GENERATIVE EMBODIED MULTIPLE BEHAVIOR CONTROL SYSTEMS FOR HUMAN-LIKE AGENTS

Anonymous authors
Paper under double-blind review

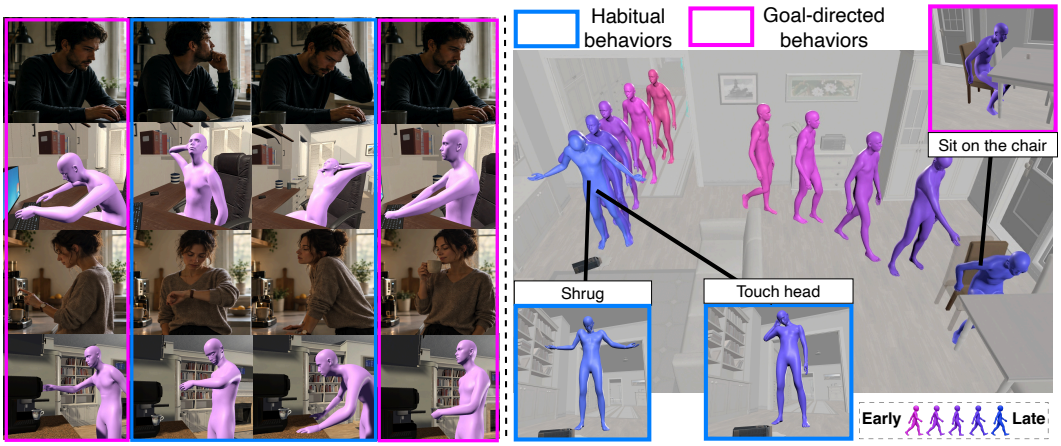


Figure 1: **Believable human-like behaviors with GEMS.** *(Left.)* Comparison between real human daily behaviors and GEMS-generated behaviors. Beyond goal-directed behaviors, GEMS generates habitual behaviors commonly observed in everyday human life but often overlooked. *(Right.)* Examples of motion sequences generated by GEMS. Our method generates diverse habitual and goal-directed behaviors and realizes them as coherent human–scene interactions in 3D environments.

ABSTRACT

Building autonomous agents that reproduce human behavior in realistic 3D environments has been a longstanding objective in AI. Cognitive neuroscience commonly believes that human behaviors are controlled by multiple control systems including goal-directed and habitual. However, existing human-like agent frameworks primarily focus on goal-directed system and habitual system is largely overlooked. In this paper, we address this gap by proposing **GEMS**, a **G**enerative **E**mbodied **M**ultiple Behavior Control **S**ystem that jointly models goal-directed and habitual behaviors for generating believable human-like behavior. GEMS employs a cognitive module to generate natural language behaviors and a motion generation module to construct them in 3D environment. The cognitive module comprises a Habitual Controller that retrieves habitual actions from habit memory in response to environmental stimuli, a Goal-directed Controller that proposes goal-directed actions for current goal, and an Arbiter that governs their influence and selects the final action, supporting diverse and naturalistic behavior patterns generation. To construct these behaviors, the motion generation module adopts a Text-to-Motion model to generate versatile and expressive 3D human–scene interactions without relying on predefined actions. Extensive evaluations, human and ablation studies demonstrate that human-likeness performance is substantially improved by GEMS. The efficacy of GEMS indicates the benefits of leveraging habitual behavior and multiple behavior control system coordination for believable embodied human-like agents design. The code is available at this link.

1 INTRODUCTION

Building human-like autonomous agents (Sloman, 1999; 2000) that can reproduce human behavior in realistic 3D environments has been a long-standing pursuit since the inception of artificial intelli-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

gence (AI) (Fetzer, 1990). The study could empower non-player game characters (Li et al., 2025), underpin human-robot interaction and cooperation, populate virtual reality communities (Zhou et al., 2026), and accelerate Embodied AI (Feng et al., 2026). With recent advances in large language models (LLMs), building embodied agents that can simulate believable (Bates et al., 1994) and coherent human-like behavior in daily life has attracted increasing attention. Existing frameworks (Park et al., 2023; Wang et al., 2025; Leng et al., 2026; Liang et al., 2025; Zhang et al., 2025; Lin et al., 2026; Li et al., 2025) typically focus on modeling goal-directed behaviors (Wood et al., 2022), which are generated by considering the outcomes that actions are likely to bring about, such as satisfying desires (Wang et al., 2025) or completing plans (Park et al., 2023).

However, these methods largely overlook habitual behavior, which accounts for a substantial proportion of everyday human behavior (Rebar et al., 2025; Wood & Rünger, 2016). Insights from cognitive psychology and neuroscience (Miller et al., 2019; Dolan & Dayan, 2013; Lee et al., 2014) suggest that human behavior is governed by multiple control systems, including goal-directed and habitual control systems. (Rebar et al., 2025; Wood & Rünger, 2016). Specifically, these systems generate behaviors through fundamentally different mechanisms. The goal-directed system generates actions that are expected to produce desired outcomes, whereas the habitual system triggers actions in response to relevant environmental stimuli. Multiple control systems jointly shape human behavior, contributing to the diverse and natural patterns observed in everyday human activity. Motivated by these observations, a pertinent question arises: *How can we integrate multiple behavioral control systems to generate believable human-like behaviors when building human-like agents?* Addressing this question poses three key challenges: (1) how to model habitual behavior; (2) how to exploit multiple control systems jointly shape the final human behavior; and (3) how to construct diverse habitual and goal-directed behaviors naturally and accurately in 3D environments.

In this work, we propose a **G**enerative **E**mbodied **M**ultiple Behavior Control **S**ystem (**GEMS**) for building human-like agent, which jointly models goal-directed and habitual behaviors, enabling believable 3D human-like behavior generation. GEMS employs a cognitive module to generate natural language behaviors and a motion generation module to construct them in 3D environment. The cognitive module reflects the multiple control mechanisms underlying human behavior through three interacting components. The **Habitual Controller** models the stimulus–response nature of habitual behavior by retrieving habitual actions associated with relevant environmental cues from a habit memory synthesized from prior empirical studies. In parallel, the **Goal-directed Controller** uses an internal world model (Wu et al., 2023) to predict action consequences, supporting goal-directed reasoning in open-ended embodied environments. The **Arbiter** dynamically governs the relative influence of each controller and determines the final action. Together, these components integrate behaviors driven by different control mechanisms in a plausible manner, thereby enabling diverse and naturalistic behavior generation.

In addition, the motion generation module performs 3D human–scene interaction generation based on the resulting natural-language behaviors. Existing language-conditioned interaction methods (Wang et al., 2022; Wu et al., 2025; Zou et al., 2026) have limited ability to cover the diverse behaviors produced by our cognitive module. We therefore adopt a scene-agnostic Text-to-Motion (T2M) model (Rempe et al., 2026) for versatile and expressive motion synthesis. Since such models lack explicit scene awareness, we further introduce keyframe-guided conditioning to impose spatial and interaction constraints from the environment. Compared with existing human-like agents that rely on predefined action spaces, GEMS supports more flexible and varied behavior generation.

We systematically compare our method against existing human-like agent frameworks (Leng et al., 2026; Lin et al., 2026; Liang et al., 2025; Wang et al., 2025; Yu et al., 2024) through comprehensive quantitative evaluations and human studies with **94 participants**, demonstrating its superiority in generating believable human-like behaviors. With extensive ablation studies (§4.4), we offer insights into the efficacy of our designs and showcase the benefits of multiple behavior control systems.

Our contributions are three-fold. (i) We introduce a real human-inspired multiple behavioral control system for believable human-like daily behavior simulation by jointly modeling goal-directed and habitual behaviors. (ii) We propose GEMS, a human-like agent framework to effectively generate human-like behaviors in 3D environments, with a Habitual Controller, a Goal-directed Controller, an Arbiter, and a 3D motion generation module. (iii) Extensive evaluation methods, human studies, and ablations demonstrate the effectiveness of GEMS and validate the benefits of introducing habitual behavior and multiple system coordination.

2 RELATED WORK

Human-like LLM agent. Human daily activity modeling supports applications in psychology, sociology, and virtual human simulation (Gärling & Garvill, 1993; Adler et al., 1987; Schmidt et al., 2024; Fraser et al., 2024). Recent advances in LLMs have enabled increasingly realistic simulations of human behavior. Existing works mainly follow two directions. One line models cognitive mechanisms behind behavior, including memory and planning (Park et al., 2023), desires (Wang et al., 2023; 2025), individual differences (He & Zhang, 2024; Li et al., 2025), and social interactions (Zhang et al., 2026; Lin et al., 2026). Another line grounds LLM-generated behaviors in 3D environments (Leng et al., 2026; Yu et al., 2024; Ye et al., 2026; Zhang et al., 2025). Most methods rely on predefined atomic actions, while ACTOR (Liang et al., 2025) uses motion clips pre-generated by T2M models but may suffer from motion transition drift. Despite these advances, existing frameworks are largely driven by explicit goals, plans or desires, following a goal-directed paradigm. In contrast, GEMS jointly models goal-directed and habitual behaviors and generates motions at runtime, enabling more diverse, natural, and believable behaviors in 3D environments.

Human behavior control mechanism study. An enduring and richly elaborated dichotomy in cognitive neuroscience is that of human behavior control mechanisms, divided into habitual and goal-directed (Dolan & Dayan, 2013; Miller et al., 2019; Wood et al., 2022). Existing computational frameworks of human behavioral control typically associate goal-directed control with model-based planning and habitual control with model-free learning (Daw et al., 2005; Dolan & Dayan, 2013; Wood & Rünger, 2016). Many works further introduce arbitration mechanisms that allocate control between these systems (Daw et al., 2005; Lee et al., 2014). Related cognitive architectures, including ACT-R, Soar, and the more recent CoALA, also employ procedural memory and model-free mechanisms related to habit learning for efficient action selection (Taatgen et al., 2008; Laird et al., 1987; Sumers et al., 2024). However, most computational frameworks have focused on explaining multiple control systems from a classical reinforcement learning (RL) perspective, with limited exploration in realistic, open-ended 3D environments. On the other hand, such multiple systems accounts are well established in cognitive neuroscience and psychology, existing work on human-like agents has rarely explored such mechanisms. To address this gap, we study multiple behavioral control system for human-like daily behavior simulation in realistic 3D environments.

3 METHODOLOGY

3.1 PROBLEM SETUP

Human-like behavior simulation. Following prior works (Wang et al., 2025; Lin et al., 2026), we formulate human-like daily behavior simulation as a sequential decision-making problem (Littman, 1996). Formally, at each timestep t, a human-like agent observes a state s_t, which includes a psychological state component s_t^{psy}. Multiple quantified factors are included in s_t^{psy}, such as stress, depletion, and cognitive load. We use b_t to denote the agent's belief state, including its personality and memory. The agent's goal I can be specified by an external instruction or determined by the agent itself based on its current state s_t. In this work, we primarily consider I specified by external instructions. The agent then generates an action a_t^G based on its goal-directed system, i.e.,

$$a_t^G \sim \text{Agent}(\cdot \mid s_t, b_t, I; p_\theta), \tag{1}$$

where p_θ denotes the underlying LLM parameterized by θ, and a_t^G is generated to pursue the current behavioral goal I. Following (Wang et al., 2025), each psychological factor of s_t^{psy} is represented on a 0–10 Likert scale and updated by an LLM after each interaction based on the previous state, executed action, and environmental feedback.

Human-like behavior simulation with multiple control systems. Unlike the conventional goal-directed setup, we extend behavior control to a multiple behavior control system formulation, where the agent can generate a goal-directed action a_t^G and a habitual action a_t^H under the goal-directed and habitual systems, respectively. This enables the agent to make decisions under the joint influence of multiple control systems inspired by human behavioral control mechanisms. The overall framework of the proposed GEMS is depicted in Fig. 2. The proposed GEMS consists of a **cognitive module** including Habitual Controller, Goal-directed Controller and Arbiter, as well as a **motion generation module** for human motion generation and object interaction in open-ended 3D environment.

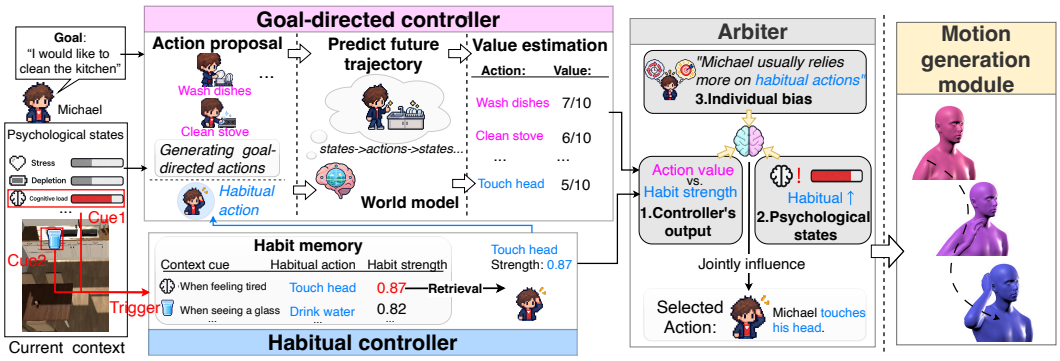


Figure 2: **Overall framework of the proposed GEMS.** The cognitive module takes the agent's goal and its current state as input: the Habitual Controller retrieves habitual actions based on cues triggered by the environment; the Goal-directed Controller proposes goal-directed actions and estimates the values of both habitual and goal-directed actions based on outcomes imagined by the world model; and the Arbiter selects an action based on both controller's output, individual bias, and psychological states. The selected action is then constructed by the motion generation module.

3.2 Cognitive Module

3.2.1 Habitual Controller

Habitual behaviors can be triggered directly by a stimulus, i.e., *a context cue* in the environment that activates the corresponding actions encoded in habit memory. To simulate this process, we synthesize a prior habit dataset from empirical sources to support personalized habit memory creation for each agent, together with a cue retrieval algorithm. This easy-to-achieve design avoids the cost of modeling long-term habit formation. At timestep t, the agent receives its current experienced context s_t, which may contain a context cue denoted as q_i^{cue} that triggers a habitual action a_i^H.

Habit memory. We construct the habit memory in three steps. 1) We model context cues. Potentially, any internal or external event may trigger habitual behavior (Gardner et al., 2016). Since the full space of possible cues cannot be exhaustively represented, we represent the cue as a multidimensional vector whose dimensions are both empirically validated and available in the agent's current context s_t, including visual episodes, psychological states, events, and active goals. For simplicity, we assume that each habitual action is associated with a single cue dimension. 2) We synthesize a prior habit dataset from existing empirical sources, including habit studies and datasets (Ersche et al., 2017; Moors & De Houwer, 2006). We extract habitual behaviors and their associated context cues from these sources and represent each instance in natural language as a cue–action pair $\langle q_i^{cue}, a_i^H \rangle$, with each cue mapped to a predefined cue dimension, yielding a habit dataset $\mathcal{D}^H$ containing 571 habitual behaviors observed in everyday human life. (see Appendix E.1 for details). 3) Finally, we initialize a personalized habit memory for each agent. Since individuals may exhibit different habits with varying strengths, we use LLMs to sample a subset of cue–action pairs from $\mathcal{D}^H$ based on the agent's persona and assign each sampled pair a habit strength $H(a_i^H) \in [0, 1]$ (see Appendix E.2). The resulting personalized habit memory can be written as $\widehat{\mathcal{D}}^H = \left\{ \left\langle q_i^{\text{cue}}, a_i^H, \right\rangle H(a_i^H) \right\}_{i=1}^{l}$ in which l represents the number of cue-action pairs available for human-like agents.

Cue retrieval algorithm. Habitual responses are triggered by recurring context cues, while stronger habits are more likely to guide behavior (Wood & Rünger, 2016). Based on these observations, we propose a two-stage retrieval algorithm that first identifies relevant context cues from $\widehat{\mathcal{D}}^H$ given s_t, and then selects the most relevant habitual action in habit memory. At the first stage, for each pair $\langle q_i^{cue}, a_i^H \rangle$ in $\widehat{\mathcal{D}}^H$, we compute the similarity C_i between the corresponding cue dimension in s_t and q_i^{cue}. We retain habits whose cue similarity exceeds a threshold κ, forming $\mathcal{E}_t = \{i \mid C_i \geq \kappa\}$. If $\mathcal{E}_t = \emptyset$, no habitual behavior is triggered and the agent relies solely on goal-directed control. Otherwise, we proceed to the second stage and compute the response score for each habit $i \in \mathcal{E}_t$ as $p_i = \frac{C_i - \kappa}{1 - \kappa} H(a_i^H)$. In step t, we denote the habitual action corresponding to the context cue q_i^{cue} as $a_{i,t}^H$. For $i \in \mathcal{E}_t$, we then select a habitual action according to the normalized distribution, i.e.,

$P(a^H_{i,t} \mid s_t, \mathcal{E}_t) = \frac{p_i}{\sum_{j \in \mathcal{E}_t} p_j}$. The two-stage design can reduce interference from strong but irrelevant habits while preserving the influence of habit strength on action selection, thereby supporting better habitual behavior generation. We compare our method with alternative strategies on our designed evaluation set (see Appendix E.3) and it achieves the best performance (Tab. 2e).

