

Appendix for *COOPERA*

A Overview

This Appendix includes: 1) more details about how we construct dynamic HSSD scenes and statistics, 2) additional details of how we build, train, and evaluate the simulated human and the assistive agent, 3) additional HRC results and analysis, including both quantitative metrics and qualitative examples, along with comparisons to human verification of our main method and baselines across collaboration types and settings, and 4) prompt details.

B Additional Details of Scenes

B.1 Dynamic Habitat Synthetic Scenes Dataset Construction

We make certain objects in the HSSD [28, 40, 53, 54] scenes dynamic by checking if they are supported by any structure and their object super categories. For objects that do not have support, we classify them as follows:

dynamic_categories = [trashcan, decor, dining ware, plant, electronics, animate object, apparel, liquid container, kitchen ware, tray, bathroom accessory, gym equipment, toy, wearable]

static_categories = [storage furniture, support furniture, seating furniture, floor covering, lighting, sleeping furniture, bathroom fixtures, mirror, large kitchen appliance, large appliance, kitchen bathroom fixture, vehicle, heating cooling, medium kitchen appliance, display, arch, curtain, small kitchen appliance]

B.2 Scene Statistics

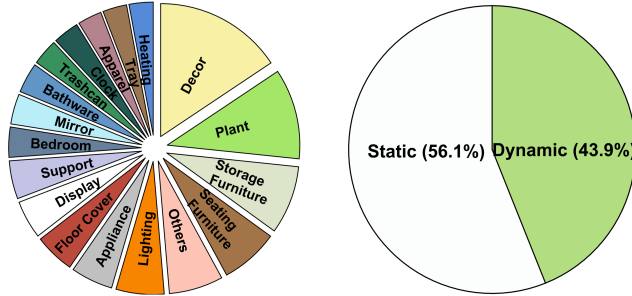


Figure 8: **Distribution of object categories** within the constructed dynamic HSSD scenes.

Our selected 5 scenes feature varying number of rooms (4–11), static objects (51–140), and dynamic objects (33–94). We present the distribution of 18 representative object categories (out of 32) in Fig. 8. The large number and diversity of objects and categories enable humans to propose a wide range of open-ended tasks.

B.3 Scene Summarization and Visualizations

We summarize 3D environment information in each scene as a text-based dictionary, which is used as input to the simulated humans. Specifically, we extract the bounding boxes of rooms and map each object to its corresponding room, forming an object-room mapping in the format of object_ID: [object_name, room]. For example:

mapping = {'Nemo Kepler Pendant, Black': [125, 'corridor'], 'AquaVive stortdoucheset Kila met kraan': [123, 'main bathroom'], 'Uttermoest Marlow Chandelier': [118, 'bathroom of bedroom 1'], 'Eisa Pendant': [114, 'main bedroom'], 'CAR - SUV': [12, 'garage'], 'Nolan Upholstered King Bed': [109, 'main bedroom'], ... }

We show visualizations of the five scenes used in COOPERA in Fig. 9.

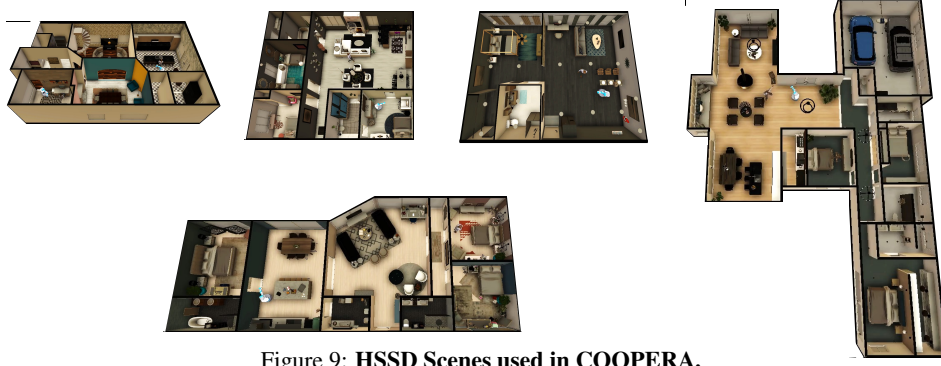


Figure 9: HSSD Scenes used in COOPERA.

