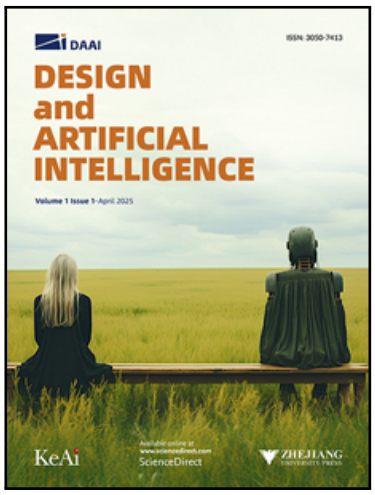
Journal Pre-proof



A Natural Behavior planner for Multi-personal Human-Robot Interaction within the Simulated Environment

Yue Chen , Pai Zheng , Zhiyuan Zhou , Chin-En Keith Soo , Haining Wang , Chunyang Yu

PII: S3050-7413(26)00001-7
DOI: https://doi.org/10.1016/j.daai.2026.100062
Reference: DAAI 100062

To appear in: *Design and Artificial Intelligence*

Received date: 2 December 2025
Revised date: 15 February 2026
Accepted date: 17 February 2026

Please cite this article as: Yue Chen , Pai Zheng , Zhiyuan Zhou , Chin-En Keith Soo , Haining Wang , Chunyang Yu , A Natural Behavior planner for Multi-personal Human-Robot Interaction within the Simulated Environment, *Design and Artificial Intelligence* (2026), doi: https://doi.org/10.1016/j.daai.2026.100062

RoboActor：

A Natural Behavior planner for Multi-personal Human-Robot Interaction within the Simulated Environment

Yue Chen[a], Pai Zheng[b], Zhiyuan Zhou[c], Chin-En Keith Soo[d], Haining Wang[e], Chunyang Yu[*]

[a] Design-AI Lab, China Academy of Art, Hangzhou, 311113, China

[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, HKSAR, China

[c] College of Design and Innovation, Intelligent Big Data Visualization Lab, Tongji University, Shanghai,200092 China

[d] Department of Design, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

[e] School of Design, Hunan University, 410012, Changsha, China

* Corresponding Author: yuchunyang@caa.edu.cn

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare no competing interests.

Data availability

All data generated or analyzed during this study are included in the published article and its supplementary information files.

Authors' contributions

Yue Chen: Conceptualization, Methodology, Software, Validation. Pai Zheng: Data curation, Writing- Original draft preparation. Zhiyuan Zhou: Visualization, Investigation, Software. Chin-En Keith Soo: Software, Validation. Haining Wang: Writing- Reviewing and Editing. Chunyang Yu: Methodology, Supervision, Project administration, Funding acquisition

RoboActor：

A Natural Behavior planner for Multi-personal Human-Robot Interaction within the Simulated Environment

Yue Chen[a], Pai Zheng[b], Zhiyuan Zhou[c], Chin-En Keith Soo[d], Haining Wang[e], Chunyang Yu[*]

[a] Design-AI Lab, China Academy of Art, Hangzhou, 311113, China

[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, HKSAR, China

[c] College of Design and Innovation, Intelligent Big Data Visualization Lab, Tongji University, Shanghai,200092 China

[d] Department of Design, University of Waikato, Hamilton 3240, New Zealand

[e] School of Design, Hunan University, 410012, Changsha, China

* Corresponding Author: yuchunyang@caa.edu.cn

***Abstract**: In recent years, diffusion models have made remarkable success in generating realistic human motions. However, existing robot pose-learning approaches are largely focused on single-task and one-to-one scenarios, failing to account for multi-person social interactions. This limitation leads to rigid, context-insensitive behaviors that are ill-suited for real-world service scenarios. Consequently, current systems often produce robotic behaviors incapable of the fluidity and responsiveness expected in human-centered environments, a shortcoming underscored by affordance theory in robotics. To address this issue, we propose RoboActor, an innovative human-robot interaction behavior planner, draws inspiration from theatrical acting to orchestrate both deliberate and automatic actions. Our framework leverages Large Language Models (LLMs) to disentangle primary command-driven tasks from secondary, context-induced subtasks. By this means, RoboActor generates lifelike and socially appropriate behaviors in multi-person settings, significantly enhancing the naturalness, engagement, and realism of service robots in everyday social applications.*

Keywords: Human-Robot Interaction; Social Robotics; Affordance Design; Large Language Model; Multi-Person Interaction

1. Introduction

Affordance theory, originally developed in ecological psychology, has been widely adopted in design disciplines to explain how the properties of objects suggest their potential uses [1]. In product and interface design, an object's shape, material, or layout inherently conveys how it can be interacted with, thereby enabling intuitive human behavior without the need for explicit instructions [2]. This perspective underscores that action arises not solely from internal cognitive rules, but from the dynamic coupling between an agent and its environment. When extended to robotics, "affordance" transcends physical manipulation to encompass social cues and contextual roles [3,4], implying that a robot should perceive not only what actions an object affords, but also how it ought to behave toward people in shared social spaces [5]. Hence, affordance provides a unified framework for grounding both functional and socially appropriate behaviors in situated perception [6].

With the deep integration of artificial intelligence (AI) and robotics, human-robot interaction (HRI) has emerged as a key enabler for service robots transitioning from controlled laboratory settings to dynamic real-world environments [7]. Systems such as restaurant servers and household companions are expected to engage with humans in natural, fluid, and socially aware ways. However, real-world deployments often involve multi-person scenarios characterized by intricate social relationships, competing intentions, and diverse contextual demands, e.g., serving multiple customers at a table or interacting with several family members during a gathering [8].

Despite these complexities, current HRI research remains largely confined to one-to-one interactions or task-centric models that assign robots narrow functional roles, yielding inflexible behaviors [9]. These approaches often overlook the "social affordances" of multi-agent settings, failing to account for how group dynamics, spatial arrangements, and object-mediated roles jointly prescribe appropriate conduct. Consequently, robots struggle to interpret implicit cues, resulting in rigid behaviors that undermine user experience. To bridge this gap, this work proposes an interaction framework grounded in affordance theory. By enabling robots to discern not only functional possibilities but also socially normative behaviors, we aim to generate adaptive, human-like interaction logic essential for multi-person environments.

Traditional human-robot interaction (HRI) approaches fall almost exclusively into two paradigms: rule-based controllable supervised interaction and unsupervised adaptive interaction grounded in autonomous learning and contextual reasoning [10]. Rule-based systems encode all behavioral contingencies into rigid "if-else" hierarchies, yielding deterministic predictability at the cost of brittleness in dynamic settings. For instance, a robot in a busy restaurant may place plates on the table without regard for who is attentive or in what emotional state [11]. Supervised-learning alternatives replace hand-coded rules with large datasets of annotated behavior-scenario pairs, yet they only partially solve the problem: annotation costs escalate rapidly, and the resulting policies remain fragile, failing when user count or behavioral nuance deviates from the labeled distribution. Current approaches face three core challenges: (1) inability to dynamically identify social foci in multi-person scenes;

(2) lack of autonomous generation of non-instructional behaviors; (3) difficulty in grounding abstract language instructions into actions that respect both physical and social affordances.

In the application of large language models (LLMs) involved in the HRI field, existing research predominantly focuses on dialogic interaction, which means enabling robots to respond to user queries through LLMs while rarely addressing physical behavior decision-making [12]. Specifically, how robots navigate physical spaces and adapt actions in multi-agent scenarios. Moreover, even when LLMs are employed for behavioral decision-making, existing studies fail to integrate natural behavioral patterns with dynamic situational contexts [13,14]. Consequently, robots remain mere task executors rather than socially conscious agents capable of meaningful interaction.