3.2.2 GOAL-DIRECTED CONTROLLER

Goal-directed behaviors are driven by considering potential action outcomes, a process commonly modeled through model-based RL (Miller et al., 2019). To extend this principle to embodied environments with open-ended natural language action spaces, our Goal-directed Controller proposes candidate actions based on the agent's current state s_t and goal I, predicts the consequences of these actions with an internal world model, and estimates their values for contributing to the goal.

Action proposal. Denote $\mathcal{A}^G_t$ as the set of actions generated by the Goal-directed Controller at step t. The agent samples N candidate actions that are likely to contribute positively to I, conditioned on s_t and b_t, i.e., $\mathcal{A}^G_t = \{a^G_{i,t}\}_{i=1}^N \sim p_\theta(\cdot \mid s_t, b_t, I)$. When a habitual action $a^H_{i,t}$ is triggered, it is considered alongside the goal-directed actions (Morris et al., 2019), yielding $\mathcal{A}_t = \mathcal{A}^G_t \cup \{a^H_{i,t}\}$.

Internal model. Humans rely on internal models to predict action consequences and adapt to dynamic environments (Yu et al., 2026). Inspired by this mechanism, we introduce a frozen LLM-based world model (Li et al., 2026; Nottingham et al., 2023) $\mathcal{M}_{p_\theta}$ to approximate the environment transition function T. Given a candidate action $a_{i,t} \in \mathcal{A}_t$, the model predicts the next state as

$$\hat{s}_{t+1} \sim \mathcal{M}_{p_\theta}(\cdot \mid s_t, a_{i,t}) \approx T(\cdot \mid s_t, a_{i,t}). \tag{2}$$

By recursively applying the world model, the agent can prospectively predict its future states following each candidate action. Further design details of the world model are provided in Appendix G.

Value estimation. The internal model enables the agent to estimate the goal-directed value of each candidate action through prospective planning. Starting from $a_{i,t} \in \mathcal{A}_t$, a planning strategy $\mathcal{P}$ simulates possible future developments and evaluates the imagined outcomes with respect to I:

$$Q(s_t, a_{i,t} \mid I) = \mathbb{E}_{\hat{\tau} \sim \mathcal{P}(\mathcal{M}_{p_\theta}; s_t, a_{i,t}, I)} \left[\mathcal{V}_{p_\theta}(\hat{\tau}, I) \right], \tag{3}$$

where $\hat{\tau}$ denotes an imagined future trajectory and $\mathcal{V}_{p_\theta}$ prompts the agent to assign a subjective score from 0 to 10 according to how well the predicted outcome satisfies I. For simplicity, we write $Q(a_{i,t})$ for $Q(s_t, a_{i,t} \mid I)$ hereafter.

3.2.3 ARBITER

The Goal-directed and Habitual Controllers may favor different actions under the same experienced context s_t. Additionally, considering individual differences and their state (i.e., environment and psychological state) variations of real human, our Arbiter jointly integrates three sources of influence: the output of each controller (Miller et al., 2019), the agent's individual control bias (Piray et al., 2016), and its current psychological state (Wood et al., 2022), for selecting the final action. This mechanism is motivated by findings in psychology and neuroscience and shows how multiple behavior control systems jointly shape the final behavior. We represent each influence as:

1) Controllers' output: For an action $a_{i,t}$, the Goal-directed Controller provides action value $Q(a_{i,t})$, while the Habitual Controller provides habit strength $H(a_{i,t})$. For actions proposed only by the Goal-directed Controller, we set $H(a_{i,t}) = 0$. A higher output value indicates stronger influence from the corresponding controller. *2) Individual control bias*: We further denote $\rho \in (0, 1)$ as the agent's baseline preference for goal-directed control, where larger values indicate that the agent relies more on goal-directed control. *3) Current psychological state*: We design a psychological state interference function $F(s^{\text{psy}}_t) \in [0, 1]$ to characterize how adverse psychological states increase reliance on habitual control. For each psychological factor in s^{psy}_t, quantified on a 0–10 scale, A higher score indicates a more negative psychological state, making $F(s^{\text{psy}}_t)$ more likely to take a larger value and thereby increasing the weight of Habitual Controller. Here, we consider three psychological factors: stress (*str.*), depletion (*dep.*) and cognitive load (*cog.*) (Wood & Rünger, 2016). $F(s^{\text{psy}}_t)$ introduces transient dynamics induced by the agent's current psychological state into the balance with the two controllers. It can be instantiated in various ways, such as using an LLM. In this work, we calibrate $F(s^{\text{psy}}_t)$ using estimation values from existing human behavioral studies. More details are provided in Appendix F.

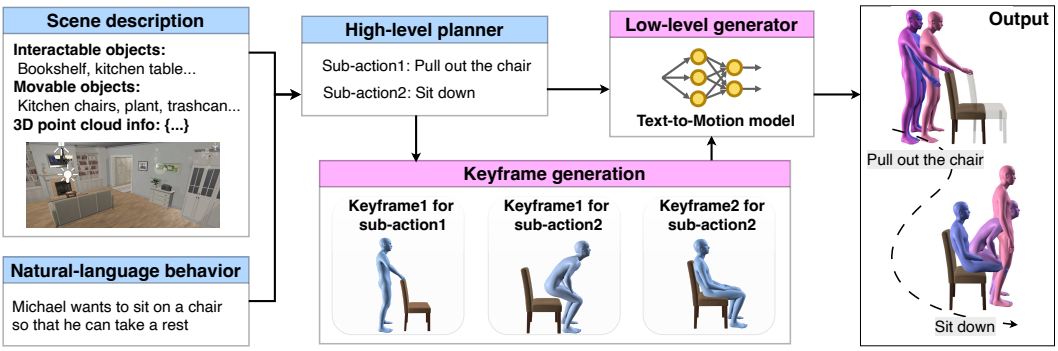


Figure 3: The motion generation module consists of two components. Given the action description in natural language from the cognitive module and the current scene information, the high-level planner produces an executable sub-action plan; sub-actions are aligned with keyframes where needed. Using this as a condition, the low-level generator uses T2M models to generate motion sequences.

Inspired by the arbitration formulation in Miller et al. (2019), we further define the overall drive $U(a_{i,t})$ of each candidate action by considering the three sources of influence together:

$$U(a_{i,t}) = w_t^G \bar{Q}(a_{i,t}) + w_t^H H(a_{i,t}), \tag{4}$$

where $w_t^G = \rho\,(1 - F(s_t^{\text{psy}}))$ denotes the joint weight of goal-directed control at time t, and the corresponding habitual control weight is given by $w_t^H = 1 - w_t^G$, and $\bar{Q}(a_{i,t}) \in [0, 1]$ denotes the rescaled goal-directed value of $Q(a_{i,t})$. Finally, the Arbiter selects the action with the highest overall drive as the final action of the human-like agent, i.e., $a_t = \arg\max_{a_{i,t} \in \mathcal{A}_t} U(a_{i,t})$. This design accounts for the intrinsic action preferences of each controller while incorporating individual differences and dynamic psychological states, thereby supporting heterogeneous decision-making and promoting coherent and natural behavior sequences.

3.3 Motion Generation Module

Unlike existing approaches (Liang et al., 2025; Leng et al., 2026) constrained by a predefined action space, our motion generation module constructs human–scene interactions in 3D environments from the natural language action a_t. It employs a High-Level Planner and a Low-Level Motion Generator.

High-Level Planner. As shown in Fig. 3, the planner augments the action description a_t with necessary execution steps required to achieve the action in the current scene. Given a_t and the current scene state, including the scene layout, objects, and the agent's states, the planner determines the desired scene state after the action is completed and, when necessary, decomposes the action into an ordered sequence of executable sub-actions and object interactions. For instance, if the agent wants to sit on a chair, the chair should be pulled out before sitting down. An LLM is used to produce this ordered decomposition, yielding a sequence of textual sub-action descriptions and matched keyframe constraints used to guide the Low-Level Motion Generator.

Low-Level Motion Generator. Given the decomposed sub-action descriptions, the generator executes them sequentially by synthesizing human motion segments and coordinating object interactions, yielding diverse human-object interaction sequences. In particular, the scene-agnostic T2M model (Rempe et al., 2026) we adopt offers expressive motion synthesis but does not directly perceive scene geometry. To address this limitation, we introduce a keyframe mechanism that bridges high-level planning and low-level motion generation by encoding planned sub-actions and scene information as 3D constraints for motion synthesis (Xie et al., 2024). Given a sub-action instruction and the perceived scene point cloud, a VLM predicts target human skeletal poses and corresponding object poses for a sparse set of keyframes. This mechanism supports flexible sub-action composition and smooth motion transitions. The T2M model then synthesizes each motion segment from the sub-action description, incorporating keyframe constraints and a planned navigation path when applicable. When the generated motion reaches a designated keyframe, the system triggers the corresponding interaction event, such as establishing a hand-object attachment when grasping an apple. Further implementation details are provided in Appendix H.

Table 1: Quantitative comparison with previous human-like agent frameworks on controlled daily-life goals with the same initialization. Mean denotes the average of Nat., Coh., and PA. Besides, Qual. and Act. have a maximum score of 5. The best results are highlighted in **bold**.

Method	Cognitive module				Motion generation module					Action type
	Nat. ↑	Coh. ↑	PA ↑	Mean. ↑	SSR ↑	GSR ↑	DTG ↓	Qual. ↑	Act. ↑	
X-VirtualHome (Leng et al., 2026)	42.26	83.81	77.52	67.86	56.67	53.33	**19.95**	1.27	2.86	Atomic
AGA (Yu et al., 2024)	56.98	53.74	69.17	59.96	51.33	26.67	19.96	1.21	1.68	Atomic
D2A (Wang et al., 2025)	63.74	76.90	75.43	72.02	–	–	–	–	–	–
ASVO (Lin et al., 2026)	42.01	46.83	27.11	38.65	–	–	–	–	–	–
ACTOR (Liang et al., 2025)	47.77	86.18	77.62	70.52	59.79	46.67	98.42	3.75	3.59	Motion clips
ACTOR† (Liang et al., 2025)	52.79	85.85	78.27	72.30	62.67	48.67	94.74	3.66	3.69	Motion clips
GEMS (ours)	**90.31**	**89.96**	**84.35**	**88.21**	**86.24**	**79.33**	22.48	**4.12**	**4.23**	**Generative**

4 EXPERIMENT

4.1 EXPERIMENTAL SETUP AND EVALUATION METRICS

To evaluate the human-like performance of GEMS, we separately evaluate its two main modules: (i) the cognitive module, by assessing the human-likeness of the behavior sequences it generates; and (ii) the motion generation module, by assessing whether the selected behaviors can be faithfully and naturally constructed in 3D environments. For the cognitive module, we follow prior LLM-based evaluation protocols (Wang et al., 2025) and use three metrics, each scored on a 0–100 scale: **Naturalness (Nat.)** for human-likeness (Wang et al., 2025), **Coherence (Coh.)** for logical consistency (Wang et al., 2025), and **Personality Alignment (PA)** for consistency with the given persona (Li et al., 2025). **Mean** denotes their average value. This evaluation protocol emulates a human user study (Cai et al., 2025). For the motion generation module, **Step Success Rate (SSR)** measures individual-step success, and **Goal Success Rate (GSR)** measures whole-plan success (Liang et al., 2025). **Distance to Goal (DTG)** measures the minimum wrist-to-goal-object distance (Diomataris et al., 2024). **Quality Score (Qual.)** measures motion naturalness, while **Action Score (Act.)** measures consistency with the target behavior description (Wang et al., 2022). The evaluation criteria and the definitions of successful task execution follow the same settings in the corresponding prior works. Details are provided in Appendix A.3.

We compare GEMS with five previous human-like agent frameworks. X-VirtualHome, AGA, and ACTOR, which support embodied behavior execution and can therefore be evaluated on both the cognitive and motion generation module metrics, whereas D2A and ASVO only focus on language behavior generation and are therefore evaluated only on the cognitive module metrics. We further evaluate a variant of ACTOR, denoted as ACTOR†, which retains the original ACTOR pipeline while using the same Habit Memory (with habitual behaviors sampled using an LLM), T2M model, and simulation platform as GEMS. For a fair comparison, we evaluate all methods on 30 shared daily life goals (such as preparing dinner and cleaning the kitchen) with identical initialization settings. Each goal runs five times, yielding 150 runs per method, and all scores are averaged over runs. Full implementation details are provided in Appendix A.2.

4.2 COMPARISON WITH STATE-OF-THE-ART METHODS

As shown in Tab. 1, ACTOR† yields only limited gains over ACTOR, suggesting that the effective triggering and arbitration of habitual behaviors are critical, thereby highlighting the importance of the Habitual Controller and Arbiter design. The limited improvement in motion metrics further emphasizes the importance of keyframe-guided motion generation. In contrast, GEMS outperforms all the baselines on the cognitive module metrics, and improves the Mean score by about 16 points over the previous best method D2A. These results demonstrate the benefit of modeling habitual behaviors and coordinating habitual and goal-directed control for believable human-like behavior. For the motion generation module, GEMS also demonstrates the best overall performance. Compared with predefined atomic actions or independent motion clips, our keyframe-guided approach generates continuous, scene-aware motions at runtime. This improves the overall quality of the generated motions. A qualitative visualization of GEMS are shown in Fig. 4, with more results in Appendix C.

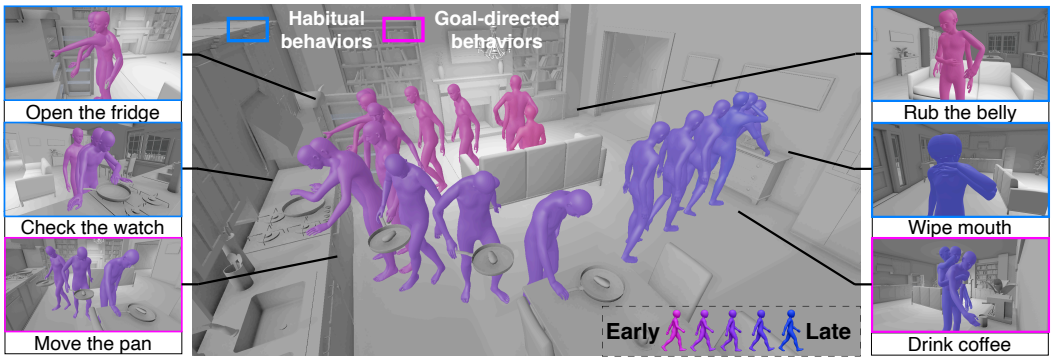


Figure 4: Qualitative example generated by GEMS for the goal "prepare dinner".

4.3 Human Study

To evaluate whether incorporating habitual behavior improves perceived human-likeness and to assess how closely GEMS approximates real human behavior, we conduct blind pairwise human evaluations through questionnaires. We evaluate the language behavioral sequences generated by the cognitive module and the motion videos generated by the motion generation module, with a total of **94** crowd-sourced participants involved in the study.

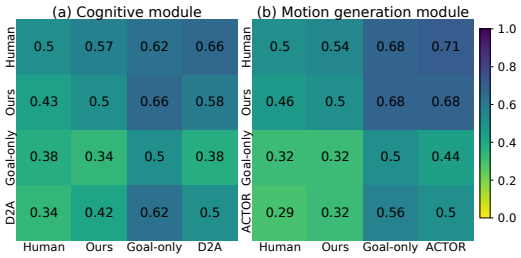


Figure 5: Key human preference results. Each entry compares the win rate of the method on the vertical axis against that on the horizontal axis, with diagonal values fixed at 0.5. Full results are provided in Appendix B.