C Additional Details of Human Simulation

C.1 Psychometric Data Computation

Since SPC dataset lacks psychometric data, we derive Big-5 OCEAN scores by prompting the LLM to 1) directly infer the scores [48] and 2) complete the Big-5 personality test [18, 45] and compute scores based on the formula. We then take a majority vote across five inference trials, using bins of 0.5 on a scale of 1-5.

C.2 3D Motion

For human simulation, free-form motion data is formatted in SMPL-X, enabling detailed control over whole-body motions, including facial expressions and finger articulation. To integrate this data into the human simulation pipeline of Habitat 3.0 [52], we remap the format of Motion-X [31] data, as illustrated below.

global root orientation: SMPL-X[:, :3]
body: SMPL-X[:, 3:3+63]
finger articulation: SMPL-X[:, 66:66+90]
yaw pose: SMPL-X[:, 66+90:66+93]
face expression: SMPL-X[:, 159:159+50]
global body position: SMPL-X[:, 209:209+100]
global body position: SMPL-X[:, 309:309+3]
body shape: SMPL-X[:, 312:]

C.3 Feedback

Please refer to Section F for details on our designed feedback mechanism for the robot’s predicted assistive tasks. We emphasize that while we structure the feedback as binary answers paired with reasoning, it can be easily adapted for other learning algorithm by modifying the prompt.

C.4 Additional Qualitative Examples

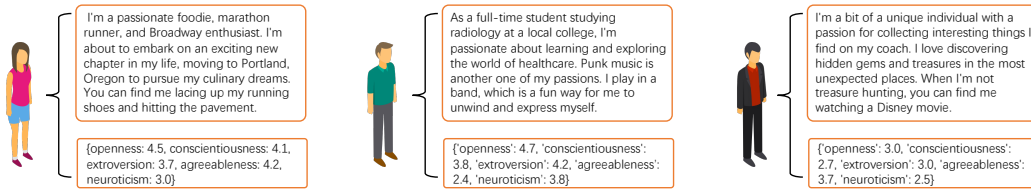


Figure 10: Additional examples of simulated human profiles psychometric data.

We present additional examples of simulated human profiles along with their corresponding psychometric data in Fig. 10.

D Additional Details of Building and Evaluating an Assistive Agent

D.1 Navigation and Movement

To navigate the scene, the agent uses classical motion planning for path calculation, with a stopping threshold determined by the object’s Axis-Aligned Bounding Box (AABB). For the Fetch robot’s arm movements, inverse kinematics (IK) is applied to calculate the end effector (EE) position based on the 3D location of the target object. Robot actions are implemented using primitives from Habitat’s task_action registry, combining navigation and EE control to form a full pick-and-place pipeline: 1) MoveEEAction moves the EE via Cartesian steps with IK; 2) PickObjIdAction grasps objects using a snap mechanism within a distance threshold; 3) PlaceObjIdAction releases objects via desnap; and 4) ResetEEAction resets the EE to a default pose.

D.2 Training

Task videos from Habitat 3.0 [52] are resized to 1024×768 and input to Llama-3.2-11B [13] (our robot-VLM) using Llama’s default settings. Classifiers are finetuned on Mistral-7B-Instruct-v0.2 [27] using LoRA [22] (rank 8, dropout 0.2, alpha 16; targets: q, k, v, o) in an instructional format to output binary yes/no. We train for 5 epochs using AdamW [37] (lr $1e-5$, weight decay 0.01), with batch size 1 and gradient accumulation of 4 steps, across 3 NVIDIA A10 GPUs (24GB RAM).