This limitation is especially pronounced in multi-person scenarios, which remain underexplored in HRI research. Most studies focus on either one-on-one interactions or solitary robot tasks, neglecting contexts where multiple humans coexist, communicate, and maintain social ties while interacting with a robot [15]. This research gap leads to robots either behaving inappropriately or simply ignoring human presence in real-world multi-person environments, which seriously violates human social expectations.

To address these gaps, we propose RoboActor, a framework that deeply integrates adaptive robot personality with unsupervised LLMs to generate context-sensitive, naturalistic interaction patterns for multi-person environments. By modeling personality as a dynamic trait that adjusts to interaction partners and situational cues, RoboActor enables robots to behave less like machines and more like responsive social actors.

In summary, the contributions of this paper mainly lie in the following three aspects:

(1) We extend affordance theory in robotics to advocate for more natural interaction of robots in multi-person environments.

(2) We propose RoboActor, an innovative HRI framework enabling robots to engage with multiple people in a natural and socially appropriate manner.

(3) We introduce a novel approach to quantitatively assess the naturalness of robot behavior in multi-person interactive settings.

2. Related Work

2.1 Affordance in HRI

The Affordance Theory was first proposed by American ecological psychologist James J. Gibson in 1977. The core idea of this theory is that affordance represents the action possibilities that the environment offers to humans [16]. This possibility is neither purely objective nor entirely subjective, but rather an attribute of the relationship between the environment and the human being ourselves. In recent decades, affordance theory has been applied to the field of design as an alternative or supplement to functional reasoning [17]. The

functionalist approach is normative, while the affordance approach emphasizes relationality and adaptability. They offer a comprehensive meaning for describing how users perceive, engage and interpret spatial configurations, whether at a small scale like a door handle or at a large scale like a museum exhibition hall [18]. Recent research highlights that affordance is not limited to technological design but also encompasses cognitive, emotional, and social dimensions.

With the development of robot technology, the affordance theory has been introduced into the field of robot perception and control to help robots understand what can be done in a specific environment [19-22]. In robotics, the model of representing affordance as an entity-behavior-effect triplet is currently widely adopted. Sahin proposed in their 2007 work to formalize an interaction event as a triplet: Entity, Behavior, Effect, corresponding respectively to perceived object or environment features, action units that robot can perform, and observable changes caused by the action [19]. In 2025, Wang J et al. explicitly juxtaposed static attributes such as appearance and sound with dynamic behaviors such as actions and language, viewing them as dual channels through which robots display their affordances to users. It is based on these cues that users form expectations of what I can do to it or what it can do to me [20]. The research team that developed ACKnowledge has for the first time proposed a framework of human-compatible affordance planning, enabling robots to autonomously engage in natural behavioral interactions beyond their core tasks. This framework allows robots to proactively insert low-cost overflow behaviors such as emotional, conversational, and postural ones including situational humor, nodding and smiling, and self-disclosure in human-robot interaction. It is adaptive and user-adaptive, supporting agents working in dynamic human environments to conduct human-compatible, affordance-based interaction planning [21]. AffordDex is a novel two-stage training framework that can intrinsically understand motion priors and object affordances, thereby enabling a universal grasping strategy. The research designed a robot object retrieval method based on affordance, developing an object retrieval model based on object use descriptions, which can assist robots in conducting natural language object retrieval in human-populated environments [22].

In conclusion, the existing interaction planning methods based on affordance, although they incorporate conventional human-robot interaction, do not address the more realistic scenario of multiple people being present. Although these methods can successfully execute low-level instructions and high-level planning, they lack a certain degree of generalization and are unable to perform behavioral interaction planning for robots in the presence of multiple people.

2.2 Human-Robot Interaction Methodology

2.2.1 Human-Robot Interaction on Supervised Learning

Supervised learning models emerged with the advancement of machine learning, where researchers train models by annotating large numbers of scene-behavior pairs such as when elderly individuals are present in a scene, the robot should slow down its speech rate [23].

However, this approach suffers from the limitations and lag of annotated data. Real-world multi-person scenarios are highly dynamic, and annotated data cannot cover all possible situations. Moreover, when new scenarios arise, models require re-annotating data to adapt, which incurs extremely high costs. For instance, in multi-person meeting scenarios, supervised models might only respond to speakers while neglecting other participants' body language and attention changes.

Most current robot learning methods rely on supervised learning frameworks, such as imitation learning and Vision-Language-Action Models (VLA). PaLM-E [24], as a large-scale embodied multi modal model, can handle various embodied reasoning tasks across multiple observation modalities and embodied models, demonstrating strong transferability. This is achieved through diverse joint training across internet-scale linguistic, visual, and visual-language domains. For instance, Unitree g1 employs video imitation to control robot movements [25], but the scarcity and cost of action-tagged data in robotics limit the generalization of learned strategies. In contrast, abundant untagged video data is easily acquired, yet converting these observations into effective strategies is still challenging. AMPLIFY [26] proposes a novel framework that encodes visual dynamics into compact discrete motion tokens derived from key point trajectories, transforming massive video data into usable training resources for Human-Robot Interaction (HRI). These video imitation, VLA and VLM-based learning methods significantly reduce the cost of traditional manually annotated datasets while enhancing action dataset diversity.

However, in real-world robotic applications, human control remains essential as robots cannot autonomously perform real-time action judgments and execution. Traditional VLM systems rely on straightforward visual-text alignment such as image-text matching or semantic action descriptions, operating at the perceptual level without generating robot actions[24-27]. The core of conventional VLA focuses on single-user, single-task action generation, which often falls short in multi-user environments and fails to meet diverse needs. While traditional VLA excels in industrial inspections and household tasks like vacuuming or object retrieval, these applications require minimal social interaction.

2.2.2 Application of LLM in Human-Robot Interaction

Nowadays large language models such as GPT-4o [28] and Claude 3 [29] have made significant progress in human-robot interaction (HRI) dialogue. These models now enable robots to understand natural language instructions, answer complex questions, and even engage in emotional communication, with real-time feedback through unsupervised commands. Current research primarily focuses on how LLMs promote robot autonomy by converting advanced natural language instructions into low-level control signals, supporting semantic planning, and enabling adaptive execution. SayTap [30] improve robot gait stability through contact patterns generated by LLMs, while TrustNavGPT [31] achieves a word error rate of 5.7% in noisy speech environments by modeling user uncertainty. Van der Geer J et al. explored using GPT-4o to directly instruct robots, while Alter3 demonstrates real-time generation of action feedback through diverse human natural language, significantly improving the generalization of rule-based robot actions [32]. The WildLMa [33] team

developed a skill-learning interface, which enables large language models to coordinate skills for robots during extended tasks.

More crucially, existing large language models have never been applied to dynamic behavioral adaptation in multi-person scenarios within HRI. While LLMs powerful contextual understanding and generative capabilities could be used to analyze complex information like social relationships and emotional states in multi-person settings, this potential remains untapped [34]. Robots in such scenarios remain one-way task executors rather than two-way social participants.

2.3 Human-Robot Interaction in Multi-person Environment

Meanwhile, the robotics community is actively exploring natural human-robot interaction [35]. For instance, SkillMimic [36] developed a unified human-robot interaction imitation framework that effectively captures diverse interaction patterns from human-robot interaction datasets. Actformer [37] utilizes Transformer encoders to alternately model temporal correlations and human interactions, achieving superior performance in generating both single-person and multi-person motion tasks while providing a comprehensive multi-person combat behavior synthesis dataset, offering valuable prior knowledge for multi-person interaction research. Samantha Regan [38] and colleagues from Carnegie Mellon University and Princeton University designed flexible agent embodiments that allow agents to control different robots when switching between scenarios, enabling multi-person interactions within a single robot. This innovation allows robots to handle individual user tasks while maintaining appropriate behavior in the presence of others. InterGen [39] introduced an efficient diffusion method enabling users to customize high-quality two-person interactions through text guidance. The team also created the InterHuman multi-modal dataset, containing approximately 107 million frames simulating various two-person interactions. Furthermore, Philipp Muller [40] exploit gaze focus during conversations, achieving eye-contact detection in natural three to four person interactions via environmental simulated cameras, though limitations persist in simulated environments. Yun and his colleagues at the University of Cambridge developed a computational model to assist social interaction robots in selecting appropriate conversation partners within multi-person interaction scenarios [41].