We further compare GEMS with a variant that uses only the Goal-directed Controller (denoted as Goal-only), other baseline methods, and real human behavior references (Yang et al., 2025; Jiang et al., 2024). In each pairwise comparison, participants are asked to select the behavior that appears more human-like. As shown in Fig. 5, GEMS receives higher preference rates than Goal-only, demonstrating that incorporating habitual behavior improves perceived human-likeness. When directly compared with real human behavior, GEMS achieves preference rates of 43.4% and 46.3% in the two evaluations, respectively, indicating that GEMS is capable of generating believable and human-like behaviors. Among all agent frameworks, GEMS achieves the highest overall preference rate. The complete study design and detailed results are provided in Appendix B.

4.4 Ablation Study

We define several additional metrics for the ablation studies. For a completed trajectory τ consisting of m execution steps, **Habit Ratio** is the fraction of habitual actions, and **Task Execution Efficiency (TE)** is $\mathrm{TE} = \frac{\mathcal{V}_{p_\theta}(\tau, I)}{m}$. **Retrieval Accuracy (Acc.)** measures correctness on the retrieval evaluation set (see Appendix E.4). Each setting in Tab. 2a, 2b, and 2f is *averaged* over 15 independent runs.

Impact of the Key Components in the Cognitive Module. Tab. 2a evaluates the effectiveness of the Habitual Controller, World Model, and Arbiter. Without the Arbiter, an LLM selects among the actions proposed by two controllers; without the World Model, an LLM directly estimates action values. Adding the Habitual Controller improves Nat., supporting our motivation to model habitual behaviors for believable human-like agents simulation. Adding the Arbiter further improves all three metrics, suggesting that generating natural and coherent human-like behavior benefits not only from incorporating habitual behaviors but also from coordinating habitual and goal-directed control.

Ablation on $F(s_t^{\mathrm{psy}})$ in Arbiter. Tab. 2b evaluates the effect of $F(s_t^{\mathrm{psy}})$ under different psychological state initializations. *Normal* sets the initial psychological state (stress, depletion, cognitive load) to $s_0^{\mathrm{psy}} = (3, 3, 3)$. *Adverse* averages the results across three settings: $(7, 3, 3)$, $(3, 7, 3)$, and

Table 2: **Ablation study.** (a) Ablation of the key components in the cognitive module. (b) Ablation of $F(s_t^{\text{psy}})$ in the Arbiter under different initializations of s_t^{psy}. Initial states s_0^{psy} include (stress, depletion, cognitive load): *Normal* uses $(3, 3, 3)$; *Adverse* denotes negative psychological states and it averages results across three settings: $(7, 3, 3)$, $(3, 7, 3)$, and $(3, 3, 7)$. (c) Comparison of alternative motion-generation methods. (d) Ablation of keyframe guidance in the motion generation module. (e) Comparison of different habit retrieval methods. (f) Effect of the coefficient ρ. Habit Ratio measures the proportion of habitual actions, while TE measures task execution efficiency.

Habit	WM	Arb.	Nat. ↑	Coh. ↑	PA ↑
			39.54	55.39	71.48
✓			61.05	74.72	81.80
	✓		52.25	84.66	73.92
✓	✓		76.17	72.67	82.83
✓	✓	✓	**88.48**	**87.38**	**85.22**

(a)

$F(s_t^{\text{psy}})$	Habit Ratio (%)		Mean. ↑
	Normal	Adverse	
	35.81	36.57	79.54
✓	40.92	**53.06**	**85.71**

(b)

Motion generation method	SSR ↑	Qual. ↑	Act. ↑
HUMANISE (Wang et al., 2022)	33.33	1.38	1.31
HOI-from-HLI (Wu et al., 2025)	37.78	2.13	1.72
InfBaGel (Zou et al., 2026)	40.00	1.76	1.69
GEMS (ours)	**95.56**	**4.23**	**4.52**

(c)

Keyframe	SSR ↑	Qual. ↑	DTG ↓
	52.22	1.75	76.80
✓	**86.24**	**4.12**	**22.48**

(d)

Retrieval method	Acc. (%) ↑
Sim.×HS	71.4
Sim.-only	73.3
GEMS (ours)	**84.7**

(e)

ρ	0.05	0.1	0.3	0.5	0.7	0.9	0.95
Habit Ratio (%)	55.62	52.27	52.94	41.83	41.65	9.46	9.92
TE ↑	0.51	0.59	0.57	0.78	0.66	**0.99**	0.81
Mean. ↑	77.37	80.53	85.87	**87.31**	84.74	78.00	75.53

(f)

$(3, 3, 7)$. Adding $F(s_t^{\text{psy}})$, the *Adverse* setting increases the habit ratio, while this shift is attenuated without it. This shows that $F(s_t^{\text{psy}})$ enables the Arbiter to dynamically balance two controllers according to psychological states (see Tab. 4c for detailed results).

Comparison of the Motion Generation Module. Tab. 2c compares alternative human–scene interaction generation methods for our motion generation module on 45 randomly sampled actions generated by the cognitive module. Existing methods achieve low SSR, highlighting their limitation in constructing the diverse behaviors in our setting and motivating the design of our module.

Impact of Keyframe Guidance. Tab. 2d evaluates keyframe guidance in the motion generation module. Introducing keyframes improves SSR and motion quality while reducing DTG, showing that keyframe constraints effectively ground the scene-agnostic T2M model in 3D interactions while improving overall motion naturalness.

Comparison of Retrieval Methods. Tab. 2e compares our two-stage retrieval with cue similarity alone (Sim.-only) and cue similarity multiplied by habit strength (Sim.×HS). Our two-stage method achieves the best performance, supporting the separation of cue matching from habit-strength weighting (see Tab. 6 for detailed results). The results in Tab. 2a further suggest that effective habit retrieval schema in the Habit Controller contributes to the human-like performance.

Influence of the Individual Bias Coefficient ρ. Tab. 2f studies the influence of ρ. Different values of ρ induce distinct behavior patterns, showing that the Arbiter provides a controllable mechanism for individual variation. Increasing ρ shifts behavior control toward goal-directed behavior, reducing the habit ratio while generally improving TE. An intermediate value of ρ achieves the best Mean score, indicating a trade-off between habit ratio and TE in achieving human-like performance. Based on these results, we recommend $\rho \in [0.3, 0.7]$ as the preferred range for persona configuration. These findings suggest that simply increasing the habit ratio does not necessarily yield more human-like performance; rather, effective coordination between the two controllers is the key.

5 Conclusion

In this work, we propose GEMS, a Generative Embodied Multiple behavioral control Systems for building human-like agent by jointly modelling goal-directed and habitual behaviors through a Habitual Controller, a Goal-directed Controller, and an Arbiter, together with a keyframe-guided motion generation module for converting natural language behavior descriptions into 3D human-object interactions. Extensive quantitative evaluations and human studies demonstrate that GEMS generates more believable human-like behaviors. Further ablation studies validate the benefits of multiple behavioral control system coordination, and keyframe-guided motion generation. We hope this work represents a step toward more natural and diverse human-like agents in interactive 3D environments and offers valuable insights for future research.

AI USE DISCLOSURE

In this work, we used generative AI for data cleaning and reformatting, as well as for synthetic data generation. In addition, we used ChatGPT for automated quantitative evaluation. We did not use generative AI tools for research idea development, experimental design, or the analysis of experimental results. All AI-assisted work was reviewed or validated by the authors. The authors take full responsibility for the final content of this work.

ETHICS STATEMENT

This work includes human-subject evaluations conducted to assess the perceived human-likeness of generated behavioral sequences and 3D human motions. The study protocol was reviewed and approved by the ethics committee of our institution prior to participant recruitment. Participants were recruited through Amazon Mechanical Turk (MTurk), and all participants provided informed consent before taking part in the study. Participation was limited to individuals aged 18 years or older. We collected demographic information, including age, gender, ethnicity, and education level, solely for descriptive characterization of the participant population. All responses were analyzed in de-identified form, and the collected data were used solely for the human evaluations reported in this work. Further details of the study protocol, participant population, and evaluation procedure are provided in Appendix B.

REFERENCES

Patricia A. Adler, Peter Adler, and Andrea Fontana. Everyday life sociology. *Annual Review of Sociology*, 13:217–235, 1987. doi: 10.1146/annurev.so.13.080187.001245.

NH Azrin and RG Nunn. Habit-reversal: a method of eliminating nervous habits and tics. *Behaviour research and therapy*, 11(4):619–628, 1973.

Joseph Bates et al. The role of emotion in believable agents. *Communications of the ACM*, 37(7): 122–125, 1994.

Shengqu Cai, Eric Ryan Chan, Yunzhi Zhang, Leonidas Guibas, Jiajun Wu, and Gordon Wetzstein. Diffusion self-distillation for zero-shot customized image generation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18434–18443. IEEE, 2025.

Nathaniel D. Daw, Yael Niv, and Peter Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711, 2005. doi: 10.1038/nn1560.

Markos Diomataris, Nikos Athanasiou, Omid Taheri, Xi Wang, Otmar Hilliges, and Michael J Black. Wandr: Intention-guided human motion generation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 927–936. IEEE, 2024.

Ray J Dolan and Peter Dayan. Goals and habits in the brain. *Neuron*, 80(2):312–325, 2013.

Karen D Ersche, Tsen-Vei Lim, Laetitia HE Ward, Trevor W Robbins, and Jan Stochl. Creature of habit: A self-report measure of habitual routines and automatic tendencies in everyday life. *Personality and Individual Differences*, 116:73–85, 2017.

Zhaohan Feng, Ruiqi Xue, Lei Yuan, Yang Yu, Ning Ding, Meiqin Liu, Bingzhao Gao, Jian Sun, Xinhu Zheng, and Gang Wang. Multi-agent embodied ai: Advances and future directions. *Science China Information Sciences*, 69(5):151202, 2026.

James H Fetzer. What is artificial intelligence? In *Artificial intelligence: Its scope and limits*, pp. 3–27. Springer, 1990.

Alan D. Fraser, Isabella Branson, Ross C. Hollett, Craig P. Speelman, and Shane L. Rogers. Do realistic avatars make virtual reality better? examining human-like avatars for vr social interactions. *Computers in Human Behavior: Artificial Humans*, 2(2):100082, 2024. doi: 10.1016/j.chbah.2024.100082.

Benjamin Gardner, L Alison Phillips, and Gaby Judah. Habitual instigation and habitual execution: Definition, measurement, and effects on behaviour frequency. *British journal of health psychology*, 21(3):613–630, 2016.

Tommy Gärling and Jörgen Garvill. Psychological explanations of participation in everyday activities. In Tommy Gärling and Reginald G. Golledge (eds.), *Behavior and Environment: Psychological and Geographical Approaches*, volume 96, pp. 270–297. Elsevier, 1993. doi: 10.1016/S0166-4115(08)60047-3.

Dejan Georgiev, Rosie Christie, Mariam Torkamani, Ruifeng Song, Patricia Limousin, and Marjan Jahanshahi. Development and validation of a daily habit scale. *Frontiers in neuroscience*, 16: 880023, 2022.

Ziyan Guo, Haoxuan Qu, Hossein Rahmani, Dewen Soh, Ping Hu, Qiuhong Ke, and Jun Liu. Tstmotion: Training-free scene-aware text-to-motion generation. In *2025 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2025.

Mohamed Hassan, Yunrong Guo, Tingwu Wang, Michael Black, Sanja Fidler, and Xue Bin Peng. Synthesizing physical character-scene interactions. In *ACM SIGGRAPH 2023 Conference Proceedings*, pp. 1–9, 2023.

Zihong He and Changwang Zhang. Afspp: Agent framework for shaping preference and personality with large language models. *arXiv preprint arXiv:2401.02870*, 2024.

Bernhard Hommel. Event files: Evidence for automatic integration of stimulus-response episodes. *Visual cognition*, 5(1-2):183–216, 1998.

Nan Jiang, Zhiyuan Zhang, Hongjie Li, Xiaoxuan Ma, Zan Wang, Yixin Chen, Tengyu Liu, Yixin Zhu, and Siyuan Huang. Scaling up dynamic human-scene interaction modeling. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1737–1747. IEEE, 2024.

Dongjae Kim, Geon Yeong Park, John P. O'Doherty, and Sang Wan Lee. Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. *Nature Communications*, 10:5738, 2019. doi: 10.1038/s41467-019-13632-1.

John E. Laird, Allen Newell, and Paul S. Rosenbloom. SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64, 1987. doi: 10.1016/0004-3702(87)90050-6.

Sang Wan Lee, Shinsuke Shimojo, and John P O'doherty. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3):687–699, 2014.

Zikang Leng, Megha Thukral, Yaqi Liu, Hrudhai Rajasekhar, Shruthi K. Hiremath, Jiaman He, and Thomas Plötz. Agentsense: Virtual sensor data generation using llm agents in simulated home environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1891–1899, 2026. doi: 10.1609/aaai.v40i3.37169. URL https://ojs.aaai.org/index.php/AAAI/article/view/37169.

Haoyang Li, Zan Wang, Wei Liang, and Yizhuo Wang. X's day: Personality-driven virtual human behavior generation. *IEEE transactions on visualization and computer graphics*, 31(5):3514–3524, 2025.

Hongwu Li, Zhenliang Zhang, and Wei Wang. Habitaction: A video dataset for human habitual behavior recognition. *arXiv preprint arXiv:2408.13463*, 2024.

Yixia Li, Hongru Wang, Peng Lai, Zhiwen Ruan, He Zhu, Youxin Zhu, Ganlong Zhao, Minda Hu, Yun Chen, Sibei Yang, et al. Bridging the agent-world gap: Text world models for llm-based agents. *arXiv preprint arXiv:2606.09032*, 2026.

Chen Liang, Wenguan Wang, and Yi Yang. Towards human-like virtual beings: Simulating human behavior in 3d scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10753–10763, 2025. doi: 10.1109/ICCV51701.2025.01001. URL https://openaccess.thecvf.com/content/ICCV2025/html/Liang_Towards_Human-like_Virtual_Beings_Simulating_Human_Behavior_in_3D_Scenes_ICCV_2025_paper.html.

Jingzhe Lin, Ceyao Zhang, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, and Fangwei Zhong. Beyond self-interest: Modeling social-oriented motivation for human-like multi-agent interactions. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, Paphos, Cyprus, 2026. International Foundation for Autonomous Agents and Multiagent Systems. doi: 10.65109/OITH7375.

Michael Lederman Littman. *Algorithms for sequential decision-making*. Brown University, 1996.

Riccardo Maramotti, Thomas Parr, Manuela Tondelli, Daniela Ballotta, Sanjay G. Manohar, Giovanna Zamboni, and Giuseppe Pagnoni. Understanding mechanisms of voluntary engagement of mental effort using active inference. *Cognitive, Affective, & Behavioral Neuroscience*, 26(2): 688–702, 2026. doi: 10.3758/s13415-026-01419-z.

Kevin J Miller, Amitai Shenhav, and Elliot A Ludvig. Habits without values. *Psychological review*, 126(2):292, 2019.

Agnes Moors and Jan De Houwer. Automaticity: a theoretical and conceptual analysis. *Psychological bulletin*, 132(2):297, 2006.

Adam Morris, Jonathan Phillips, Karen Huang, and Fiery Cushman. Generating options and choosing between them rely on distinct forms of value representation. 2019.

David T Neal, Wendy Wood, and Aimee Drolet. How do people adhere to goals when willpower is low? the profits (and pitfalls) of strong habits. *Journal of Personality and Social Psychology*, 104 (6):959, 2013.

Kolby Nottingham, Prithviraj Ammanabrolu, Alane Suhr, Yejin Choi, Hannaneh Hajishirzi, Sameer Singh, and Roy Fox. Do embodied agents dream of pixelated sheep: Embodied decision making using language guided world modelling. In *International Conference on Machine Learning*, pp. 26311–26325. PMLR, 2023.

A. Ross Otto, Samuel J. Gershman, Arthur B. Markman, and Nathaniel D. Daw. The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, 24(5):751–761, 2013. doi: 10.1177/0956797612463080.

Hio Tong Pang, Feride Canarslan, and Mingyuan Chu. Individual differences in conversational self-touch frequency correlate with state anxiety. *Journal of Nonverbal Behavior*, 46(3):299–319, 2022.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, 2023. doi: 10.1145/3586183.3606763. URL https://doi.org/10.1145/3586183.3606763.

Payam Piray, Ivan Toni, and Roshan Cools. Human choice strategy varies with anatomical projections from ventromedial prefrontal cortex to medial striatum. *The Journal of Neuroscience*, 36 (10):2857–2867, 2016.

Ruyi Qiu, Malte Möller, Iring Koch, Christian Frings, and Susanne Mayr. The influence of event segmentation by context on stimulus–response binding. *Journal of Experimental Psychology: Human Perception and Performance*, 49(3):355, 2023.