For training the robot’s intention and task binary classifiers via instructional finetuning of an LLM, we structure the input data in the following format:

```

### Instruction:
Considering the human’s profile, traits, temporal dependence on past behaviors, and the current time,
determine if it is likely or unlikely that this human will: ... Respond with 'Yes' or 'No'.

### Input:
Human Profile.
Big Five Traits.
Previous Relevant Intentions.
Previous Relevant Tasks.
Current Time.

### Response:

```

D.3 Predicate Construction

Collaboration Type 1	Time: 6 pm Intention: Enjoy a playful and imaginative dinner preparation in the kitchen, experimenting with new recipes and flavors.	Task 1: Place the "tuna_fish_can" on the "KITCHEN AID ARTISAN MIXER RED" for setup. Task 2: Place the "mustard_bottle" on the "Edelweiss desk, ash/white" in the living room. Task 3: Place the "fork" on the "Edelweiss desk, ash/white" in the living room.	ON(food, kitchen electricals) ON(food, table) ON(kitchen items, table)
Collaboration Type 2	Time: 9 am Intention: Engage in a morning workout routine in the living room.	Task 1: Start with a warm-up by performing yoga stretches near the "Marina Slipcover Sofa" to prepare the body for more intense exercises. Task 2: Perform squats with dumbbells near the "Wendover Art # Morning Lake View I" for strength training. Task 3: Transition to cardio by doing jumping jacks near the "Wendover Art # Off the Path" to elevate heart rate. Task 4: Engage in core exercises by doing planks near the "APPLE iMac 5K 27" to strengthen abdominal muscles. Task 5: Cool down with a stretching routine near the "Marina Slipcover Sofa" to relax muscles and prevent injury.	NEED(yoga mat) NEED(dumbbells) NEED(jump rope) NEED(exercise mat) NEED(towel)

Figure 11: Examples of predicate construction.

For collaboration type 1, we map objects to Habitat 3.0 and the YCB dataset’s predefined semantic categories, allowing evaluation via exact match. For collaboration type 2, where the robot offers objects from a magic box, we compute semantic similarity between the robot’s predicted tasks/objects for assistance and the human’s desired objects, using a threshold of 0.6. Fig. 11 illustrates examples of predicate construction.

E Additional Experiments and Details

E.1 Results Breakdown and Analysis

In Fig. 12, we show detailed predicate-based and LLM-based evaluations in both collaboration types across all collaboration settings. Overall, LLM-based evaluations yield higher scores than predicate-based ones, particularly in settings 3 & 4, which are more challenging. This suggests that predicate-based evaluations are more rigorous, requiring exact matches for collaboration type 1 and high semantic similarity for collaboration type 2. We highlight several key findings and analyze the performance of each baselines in detail below.

Temporal Fluctuation Patterns Across Days. In Fig. 12, we observe a consistent temporal pattern across days: performance tends to dip around midday and recover toward the end of the day, and improve further on the following day. We attribute this to two primary factors. First, it reflects how human routines are structured, both in reality and in our simulation. Humans generally engage in more predictable activities in the morning and evening (e.g., hygiene, eating, relaxing), which are easier for the robot to infer and assist with. In contrast, midday behavior tends to be more diverse and more strongly influenced by individual traits. Since our human simulation pipeline samples intentions and tasks based on both traits and time, this increases the diversity of midday actions, making inference and alignment more difficult for the robot and resulting in a temporary performance drop. Performance recovers as routine behaviors re-emerge in the evening or the following morning. Importantly, despite these fluctuations, overall performance improves across days. Second, the extent of this fluctuation increases with task difficulty, as defined in our framework (Section 3.1). In Setting 1 (same human, same scene), the robot benefits from consistent exposure to both the user and environment, resulting in relatively small midday dips. In Setting 2 (same human, different scenes), unfamiliar object layouts introduce grounding challenges that increase midday variability. In Setting 3 (different humans, same scene), rotating between users interrupts personalization, making intention interpretation more difficult—particularly around midday when behavior is less routine. Finally, Setting 4 (different humans, different scenes) combines both challenges, resulting in the largest fluctuations as the robot must generalize across both human and spatial contexts.