To date, academia has largely overlooked multi-person human-robot interaction (HRI) research. Existing studies on multi-agent interactions predominantly examine interactions between multiple robots, rather than between human groups and individual robots [42]. In the limited literature, researchers either reduce multi-person scenarios to separate single-person interactions where robots sequentially interact with individuals while ignoring social connections or assign robots to non-social tasks such as cargo transportation in crowded areas [43-47]. This research gap leaves robots completely lacking in social awareness and behavioural adaptability within real-world social environments [48,49]. For instance, during group gatherings, robots might only respond to the nearest individual while ignoring others' interaction needs, which starkly contradicts human social norms [50]. In summary, while HRI

has advanced in single-user interaction and task execution, robots still lack an affordance-theory-grounded, LLM-enhanced adaptive behavior planning capability for real-world multi-person social scenarios [51-54]. The proposed RoboActor framework aims to fill this critical gap.

Therefore, this article aims to, based on the affordance theory, utilize the automatically recognized E-B-E triplets' associative ability in multi-person environments to enable robots to perform adaptive and flexible autonomous multi-person interaction behaviors beyond the designated tasks.

3. Method: RoboActor

As previously discussed, to enable service robots to exhibit more natural behaviors in multi-person scenarios, they must be capable of considering the awareness of multiple participants present [55]. Much like a skilled actor who considers various factors during performance, the robot should immerse itself in the environment rather than merely completing tasks [56]. This requires the robot to perform comprehensive visual detection of the surrounding environment, identify key entities within the space, and execute dynamic path planning optimized for multi-agent coordination. Therefore, to better achieve natural interaction between robots and humans in multi-person environments, the overall framework of RoboActor comprises three components: (i) Structured information input layer; (ii) LLM semantic analysis layer; (iii) Behavioral planning layer

To enable seamless multi-person interaction, RoboActor integrates a three-layer architecture that eliminates redundancy while ensuring technical precision. The structured input layer captures scene context through Entity-Behavior-Effect (E-B-E) triplets, providing LLMs with grounded semantic inputs for accurate task planning. The LLM semantic analysis layer then dynamically infers contextual relationships and generates action-aware task sequences from this structured data. Finally, the behavioral planning layer translates these sequences into step-by-step executable actions with affordance-aware motion trajectories, ensuring robots move beyond task execution to socially adaptive engagement in dynamic group settings. From contextual perception to embodied action, this integrated flow enables natural, context-sensitive interaction without redundant technical phrasing

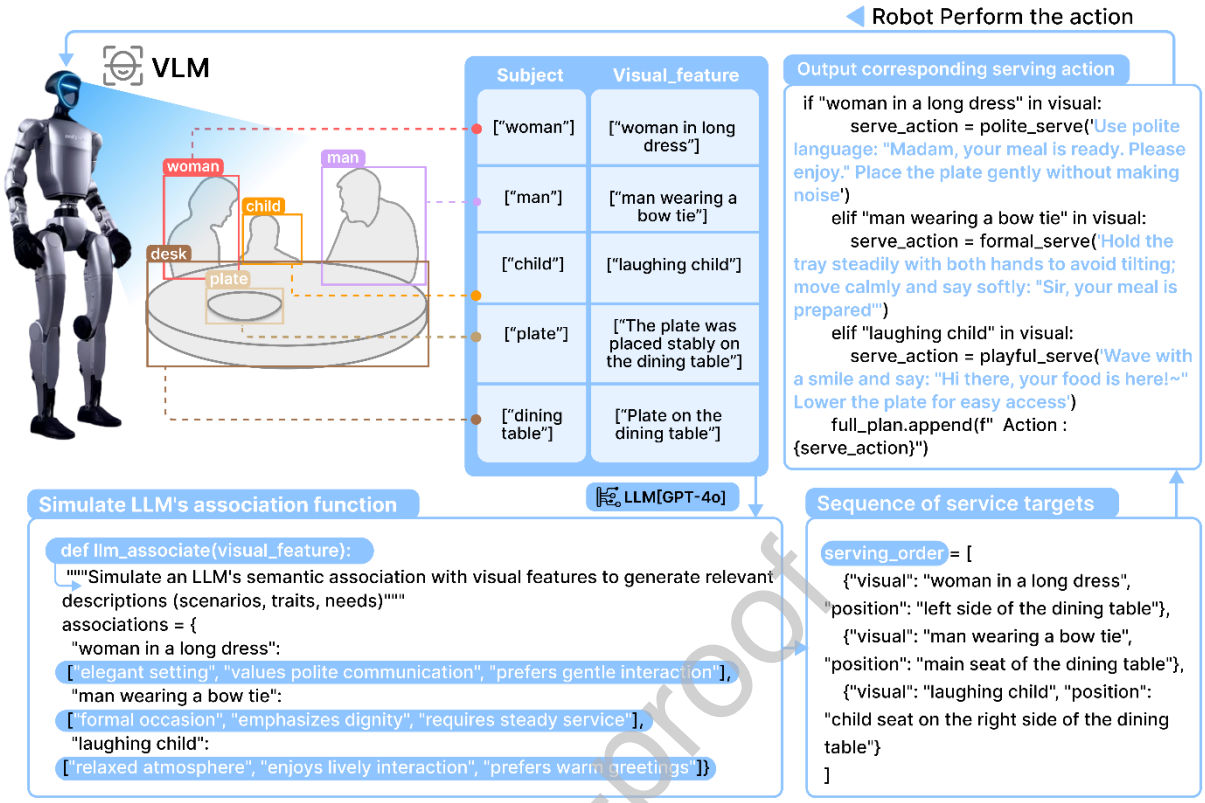


Figure 1 Framework of RoboActor

3.1 Structured Information Input Layer

To ensure comprehensive information acquisition by service robots, RoboActor collects visual data exclusively from MuJoCo's built-in camera system (resolution: 1280×720, 30Hz) within a defined scope. The system integrates a unified computational framework combining perception, natural language understanding, and embodied actions through camera inputs.

The input comprises two components: a specified task such as serving the dishes to the table, and visual scene perception captured in real-time by a virtual camera within the simulated environment. During processing, the Visual Language Model first analyzes the image data to identify key elements in the scene, including different types of people such as women, men, and children with further identification of age and emotional tendencies, objects such as plates, tables, and dishes, and environmental characteristics such as whether the scene is a restaurant, home, or meeting room.

Next, the LLM intervenes. Leveraging its contextual reasoning capabilities, the RoboActor system analyzes multiple semantic labels identified by the robot's virtual camera to form multiple triples. It then combines these with the Scene Element Tag output from the Visual Language Model to generate behavior instructions without supervision. These instructions are categorized into Permitted Actions and Prohibited Actions. Permitted Actions refer to movements the robot can execute in the current scenario, such as smiling at a child or

gently placing a plate. Prohibited Actions explicitly forbid certain behaviors like breaking a plate, attacking humans, or abruptly stopping tasks, ensuring safe and reasonable interactions.