Amanda L Rebar, Grace Vincent, Katya Kovac Le Cornu, and Benjamin Gardner. How habitual is everyday life? an ecological momentary assessment study. *Psychology & health*, pp. 1–26, 2025.

Davis Rempe, Mathis Petrovich, Ye Yuan, Haotian Zhang, Xue Bin Peng, Yifeng Jiang, Tingwu Wang, Umar Iqbal, David Minor, Michael de Ruyter, et al. Kimodo: Scaling controllable human motion generation. *arXiv preprint arXiv:2603.15546*, 2026.

Mauricio GC Resende and Celso C Ribeiro. Grasp: Greedy randomized adaptive search procedures. In *Search methodologies: introductory tutorials in optimization and decision support techniques*, pp. 287–312. Springer, 2013.

Susanne Schmidt, Ipek Köysürenbars, and Frank Steinicke. Frankenstein's monster in the metaverse: User interaction with customized virtual agents. *IEEE Transactions on Visualization and Computer Graphics*, 30(11):7162–7171, 2024. doi: 10.1109/TVCG.2024.3456205.

Darryl W Schneider and Gordon D Logan. Selecting a response in task switching: testing a model of compound cue retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1):122, 2009.

Lars Schwabe and Oliver T Wolf. Stress increases behavioral resistance to extinction. *Psychoneuroendocrinology*, 36(9):1287–1293, 2011.

Aaron Sloman. What sort of architecture is required for a human-like agent? In *Foundations of rational agency*, pp. 35–52. Springer, 1999.

Aaron Sloman. Architectural requirements for human-like agents both natural and artificial. *Human cognition and social agent technology*, pp. 163–195, 2000.

Theodore R. Sumers, Shunyu Yao, Karthik R. Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=1i6ZCyf1QJ.

Maciej Świechowski, Konrad Godlewski, Bartosz Sawicki, and Jacek Mańdziuk. Monte carlo tree search: a review of recent modifications and applications: M. świechowski et al. *Artificial Intelligence Review*, 56(3):2497–2562, 2023.

Niels A. Taatgen, David Huss, Daniel Dickison, and John R. Anderson. The acquisition of robust and flexible cognitive skills. *Journal of Experimental Psychology: General*, 137(3):548–565, 2008. doi: 10.1037/0096-3445.137.3.548.

Yiding Wang, Yuxuan Chen, Fangwei Zhong, Long Ma, and Yizhou Wang. Simulating human-like daily activities with desire-driven autonomy. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/513cb685f67550dbd133b81a7a24249f-Abstract-Conference.html.

Zan Wang, Yixin Chen, Tengyu Liu, Yixin Zhu, Wei Liang, and Siyuan Huang. Humanise: Language-conditioned human motion generation in 3d scenes. *Advances in Neural Information Processing Systems*, 35:14959–14971, 2022.

Zhilin Wang, Yu Ying Chiu, and Yu Cheung Chiu. Humanoid agents: Platform for simulating human-like generative agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 167–176, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-demo.15. URL https://aclanthology.org/2023.emnlp-demo.15/.

Florian Waszak and Vanida Pholulamdeth. Episodic s–r bindings and emotion: About the influence of positive and negative action effects on stimulus–response associations. *Experimental Brain Research*, 194(3):489–494, 2009.

Wendy Wood and Dennis Rünger. Psychology of habit. *Annual review of psychology*, 67(1):289–314, 2016.

Wendy Wood, Asaf Mazar, and David T Neal. Habits and goals in human behavior: Separate but interacting systems. *Perspectives on Psychological Science*, 17(2):590–605, 2022.

Jialong Wu, Haoyu Ma, Chaoyi Deng, and Mingsheng Long. Pre-training contextualized world models with in-the-wild videos for reinforcement learning. *Advances in Neural Information Processing Systems*, 36:39719–39743, 2023.

Zhen Wu, Jiaman Li, Pei Xu, and C Karen Liu. Human-object interaction from human-level instructions. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11176–11186. IEEE, 2025.

Yiming Xie, Varun Jampani, Lei Zhong, Deqing Sun, and Huaizu Jiang. Omnicontrol: Control any joint at any time for human motion generation. In *International Conference on Learning Representations*, volume 2024, pp. 28176–28194, 2024.

Jingkang Yang, Shuai Liu, Hongming Guo, Yuhao Dong, Xiamengwei Zhang, Sicheng Zhang, Pengyun Wang, Zitang Zhou, Binzhu Xie, Ziyue Wang, et al. Egolife: Towards egocentric life assistant. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28885–28900. IEEE, 2025.

Hang Ye, Xiaoxuan Ma, Fan Lu, Wayne Wu, Kwan-Yee Lin, and Yizhou Wang. Visually-grounded humanoid agents. *arXiv preprint arXiv:2604.08509*, 2026. doi: 10.48550/arXiv.2604.08509. URL https://arxiv.org/abs/2604.08509.

Xiao Yu, Baolin Peng, Ruize Xu, Yelong Shen, Pengcheng He, Suman Nath, Nikhil Singh, Jiangfeng Gao, and Zhou Yu. Reinforcement world model learning for llm-based agents. *arXiv preprint arXiv:2602.05842*, 2026.

Yangbin Yu, Qin Zhang, Junyou Li, Qiang Fu, and Deheng Ye. Affordable generative agents. *Transactions on Machine Learning Research*, 2024.

Hongxin Zhang, Zheyuan Zhang, Zeyuan Wang, Zunzhe Zhang, Lixing Fang, Qinhong Zhou, and Chuang Gan. Ella: Embodied social agents with lifelong memory. *arXiv preprint arXiv:2506.24019*, 2025. doi: 10.48550/arXiv.2506.24019. URL https://arxiv.org/abs/2506.24019.

TianZe Zhang, Sirui Sun, Yuhang Xie, Xin Zhang, Zhiqiang Wu, and Guojie Song. Context-value-action architecture for value-driven large language model agents. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 5049–5073, 2026.

Zirui Zhao, Wee Sun Lee, and David Hsu. Large language models as commonsense knowledge for large-scale task planning. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Qinhong Zhou, Hongxin Zhang, Xiangye Lin, Zheyuan Zhang, Yutian Chen, Wenjun Liu, Zunzhe Zhang, Sunli Chen, Lixing Fang, Qiushi Lyu, et al. Virtual community: An open world for humans, robots, and society. In *International Conference on Learning Representations*, volume 2026, pp. 63971–64010, 2026.

Yude Zou, Junji Gong, Xing Gao, Zixuan Li, Tianxing Chen, and Guanjie Zheng. Infbagel: Human-object-scene interaction generation with dynamic perception and iterative refinement. *arXiv preprint arXiv:2604.04843*, 2026.

Table 3: Quantitative comparison with previous human-like agent frameworks under long-horizon daily simulation.

Method	Cognitive module				Motion generation module					Action type
	Nat. ↑	Coh. ↑	PA ↑	Mean. ↑	SSR ↑	GSR ↑	DTG ↓	Qual. ↑	Act. ↑	
X-VirtualHome (Leng et al., 2026)	45.18	84.62	78.10	69.30	58.11	54.22	**19.73**	1.31	2.91	Atomic
AGA (Yu et al., 2024)	59.42	55.87	70.31	61.87	52.40	28.11	19.81	1.24	1.71	Atomic
ACTOR (Liang et al., 2025)	53.84	87.21	79.05	73.37	61.08	47.15	96.31	3.78	3.62	Motion clips
ACTOR† (Liang et al., 2025)	57.12	87.66	80.13	74.97	64.02	50.21	91.76	3.70	3.73	Motion clips
Ours	**87.19**	**88.34**	**85.01**	**86.85**	**79.11**	**70.24**	21.87	**4.15**	**4.25**	**Generative**

Table 4: **Additional ablation studies.** (a) Comparison of different planning algorithms. (b) Comparison of different LLMs, with GPT-5.5 used by default. (c) Effect of $F(s_t^{\text{psy}})$ under different psychological states with $\rho = 0.5$.

Algorithm	TE ↑	Mean ↑
Greedy	0.69	**89.97**
Beam	0.71	88.23
MCTS	**0.79**	88.19

(a)

LLM	Mean ↑	SSR ↑	GSR ↑
DeepSeek v4.1 Flash	**89.33**	86.47	80.00
GPT-5.5	88.37	88.12	**86.67**
GPT-5.6	87.09	**89.95**	73.33

(b)

Condition	(*str.*, *dep.*, *cog.*)	$F(s_t^{\text{psy}})$	Habit Ratio (%)	Mean ↑
Stress	$(7, 3, 3)$		38.76	76.48
		✓	50.84	87.56
Depletion	$(3, 7, 3)$		31.80	80.37
		✓	52.31	82.90
Cog. Load	$(3, 3, 7)$		39.15	79.39
		✓	56.03	88.04
Normal	$(3, 3, 3)$		35.81	81.92
		✓	40.92	84.34
Ref. from Tab. 2f	$(3, 3, 3)$	$\rho = 0.05$	55.62	77.37

(c)

A Experiment details

A.1 Additional ablation studies

Comparison of Planning Algorithms. Tab. 4a compares different planning algorithms for the planning procedure $\mathcal{P}$. Different planning strategies can instantiate $\mathcal{P}$, including greedy search (Resende & Ribeiro, 2013) and Monte Carlo tree search (MCTS) (Świechowski et al., 2023). MCTS achieves the highest TE, while Greedy obtains the highest Mean score. The three planning algorithms achieve comparable Mean scores, indicating that GEMS is relatively robust to the choice of planning strategy. Meanwhile, the higher TE achieved by MCTS suggests that stronger prospective search can improve goal-directed task execution efficiency.

Comparison of LLMs. Tab. 4b evaluates GEMS with different LLM backbones. DeepSeek achieves the highest Mean score, GPT-5.6 achieves the highest SSR, and GPT-5.5 achieves the highest GSR. The three LLM backbones yield comparable overall performance across the cognitive and motion-generation metrics, suggesting that the effectiveness of GEMS is not specific to a particular LLM backbone. We use GPT-5.5 as the default LLM in our experiments.

Effect of $F(s_t^{\text{psy}})$. Tab. 4c further examines the effect of the psychological state interference function $F(s_t^{\text{psy}})$ under different psychological states. Under each psychological-state initialization, enabling $F(s_t^{\text{psy}})$ yields both a higher habit ratio and a higher Mean score than its ablated counterpart. These paired ablations support the effectiveness of $F(s_t^{\text{psy}})$ in strengthening state-dependent habitual responding and improving evaluated behavioral quality. As a complementary reference, the adverse cognitive-load setting achieves a higher Mean score than the low-ρ setting in Tab. 2f, despite their similar habit ratios of around 55%. This comparison further suggests that overall habit frequency alone does not adequately characterize the quality of generated behavior.

A.2 Implementation details

Implementation setup. We use GPT-5.5 as the default LLM for GEMS and all cognitive baselines, ensuring that all methods are compared under the same LLM backbone. We use GPT-5.6 as the automatic evaluator and DeepSeek V4.1 Flash as the VLM. In addition, we use qwen3-embedding as the embedding model. The default controller-bias coefficient is set to $\rho = 0.5$, and the initial

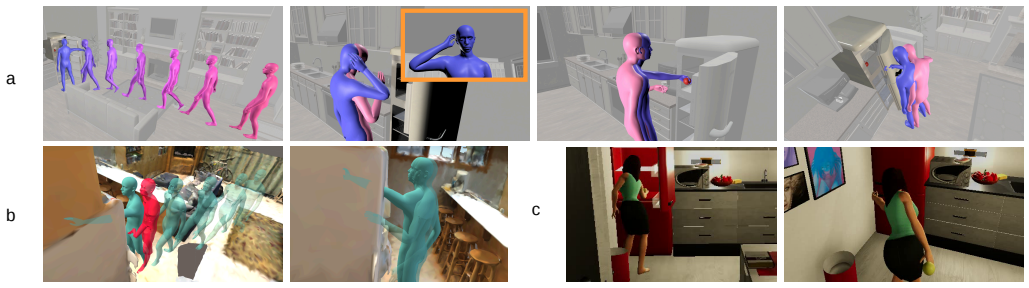


Figure 6: Qualitative comparison under the goal "find something to eat in the fridge." We compare (a) Ours, (b) ACTOR, and (c) X-VirtualHome. While (b)–(c) are largely task-oriented, ours exhibits richer habitual behaviors and appears more human-like. Red frames indicate skeletal drift.

psychological state is $(3, 3, 3)$, corresponding to stress, depletion, and cognitive load, respectively. Action sample window N is set to 3. For the planning procedure $\mathcal{P}$, we use a rollout horizon of 30 steps. For motion generation, we use Kimodo (Rempe et al., 2026) as the T2M model. For ACTOR†, we follow the original ACTOR motion-clip sampling procedure, with the motion clips sampled from motions generated by Kimodo to ensure the same underlying motion-generation capability for comparison.

Controlled-goal evaluation. We first evaluate all methods under a controlled-goal setting. All methods use the same persona, initial state, current time, and recent behavior history. Since different methods may vary in the range of activities they can generate or execute, we construct a shared daily-activity pool based on the semantic overlap among activities supported by different methods, including tasks such as preparing breakfast, cleaning the kitchen, and working from home. We then randomly sample 30 goals from this pool as the common test set for all methods.

For methods that support explicit task specification, including ACTOR, AGA, and X-VirtualHome, we directly provide the corresponding high-level goal. For desire-driven methods such as D2A and ASVO, we set the desire state associated with the target goal such that a semantically aligned behavior is triggered through the method's original decision-making mechanism. Once the target goal is explicitly provided or triggered through the desire state, we record the resulting behavior sequence until goal termination and evaluate only this goal-conditioned segment.

For each goal, each method is independently run five times, resulting in 150 behavior sequences per method. Since different human-like agent frameworks employ different cognitive modules and action representations, their generated behavior sequences may naturally differ in both action content and sequence length, even under the same goal and initialization. For each run, we use the behavior sequence produced by each method's own cognitive module as the target behavior description for its corresponding motion-generation module, and evaluate whether the system can faithfully realize its cognitive decisions as 3D behaviors. Specifically, SSR measures the fraction of successfully executed individual actions in the generated behavior sequence, GSR measures whether the entire generated behavior sequence is successfully completed, DTG measures the minimum wrist-to-goal distance during execution, and Qual. and Act. evaluate motion naturalness and consistency with the target behavior description, respectively.

This evaluation protocol is important because existing human-like agent frameworks differ substantially in their executable action spaces and action representations. For example, VirtualHome-based methods such as X-VirtualHome and AGA rely on predefined atomic action sets, while ACTOR relies on a library of pre-collected motion clips. These methods therefore generate, validate, and execute behaviors within the action spaces natively supported by their respective frameworks. Requiring all methods to execute an identical set of open-ended human-level action instructions would additionally penalize methods that cannot represent such actions due to their original design constraints. Conversely, restricting the evaluation to the intersection of predefined actions supported by all methods would artificially diminish the action-space extensibility that generative motion models are specifically intended to provide.

Long-horizon evaluation. We additionally evaluate the methods under long-horizon daily simulation in Tab. 3. In this setting, we control only the persona and initial state across methods. For

ACTOR and GEMS, which do not include a complete daily-schedule generation module, we construct a shared pool of high-level daily plans generated by the other methods and randomly sample a complete daily plan as input. ACTOR and GEMS then sequentially execute the activities specified in the sampled plan. Each simulation starts from the initial state of the day and continues until all scheduled daily activities are completed.

A.3 METRIC DETAILS.

For the cognitive module, we follow prior LLM-based evaluation protocols and report GPT-based scores on three dimensions. **Naturalness (Nat.)** measures the degree to which the generated behavior resembles real human behavior (Wang et al., 2025). A high Naturalness score indicates that the agent exhibits realistic human behavioral rhythms, rather than overly regular or mechanical patterns. **Coherence (Coh.)** measures the logical and temporal consistency of the behavior sequence (Wang et al., 2025). A high Coherence score indicates that consecutive behaviors follow plausible causal relationships, with intermediate actions and contextual changes forming a convincing and coherent behavioral process. **Personality Alignment (PA)** measures the consistency between the generated behavior and the given persona (Li et al., 2025). A high Personality Alignment score indicates that the agent's behavioral choices consistently reflect persona-specific preferences and tendencies, leading to distinguishable behavioral patterns across different personas. **Mean** denotes the average of the three scores.