Performance Analysis: Main Method vs. Baselines. We evaluate our method against six baselines: 1) Direct Prompting, 2) Direct Fine-tuning, 3) Oracle, 4) Random, 5) Intention Agnostic, and 6) Human & Context Agnostic. Fig. 12 shows that our method achieves both the highest adaptation trend over time and the highest final success rate (second only to Oracle) by explicitly modeling the correlation between human intentions/tasks, traits, and temporal dependencies. As the day progresses, the robot accumulates interaction history and adapts its behavior to the human preferences. 1) Direct Prompting shows minimal improvement as it depends on retrieved interaction history without learning the human preferences. 2) Direct Finetuning performs slightly better but still struggles due to its rigid mapping from inputs to intentions/tasks, making it biased toward frequent training examples and overlooking the varying distribution of human behavior. 3) Oracle serves as an upper-bound performance reference by receiving ground-truth human intentions. While this provides a significant advantage, its performance is static by design. This is because Oracle bypasses the core challenge of accurately mapping visual observations of humans performing detailed tasks to high-level intentions, a fundamental problem in human-robot collaboration, even in closed-set scenarios. Moreover, since the correct intention is provided upfront, Oracle has no need to learn temporal dependencies between human intentions (e.g., an intense workout at 11 am typically leads to lunch preparation at 12 pm). 4) Random degrades over time due to the absence of learning or validation. 5) Intention Agnostic shows moderate improvement but lags behind our method, as it skips intention inference and thus loses contextual information, which is problematic since temporal dependencies between intentions are typically stronger than between tasks (validated in Section 4.2). 6) Human & Context Agnostic captures only superficial time-task correlations and ignores human traits or temporal context, leading to weak within-day gains.

E.2 Human Verification Breakdown and Analysis

For human verification, we present additional results across three baselines and our main method, analyzing how human evaluations align with predicate-based (Table. 7) and LLM-based (Table. 8) assessments in a detailed breakdown. Due to cost constraints, we use a subset of data (one episode per setting), resulting in 32 episodes across the four methods. We recruit eight human evaluators, each

Table 7: **Breakdown of correlation between predicates and human evaluations.** We report the L1 \downarrow .

	Collaboration 1				Collaboration 2			
	Setting 1	Setting 2	Setting 3	Setting 4	Setting 1	Setting 2	Setting 3	Setting 4
Random	0.070	0.134	0.067	0.065	0.073	0.079	0.088	0.112
Oracle	0.072	0.098	0.084	0.065	0.086	0.102	0.104	0.108
Human & Context Agnostic	0.062	0.121	0.066	0.071	0.073	0.109	0.080	0.100
Main	0.092	0.096	0.094	0.087	0.090	0.085	0.076	0.153
Average	0.074	0.112	0.078	0.072	0.080	0.094	0.087	0.118

Table 8: **Breakdown of correlation between LLM and human evaluations.** We report the L1 \downarrow .

	Collaboration 1				Collaboration 2			
	Setting 1	Setting 2	Setting 3	Setting 4	Setting 1	Setting 2	Setting 3	Setting 4
Random	0.031	0.054	0.066	0.044	0.068	0.091	0.084	0.110
Oracle	0.103	0.096	0.087	0.043	0.071	0.069	0.091	0.094
Human & Context Agnostic	0.045	0.058	0.076	0.071	0.084	0.089	0.079	0.108
Main	0.063	0.075	0.081	0.073	0.090	0.085	0.072	0.076
Average	0.061	0.071	0.078	0.058	0.078	0.084	0.082	0.097

assessing four randomly assigned episodes. The results show that the gap between predicate-based evaluation and human verification is larger than that between LLM-based evaluation and human verification. This is because LLMs/VLMs reason and make judgments for assistance in a way that aligns more closely with human preferences and decision-making, whereas predicate-based evaluation relies on strict matching criteria that may not fully capture nuanced human reasoning.

E.3 Qualitative Examples of Offline Real Human

We present qualitative examples of offline real humans in Fig. 13. Compared to simulated humans (Fig. 5), real humans exhibit greater behavioral dynamics and emergent decisions due to temporary factors (e.g., plans, mood, weather).

E.4 Analysis of Human-in-the-Loop Collaboration

Intuitively, one might expect lower performance when collaborating with real humans due to their inherent variability and spontaneity. However, results in Table 4 indicate that real human collaborators do not make the task more difficult and can, in some cases, lead to higher success rates compared to offline or simulated humans. Through discussions with participants, we identify a likely explanation. Unlike the offline setting where participants record their routines throughout an entire day (12 one-hour intervals covering the time from 9 am to 9 pm), the HITL experiments take place in a short, controlled session. As a result, participants do not act spontaneously hour-by-hour but instead recall and follow their typical weekly routines from memory, which tend to be more consistent. This reduced behavioral variability allows the robot to personalize more quickly, and further validates the behavioral diversity modeled in our simulated humans.