3.2 LLM Semantic Analysis Layer: Implementation of the "Actor" Attribute

The LLM semantic analysis layer serves as the primary component implementing the Actor workflow in RoboActor. In multi user human robot interaction scenarios such as service robots simultaneously serving multiple users with distinct characteristics robots must accurately adapt to complex social environments and user needs. Traditional action decision systems struggle to capture implicit social rules like avoid collisions with long skirts or slow movements for children that lack explicit definitions. This research focuses on leveraging the core strengths of LLM in massive social knowledge reserves and advanced logical reasoning. It systematically transforms structured descriptions containing user characteristics such as age attire environmental information such as interaction space and task objectives such as file delivery into an entity relationship social constraint triplet format. This transformation process explicitizes and standardizes implicit social rules providing clear executable semantic constraints for robot action decision modules ultimately enhancing human robot interaction safety adaptability and user experience. To ensure RoboActor generated triplets meet technical requirements for robot motion control the design includes two key components, i.e., Prompt Engineering Design and LLM configuration.

(1) Prompt Engineering Design: Prompt engineering design adopts an integrated architecture composed of four collaborative modules: task definition, constraint conditions, few-shot examples, and input/output format. These modules operate independently while maintaining mutual linkages to ensure coherent system behavior.

(i) Task Definition: The core objective is to extract quantifiable social elements from structured data captured by the camera, while avoiding irrelevant information generation. Additionally, it requires incorporating contextual knowledge. Specifically, the role must integrate social etiquette norms and ergonomic principles to translate natural language interaction requirements into technical action constraints.

(ii) Constraints: To ensure the comprehensiveness and implement ability of social constraints, it is explicitly required that the constraints must cover four core dimensions, and each dimension must meet the hard requirement of "convertible to robot action parameters".

(iii) Safety constraints: focus on avoiding personal injury and damage to property, such as contact force threshold and collision risk avoidance indicators;

(iv) Physical adaptation constraints: adapt to user physiological characteristics, clothing characteristics and environmental physical constraints, such as the avoidance distance for users with long skirts and the movement amplitude limit for children;

(v) Task execution constraints: ensure the efficient and accurate completion of core service tasks, such as document placement intensity and item delivery accuracy indicators;

(vi) Social interaction constraints: comply with daily social etiquette norms, such as interaction distance and action speed, which should take into account the psychological comfort of users.

The essence of flexible personality lies in enabling robots to adapt behaviors to different objects like humans, a capability achieved through the integration of Etag and LLM. By learning from vast amounts of natural language data, LLM has mastered the implicit logic of object-appropriate behaviors in human society. For instance, it understands that children require more affectionate and gentle actions, elders demand polite and respectful postures, and fragile items like plates require careful handling. When VLM identifies scene elements and generates Etag tags such as child, woman, or plate, LLM applies these implicit rules to autonomously produce adaptive behavioral instructions.

In a multi-person dining scenario, when VLM detects child and woman with a plate present, LLM might generate the following sequence: first, smile at the child and softly say this is your meal, then gently place the plate on the child's table, turn to the woman, politely say please enjoy your meal, and maintain the smile. This behavioral sequence perfectly mirrors human social logic in similar situations, making the robot's personality lively and adaptable rather than mechanically repeating the same actions.

(2) LLM Configuration: After comprehensive evaluation of mainstream large language models including GPT-4 Turbo, Claude 3, and Gemini Pro, we ultimately selected the GPT-4 Turbo model. This model demonstrates superior accuracy in complex semantic understanding and constrained generation, while supporting long-text input for structured description analysis across multiple users and scenarios. The parameter configuration includes: temperature set to 0.2, where lower values reduce randomness and enhance determinism, aligning with constrained quantization requirements; top-p set to 0.8, a kernel sampling parameter controlling diversity to prevent constraint loss due to over-constraint; and maximum output tokens capped at 2000 to ensure sufficient triple generation without redundancy.

3.3 Behavior Planning Layer

RoboActor's unsupervised behavior planning capability fundamentally distinguishes it from traditional supervised models in multi-person environments. Traditional supervised robot models require extensive manually annotated behavior data for basic action decisions [56]. When serving multiple people with different identities developers must input specific scenario-based guidelines [57]. This process demands significant manual effort and suffers from annotator subjectivity leading to biased decisions in complex scenarios [58]. RoboActor eliminates reliance on manual annotation by leveraging large language models' semantic understanding and knowledge integration capabilities. It achieves unsupervised interpretation of natural language and human social behavior patterns through the model's built-in social knowledge reserve. For unfamiliar interaction scenarios the large language model automatically generates socially appropriate behavior guidelines by analyzing implied social norms in text without additional developer rules. This breakthrough delivers dual value

eliminating huge time and labor costs in data annotation while enabling strong scene generalization. The robot adapts interaction styles dynamically to new user identity labels.

Furthermore, the multi-focal behavior planning mechanism endows RoboActor with rich diversity in its actions, completely breaking the limitation of the single action mode of traditional robots. The core of this planning mechanism is that the robot parses task requirements from multiple dimensions instead of being confined to a single functional goal. Taking the common dining service task as an example, traditional robots usually only focus on the basic goal of deliver dishes to the designated location. In contrast, RoboActor can generate multiple complete behavior plans with distinct focuses, each corresponding to a different service value orientation. Figure 2 below shows an example how RoboActor plan its multifocal actions autonomously in a typical scene of restaurant service with multiple users.

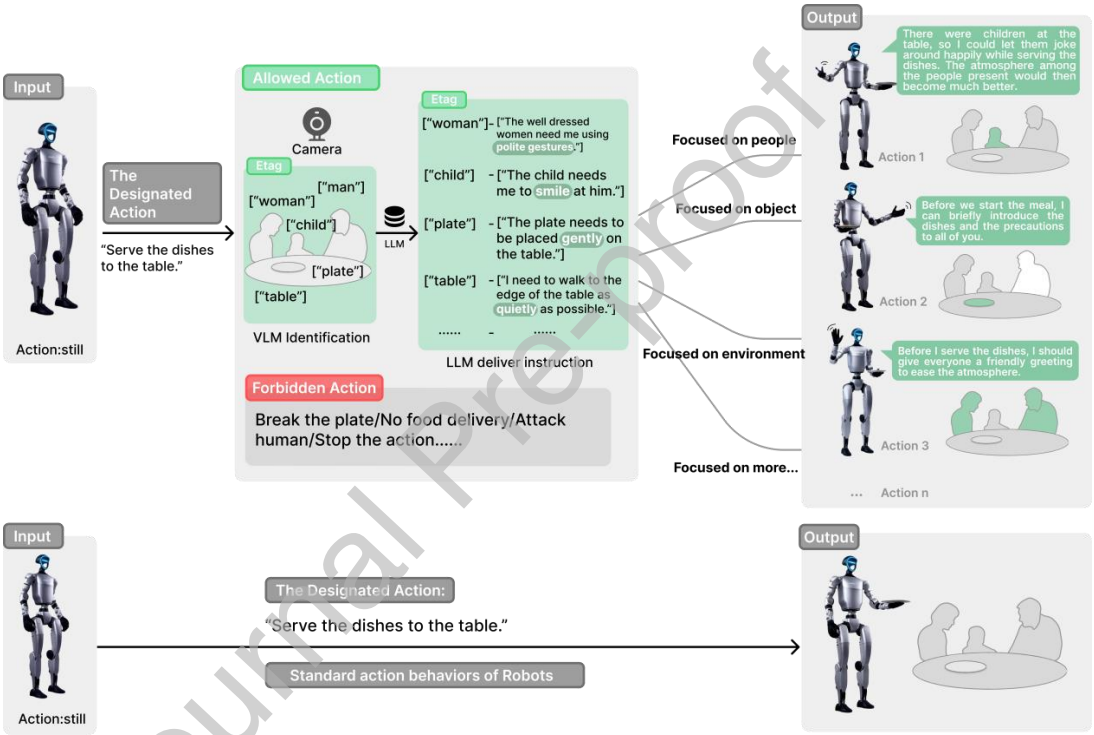


Figure 2 RoboActor's Planning for Different Person Target

Plan A takes interpersonal interaction as its core focus and strives to make every diner feel fully noticed and respected. During execution, the robot actively establishes gentle interactions with diners. For example, when approaching children, it will lower its body height appropriately and ask in a lively tone if they need children's tableware. When facing elderly diners, it will slow down its speech rate and clearly explain the names and temperatures of the dishes. When communicating with young people, it can even provide simple matching suggestions based on current dietary trends. Throughout the process, it maintains natural simulated eye contact, filling the service experience with warmth.