LLM-based evaluation protocols have been widely used in generative research as a scalable proxy for human user studies (Cai et al., 2025). The sequence-quality ranking obtained from the human subjective evaluation in Fig. 7 is consistent with the ranking reported above, further supporting the validity of our evaluation protocol.

For the motion generation module, we evaluate both task execution and motion quality. **Step Success Rate (SSR)** measures the percentage of steps for which the agent successfully achieves the corresponding step objective, based on a predefined contact-distance threshold. For example, a lying-down action is considered successful if both the hip and head joints of the human agent are within 30 cm of the target location (Liang et al., 2025; Hassan et al., 2023). **Goal Success Rate (GSR)** measures the percentage of plans in which all steps are successfully executed (Liang et al., 2025). **Distance to Goal (DTG)** measures the minimum wrist-to-goal distance during execution (Diomataris et al., 2024). For motion quality, given the rendered motion and its corresponding language description, human annotators rate each sample on a 1–5 scale in terms of (i) **Quality Score (Qual.)**, which evaluates the naturalness and plausibility of the generated motion in the given scene, and (ii) **Action Score (Act.)**, which evaluates how well the generated motion matches the action specified by the language description (Wang et al., 2022). Higher scores indicate better motion quality and stronger consistency with the scene and language description. Each sample is independently rated by three human annotators, and we report the average score across the three ratings.

B HUMAN STUDY

B.1 HUMAN EVALUATION

Ethics and Participants. The study protocol was reviewed and approved by the ethics committee of our institution. We recruited participants through Amazon Mechanical Turk (MTurk), an online crowdsourcing platform. All participants provided informed consent before taking part in the study. We obtained 53 valid responses for the Cognitive study and 41 valid responses for the 3D study. All respondents reported an age of 18 years or older. We additionally collected demographic information, including age, gender, ethnicity, and education level. All responses were analyzed in de-identified form.

For the Cognitive study, the median reported age group was 25–29 years. Among the 53 respondents, 27 identified as male, 14 as female, 6 as non-binary, and 6 preferred not to disclose their gender. Regarding education, 24 respondents reported a postgraduate degree, 17 an undergraduate degree, 5 an A-Level, vocational, or technical qualification, and 3 a secondary-school or GCSE-level qualification; 4 preferred not to disclose their education level. In terms of ethnicity, 26 respondents identified as White, 9 as Black, Black British, Caribbean, or African, 8 as Asian or Asian British, 6

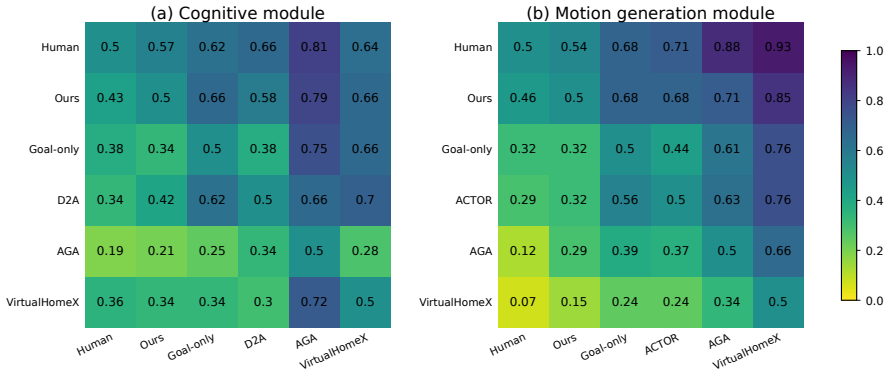


Figure 7: Human preference results from the subjective evaluation. The figure shows pairwise human-likeness comparisons for the Cognitive Behavior Simulation and 3D Motion Generation settings. Our full multiple controller system achieves preference scores close to real human behavior and consistently outperforms the Goal-only variant.

as Mixed or Multiple ethnic groups, and 2 as another ethnic group; 2 preferred not to disclose their ethnicity.

For the 3D study, the median reported age group was also 25–29 years. Among the 41 respondents, 24 identified as male, 13 as female, 3 as non-binary, and 1 preferred not to disclose their gender. Regarding education, 17 respondents reported an undergraduate degree, 14 a postgraduate degree, 6 an A-Level, vocational, or technical qualification, and 1 a secondary-school or GCSE-level qualification; 3 preferred not to disclose their education level. In terms of ethnicity, 18 respondents identified as White, 10 as Black, Black British, Caribbean, or African, 6 as Asian or Asian British, 6 as Mixed or Multiple ethnic groups, and 1 as another ethnic group.

Evaluation Protocol. We conduct two complementary human studies corresponding to the two stages of our evaluation. The Cognitive setting evaluates the perceived human-likeness of behavioral sequences, while the 3D setting evaluates the perceived human-likeness of generated motion videos. In the Cognitive setting, we compare GEMS, Goal-only, D2A, AGA, X-VirtualHome, and real human behavior. In the 3D setting, we compare GEMS, Goal-only, ACTOR, AGA, X-VirtualHome, and real human behavior. Goal-only is an ablated variant of GEMS that retains only the Goal-directed Controller. Human references are selected from EgoLife (Yang et al., 2025) and TRUMANS (Jiang et al., 2024).

For each setting, we perform all pairwise comparisons among the six behavior sources, resulting in 15 method pairs. For each method pair, we randomly sample 10 runs to construct the comparison pool. Method identities are hidden from participants. For each method pair, participants select the behavior they perceive as more human-like. Each valid response contains one judgment for each of the 15 method pairs.

For each behavior source, we select 10 behavioral samples for evaluation, resulting in 60 samples for each setting. Each questionnaire presents samples from the six behavior sources and evaluates all pairwise combinations among them. This produces $\binom{6}{2} = 15$ pairwise comparisons per questionnaire. The order and identities of the methods are hidden from participants, and participants are asked to select which of the two presented behaviors appears more human-like. Thus, each valid response provides one judgment for each of the 15 method pairs.

For each method pair, we compute the pairwise preference rate as the fraction of participants preferring each method. We further compute the overall preference rate of each method by aggregating its wins across its five pairwise comparisons with the other behavior sources.

Effect of Habitual Control. We assess the contribution of habitual control through a direct comparison between GEMS and Goal-only. In the Cognitive setting, GEMS is preferred by 35 of 53 participants (66.0%). A two-sided exact binomial test against the null hypothesis of equal prefer-

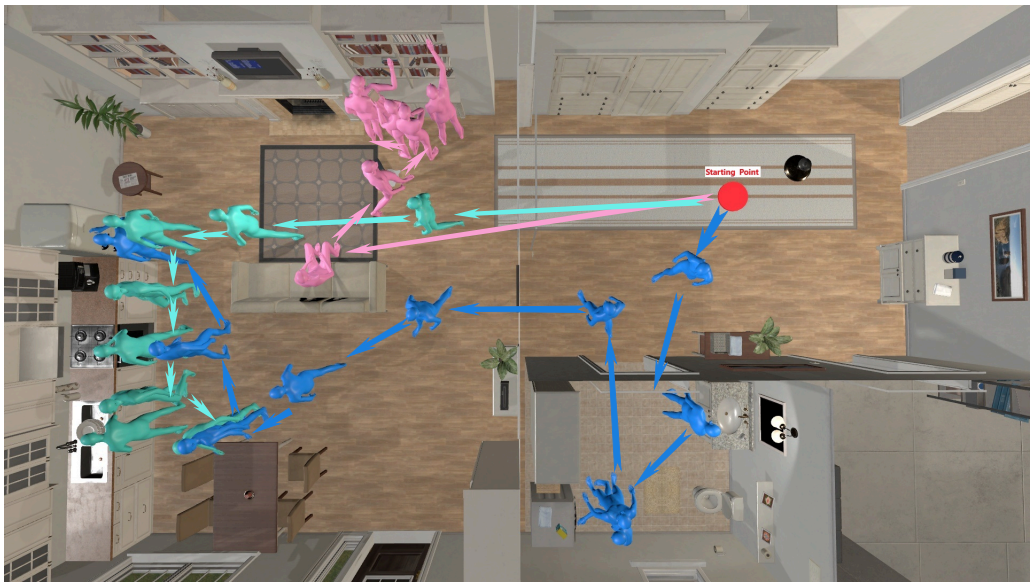


Figure 8: Diverse behavioral trajectories generated under the same goal. All three agents are assigned the goal of *"deep clean the house"* and are initialized at the same spatial location, while differing in their personas, habit memories, and psychological states. The spatial trajectories correspond to Adrian (colored green with $\rho = 0.5$), Elena (colored blue with $\rho = 0.6$), and Michael (colored pink with $\rho = 0.4$). For visual clarity, only the first six actions of each agent and their corresponding key action frames are shown. The examples also illustrate cooperation and competition between habitual and goal-directed control.

ence yields $p = 0.027$. In the 3D setting, GEMS is preferred by 28 of 41 participants (68.3%), with $p = 0.028$ under the same test. These results show a consistent preference for GEMS over its Goal-only ablation in both behavioral sequences and generated 3D motions, supporting the contribution of habitual control to perceived human-likeness.

Comparison with Real Human Behavior. GEMS receives preference rates of 43.4% and 46.3% when directly compared with human references.

In the Cognitive setting, Human achieves an overall preference rate of 66.0%, followed by GEMS at 62.6%, D2A at 54.7%, Goal-only at 50.2%, X-VirtualHome at 41.1%, and AGA at 25.3%. In the 3D setting, the corresponding rates are 74.6% for Human, 67.8% for GEMS, 51.2% for ACTOR, 48.8% for Goal-only, 36.6% for AGA, and 21.0% for X-VirtualHome. Among the simulated methods evaluated in each setting, GEMS achieves the highest overall preference rate. Together, these results indicate that GEMS can generate behaviors that are perceived as believable and human-like by participants in both behavioral-sequence and 3D-motion evaluations.

Consistency with Quantitative Evaluation. We further examine whether the human evaluation is consistent with the quantitative evaluation metrics. In the Cognitive setting, the human-preference ordering of GEMS, D2A, X-VirtualHome, and AGA exactly matches the ordering obtained using the Mean score, yielding a Spearman rank correlation of $r_s = 1.0$.

In the 3D setting, the human-preference ordering of GEMS, ACTOR, AGA, and X-VirtualHome is GEMS $>$ ACTOR $>$ AGA $>$ X-VirtualHome, and shows strong rank consistency with SSR, Qual., and Act., with $r_s = 0.8$. These results indicate that the quantitative metrics are broadly aligned with human judgments of behavioral and motion human-likeness.

C CASE STUDY: DIVERSE BEHAVIORAL PATTERNS.

We further conduct a case study to demonstrate how the proposed method generates diverse behavioral patterns under the same high-level goal. As shown in Fig. 8, three agents are assigned the same goal, *"deep clean the house"*, while having different personas, habitual associations, and internal psychological states. For visualization, we only show the first six actions of each agent and their corresponding execution trajectories.

Adrian (Green). Adrian has $\rho = 0.5$, and his initial psychological state is set to 3. At the beginning of the task, he notices the refrigerator in his visual field and chooses to check it. This action is supported by both the habitual and goal-directed systems. Checking the refrigerator is relevant to the current cleaning goal and therefore receives a relatively high action value. Meanwhile, the cleaning context also activates one of Adrian's habitual behaviors. He then cleans the stove, which is mainly goal-directed. After completing this action, he adjusts his glasses as a habitual behavior. He next wipes the dishes. Since his hands become wet after this action, the resulting cue triggers another habit, causing him to casually fix his hair. He then returns to the cleaning task and wipes the table.

Elena (Blue). Elena has $\rho = 0.6$, and her initial psychological state is also set to 3. She exhibits a different action sequence. She first cleans the sink and then hangs up the towel. Both actions directly contribute to the current cleaning goal. As her internal Dep. state increases, she stretches her shoulders, which is selected as a habitual action. She then wipes the table and finally checks the refrigerator, organizes it. Compared with Adrian, Elena shows a more continuous goal-directed trajectory during the first six actions, with fewer interruptions caused by habitual behaviors.

Michael (Pink). Michael has $\rho = 0.4$, while his initial Dep. state is set to 5. This higher internal state leads to a substantially different trajectory. Although the same cleaning goal is active, he first goes to the sofa and sits down for a short rest. This habitual action is triggered by his adverse Dep. state. He then gets up and starts organizing the bookshelf. During this task, he notices a book and starts reading it, as the book acts as a contextual cue for one of his habitual behaviors. Reading further increases his cognitive load, which subsequently triggers another habitual action, touching his nose. Michael then resumes the task and continues organizing the bookshelf. Finally, the adverse Dep. state triggers another habitual response, and he sighs.

These examples show that GEMS can generate substantially different behavioral patterns. The case study also illustrates the cooperation and competition between the habitual and goal-directed systems. Adrian's first action provides an example of cooperation. Checking the refrigerator contributes to the cleaning goal and therefore receives support from the goal-directed controller. At the same time, the cleaning context activates the same action through Adrian's habitual controller. Both controllers therefore support the same behavior, increasing its likelihood of being selected. Michael's first action provides an example of competition. Sitting on the sofa provides little immediate contribution to the cleaning goal and therefore has a relatively low goal-directed value. However, Michael's high Dep. state strongly activates the corresponding habitual response. The habitual controller consequently exerts a stronger influence on action selection at this moment, leading Michael to rest before continuing the task. Afterward, the goal-directed system drives him back to organizing the bookshelf.

D ADDITIONAL DISCUSSIONS

D.1 HUMAN-LIKE LLM AGENT

Human daily activity modeling supports a wide range of applications, from understanding everyday decision-making in psychology (Gärling & Garvill, 1993) and social interaction and structures in sociology (Adler et al., 1987), to creating realistic virtual humans for metaverse and VR applications (Schmidt et al., 2024; Fraser et al., 2024). However, collecting real-world behavioral data is costly and raises privacy concerns. Recent advances in LLMs have therefore spurred growing interest in human behavior simulation. The pioneering Generative Agents (Park et al., 2023) introduced memory, reflection, and planning to simulate believable human behaviors. Building on

this, Humanoid Agents (Wang et al., 2023) incorporated desire-driven processes, while AFSPP (He & Zhang, 2024) introduced agent preferences and personalities to capture individual differences. More recent works model action selection from different psychological perspectives: CVA (Zhang et al., 2026) uses a human-data-trained Value Verifier, D2A (Wang et al., 2025) adopts desire-driven behavior based on the Theory of Needs, and ASVO (Lin et al., 2026) incorporates Social Value Orientation for multi-agent interactions. For 3D realization, X-VirtualHome (Leng et al., 2026) generates persona-conditioned routines with LLM-based planning, while AGA (Yu et al., 2024) grounds high-level instructions into VirtualHome's predefined action space. X's Day (Li et al., 2025) models personality-driven long-term behaviors, Visually-grounded Humanoid Agents (Ye et al., 2026) grounds LLM-driven behaviors through visual perception and spatial planning, and Ella (Zhang et al., 2025) introduces long-term multimodal memory for lifelong embodied behavior. Most of these frameworks realize high-level behaviors through predefined atomic actions. In contrast, ACTOR (Liang et al., 2025) matches planned behaviors with pre-collected motion clips and uses Transformer-based motion completion to connect them, which can introduce spatial misalignment and skeleton drift during transitions. Note that existing methods are predominantly driven by explicit plans, desires, or events, resulting in a goal-directed behavior generation paradigm. Habitual behaviors, however, are often cue-triggered, non-optimal, and highly individualized, accounting for a major portion of everyday human behavior. Modeling habits can therefore enable more natural and personalized agents and benefit human behavior understanding and recognition.

D.2 Human behavior control study

Goal-directed and habitual control have long been viewed as two ends of a spectrum of human behavioral control (Dolan & Dayan, 2013). The same action may arise from goal-directed control, habitual control, or their joint influence (Miller et al., 2019; Wood et al., 2022), making the underlying source of control difficult to infer from overt behavior alone. Moreover, habits are often executed automatically with limited deliberation, while their underlying triggers may escape conscious awareness, leading their influence on behavior to be underestimated (Wood & Rünger, 2016). Experimental studies commonly distinguish the two forms of control using outcome devaluation, which tests sensitivity to current outcome value (Dolan & Dayan, 2013). In cognitive neuroscience, human behavioral control is often modeled through reinforcement learning: goal-directed control is associated with model-based planning, whereas habitual control is associated with model-free learning (Daw et al., 2005; Dolan & Dayan, 2013; Wood & Rünger, 2016). Early dual-system models emphasized competition between the two controllers, while later work introduced arbitration mechanisms that dynamically balance their influence based on relative reliability (Daw et al., 2005; Lee et al., 2014). Related cognitive architectures exploit proceduralized responses for efficient action selection: ACT-R enables perceptual cues to trigger productions after practice (Taatgen et al., 2008), Soar represents procedural knowledge with production rules (Laird et al., 1987), and CoALA extends these principles to language agents (Sumers et al., 2024). More recent accounts question the direct equivalence between habits and model-free RL, instead linking habitual behavior to habit strength shaped through repetition (Miller et al., 2019). However, existing studies mainly focus on habit formation and control allocation in constrained tasks, leaving contextual habit triggering and state-dependent shifts in realistic environments underexplored (Wood & Rünger, 2016; Neal et al., 2013). In contrast, we study multiple behavioral control system for human-like daily behavior simulation in realistic 3D environments.