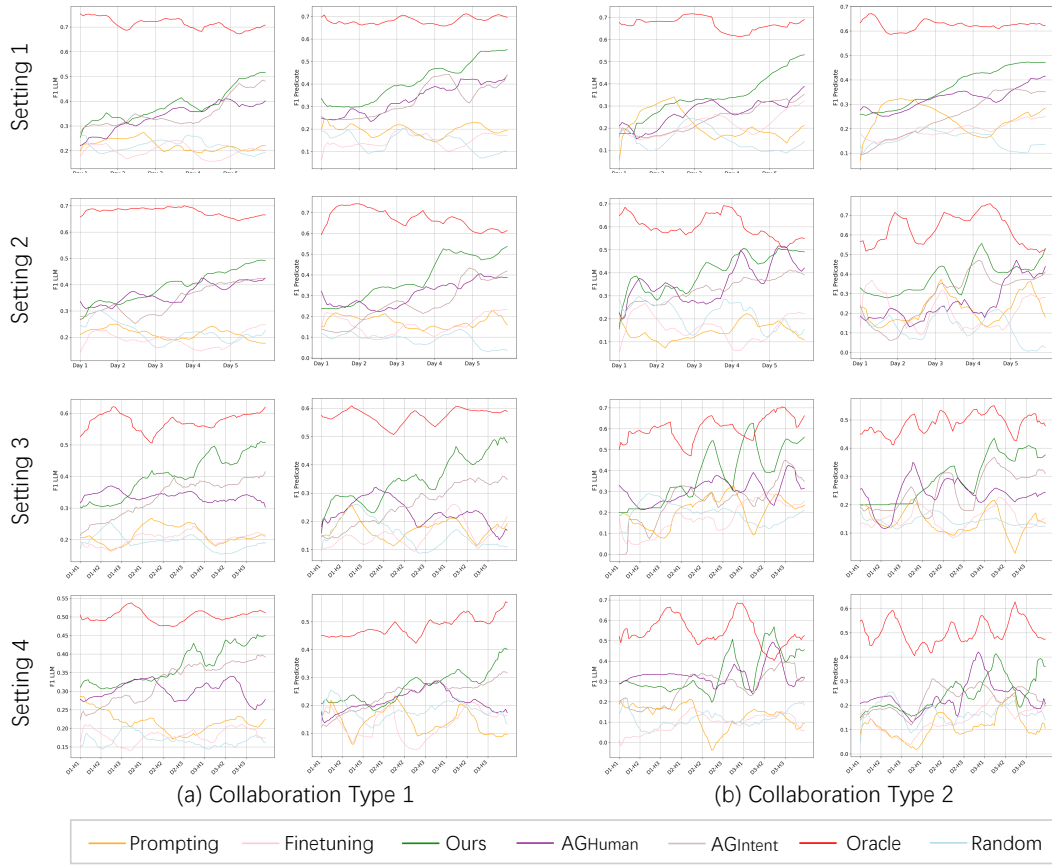


Figure 12: **Breakdown results of HRC.** We compare our main method with the baselines across days.