Plan B takes the scientificity and aesthetics of item placement as its primary goal, focusing on enhancing the visual experience and safety of dining. The robot plans the optimal placement area based on the size and shape of the dining table. It places spill-prone dishes

such as soups and hot dishes on the inner side of the table, while cold dishes and tableware are placed on the outer side for easy access. Dishes of the same type are placed together, with reasonable partitioning between staple foods and side dishes. A uniform spacing of about three centimeters is maintained between tableware to avoid collisions while ensuring neatness and order. For dishes with soup, it also proactively equips non-slip placemats to ensure safety and stability during the dining process.

Plan C takes maintaining environmental harmony as its core principle, minimizing the interference of service behaviors on the dining atmosphere. The robot uses built-in sound sensors to judge the on-site communication status. When it detects that multiple people are engaged in intensive conversation, it automatically switches to silent mode, turns off its body prompt sound, and completes serving with slow and gentle movements. If it finds that guests at adjacent tables are having a quiet conversation, it will deliberately adjust its walking route to bypass the side of the conversation area, avoiding blocking the view or interrupting the dialogue with movement noises. When it needs to ask about requirements, it will choose a gap in the conversation to speak softly, ensuring that its service behaviors are perfectly integrated with the on-site environment.

This multi-dimensional behavior planning capability fundamentally changes RoboActor's role positioning in multi-person scenarios. It is no longer just a single-function tool that can only execute fixed instructions, but a social participant that can flexibly adjust its behavior style according to scenario needs. This also makes human-robot interaction more natural, considerate, and better aligned with the complex needs of human social scenarios.

4. Case Study

4.1 Experimental Settings and Comparison Methods

To validate RoboActor's effectiveness, the experiment simulated a classic service robot scenario in a group dining virtual setting. All experiments were conducted in the environment: MuJoCo simulation system (v3.4.0), Ubuntu 20.04 LTS, Python 3.10.0 and ROS Noetic. Hardware: Intel Core i7-11800H CPU, NVIDIA RTX 4070 GPU, 32GB RAM. A virtual restaurant space was created with 3-5 virtual human characters, including children, adult women and adult men, representing diverse demographics, mimicking real-world dining environments. The virtual robot's task was to serve virtual dishes sequentially to virtual tables and interact naturally with the dining group. All experiments were conducted in MuJoCo physics simulation system, which efficiently simulates robot dynamics and multi-agent interactions.

This experiment establishes two control groups to highlight the advantages of RoboActor:

(1) **Control Group 1:** Rule-based Virtual Robot System. This system operates entirely through pre-set rules. The simplest command is: the service robot approaches the dining table,

places the virtual dish in the center, and outputs the text or voice message “Please enjoy your meal slowly”. The rules are fixed and lack any social adaptation.
(2) **Control Group 2:** Supervised LLM-HRI Virtual System. This system trains an LLM to generate behaviors using labeled single-person virtual scene-behavior data, only supporting one-to-one interaction modes. In multi-person virtual scenarios, the robot defaults to treating one virtual human character as the primary interaction partner while ignoring the presence of other virtual human characters.

4.2 Evaluation indicators

To comprehensively evaluate the “naturalness” of virtual robot behavior, the experiment designed indicators from two dimensions: subjective and objective.

(1) Subjective Metrics:

We recruited 20 to 30 human participants with diverse backgrounds, including general users, robotics researchers, and psychology scholars, to view simulated interactions of virtual robots in three systems: RoboActor, rule-based system, and supervised LLM system. Participants rated naturalness of behavior, social appropriateness, and behavioral preference on a 1-to-5 scale, with 5 being the highest, indicating behavior fully aligning with human expectations. Additionally, we posed the open-ended question to collect qualitative feedback: “Which aspects of the virtual robot’s behavior do you find most natural or least natural?”

(2) Objective indicators:

(i) Interaction Engagement: We define this metric as the total duration of reciprocal social exchanges such as robot speaks and human responds with smile or speech. It reflects the level of mutual attention rather than mere co-presence.

(ii) Behavior Diversity: Measured by the number of distinct behavior types executed such as greeting, bowing, adjusting plate position, silent waiting, normalized by total interaction time.

(iii) Interaction Fluency: Count the number of effective social interactions between virtual robots and virtual human characters during the statistical task. Effective interaction refers to the virtual human character's preset positive feedback to the robot's behavior, such as smile animation response and preset language communication.

4.3 Experimental Procedure

The experiment is divided into three stages and is completed in a simulated environment:

(1) Preparation phase: Train all human subjects on experimental tasks and scoring criteria; debug virtual robots, human avatars, and environments in the simulation platform to

ensure consistent simulation standards across three systems: RoboActor, rule-based system and supervised LLM system.

(2) Execution phase: The virtual robot performs the 'multi-person dish service' simulation task under three system configurations, with each task repeated five times to eliminate randomness. The entire process is recorded using the simulation platform's screen recording function, capturing the virtual robot's actions and the preset reactions of the virtual human character, such as facial animations and speech output.

(3) Analysis Phase: Collect subjective rating forms and open-ended feedback. Conduct statistical processing of raw data for objective indicators. Perform t-test analysis to evaluate significant differences between RoboActor and control group. Use ANOVA to explore the impact of different virtual populations on virtual robot behavior.

5. Results and Discussion

5.1 Experimental Results

From the perspective of interaction engagement, Table 1 shows a stark contrast among the three systems. The rule-based virtual robot performed actions for sixty seconds but generated no interaction time with any role, confirming its complete absence of social responsiveness. The supervised LLM HRI system extended total behavior duration to eighty-four seconds yet engaged only briefly with role A for two seconds and ignored roles B and C entirely. In comparison, RoboActor achieved ninety-five seconds of total activity and distributed meaningful interaction time across all three roles—five seconds with the child, seven seconds with the adult woman, and five seconds with the adult man—demonstrating balanced attention allocation in multi person settings.

Robot Type	Total duration	Role A's interaction	Role B's interaction	Role C's interaction
Rule-based	60	0	0	0
LLM-HRI	84	2	0	0
RoboActor	**95**	**5**	**7**	**5**

Table 1 Experimental Results from Interaction Engagement Perspective

Regarding behavior diversity, Table 2 reveals the qualitative richness of social responses. The rule-based system exhibited no interactive behaviors toward any role. The supervised LLM HRI system produced only a single behavior—face to face looking—limited to role A. By contrast, RoboActor executed distinct, role appropriate actions: it exchanged greetings and maintained mutual eye contact with the child, performed a respectful bow toward the adult woman, and provided detailed explanations about the dish to the adult man. This variation confirms that RoboActor generates context sensitive behaviors rather than repeating generic scripts.

Robot Type	Interact with Role A	Interact with Role B	Interact with Role C
Rule-based virtual robot	not have	N/A	N/A
Supervised LLM-HRI virtual robot	look face to face	N/A	N/A
***RoboActor virtual robot**	Look at each other and say hello	Bow	Dish details

Table 2 Experimental Results from Behavior Diversity Perspective (Objective Indicator (ii))

In terms of interaction fluency, Table 3 quantifies the effectiveness of social exchanges through triggered feedback. The rule-based robot failed to elicit any positive response from virtual humans, resulting in zero effective interactions. The supervised LLM HRI system succeeded in triggering one instance of feedback from its designated primary partner. RoboActor, however, generated three effective interactions, each corresponding to a different role, with virtual humans responding through smiles, verbal acknowledgments, or attentive gestures. This indicates that RoboActor's behaviors not only reach multiple participants but also successfully close the interaction loop by evoking socially expected reactions.