D.3 Underestimation of Habitual Influences

Habitual influences on behavior may be systematically underestimated because people have limited introspective access to the mechanisms that trigger habits. Wood & Rünger (2016) argue that people are often aware of performing a habitual response, yet remain largely unaware of the contextual cues that activate it. As a result, explanations for habitual behavior can become largely post hoc: people observe what they repeatedly do and infer the goals or intentions that might have caused it. Indeed, people with stronger habits reported greater certainty about their intentions and perceived their behavior as more goal guided, even though intentions were actually poorer predictors of strongly habitual behavior. This tendency may be strengthened by the intuitive appeal of goal-based explanations. Wood et al. (2022) note that people spontaneously infer goals behind behavior and generally favor goal explanations over alternative causes. Their coffee-drinking study provides a clear example. Although habit and fatigue had comparably sized effects on actual coffee consumption, participants

attributed their coffee drinking much more strongly to fatigue than to habit. This underestimation of habit persisted even when participants were financially incentivized to make accurate judgments. Together, these findings suggest that people may experience habitual behavior through a goal-based narrative: because the automatic context–response process is relatively inaccessible to introspection, people infer a plausible motive after the behavior occurs. Consequently, they may overestimate the role of goal pursuit while underestimating the contribution of habit to everyday action.

D.4 Overall discussion

This work aims to develop a human control mechanism inspired framework for generating more believable human-like behavior. Believability relies on sustaining an "illusion of life," in which an agent exhibits stable, natural, and individually distinctive behavioral patterns (Park et al., 2023; Bates et al., 1994). To support this goal, we draw on research on goal-directed and habitual behavior to derive a multiple control system design.

Human action control is commonly described as involving interacting goal-directed and habitual systems. These systems rely on partially distinct psychological and neural mechanisms while jointly shaping behavior. Goal-directed control is more strongly associated with prefrontal and anterior striatal regions, whereas habitual control is linked to posterior striatal regions, particularly the posterior putamen (Daw et al., 2005). Computationally, goal-directed control is often mapped onto model-based control, while habitual control is mapped onto model-free control. Model-based control uses an internal model of the environment to predict future states and outcomes, whereas model-free control relies on cached action values learned through experience. Daw et al. (2005) further proposed an arbitration mechanism that dynamically allocates control according to the relative reliability of the two systems. This motivates our architecture consisting of a goal controller, a habit controller, and an arbiter. For the habit controller, we further follow value-free account of habit. Standard model-free reinforcement learning still represents actions through expected future reward. Miller et al. (2019) instead argue that habits can arise from repeated actions without explicitly encoding their outcome values. Accordingly, our framework represents habitual tendencies through habit strength rather than model-free cached value.

Classical RL models nevertheless remain difficult to apply directly to open-world embodied environments, where states and actions are not fixed and are often represented through natural language. We therefore extend the multiple system architecture to an LLM-based agent and incorporate four additional principles from the psychology of habit:

Context-driven habit retrieval. Habits can be rapidly triggered by familiar contextual cues and can be activated independently of current goals (Wood & Rünger, 2016). We therefore explicitly retrieve habitual actions from the current context. *Asymmetric interaction between habits and goals.* Habitual responses are activated quickly and can narrow the space of candidate actions, whereas goal-directed control requires additional evaluation of possible outcomes (Wood & Rünger, 2016). We therefore allow the habit controller to propose candidate actions that are subsequently evaluated by the goal controller. *Dynamic modulation by psychological state.* Time pressure, distraction, and cognitive load can weaken goal-directed control and increase reliance on habits. Habit slips illustrate this imbalance, as previously learned responses may remain active even after current goals have changed (Wood et al., 2022). We abstract these factors into a current psychological state, which dynamically modulates the relative influence of the two controllers. *Stable individual differences in control bias.* Individuals differ in their tendency to rely on goal-directed versus habitual control. We represent this stable variation using an individual control bias ρ, allowing different agents to exhibit distinct control tendencies under the same environmental and psychological conditions (Lee et al., 2014).

In summary, this work aims to bring a multiple control perspective to the human-like agent community. We believe we provide useful insights for future research on believable and human-like agent behavior.

E HABIT CONTROLLER

E.1 HABIT DATASET CONSTRUCTION

We construct the Habit Dataset to address the difficulty of distinguishing habitual behavior from goal-directed behavior based solely on observation, as discussed in D.3. People often have limited awareness of the automatic mechanisms that trigger habitual behavior and tend to attribute repeated actions to goals or intentions. As a result, inferring habits directly from observed daily behavior can lead to attribution ambiguity. To reduce this ambiguity, we collect behavioral instances from empirical sources that explicitly investigate habitual behavior, including habit scales, behavioral studies, and publicly available datasets, and further extract their associated context cues. This design provides clearer empirical evidence that the collected behaviors are habit-related and improves the controllability and reliability of habit influence in the subsequent generation of daily human behavior.

Context Cue Representation. Potentially, any internal or external event may trigger habitual behavior (Gardner et al., 2016). Since the full space of possible cues cannot be exhaustively represented, we consider four cue dimensions that are both empirically supported and available in the agent's current context: visual episodes (Hommel, 1998), psychological states (Waszak & Pholulamdeth, 2009), events (Qiu et al., 2023), and active goals (Schneider & Logan, 2009). Each cue is represented in natural language and assigned to one of these dimensions. Although a habitual response may depend on multiple contextual factors, we use one primary cue dimension for each habit instance to obtain a simple and interpretable representation.

Habit Dataset Overview. We construct a dataset of daily habitual behaviors and their associated context cues from existing empirical evidence, including habit scales, behavioral studies, and publicly available habit-related datasets (Ersche et al., 2017; Moors & De Houwer, 2006; Georgiev et al., 2022; Li et al., 2024; Pang et al., 2022; Azrin & Nunn, 1973; Wood & Rünger, 2016). Each instance is represented as a cue–action pair $\langle q_i^{cue}, a_i^H \rangle$. After extraction, standardization, deduplication, and manual validation, the resulting dataset contains 571 habitual behavior instances. The following paragraphs describe the construction procedure in detail.

Data Collection. We collect raw habitual behavior data from existing sources of empirical evidence, including habit scales, empirical psychological studies, and publicly available habitual behavior datasets. All included sources explicitly describe or study habitual behaviors in daily life and provide contextual information about the cues under which these habits are expressed or triggered.

The collected behaviors mainly cover two categories considered in our setting: object-interaction habits and self-touch habits. Object-interaction habits include, for example, eating a visible snack, drinking a beverage under a particular psychological state, or checking a watch while waiting. Self-touch habits include behaviors such as scratching one's head under high cognitive load or wiping one's mouth after drinking.

Different sources provide information at different levels of granularity, including habitual actions, behavioral contexts, experimental descriptions, and video annotations. We retain the original behavior descriptions, contextual information, and source information for traceability.

LLM-based Information Extraction. The raw source materials come from multiple sources and are presented in heterogeneous forms, including paper text, tables, and dataset annotations. We therefore use an LLM to extract the habitual action, context cue, and corresponding source evidence, and convert them into a unified structured representation.

Example – Raw extraction

```
Source text:  ''anxiety increases the frequency of
self-touch''
Context cue:  feeling anxious
Habitual action:  self-touch
```

Standardization and Cleaning. We map each extracted cue to one of the four predefined cue dimensions used in our framework. These cue dimensions have been extensively studied as contex-

tual factors associated with habitual or stimulus–response behavior (Gardner et al., 2016; Wood & Rünger, 2016), including **visual episodes** (Hommel, 1998), such as observing that a glass is empty and being cued to refill it; **psychological states** (Waszak & Pholulamdeth, 2009), such as experiencing high cognitive load and being cued to touch one's head; **events** (Qiu et al., 2023), such as finishing a drink and being cued to wipe one's mouth; and **active goals** (Schneider & Logan, 2009), such as having the current goal of cleaning the kitchen and being cued to take out the trash. Both context cues and habitual actions are then normalized into natural-language descriptions. For instances with missing cues, annotators inspect the original paper, video content, dataset description, and related evidence to determine whether a supported triggering cue can be identified. Instances without reliable cue evidence are removed. We also remove instances whose cues cannot be reliably mapped to one of the four predefined cue dimensions. We further apply missing-field filtering, language normalization, and semantic deduplication to obtain standardized context cue–habitual action pairs.

Our primary objective is to study whether introducing habitual control improves the human-likeness of generated behavior. We therefore adopt a simplified representation in which each habit instance retains one primary cue dimension. Despite this simplification, the resulting framework produces clear improvements in human-likeness in our experiments, supporting the effectiveness of the proposed framework. Modeling compound context cues across multiple dimensions is left to future work for finer-grained habit triggering.

Example – Standardized pair

```
Cue dimension:  psychological state
Context cue:  ``The person feels anxious.''
Habitual action:  ``Engage in self-touch.''
```

Quality Validation. Because the processed dataset is moderate in size, we manually validate all processed instances and remove inappropriate or insufficiently supported instances. The inspection considers three aspects: (1) whether the habitual action is supported by the original habit-related source; (2) whether the extracted cue–action association is consistent with the source evidence; and (3) whether the context cue is assigned to the correct cue dimension.

After this process, we obtain 571 valid context cue–habitual action instances and use them to construct the global habit dataset $\mathcal{D}^H$.

Data examples

```
ContextCue(visual_episode = ``A glass of water is visible.'')
habitual_action = ``Drink water.''

ContextCue(psychological_state = ``The person feels
anxious.'')
habitual_action = ``Engage in self-touch.''
```

E.2 Personalized Habit Memory Initialization

Habit memory is inherently dynamic. A habit may gradually weaken or even be forgotten due to interference from external factors, while others may be further reinforced through long-term repetition. However, this work focuses on fine-grained human behavior simulation over relatively short time horizons, typically within a single day. Therefore, habit memory can be reasonably assumed to remain stable over the simulated time period.

With a convincing Habit Dataset, we can provide agents with daily habitual behaviors that are observed in real human life. This reduces the risk that our framework confuses habitual behavior with goal-directed behavior during generation. Specifically, we sample a subset of habitual behaviors from the Habit Dataset for each agent and assign a habit strength to each behavior.

It is important to clarify that habits are typically acquired gradually through repeated practice over an extended period of time. Habit formation is therefore a long-term learning process, and this process has already been extensively studied in existing computational frameworks (Miller et al.,

2019). Our focus is different. We investigate the role of habitual behavior as a behavioral driver in human-like behavior generation, rather than re-modeling the full process of habit formation. Therefore, we use an LLM to assign habitual behaviors and their corresponding strengths to each agent as an approximation of an already established habit state. This design is sufficient for studying whether incorporating habitual behavior can improve the believability and human-likeness of generated behavior. The effectiveness of this design is further supported by our experimental results.

Persona Construction. To model long-term individual differences, we construct a structured persona for each agent, including demographic attributes, personality, occupation, and long-term background. We also initialize the agent's momentary psychological states, including stress, depletion, and cognitive load, which are subsequently used for behavioral control.

Example persona

Michael Anderson (Michael), 37, male.
Personality: Gentle, responsible, cautious, pragmatic, stability-oriented, conflict-avoidant, and mildly anxious; sensitive to risk and criticism.
Occupation: Project Coordinator in the City Planning Department of Portland, Oregon, responsible for community communication, resident feedback, and cross-department coordination.
Initial Internal State: Stress 3/10; Depletion 5/10; Cognitive Load 3/10.
Background: Raised in a middle-class family near Portland and trained in geography and urban studies. He values family security, health, savings, and long-term stability, and generally prefers reliable, low-risk choices.

Persona-conditioned Habit Sampling.

Given an agent persona, we construct its personalized habit memory from the global habit dataset $\mathcal{D}^H$. Since $\mathcal{D}^H$ contains 571 candidate context cue–habitual action pairs, we directly provide the agent's long-term persona together with all candidate pairs to an LLM. The LLM selects 30 habit instances for each agent. All agents use the same habit memory capacity.

We use the following prompt:

You will be given a persona profile and a set of candidate context cue–habitual action pairs. Select the habits that are most appropriate for the long-term characteristics of this persona.
When selecting habits, consider the following three aspects:
1. Consistency: The selected habits should be consistent with the persona's personality, experience, lifestyle, and long-term behavioral tendencies, and should not clearly conflict with the persona description.
2. Plausibility: The selected behaviors should be habits that the persona could plausibly have formed through repeated experience over time. Do not assume that a persona has a habit simply because a context cue resembles its current state.
3. Diversity: The selected habits should cover different types of daily behaviors and context cues, while avoiding semantically redundant behaviors or excessive concentration in the same type of situation.
Select the most suitable habits from the candidate set according to the above criteria. Return only the IDs of the selected candidates. Do not modify the context cue or habitual action, and do not generate habits that are not present in the candidate set.

Habit Strength Assignment. For the 30 selected habit instances, we further assign an agent-specific habit strength $H(a_i^H) \in [0, 1]$ to each habitual action a_i^H. Habit strength is a dimensionless normalized scalar that represents the relative strength of the context–response association. We condition this estimation on the agent's long-term personality, experience, and living environment.

We use the following prompt:

Table 5: Comparison between LLM-based and human-annotated habit initialization for Alice. Mean denotes the average of Nat., Coh., and PA.

Habit initialization	Nat. ↑	Coh. ↑	PA ↑	Mean ↑
LLM-based	87.5	88.2	85.8	87.2
Human-annotated	86.3	89.1	87.7	87.7

Given a persona's personality, long-term experience, and living environment, together with a set of context cue–habitual action pairs, estimate a habit strength $H(a_i^H)$ between 0 and 1 for each habit.
Habit strength represents the relative strength of the context–response association. Consider the following factors:
1. Repetition: How likely the behavior is to have been repeatedly performed in the persona's long-term daily life.
2. Context Stability: How consistently and repeatedly the corresponding context cue is likely to occur in the persona's life.
3. Long-term Consistency: How consistent the habit is with the persona's personality, experience, occupation, lifestyle, and long-term behavioral tendencies.
Use the following scale:
0.0–0.4: very weak association;
0.4–0.7: moderate association;
0.7–0.9: stable habit;
0.9–1.0: strong habit.

After habit sampling and strength assignment, we obtain the personalized habit memory $\widehat{\mathcal{D}}^H$ for each agent.

Habit memory examples

```
HabitMemoryItem(
context_cue = ContextCue(
visual_episode = ''A glass of water is visible.''
),
habitual_action = ''Drink water.'',
habit_strength = 0.80
)

HabitMemoryItem(
context_cue = ContextCue(
psychological_state = ''The person feels anxious.''
),
habitual_action = ''Engage in self-touch.'',
habit_strength = 0.92
)
```

To further validate the LLM-based habit initialization scheme, we additionally construct a manually designed agent, Alice. Three volunteers jointly determine Alice's habit memory and the corresponding habit strengths based on her persona, which we use as the Human-annotated initialization. We compare this initialization with the LLM-based scheme adopted in our framework. As shown in Tab. 5, the two initialization strategies achieve comparable performance in Naturalness, Coherence, and Personality Alignment. These results indicate that the LLM-based habit initialization scheme is a reasonable approach for personalized habit assignment and habit-strength estimation.

E.3 Habit Retrieval Evaluation Set

To evaluate the cue retrieval mechanism of the Habitual Controller, we construct an agent-specific retrieval evaluation set based on each personalized habit memory $\hat{D}^H$. Each agent contains 30 habit

memories. Each memory consists of a context cue q_i^{cue}, a habitual action a_i^H, and its habit strength $H(a_i^H)$. Since different agents have different personalized habit memories, we construct the retrieval evaluation set independently for each agent.