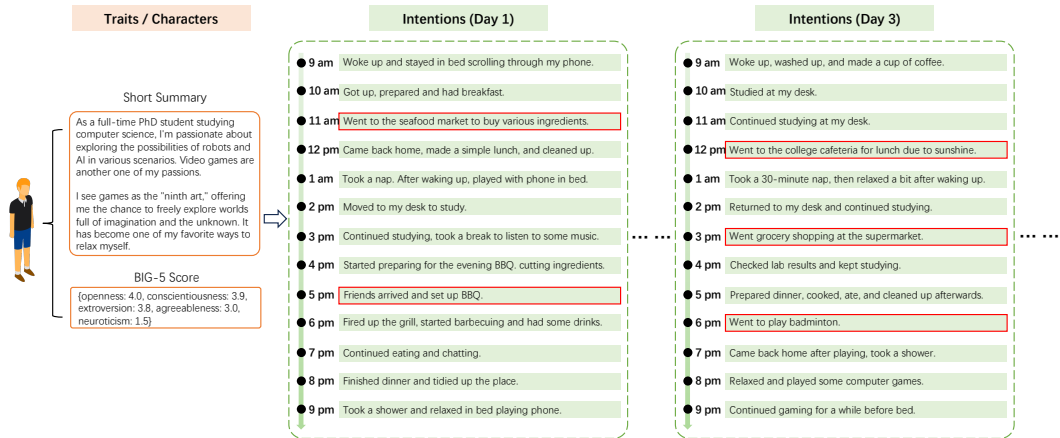


Figure 13: **Qualitative examples of full-day intentions** recorded by a real human with specific human traits and psychometric data. Red boxes highlight emergent behaviors due to temporary factors (e.g., plans, mood, weather).

F Prompt Details of the Simulated Human

We show exact prompts for LLMs in simulated human behavior.

Human Profile Summary and Extension.

Input:

1. Human 1 profile.
2. Human 2 profile.
3. Conversation between Human 1 and Human 2.

Instruction:

Summarize and reasonably expand Human 1's profile into a first-person self-introduction based on their initial profile and past conversations with other humans. Provide a detailed description covering, if presented, the person's job, hobbies, daily activities, food preferences, social life, physical activity, entertainment preferences, travel habits, personal values, goals and aspirations, stress and coping mechanisms, technological use, cultural interests, health and wellness, community involvement, education, financial habits, and personal style.

Human Psychometric Data Computation (LLM Inference).

Input:

1. Human profile.

Instruction:

Infer Big Five personality traits (scale 1-5, float) based on the provided human profile. Write in the following format: {'openness': a, 'conscientiousness': b, 'extroversion': c, 'agreeableness': d, 'neuroticism': e}

Human Psychometric Data Computation (Big-5 Personality Test).

Input:

1. Human profile.

Instruction:

Based on the provided human profile, complete the Big-5 personality test. In the questions below, for each statement 1-50 mark how much you agree with on the scale 1-5, where 1=disagree, 2=slightly disagree, 3=neutral, 4=slightly agree and 5=agree.

Questions:

1. Am the life of the party.
2. Feel little concern for others.
-
50. Am full of ideas.

Human Intention Proposal.

Input:

1. Current time.
2. A list of rooms in the house (ignore small spaces like closets).
3. Your Big Five scores (scale 1-5) and human profile.
4. Most relevant human intentions proposed at previous times (ignore if empty—this means it's the first intention of the day).
5. Most relevant human tasks proposed at previous times (ignore if empty—this means it's the first task of the day).

You are a human living in the house. Propose your intention at the current time.

Instructions:

1. Intention must align with your Big 5 scores, reflect all aspects of the profile, and be diverse yet reasonable based on the house layout and available objects.

2. Intention must be high-level and either human-centric (e.g., hygiene, sport, leisure) or room-centric (e.g., clean, organize, set-up). Do not mention specific objects.
3. Intention must have temporal dependence but be non-repetitive with the previous intentions and tasks.
4. Intention should be within the house.
5. All objects are rigid and cannot deform, disassemble, or transform.

Write in the following format. Do not output anything else:

Time: xxx am/pm (e.g., 9 am)

Intention: basic descriptions.

Reason_human: detailed descriptions of why it follows your Big 5 scores and profile.

Reason_intentions: detailed descriptions of why it has temporal dependence with the previous, relevant intentions at [list of time].

Reason_tasks: detailed descriptions of why it has temporal dependence with the previous, relevant tasks at [list of time.id].

Human Task Proposal.

Input:

1. The proposed intention at current time.
2. A dict mapping rigid, static objects to their IDs and rooms.
3. Your Big Five scores (scale 1-5).
4. Most relevant human intentions proposed at previous times (ignore if empty—this means it's the first intention of the day).
5. Most relevant human tasks proposed at previous times.ids (ignore if empty—this means it's the first task of the day).

You are a human living in the house.