Robot Type	Total Effective Interactions	Qualitative Description

Rule-based virtual robot	0	No social feedback triggered from virtual humans
Supervised LLM-HRI virtual robot	1	Single positive feedback from primary interaction partner
***RoboActor virtual robot**	3	Multiple positive feedback instances across all roles

Table 3 Experimental Results from Interaction Fluency Perspective ((Objective Indicator (iii))

These objective indicators collectively demonstrate that RoboActor transcends the binary limitation of rule-based rigidity and supervised learning narrowness. By leveraging unsupervised LLM reasoning for social affordance perception and adaptive personality generation, the system achieves both breadth of role coverage and depth of behavioral appropriateness, addressing the core challenges of multi-person HRI identified in current literature.

Figure 3 illustrates interaction sequences in a multi-person dining scenario revealing key differences in robot behavior. The rule-based robot executes unidirectional actions toward the child while ignoring others, demonstrating a rigid socially unaware approach. The supervised LLM-HRI robot interacts sequentially with child then adult man then adult woman, lacking contextual adaptation. RoboActor follows natural social norms by engaging adult woman first then adult man then child, aligning with formal settings where adults are addressed before children. Each RoboActor interaction involves bidirectional responsiveness, demonstrating simultaneous perception and reaction to multiple social cues. The figure provides visual evidence that RoboActor achieves higher social intelligence through role awareness contextual reasoning and adaptive interaction planning.

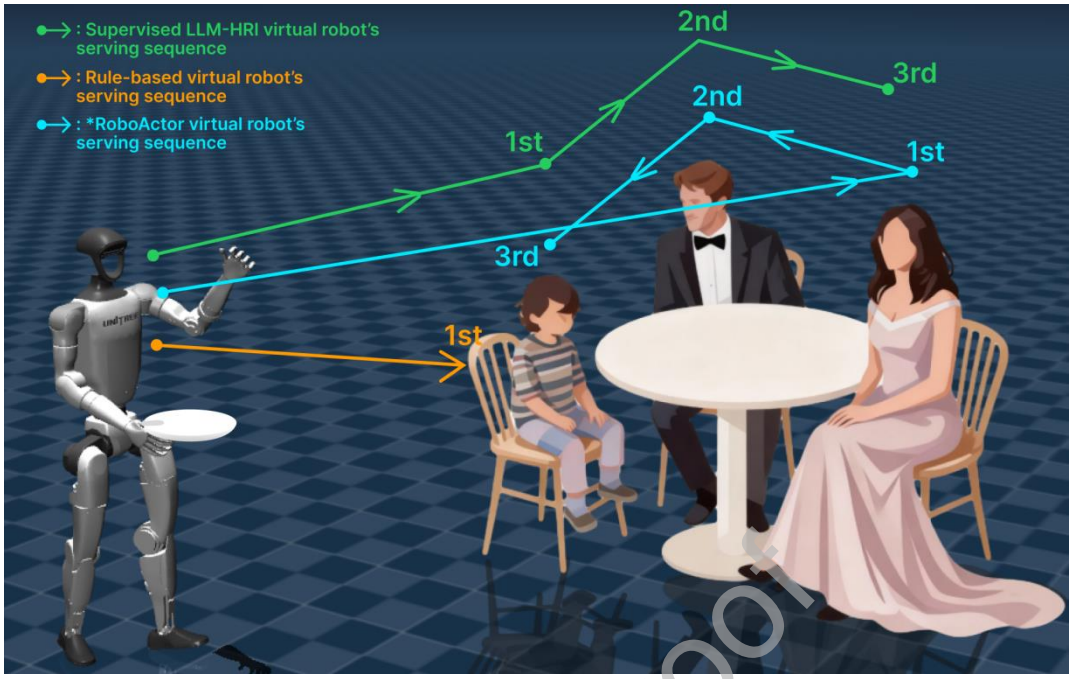


Figure 3 RoboActor's Serving Sequence in Multi-person Environment

As shown in Figure 4 below, RoboActor generates role-sensitive interaction behaviors by aligning physical gestures with socially appropriate verbal expressions. When engaging the adult woman, it performs a respectful bow, reducing its torso angle from 55 degrees to 10 degrees while delivering a courteous greeting. In contrast, its interaction with the child features an upward arm sweep from 30 to 135 degrees, paired with warm and playful speech. For the adult man, RoboActor adopts a neutral posture, adjusting its elbow angle from 90 to 65 degrees while clearly presenting the dish with a professional introduction. This integration of embodied action and contextual language reflects its capacity to perceive and respond to social affordances across diverse user roles.

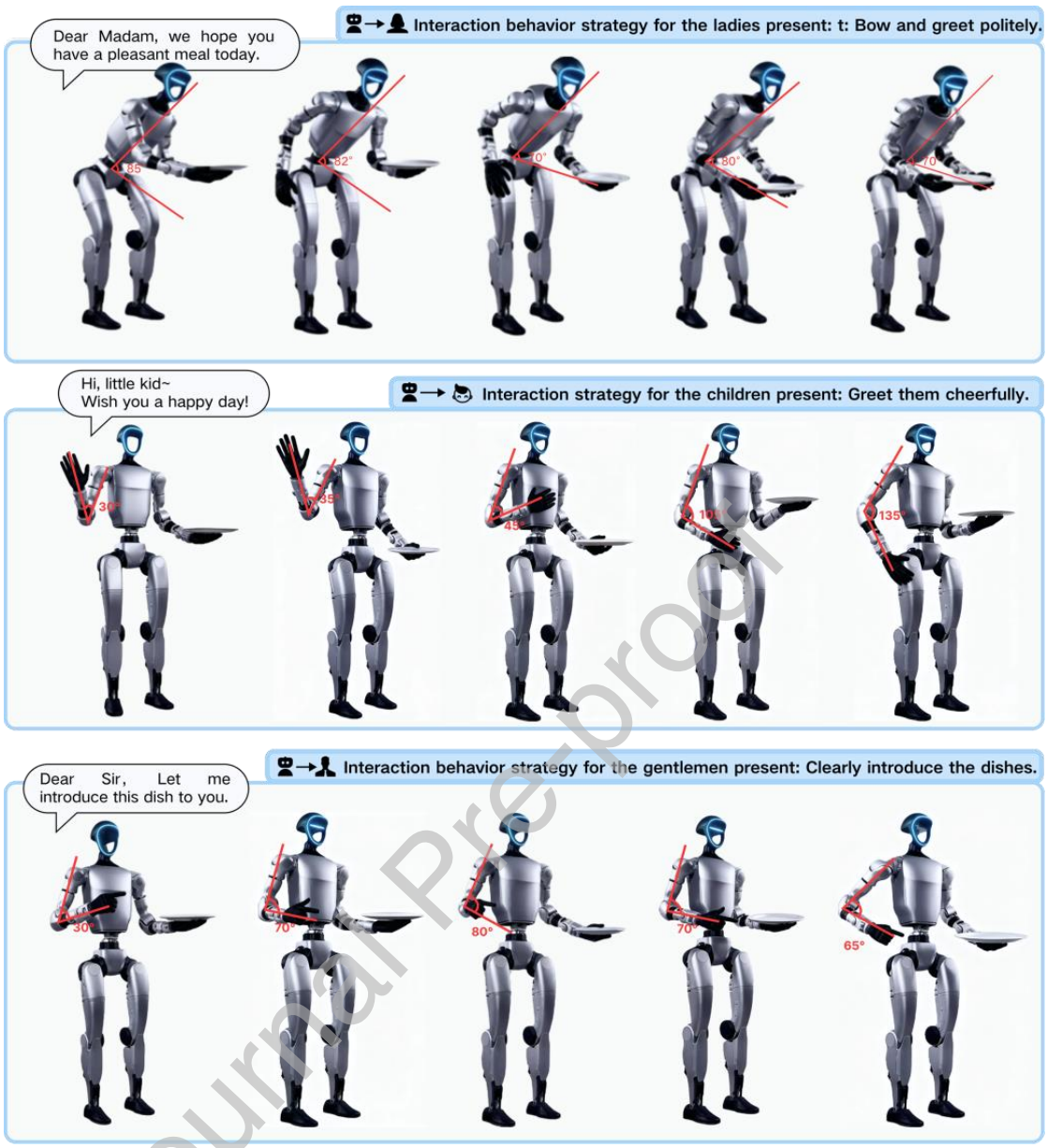


Figure 4 RoboActor's Specific Action Performance towards Role A, B, C in Multi-person Environment