For each habit, we construct four types of retrieval cases. Each case has one ground-truth retrieval label, which can be either a target habitual action or $\emptyset$, indicating that no habitual action should be retrieved. *Direct Match* provides a context that directly matches the cue of the target habit. It tests whether the retrieval algorithm can identify the correct habit under a clear triggering context. *Boundary Reversal* changes a key contextual attribute, such as object state, event timing, or subject identity. This creates a more confusing context and tests whether the algorithm can use the key contextual information to identify the target habit. *Partial Transfer* preserves part of the semantic or categorical similarity while changing the location, object, state, intensity, or event source. It tests whether the algorithm can distinguish the target habit from other similar habits under partial contextual changes. *Hard Negative* introduces strong distractors that are semantically similar to existing habit cues or contain misleading contextual information, while not satisfying the actual triggering conditions of any habit. Examples include objects shown only in photographs, descriptions of psychological states, or objects with similar but different states. These cases test whether the retrieval algorithm can correctly reject misleading cues and avoid triggering any habitual action under highly confusing contexts.

For each retrieval case, the ground-truth label is the corresponding target habitual action, or $\emptyset$ for a hard-negative case. A retrieval trial is considered correct if its output matches the ground-truth label; otherwise, it is considered incorrect. Therefore, the evaluation directly measures whether the retrieval algorithm can identify the correct habitual behavior or correctly return no habitual action under different types of contextual variation and interference.

This process produces $30 \times 4 = 120$ base retrieval cases for each agent. Among these 120 base retrieval cases, we deliberately construct a subset of challenging cases in which multiple candidate habits have high cue compatibility with the current context. Some cases are also designed such that cue compatibility and habit strength favor different candidate habits. These challenging constructions test whether the retrieval algorithm is affected by strong distractors, high habit strength, or superficial semantic similarity.

To evaluate robustness to linguistic variation, we generate four linguistic variants for each base case. These variants modify the wording, temporal descriptions, and irrelevant background information while keeping the same ground-truth retrieval label. Each agent therefore contains $120 \times 4 = 480$ retrieval cases. We independently construct the evaluation sets for three agents, resulting in 1,440 retrieval cases in total. We randomly inspect 10% of the generated cases to verify the context descriptions, case types, and ground-truth retrieval labels.

For evaluation, we split the 120 base cases of each agent into validation and test sets using a 60%/40% split. All four linguistic variants derived from the same base case are assigned to the same split, resulting in 288 validation cases and 192 test cases per agent. The validation set is used to select the retrieval threshold κ, while the test set is used only for reporting the final retrieval performance.

E.4 Retrieval Algorithm Evaluation

Habit retrieval and selection are mainly influenced by two factors: the compatibility between the current context and the stored context cue, and the strength of the habit itself. We use cue similarity to characterize context–cue compatibility, which measures how well the current context matches the stored cue of a habit. Habit strength characterizes the strength of the stimulus–response (S–R) association. A stronger S–R association indicates a higher tendency for the corresponding habitual response to occur and guide behavior. Based on these two factors, we compare three retrieval strategies to examine their different roles in habit retrieval and selection.

We first introduce *Similarity-only* as a baseline. This method performs both retrieval and selection solely according to cue similarity, allowing us to isolate the contribution of habit strength. We then consider a direct combination of the two factors, denoted as *Similarity* $\times$ *Habit Strength*, which multiplies cue similarity by habit strength before retrieval. In this formulation, cue compatibility and habit strength jointly affect habit activation from the beginning. However, a high habit strength

may compensate for relatively weak cue compatibility, while a low habit strength may prevent a well-matched cue from being retrieved.

Motivated by the observation that habitual responses are triggered by recurring context cues, while stronger habits are more likely to guide behavior (Wood & Rünger, 2016), we further introduce *Two-stage Cue Retrieval*. The first stage uses only cue similarity to identify habits that are sufficiently compatible with the current context and constructs the set of activated habits. The second stage incorporates habit strength to weight and select among these activated habits. This design separates context–cue matching from the effect of habit strength. Habit strength therefore does not interfere with the initial cue matching, while stronger habits remain more likely to be selected after activation.

We next describe the three retrieval strategies in detail. Given the current context s_t and the personalized habit memory $\hat{D}^H$, the retrieval algorithm compares the current context with the stored context cues and returns either a corresponding habitual action $a_{i,t}^H$ or no habitual action if no habit is retrieved.

We consider four types of context cues for habit triggering: visual episodes, psychological states, events, and active goals. Each stored context cue q_i^{cue} belongs to one of these cue dimensions. For habit i, we use $c_{i,t}$ to denote the information extracted from the current context s_t that belongs to the same cue dimension as q_i^{cue}. For example, if q_i^{cue} corresponds to a psychological state, $c_{i,t}$ represents the psychological-state information in the current context.

We encode $c_{i,t}$ and q_i^{cue} into the same embedding space, obtaining vector representations $\mathbf{c}_{i,t}$ and $\mathbf{q}_i^{\text{cue}}$. Their cue similarity is measured using cosine similarity, i.e., $C_i = \frac{\mathbf{c}_{i,t}^\top \mathbf{q}_i^{\text{cue}}}{\|\mathbf{c}_{i,t}\|_2 \|\mathbf{q}_i^{\text{cue}}\|_2}$. A larger C_i indicates higher compatibility between the current context and the stored cue of habit i. All three retrieval strategies use C_i, but differ in whether and when habit strength $H(a_i^H)$ contributes to retrieval and selection.

Similarity-only. This baseline relies solely on cue similarity for both retrieval and selection. We retain all habits whose cue similarity is at least the threshold κ, forming the activated set $\mathcal{E}_t = \{i \mid C_i \geq \kappa\}$. If $\mathcal{E}_t = \emptyset$, no habitual action is retrieved. For each habit $i \in \mathcal{E}_t$, we compute its response score as $p_i = \frac{C_i - \kappa}{1-\kappa}$. The response scores are normalized into selection probabilities, i.e., $P(a_{i,t}^H \mid s_t, \mathcal{E}_t) = \frac{p_i}{\sum_{j \in \mathcal{E}_t} p_j}$. A habitual action is then sampled according to this distribution. This method does not use habit strength during either retrieval or selection.

Similarity $\times$ Habit Strength. This baseline directly combines cue similarity and habit strength before retrieval. For each habit i, we first compute the joint score $S_i = C_i H(a_i^H)$. We then construct the activated set as $\mathcal{E}_t = \{i \mid S_i \geq \kappa\}$. If $\mathcal{E}_t = \emptyset$, no habitual action is retrieved. For each $i \in \mathcal{E}_t$, the response score is $p_i = \frac{S_i - \kappa}{1-\kappa}$, which is normalized as $P(a_{i,t}^H \mid s_t, \mathcal{E}_t) = \frac{p_i}{\sum_{j \in \mathcal{E}_t} p_j}$. A habitual action is then sampled according to this distribution. In this strategy, habit strength directly contributes to the retrieval score and therefore also affects whether a habit enters the activated set.

Two-stage Cue Retrieval (Ours). Our method separates cue matching from the effect of habit strength. In the first stage, we use only cue similarity to construct the activated set: $\mathcal{E}_t = \{i \mid C_i \geq \kappa\}$. If $\mathcal{E}_t = \emptyset$, no habitual action is retrieved. In the second stage, habit strength is introduced only for habits that have already entered $\mathcal{E}_t$. For each $i \in \mathcal{E}_t$, we compute $p_i = \frac{C_i - \kappa}{1-\kappa} H(a_i^H)$, and normalize the response scores as $P(a_{i,t}^H \mid s_t, \mathcal{E}_t) = \frac{p_i}{\sum_{j \in \mathcal{E}_t} p_j}$. A habitual action is then sampled according to this distribution. This two-stage design prevents habit strength from interfering with context–cue matching, while still allowing habit strength to influence the subsequent action selection. As a result, habits with stronger S–R associations have a higher tendency to be selected once they have been activated.

Retrieval Accuracy. We evaluate the three retrieval strategies using the agent-specific validation and test sets described above. Each retrieval case has one ground-truth label, which can be either a habitual action or no habitual action. For agent j, let $a_{j,n}^{H,*}$ denote the ground-truth label of the n-th retrieval case, where $a_{j,n}^{H,*} = \emptyset$ if no habitual action should be retrieved. Since habitual action selection is stochastic, for each retrieval strategy and each evaluated threshold κ, we independently

Table 6: Validation-selected retrieval thresholds and held-out test performance. For each agent and each retrieval strategy, κ is selected on the validation set and fixed for test evaluation.

	Ours			Sim.×HS			Sim.-only		
Agent	Val.	κ^*	Test	Val.	κ^*	Test	Val.	κ^*	Test
Michael	**82.8**	0.55	**81.2**	71.0	0.80	71.6	73.0	0.85	72.1
Adrian	**85.4**	0.55	**82.1**	68.7	0.70	66.9	71.1	0.90	70.4
Elena	**93.1**	0.45	**90.7**	77.1	0.75	75.8	78.6	0.85	77.3
Average	**87.1**		**84.7**	72.3		71.4	74.2		73.3

sample the retrieval outcome 10 times for each case. Let $\hat{a}^H_{j,n,r}(\kappa)$ denote the output returned in the r-th retrieval trial, where $r = 1, \ldots, 10$. If the algorithm retrieves no habitual action in a trial, we set $\hat{a}^H_{j,n,r}(\kappa) = \emptyset$. Each retrieval trial is counted as correct when its output matches the ground-truth label; otherwise, it is counted as incorrect. Retrieval Accuracy is computed by averaging correctness over the 10 trials for each case and then over all cases.

For both the validation and test sets, Retrieval Accuracy is computed in the same manner. Each agent contains 288 validation cases and 192 held-out test cases. For each retrieval strategy, we first evaluate Retrieval Accuracy under different values of κ ($\kappa < 1$) on the validation set and select the threshold that gives the highest validation accuracy for each agent. Since each agent has a different personalized habit memory, the selected retrieval threshold can differ across agents. After selecting the threshold, we keep it fixed and evaluate the corresponding retrieval strategy on the held-out test set. The Retrieval Acc. reported in the main paper is the average of the test-set Retrieval Accuracy across the three agents at their validation-selected thresholds.

Tab. 6 reports the best validation accuracy, the corresponding selected threshold κ^*, and the held-out test accuracy for each agent and retrieval strategy. We search κ with an interval of 0.05 on the validation set and then fix the selected κ^* for evaluation on the test set. Across the three agents, our Two-stage Cue Retrieval selects consistently lower thresholds than the two baselines. This pattern is consistent with our two-stage design, where κ is applied only to cue similarity during habit activation, while habit strength is introduced only after a habit has been activated. This separation prevents habit strength from interfering with the initial cue-matching stage and allows the threshold to focus on whether the contextual cue is sufficiently compatible. With the validation-selected thresholds fixed, our method achieves an average test accuracy of 84.7%, compared with 71.4% for Sim. $\times$ Habit Strength and 73.3% for Similarity-only. The final results are reported in Tab. 2e.

F Design of the Psychological state function in the arbiter

The effects of momentary psychological states on the balance between goal-directed and habitual control are complex. Existing computational models have mainly focused on how factors such as controller reliability, uncertainty, and task complexity influence the allocation of behavioral control, while relatively little work has examined how subjective human psychological states can be mapped to quantitative modulation of these controllers (Lee et al., 2014; Kim et al., 2019). We introduce the psychological state interference function to capture the dynamic influence of different psychological states on goal-directed and habitual control. The purpose of this mechanism is to increase the state dependence and behavioral dynamics of human-like agents, thereby improving the believability and human-likeness of their generated behavior. Existing controlled human studies generally show that adverse stress, self-control depletion, or cognitive load is associated with a marked shift in behavioral control, characterized by reduced goal-directed control and stronger habitual responding. Motivated by this empirical trend, we design psychological-state interference, once activated, to induce a behaviorally meaningful shift toward habitual control. Precisely modeling the mechanisms through which individual psychological factors affect behavioral control is beyond the scope of this work. Instead, inspired by the overall directional trends reported in controlled human studies, we introduce a heuristic mechanism that provides a coarse-grained approximation of these state-dependent shifts in behavioral control.

At each time step t, the agent maintain s a psychological state $s_t^{\text{psy}} = [s_t^{\text{stress}}, s_t^{\text{depletion}}, s_t^{\text{load}}]$, consisting of stress, self-control depletion, and cognitive load. Each factor is represented by an LLM-assigned subjective score on an 11-point Likert-style scale from 0 to 10, based on the agent's current state and recent interactions. This representation provides a coarse-grained semantic estimate of the agent's momentary psychological state rather than a standardized measurement from a specific psychological questionnaire. We use a score of 3 as a reference for a typical psychological state and a score of 7 as an adverse state reference, whose intensity is approximately comparable to the manipulated conditions in the behavioral studies considered below. These values serve as semantic reference points for the state-dependent gate and do not impose an exact baseline-to-adverse control ratio. Scores below 3 indicate weaker interference, whereas scores above 7 indicate increasingly stronger psychological interference.

We consider stress, self-control depletion, and cognitive load because they are closely related to behavioral-control allocation. First, these factors have been extensively studied in psychology and cognitive science in relation to shifts between flexible, deliberative goal-directed control and more automatic or habitual responding. Second, they provide suitable empirical references for parameterization. Controlled studies manipulate these factors individually and compare behavioral changes between baseline and elevated conditions, while related studies provide subjective measures for interpreting psychological-state intensity and behavioral or computational measures that reflect changes in goal-directed and habitual control. This allows us to establish an empirical correspondence between changes in individual psychological factors and changes in behavioral-control allocation. For some factors, subjective-state measurements and behavioral-control effects are obtained from different but related studies; we therefore use them separately for state-semantic interpretation and control-effect scaling. Because these factors are examined under heterogeneous experimental paradigms, the reported behavioral measures and absolute control ratios cannot be directly transferred across studies. For each factor $k \in \{\text{stress}, \text{depletion}, \text{load}\}$, we instead estimate the relative suppression of goal-directed control within each study as $\beta_k = 1 - G_k^{\text{high}}/G_k^{\text{base}}$, where G_k^{base} and G_k^{high} denote empirical proxies for goal-directed control under the baseline and elevated conditions, respectively.

Stress. Schwabe and Wolf exposed participants to either a Socially Evaluated Cold Pressor Test or a control condition before an instrumental learning task with extinction blocks (Schwabe & Wolf, 2011). Subjective stressfulness increased from 3.8 to 47.2 on the original 0–100 scale. Stress did not impair overall instrumental learning, but stressed participants persisted more strongly in previously rewarded actions during extinction, indicating reduced sensitivity to changes in action–outcome contingencies. Based on this behavioral persistence, we approximate the Habit:Goal balance as 38:62 $\rightarrow$ 66:34, which gives $\beta_{\text{stress}} = 1 - 0.34/0.62 \approx 0.45$. These Habit:Goal ratios are behavioral proxies rather than controller weights directly estimated in the original study.

Depletion. Neal et al. studied how reduced self-control affects the expression of strong habits (Neal et al., 2013). In Study 4 ($N = 134$), participants completed either a depletion manipulation or a control writing task. On 1–7 subjective scales, ratings of feeling drained increased from 1.99 to 2.58, while mental exhaustion increased from 1.94 to 2.64. Depletion also increased the probability of choosing strongly habitual options by approximately 31–32 percentage points, while weak habits changed by less than one percentage point. We therefore use the approximate behavioral proxy 16:84 $\rightarrow$ 48:52, yielding $\beta_{\text{depletion}} = 1 - 0.52/0.84 \approx 0.38$.

Cognitive load. Otto et al. manipulated working-memory availability by adding a concurrent numerical Stroop task to a two-stage reinforcement-learning task (Otto et al., 2013). Their computational model explicitly mixed model-based and model-free action values using a fitted mixing weight. The model-based contribution was approximately 0.45 when cognitive resources were relatively available and decreased to approximately 0.27 under concurrent or recent working-memory load. Using model-based and model-free control as proxies for goal-directed and habit-like control, respectively, gives 55:45 $\rightarrow$ 73:27, and therefore $\beta_{\text{load}} = 1 - 0.27/0.45 \approx 0.40$. We note that model-free reinforcement learning is not strictly equivalent to habitual control and use it only as a computational proxy. Otto et al. did not report subjective cognitive-load ratings. We therefore additionally use the subjective workload measurements of Maramotti et al. as a semantic reference for this dimension (Maramotti et al., 2026). In their Stroop experiment, NASA-TLX Mental Demand increased from 5.15 in the RELAX condition to 6.57 in the EXERT condition on a 0–10 scale.

Table 7: Empirical effect-size references for parameterizing the psychological state gate. Habit:Goal ratios are behavioral or computational proxies.