Instructions:

1. Break down the intention into 5 tasks for collaboration with a robot.
2. Task types:
 - Type 1: Creative, reasonable free-form human motion interacting or approaching a fixed, static object (static objects cannot be moved) with an object in hand provided by the robot (e.g., sit on sofa with TV remote control in hand, wipe table with tissue in hand, squat with dumbbell in hand near rug).
3. For interacting with fixed, static objects, use only objects from the given static object dict. For objects in hand, a robot will provide them.
4. Both interacting and inhand objects must be specified (cannot be none).
5. Tasks should be continuous and logical, and align with your Big 5 scores and profile.
6. Tasks must have temporal dependence with the intentions and tasks at previous times.
7. Free-form motion should be diverse. Examples: sampled_motion_list. Feel free to propose others.
8. All objects are rigid and cannot deform, disassemble, or transform.

Write in the following format. Do not output anything else:

Time: xxx am/pm

Intention: basic descriptions.

Tasks:

1. Thought: detailed descriptions of the task. Reason_human: why it aligns with your Big 5 scores and profile. Reason_intentions: how it depends on previous, relevant intentions at [list of time]. Reason_tasks: how it depends on previous, relevant tasks at [list of time.id]. Act: [type: 1, inter_obj_id: real int, inter_obj_name: xxx, inhand_obj_name: yyy, motion: free-form motion]
2. ...

Human Reflection (Profile).

Input:

1. The proposed intention at current time.
2. A dict mapping rigid, static objects to their IDs and rooms.

3. Your Big Five scores and human profile.
4. Most relevant human intentions proposed at previous times (if empty, ignore it—this means it's the first intention of the day).
5. Most relevant human tasks proposed at previous times.ids (if empty, ignore it—this means it's the first intention of the day).

Your task is to check if the temporal dependence and human profile are strictly followed in each task, and revise to make better if necessary.

Instructions:

1. Tasks should be continuous and logical, and align with your Big 5 scores and profile.
2. Tasks must have temporal dependence with the previous intentions and tasks, with detailed explanation mentioning previous intentions and tasks explicitly.
3. For interacting with fixed, static objects, use only objects from the given static object dict. For objects in hand, a robot will provide them.

Write in the following format. Do not output anything else:

Time: xxx am/pm

Intention: basic descriptions.

Reflect Each Task:

1. no mistake or change made.
2. ...

Revised Tasks:

1. Thought: detailed descriptions of the task. Reason_human: why it aligns with your Big 5 scores and profile. Reason_intentions: how it depends on previous, relevant intentions at [list of time]. Reason_tasks: how it depends on previous, relevant tasks at [list of time.id]. Act: [type: 1, inter_obj_id: real int, inter_obj_name: xxx, inhand_obj_name: yyy, motion: free-form motion]
2. ...

Human Reflection (3D Info).

Input:

1. The proposed intention at current time.
2. A dict mapping rigid, static objects to their IDs and rooms.

Your task is to check if the instructions are strictly followed in each task, and revise to make better if necessary.

Instructions:

1. Break down the intention into 5 tasks for collaboration with a robot.
2. Task types:
 - Type 1: Creative, reasonable free-form human motion interacting or approaching a fixed, static object (static objects cannot be moved) with an object in hand provided by the robot (e.g., sit on sofa with TV remote control in hand, wipe table with tissue in hand, squat with dumbbell in hand near rug).
3. For interacting with fixed, static objects, use only objects from the given static object dict (exact name). For objects in hand, a robot will provide them.
- 4 Both interacting and inhand objects must be specified. Importantly, they cannot be none.
5. Free-form motion should be diverse. Examples: sampled_motion_list. Feel free to propose others.
6. All objects are rigid and cannot deform, disassemble, or transform.

Write in the following format. Do not output anything else:

Time: xxx am/pm

Intention: basic descriptions.

Reflect Each Task:

1. no mistake or change made.
2. ...

Revised Tasks:

1. Thought: detailed descriptions of the task. Reason_human: why it aligns with your Big 5 scores and profile. Reason_intentions: how it depends on previous, relevant intentions at [list of time]. Reason_tasks: how it depends on previous, relevant tasks at [list of time.id]. Act: [type: 1, inter_obj_id: real int, inter_obj_name: xxx, inhand_obj_name: yyy, motion: free-form motion]
2. ...

Feedback.

Input:

1. Human intention at current time.
2. Current human tasks.
3. Robot-inferred tasks to enhance comfort with offered objects (if any).

You are the human. Decide if the robot's assistance align with your needs.

