(1):1–21, 2022. 2

[56] S. Sitanskiy, L. Sebastia, and E. Onaindia. Learning and identifying human behaviour using decision trees. In *IEEE International Conference on Web Intelligence and Intelligent Agent Technology*, 2023. 2

[57] M. Soliman, A. Pesyridis, D. Dalaymani-Zad, M. Gronfula, and M. Kourmpetis. The application of virtual reality in engineering education. *Applied Sciences*, 11(6):2879, 2021. 1

[58] S. Sonlu, U. Güdükbay, and F. Durupinar. A conversational agent framework with multi-modal personality expression. *ACM Transactions on Graphics (TOG)*, 40(1):1–16, 2021. 3

[59] C. J. Soto and O. P. John. 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, 2017. 6

[60] Z. Su, Q. Fan, X. Chen, O. Van Kaick, H. Huang, and R. Hu. Scene-aware activity program generation with language guidance. *ACM Transactions on Graphics (TOG)*, 42(6), 2023. 1, 2

[61] S. Thomas, Y. Ferstl, R. McDonnell, and C. Ennis. Investigating how speech and animation realism influence the perceived personality of virtual characters and agents. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022. 3

[62] R. S. Verhagen, M. A. Neerincx, C. Parlar, M. Vogel, and M. L. Tielman. Personalized agent explanations for human-agent teamwork: Adapting explanations to user trust, workload, and performance. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2023. 1

[63] K. Viard, M. P. Fanti, G. Faraut, and J.-J. Lesage. An event-based approach for discovering activities of daily living by hidden markov models. In *International Conference on Ubiquitous Computing and Communications and International Symposium on Cyberspace and Security*, 2016. 2

[64] S. T. Völkel, R. Schödel, D. Buschek, C. Stachl, Q. Au, B. Bischl, M. Bühner, and H. Hussmann. Opportunities and challenges of utilizing personality traits for personalization in hci. *Personalized Human-Computer Interaction*, 31, 2019. 4

[65] L. Wang, X. Zhao, J. Sun, Y. Zhang, H. Zhang, T. Yu, and Y. Liu. Styleavatar: Real-time photo-realistic portrait avatar from a single video. In *ACM SIGGRAPH Conference Proceedings*, 2023. 3

[66] Z. Wang, Y. Chen, B. Jia, P. Li, J. Zhang, J. Zhang, T. Liu, Y. Zhu, W. Liang, and S. Huang. Move as you say interact as you can: Language-guided human motion generation with scene affordance. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1

[67] Z. Wang, Y. Chen, T. Liu, Y. Zhu, W. Liang, and S. Huang. Humanise: Language-conditioned human motion generation in 3d scenes. *Advances in Neural Information Processing Systems (NIPS)*, 35:14959–14971, 2022. 9

[68] Z. Wang, Y. Y. Chiu, and Y. C. Chiu. Humanoid agents: Platform for simulating human-like generative agents. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations*, 2023. 2, 3

[69] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems (NIPS)*, 35:24824–24837, 2022. 4

[70] Z. Xi, W. Chen, X. Guo, W. He, Y. Ding, B. Hong, M. Zhang, J. Wang, S. Jin, E. Zhou, et al. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*, 2023. 2

[71] S. Xu, X. Zhang, and L. Qin. Eduagent: Generative student agents in learning. *arXiv preprint arXiv:2404.07963*, 2024. 2

[72] Y. Yuan, H. Wang, J. Ding, D. Jin, and Y. Li. Learning to simulate daily activities via modeling dynamic human needs. In *Proceedings of the ACM Web Conference (WWW)*, 2023. 3, 4, 5

[73] Y. Zhang and S.-B. Tsai. Application of adaptive virtual reality with ai-enabled techniques in modern sports training. *Mobile Information Systems*, 2021(1):6067678, 2021. 1

[74] Y. Zheng, Y. Hicks, D. Marshall, and D. Cosker. Real-time generation of interactive virtual human behaviours. In *International Conference on Computer Vision and Computer Graphics Theory and Applications*, 2009. 2