5.2 Result Analysis

RoboActor's performance stems from its' trinity 'design advantages: The LLM's unsupervised generation capability frees it from the constraints of manually labeled data, enabling autonomous comprehension of complex multi-person virtual scenarios; VLM's scene perception ensures Etag accuracy, providing reliable behavioral adaptation for LLM generation; the flexible personality design aligns the robot's behavioral logic with human social habits, thereby comprehensively outperforming the control group in both subjective and objective metrics.

In contrast, the control group's rule-based system, constrained by rigid behavioral logic, demonstrated poor social awareness and adaptability in multi-user virtual environments, resulting in subpar performance across all metrics. The supervised LLM system, relying on single-user labeled data, was limited to' single-object 'interactions in group scenarios, failing to grasp the essence of' social group dynamics,' and consequently underperformed RoboActor.

Experimental results demonstrate that RoboActor outperforms baseline methods across all ask success rate, natural interaction, and real-time performance without requiring customized multi-user interaction datasets. By integrating social knowledge from public datasets, LLM-driven decision logic, dynamic priority ordering, and unsupervised calibration, this framework effectively addresses core challenges in multi-user human-computer interaction, establishing a new paradigm for natural and efficient human-computer interaction in complex multi-person social environments.

6. Conclusion

The RoboActor framework proposed in this paper represents a groundbreaking exploration in the field of multi-agent HRI RoboActor pioneers a deep integration of flexible personality with unsupervised large language models, enabling virtual robots to exhibit human-like dynamic behavioral adaptation in multi-agent simulation scenarios. This innovation not only overcomes the limitations of traditional HRI methods characterized by rigid behaviors and scenario constraints, but also provides a novel approach for deploying service robots in real-world social contexts. Robots are no longer mere tools performing fixed tasks, but rather social participants capable of integrating into human groups. The paper also validates its significant advantages in behavioral naturalness and social adaptability through simulation experiments.

RoboActor fills the research gap in the natural generation of autonomous motion and posture for service humanoid robots in multi-person interaction scenarios in social settings. However, in industrial factory and other scenarios, it still requires fine-tuning and adaptation to corresponding datasets. The research group plans to extend the application of the RoboActor framework to industrial and other scenarios in the future, with the aim of achieving higher generalization and enabling robots to have more natural collaborative and interactive work with humans in factories.

Beyond academic contribution, this work offers practical guidance for commercial deployment of service robots. By enabling natural multi-person interaction, RoboActor can significantly enhance user satisfaction and reduce labor costs, making it suitable for high-interaction settings such as restaurants, hospitals, and eldercare facilities. From a management perspective, this framework introduces new paradigms for human-robot team coordination and workflow optimization. Managers can leverage RoboActor's adaptive personalities to dynamically allocate robotic agents based on real-time interaction demands, thereby improving operational efficiency and service quality. Additionally, the reduction in training overhead and the flexibility of unsupervised adaptation present compelling cost-benefit

advantages for organizational adoption, supporting scalable deployment strategies across diverse service sectors. Future work includes industry collaboration for real-world pilot validation.

References

[1] J. J. Gibson, The Ecological Approach to Visual Perception, 1979.

[2] Priya K, Pillai J, Shende A. Human interaction with the physical world: a brief review of studies on affordances[J]. Proceedings of the Design Society, 2024, 4: 105-114.

[3] Steffen J H, Gaskin J E, Meservy T O. Framework of affordances for virtual reality and augmented reality. Journal of Management Information Systems, 2019, 36(3): 683–729. https://doi.org/10.1080/07421222.2019.1628877

[4] Ardón P, Pairet È, Lohan K S, et al. Building affordance relations for robotic agents-a review[J]. arXiv preprint arXiv:2105.06706, 2021.

[5] Halilovic A, Krivic S. Affordance-Based Explanations of Robot Navigation[C]//2025 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2025: 13523-13529.

[6] Yang X, Ji Z, Wu J, et al. Recent advances of deep robotic affordance learning: a reinforcement learning perspective[J]. IEEE Transactions on Cognitive and Developmental Systems, 2023, 15(3): 1139-1149.

[7] Suzuki R, Karim A, Xia T, et al. Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022: 1–33.

[8] Zaballa K N H, Cameron L D, Lugo A S. Human-robot interactions design for interview process: Needs-affordances-features perspective[C]//International Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2021: 645-655.

[9] Liu P, Glas D F, Kanda T, et al. Data-driven HRI: Learning social behaviors by example from human–human interaction[J]. IEEE Transactions on Robotics, 2016, 32(4): 988-1008.

[10] Irfan B, Kennedy J, Lemaignan S, et al. Coffee with a hint of data: Towards using data-driven approaches in personalised long-term interactions[J]. Frontiers in Robotics and AI, 2021, 8: 726396

[11] Reimann M M, Kunneman F A, Oertel C, et al. A survey on dialogue management in human-robot interaction[J]. ACM Transactions on Human-Robot Interaction, 2024, 13(2): 1-22.

[12] Jeong H, Lee H, Kim C, et al. A survey of robot intelligence with large language models[J]. Applied Sciences, 2024, 14(19): 8868.

[13] Ravishankar J, Doering M, Kanda T. Zero-shot learning to enable error awareness in data-driven hri[C]//Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. 2024: 592-601.

[14] Zhao W X, Zhou K, Li J, et al. A survey of large language models[J]. arXiv preprint arXiv:2303.18223, 2023, 1(2).

[15] Nasiriany S, Kirmani S, Ding T, et al. Rt-affordance: Affordances are versatile intermediate representations for robot manipulation[C]//2025 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2025: 8249-8257.

[16] Myers H E W, Mabey C S, Burleson G. Applying affordance theory to improve design practices for physical, mental, and social health outcomes[C]//Proceedings of the ASME Design Engineering Technical Conference. New York: ASME, 2024.

[17] Kannengiesser U, Gero J S. A process framework of affordances in design[J]. Design Issues, 2020, 28(1): 50-62.

[18] Maramis Y N, Kesuma S A, Syarif F, et al. A Systematic Review of Affordance Theory in Digital Transformation Research: Evidence from 2020–2025[J]. Factory Jurnal Industri, Manajemen dan Rekayasa Sistem Industri, 2025, 4(2): 258-269.

[19] Sahin K, Gokce S, Ozkan M, et al. A Framework for Modeling Human-Robot Interaction Events. Proceedings of the IEEE International Conference on Robotics and Automation 2007: 1243-1248.

[20] Wang J, Li W, Wu Y, et al. Affordance Benchmark for MLLMs[J]. arXiv preprint arXiv:2506.00893, 2025.

[21] Pan Z, Zhang X, Li Z, et al. ACKnowledge: A Computational Framework for Human Compatible Affordance-based Interaction Planning in Real-world Contexts[C]//Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025: 1-20.

[22] Zhao H, Zhuang L, Zhao X, et al. Towards affordance-aware robotic dexterous grasping with human-like priors. arXiv preprint arXiv:2508.08896, 2025.

[23] Sun N, Mao B, Li Y, et al. Assistantx: An llm-powered proactive assistant in collaborative human-populated environment. arXiv preprint arXiv:2409.17655, 2024.