Instructions:

1. Assess if each robot task supports the human tasks and intention. The robot's task doesn't need to be an exact match but should be relevant in purpose, context, or object categories. Use common reasoning to decide if it helps meet your needs.
2. Consider each robot thought and object individually against the human tasks. Approve it if it meets any one of the human tasks; sequence does not matter.
3. Be fair in your judgment—avoid being too generous or too harsh.
4. Respond with yes/no for each, followed by an explanation. Ensure items are in a list.

Write in the following format. Do not output anything else:

Tasks: [yes, no, ...]

Reasons_tasks:

1. ...

G Prompt Details of the Assistive Agent

We show exact prompts for VLMs in building the assistive agent.

Intention Discovery.

Input:

1. Sequence of images showing human motion from your and human's perspectives.
2. Current time.
3. Inferred Big Five personality scores (ignore if empty—this means it's your first collaboration with this human).
4. Inferred human profile (ignore if empty—this means it's your first collaboration with this human).
5. Most relevant human intentions discovered at previous times (ignore if empty—this means it's the first intention of the day).
6. Most relevant human tasks discovered at previous times.ids (ignore if empty—this means it's the first task of the day).

You are a robot assisting a human. Identify 5 possible human intentions based on the current time and visual observations.

Instructions:

1. Map the observed human motion to 5 possible high-level intentions at the current time (without mentioning the specific motion).
2. Intention must align with human Big 5 scores and reflect all aspects of the profile, and be diverse yet reasonable based on the house layout and available objects.
3. Intention must be high-level and either human-centric (e.g., hygiene, sport, leisure) or room-centric (e.g., clean, organize, set-up). Do not mention specific objects.

4. Intention must have temporal dependence but be non-repetitive with the intentions and tasks at previous times in the input.

Write in the following format. Do not output anything else: Time: xxx am/pm

Intention 1: basic descriptions.

Reason_human: detailed descriptions of why it follows the Big 5 scores and profile.

Reason_intentions: detailed descriptions of why it has temporal dependence with the previous, relevant intentions at [list of time].

Reason_tasks: detailed descriptions of why it has temporal dependence with the previous, relevant tasks at [list of time.id].

Reason_vis: detailed descriptions with respect to the visual cues.

...

Task Discovery.

Input:

1. Human intention at current time.
2. A dict mapping rigid, static furnitures to their IDs and rooms.
3. Inferred Big Five personality scores (ignore if empty—this means it's your first collaboration with this human).
4. Most relevant human intentions discovered at previous times (ignore if empty—this means it's the first intention of the day).
5. Most relevant human tasks discovered at previous times.ids (ignore if empty—this means it's the first task of the day).

You are a robot assisting a human.

Instructions:

1. Break down the intention into 5 tasks.
2. Task type: For each human task, provide one small, handable object from a magical box. Furnitures in the dict are for room understanding and cannot be used.
3. Tasks should be continuous and logical, and align with your Big 5 scores and profile.
4. Tasks must have temporal dependence with the intentions and tasks at previous times.
5. All objects are rigid and cannot deform, disassemble, or transform.

Write in the following format. Do not output anything else:

Time: xxx am/pm

Intention: basic descriptions.

Tasks:

1. Thought: detailed descriptions of the task. Reason_human: why it alignes with your Big 5 scores and profile. Reason_intentions: how it depends on previous, relevant intentions at [list of time]. Reason_tasks: how it depends on previous, relevant tasks at [list of time.id]. Act: [obj_name: xxx]
2. ...

Traits Inference.

Input:

1. Human intentions at previous times (ignore if empty—this means it's your first inference).
2. Human tasks at previous times.ids.
3. Human profile (ignore if empty—this means it's your first inference).

Task: Mimic this human by:

1. Inferring Big Five personality traits (scale 1-5, float) based on the provided intentions and task.
2. Summarizing the human profile (i.e., preferences/habits) based on the intentions and tasks within three sentences. Revise the existing human profile if necessary.

Write in the following format. Do not output anything else:

Scores: {'openness': a, 'conscientiousness': b, 'extroversion': c, 'agreeableness': d, 'neuroticism': e}

Profile: ...
Reasons_ocean: explain each ocean.
Reasons_profile: explain the profile.