Factor	Habit:Goal (Normal → Elevated)	β_k
Stress	38:62 → 66:34	0.45
Depletion	16:84 → 48:52	0.38
Cognitive load	55:45 → 73:27	0.40

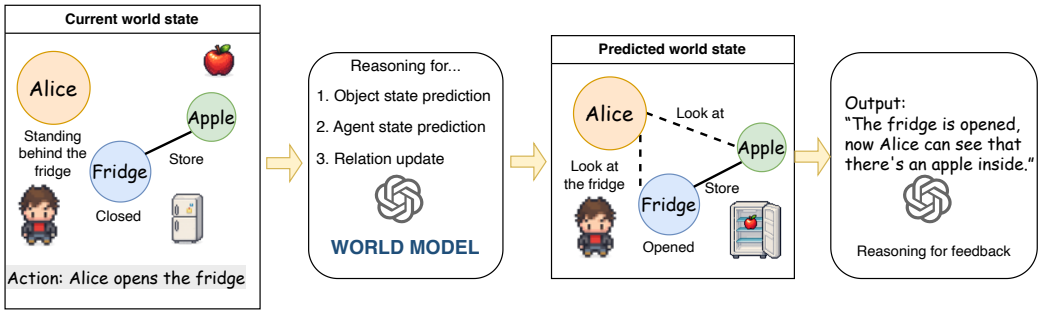


Figure 9: The framework of the context-based world model

This result is used only to inform the semantic interpretation of subjective cognitive load, while the controller shift is obtained from Otto et al. The resulting empirical effect-size references are summarized below:

State-dependent gate activation. Psychological factors do not deterministically affect behavioral control at every time step. Instead, their quantified scores control the probability that the corresponding interference becomes behaviorally active. For each factor k, we define a state-dependent activation probability $p_k^{\text{on}}(s) = \sigma(0.55(s-6))$, where $\sigma(\cdot)$ is the sigmoid function. This gives a small but non-zero activation probability at low state levels, e.g., $p^{\text{on}}(0) \approx 0.04$ and $p^{\text{on}}(3) \approx 0.16$, while the probability increases to $p^{\text{on}}(7) \approx 0.63$ and $p^{\text{on}}(10) \approx 0.90$ as the psychological-factor score increases. At each time step, we sample $z_{t,k} \sim \text{Bernoulli}(p_k^{\text{on}}(s_{t,k}))$, where $z_{t,k} = 1$ indicates that the interference associated with factor k becomes behaviorally active at the current decision step. Since β_k is estimated from average human behavioral effects, we use it as an effect-size reference to set the conditional suppression magnitude so that the expected suppression at $s = 7$ matches β_k.

Specifically, we set $\gamma_k = \beta_k / p_k^{\text{on}}(7)$ and define $f_{t,k} = z_{t,k}\gamma_k$, such that $\mathbb{E}[f_{t,k} \mid s = 7] = \beta_k$. With $p^{\text{on}}(7) \approx 0.63$, this gives $\gamma_{\text{stress}} \approx 0.71$, $\gamma_{\text{depletion}} \approx 0.60$, and $\gamma_{\text{load}} \approx 0.63$. Existing studies largely manipulate these psychological factors independently and provide limited evidence about their joint effects. We therefore avoid assuming additive interactions and conservatively define the overall interference by the strongest active factor, i.e., $F(s_t^{\text{psy}}) = \max_{k \in \{\text{stress,depletion,load}\}} f_{t,k}$. The empirical evidence determines the expected suppression magnitude at the elevated-state reference, while the quantified psychological state score controls the probability that this suppression becomes behaviorally active. Consequently, a higher psychological-factor score produces stronger expected interference on goal-directed control, consistent with the definition of $F(s_t^{\text{psy}})$ in §3.2.3.

G World Model

Inspired by (Zhao et al., 2023; Li et al., 2026), we introduce a context-based world model to approximate state transitions for the human-like agent. Specifically, we use a frozen LLM to predict possible next states and action consequences based on the current world state and the agent's action.

As shown in Fig. 9, we represent the world as a graph. Entities in the environment, including the agent and objects, are represented as nodes, while their states are represented as node attributes. Relations and interactions between entities are represented as edges. In this way, each node–edge–node tuple represents a factual relation in the current world state.

At each state transition, the LLM predicts changes in node attributes and relations between entities. These changes represent how the world may evolve after an action is executed. The predicted updates are then organized into a structured JSON format and applied to the current world representation to obtain the predicted next state.

The context-based world model has several limitations. First, its predictions depend on the domain knowledge of the underlying LLM, and prediction errors may accumulate over long trajectories. Second, the frozen LLM does not dynamically update its world knowledge during interaction. However, our setting mainly focuses on daily-life activities, for which LLMs generally provide sufficient prior knowledge to predict common action consequences. We therefore adopt a simple frozen LLM as the agent's internal model. This design also reflects a property of human internal models, whose predictions are similarly constrained by prior knowledge and experience.

H TEXT TO MOTION BASED 3D EMBODIED MODULE

3D Module Overview The 3D module translates the natural language action description into human motion within an interactive scene. Following the training-free scene-aware Text-to-Motion paradigm (Guo et al., 2025), the module grounds high-level behaviors in 3D scenes through inference-time planning and spatial constraints without retraining the underlying motion generator. The module comprises a high-level planner, keyframe constraints, motion generation, and 3D execution. Given an action description and the current character avatar and scene state, the high-level planner organizes the intended behavior into an ordered sequence of executable actions.

Based on factors such as action complexity, interaction detail, and spatial location, the planned actions are categorized into two types: actions that require spatial guidance and standard actions that do not require guidance. For those actions requiring spatial guidance, the high-level planner provides sparse keyframe constraints that specify relevant body configurations and interaction relationships. These constraints guide motion generation alongside the action descriptions. Actions that do not require spatial guidance are generated directly from the motion generator.

The generated motions along with associated object interactions are executed in the 3D simulation module. The character avatar and scene states are carried forward across successive actions to support multi-step behavior. The following subsections describe the execution setup and examine the contribution of keyframe constraints through qualitative ablations.

Scene Representation and State Management The 3D environment is represented by a structured scene description containing object identities, spatial information, and interaction states. Object identities provide consistent references across task planning and execution, while spatial information describes the locations and geometric properties of scene objects. Together with the character avatar's current position and pose, this representation provides the context for grounding behavioral descriptions in the environment.

The scene information is refreshed at each task initialization. The information is sent to the cognitive module and high-level planner for the planning and action generation process. The module maintains both the character avatar state and the object states throughout execution. The character state includes the current body configuration and ongoing interactions, such as facing direction or holding an object. The object state includes changes resulting from these interactions, such as an updated position or an opened door. Subsequent actions receive the preceding action's terminal state, allowing them to continue from the established configuration.

During continuous execution, the framework carries forward the accumulated interaction state and incorporates updates from the scene. The 3D simulation module provides scene data and executes the corresponding motions and object interactions, while the planner and cognitive module use this information to prepare subsequent actions.

Semantic Task Decomposition and Execution Planning Given an action description, the character avatar state, and the environment states, the high-level planner decomposes the coarse-grained intended behavior into an ordered sequence of executable actions. For each executable action, the planner clarifies how to perform it, identifies relevant interaction objects, and specifies their expected outcomes.

The planner considers spatial relationships and the character avatar's current states when planning the action sequence. Movement actions are included when needed to reach an interaction location, while actions that can be performed from the current position retain that position where appropriate. Affordance is also considered when determining whether preparatory actions are needed. For example, pulling out the chair will be planned when there is not enough sitting space under the table. Such adjustments remain part of semantic planning, ensuring that the resulting sequence forms a coherent and spatially feasible course of action.

Execution planning translates the resulting action sequence into inputs for the Low-Level Motion Generator. It aligns action descriptions with applicable keyframe constraints, assigns their timing, and schedules interaction events. Sequences exceeding the motion generator's per-call duration are divided into consecutive generation batches while preserving the planned action order. Boundary states are carried across batches to support continuity.

Keyframe Constraints for Scene-Grounded Motion In the current stage, we use NVIDIA Kimodo Rempe et al. (2026) as the Low-Level Motion Generator. Although Kimodo provides stable and high-quality action generation capabilities, it lacks the ability to perceive scene information. In order to generate the action that can precisely interact with objects in the scene, we use keyframe constraints to guide Kimodo to generate scene-grounded motions.

Given an intended behavior description and the current environment states, which include the character avatar state, the keyframe generation module produces sparse constraints that guide the action generator in generating scene-matched actions. For example, opening a door requires placing the hand at the handle's spatial coordinates, while sitting requires positioning the body over the seat in an appropriate posture and orientation. Constraints also specify the avatar's facing direction relative to a target object. In general, generated keyframes provide spatial reference states for the intended behavior, while the low-level motion generator produces the movement between them.

The resulting keyframes are aligned with the corresponding action descriptions during execution planning and jointly supplied to the motion generator. For continuous tasks, boundary states are propagated across successive actions to support continuity.

Keyframe Constraints Study To better understand the contribution of keyframe constraints, we conduct a keyframe ablation study on a set of common daily activities. Each task is evaluated using both the full module and an ablated variant in which the keyframe constraint module is disabled, while the low-level motion generator and action descriptions remain unchanged. In this ablation, basic navigation and positional guidance are retained, including the coordinates of the target object or support surface and the associated body-joint relationships. Only the additional keyframe-level pose and interaction constraints are removed: the ablated variant receives no explicit keyframe poses, temporally aligned contact states, or keyframe-level guidance for character position and orientation. This setting allows us to isolate the contribution of keyframe constraints while retaining the minimum information required for scene-aware motion generation.

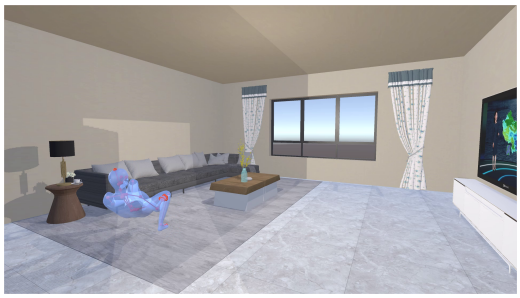
From the complete set of generated results, we select four representative cases for qualitative visualization: (1) sitting on a sofa while watching television and subsequently standing up to cheer, (2) sitting at a dining table, (3) performing jumping jacks in a suitable open area in the living room, and (4) walking toward a toilet while holding the abdomen in pain and then sitting on the toilet. These cases cover different forms of spatial guidance, including body orientation, placement relative to scene objects, positioning in an open area, transitions between supported and standing poses, and locomotion accompanied by a specified body gesture. The first and fourth cases are each presented using four snapshots to illustrate multiple stages of the behavior, while the second and third cases are shown as paired comparisons. Together, these examples provide a qualitative comparison of the motions generated with and without keyframe constraints.

Scenario 1: [Sitting on a sofa while watching television and subsequently standing up to cheer].

As shown in Figs. 10 and 11, in the ablated setting, navigation brings the character to the vicinity of the sofa, but the character remains too far away to establish physical support from the seat. Without keyframe guidance describing the spatial relationship between the body and the sofa geometry, the generator produces an implausible sitting–reclining pose: the feet remain on the floor, while the



(a) Full Module



(b) Without keyframe constraints

Figure 10: Ablation task: Sitting on the sofa while watching television



(a) Full Module



(b) Without keyframe constraints

Figure 11: Ablation task: standing up to cheer

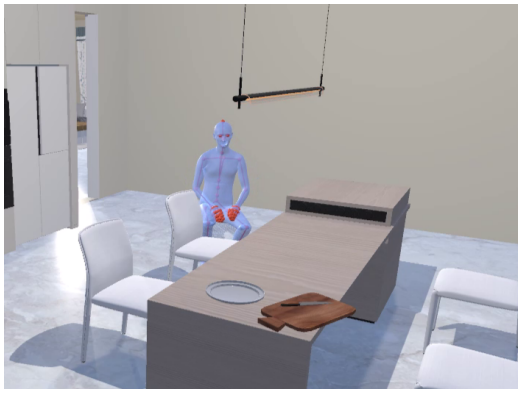
upper body assumes a reclined posture without being supported by the sofa. The character's gaze is also not aligned with the television. By comparison, the full module places the character on the sofa with an appropriate supported pose and viewing direction. The subsequent self-interaction component of cheering is successfully generated in both settings.

Scenario 2: [Sitting at a dining table].

In the second comparison, as can be seen from Fig. 12, the navigation component successfully brings the character into the vicinity of the target in both settings. However, target proximity alone is insufficient to establish a valid sitting configuration. Without keyframe constraints, the ablated model lacks the scene-level geometric guidance needed to interpret the chair's supporting region and functional orientation. Even when the coordinates of the sittable surface are provided, the



(a) Full Module



(b) Without keyframe constraints

Figure 12: Ablation task: Sitting in front of the kitchen table



(a) Full Module



(b) Without keyframe constraints

Figure 13: Ablation task: Performing jumping jacks in a suitable open area in the living room



(a) Full Module



(b) Without keyframe constraints

Figure 14: Ablation task: Walking toward a toilet while holding the abdomen in pain

character places the pelvis beside the supporting region and adopts a sitting direction unrelated to the orientation of the chair. This results in a partially unsupported and spatially inconsistent sitting pose. In contrast, the full module aligns both the body position and orientation with the seat, producing a properly supported sitting pose that is consistent with the chair and its surrounding scene.

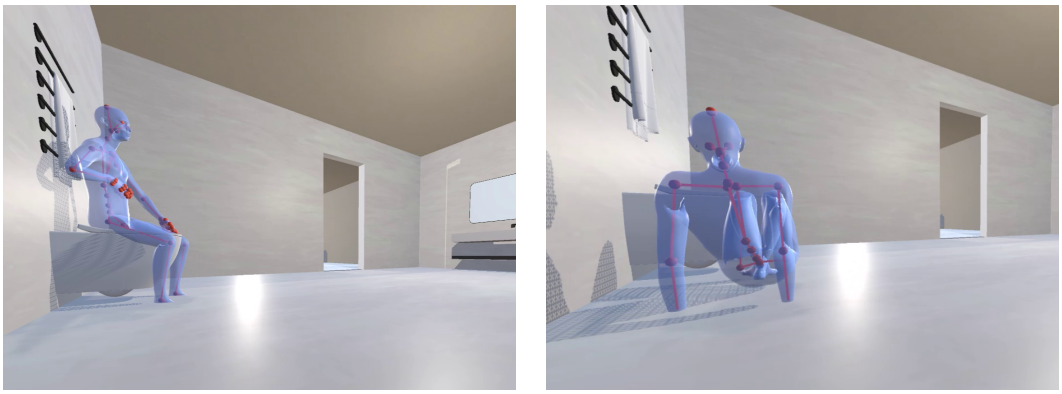
Scenario 3: [Performing jumping jacks in a suitable open area in the living room].

As shown in Fig. 13, both settings correctly interpret the non-object-specific navigation target, identify a suitable open area in the scene, and successfully navigate the character to the selected location. After reaching the open area, both the full module and the ablated variant successfully generate the requested jumping-jack motion.

Scenario 4: [Walking toward a toilet while holding the abdomen in pain and then sitting on the toilet].

The fourth case, as shown in Figs. 14 and 15, walking toward a toilet while holding the abdomen in pain, differs from the preceding examples because the specified body configuration must be maintained throughout locomotion. Without keyframe guidance, Kimodo produces a plausible but generic walking pattern, in which the abdomen-holding component is no longer visibly expressed. The full module supplements the action description with spatial pose guidance, enabling the requested body configuration to remain identifiable during walking. In the subsequent sitting stage, the ablated variant fails to establish the required spatial relationship with the toilet and lowers the character directly onto the floor. In contrast, the full module positions the character relative to the toilet and completes the intended sitting action.

Keyframe Constraints Study Conclusion Across the four scenarios, disabling keyframe constraints does not noticeably affect the navigation component: the character can still identify the intended region and move to the vicinity of the target. However, the ablated module largely loses



(a) Full Module (b) Without keyframe constraints

Figure 15: Ablation task: Sitting on the toilet

the ability to establish reliable interactions with external objects, particularly when the action requires precise spatial relationships among the character, the target object, and the surrounding scene. Self-interaction actions are generally less affected because they depend less on external geometry. Nevertheless, when an action description does not reliably elicit the requested motion pattern from the low-level generator, the ablated module may fall back to a more generic action. The full module provides more consistent scene grounding and better preserves the intended action characteristics in these cases. Its motion quality, however, remains partially dependent on the quality of the generated keyframes; strongly constrained poses, such as some sitting configurations, may occasionally appear slightly less natural.