[24] Driess D, Xia F, Sajjadi M S M, et al. Palm-e: An embodied multi modal language model. 2023.

[25] Wu P, Escontrela A, Hafner D, Wu P, Escontrela A, Hafner D, et al. Daydreamer: World models for physical robot learning[C]//Conference on robot learning. PMLR, 2023: 2226-2240.

[26] Collins J A, Cheng L, Aneja K, Collins J A, Cheng L, Aneja K, et al. AMPLIFY: Actionless Motion Priors for Robot Learning from Videos[J]. arXiv preprint arXiv:2506.14198, 2025.

[27] Sonoda Y, Kurokawa R, Nakamura Y, et al. Diagnostic performances of gpt-4o, claude 3 opus, and gemini 1.5 pro in "diagnosis please" cases[J]. Japanese journal of radiology, 2024, 42(11): 1231-1235.

[28] Hurst A, Lerer A, Goucher A P, et al. Gpt-4o system card[J]. arXiv preprint arXiv:2410.21276, 2024.

[29] Anthropic. The Claude 3 model family: Opus, Sonnet, Haiku. 2024.

[30] Tang Y, Yu W, Tan J, et al. Saytap: Language to quadrupedal locomotion[J]. arXiv preprint arXiv:2306.07580, 2023.

[31] Sun, X., Zhang, Y., Tang, X., Bedi, A.S., & Bera, A. (2024). TrustNavGPT: Modeling Uncertainty to Improve Trustworthiness of Audio-Guided LLM-Based Robot Navigation. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 8794-8801.

[32] Yoshida T, Masumori A, Ikegami T. From text to motion: grounding gpt-4 in a humanoid robot "alter3" [J]. Frontiers in Robotics and AI, 2025, 12: 1581110.

[33] Qiu R Z, Song Y, Peng X, et al. Wildlma: Long horizon loco-manipulation in the wild[C]//2025 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2025: 10011-10019.

[34] Xu S, Wei Y, Zheng P, et al. LLM enabled generative collaborative design in a mixed reality environment[J]. Journal of Manufacturing Systems, 2024, 74: 703-715.

[35] Suomalainen M, Karayiannidis Y, Kyrki V. A survey of robot manipulation in contact[J]. Robotics and Autonomous Systems, 2022, 156: 104224.

[36] Wang Y, Zhao Q, Yu R, et al. Skillmimic: Learning basketball interaction skills from demonstrations[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 17540-17549.

[37] Xu L, Song Z, Wang D, Xu L, Song Z, Wang D, et al. Actformer: A gan-based transformer towards general action-conditioned 3d human motion generation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 2228-2238.

[38] Reig S, Luria M, Wang J Z, et al. Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot. Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 2020: 289–297.

[39] Liang H, Zhang W, Li W, Liang H, Zhang W, Li W, et al. Intergen: Diffusion-based multi-human motion generation under complex interactions[J]. International Journal of Computer Vision, 2024, 132(9): 3463-3483.

[40] Müller P, Huang M X, Zhang X, et al. Robust eye contact detection in natural multi-person interactions using gaze and speaking behaviour. Proceedings of the 2018 ACM ETRA Symposium, 2018: 1–10.

[41] Yun, S.-S. (2017). A gaze control of socially interactive robots in multiple-person interaction. Robotica, 35(11), 2122–2138. doi:10.1017/S0263574716000722

[42] Sontakke S, Zhang J, Arnold S, Sontakke S, Zhang J, Arnold S, et al. Roboclip: One demonstration is enough to learn robot policies[J]. Advances in Neural Information Processing Systems, 2023, 36: 55681-55693.

[43] Li X, Liu M, Zhang H, Li X, Liu M, Zhang H, et al. Vision-language foundation models as effective robot imitators[J]. arXiv preprint arXiv:2311.01378, 2023.

[44] Shah D, Sridhar A, Bhorkar A, Shah D, Sridhar A, Bhorkar A, et al. Gnm: A general navigation model to drive any robot[J]. arXiv preprint arXiv:2210.03370, 2022.

[45] Walke H R, Black K, Zhao T Z, Walke H R, Black K, Zhao T Z, et al. Bridgedata v2: A dataset for robot learning at scale[C]//Conference on Robot Learning. PMLR, 2023: 1723-1736.

[46] Black K, Brown N, Driess D, et al: A Vision-Language-Action Flow Model for General Robot Control[J]. arXiv preprint arXiv:2410.24164, 2024.

[47] Dahiya A, Aroyo A M, Dautenhahn K, Dahiya A, Aroyo A M, Dautenhahn K, et al. A survey of multi-agent human–robot interaction systems[J]. Robotics and Autonomous Systems, 2023, 161: 104335.

[48] Nair S, Rajeswaran A, Kumar V, Nair S, Rajeswaran A, Kumar V, et al. R3m: A universal visual representation for robot manipulation[J]. arXiv preprint arXiv:2203.12601, 2022.

[49] Liu Z, Bahety A, Song S. Reflect: Summarizing robot experiences for failure explanation and correction[J]. arXiv preprint arXiv:2306.15724, 2023.

[50] Rodriguez-Guerra D, Sorrosal G, Cabanes I, Rodriguez-Guerra D, Sorrosal G, Cabanes I, et al. Human-robot interaction review: Challenges and solutions for modern industrial environments[J]. Ieee Access, 2021, 9: 108557-108578.

[51] Kim S, Kim S. Working with robots: human resource development considerations in human–robot interaction[J]. Human Resource Development Review, 2022, 21(1): 48-74.

[52] Winkle K, McMillan D, Arnelid M, Winkle K, McMillan D, Arnelid M, et al. Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical HRI[C]//Proceedings of the 2023 ACM/IEEE international conference on human-robot interaction. 2023: 72-82.

[53] Kim H, So K K F, Wirtz J. Service robots: Applying social exchange theory to better understand human–robot interactions[J]. Tourism Management, 2022, 92: 104537.

[54] Wang T, Zheng P, Li S, Wang T, Zheng P, Li S, et al. multi modal human–robot interaction for human-centric smart manufacturing: a survey[J]. Advanced Intelligent Systems, 2024, 6(3): 2300359.

[55] Gasteiger N, Hellou M, Ahn H S. Factors for personalization and localization to optimize human–robot interaction: A literature review[J]. International Journal of Social Robotics, 2023, 15(4): 689-701.

[56] Song C S, Kim Y K. The role of the human-robot interaction in consumers' acceptance of humanoid retail service robots[J]. Journal of Business Research, 2022, 146(2): 489-503.

[57] Tunyasuvunakool S, Kabra R, Wäldchen S, et al. MuJoCo: A general-purpose physics simulator for robotics and beyond. arXiv preprint arXiv:2207.04524, 2022.

[58] Alam A. Social robots in education for long-term human-robot interaction: socially supportive behaviour of robotic tutor for creating robo-tangible learning environment in a guided discovery learning interaction. ECS Transactions, 2022, 107(1): 12389.

CRediT Author Statement

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We futher confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He is responsible for communicating with the other authors about progress,submissions of revisions and final approval of proofs.

All authors as follows:

Yue Chen: Conceptualization, Methodology, Software, Validation

Pai Zheng.: Data curation, Writing- Original draft preparation

Zhiyuan Zhou: Visualization, Investigation, Software

Chin-En Keith Soo: Software, Validation

Haining Wang: Writing- Reviewing and Editing

Chunyang Yu: Methodology, Supervision, Project administration, Funding acquisition

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author *Click here to enter your name* is Choose an item for *Click here to enter the journal's name* and was not involved in the editorial review or the decision to publish this article.

☐The authors declare the following financial interests (e.g., any funding for the research project)/personal relationships (e.g., the author is an employee of a profitable company) which may be considered as potential competing interests:

Click here to enter your full declaration