

ソフトバンク株式会社 CS室 ロボット研究開発ご紹介

ソフトバンク株式会社
Chief Scientist室
沼井 隆晃



1. Chief Scientist 室 紹介
2. SOAR概要
3. RoboCup@Home 2024タスクでの ChatGPT活用
4. まとめ・課題

● デューデリジェンス業務

- 投資先となる企業の技術的価値を判断
- 理論的な知識、実装レベルの技術が必要

● 研究開発業務

- 技術調査・デバイス検証
(ロボット、無線、VR、etc.)
- 競技会・展示会の視察

● ロボットチーム

- 具体的な課題(技術・価格等)調査
のためロボット開発

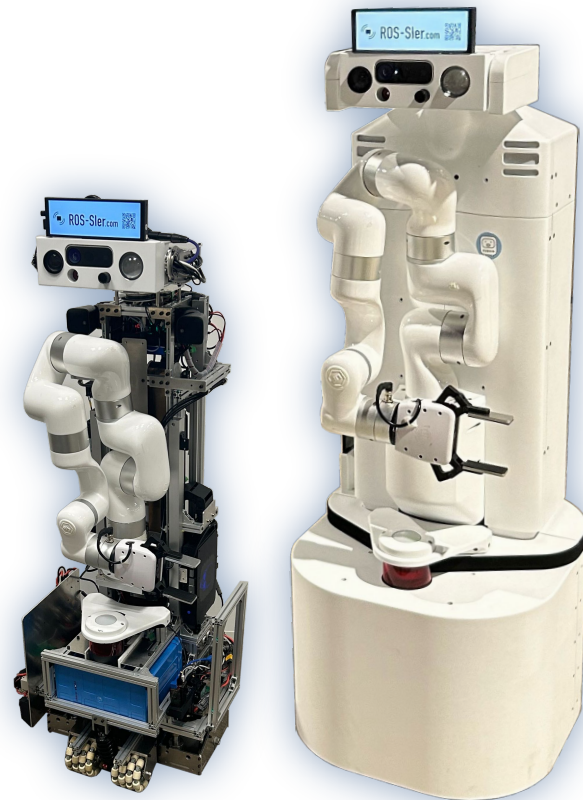


● 目的

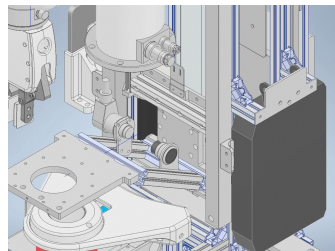
- 試作を通じた技術的課題の抽出(デューデリジェンス業務)
- 市販品の組み合わせで生産性のあるタスクを行う
 - 安価な部品で構成
- ユーザが改造できるオープンハード・オープンソース
 - モジュールごとのカスタマイズ

● 想定タスク・機能

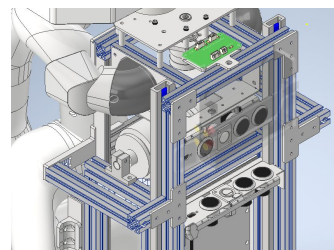
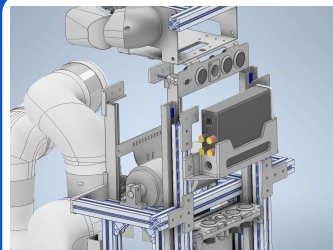
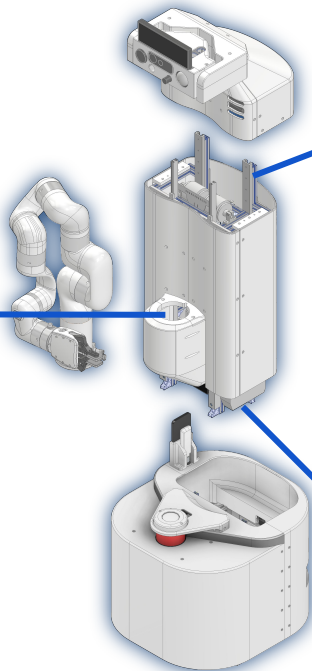
- エレベータでの社内巡回
- コンビニでの商品購入
- 音声認識・発話・道案内
- 自動充電



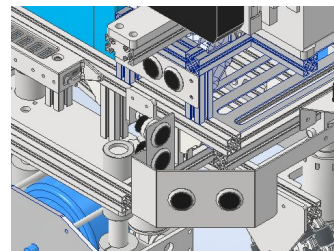
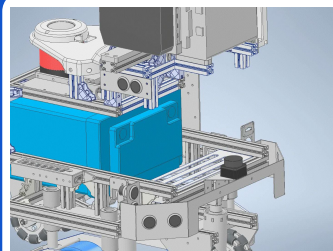
各モジュールにコネクタを装備し
個別の動作確認や組み換えが可能



アーム - 胴部



頭部 - 胴部



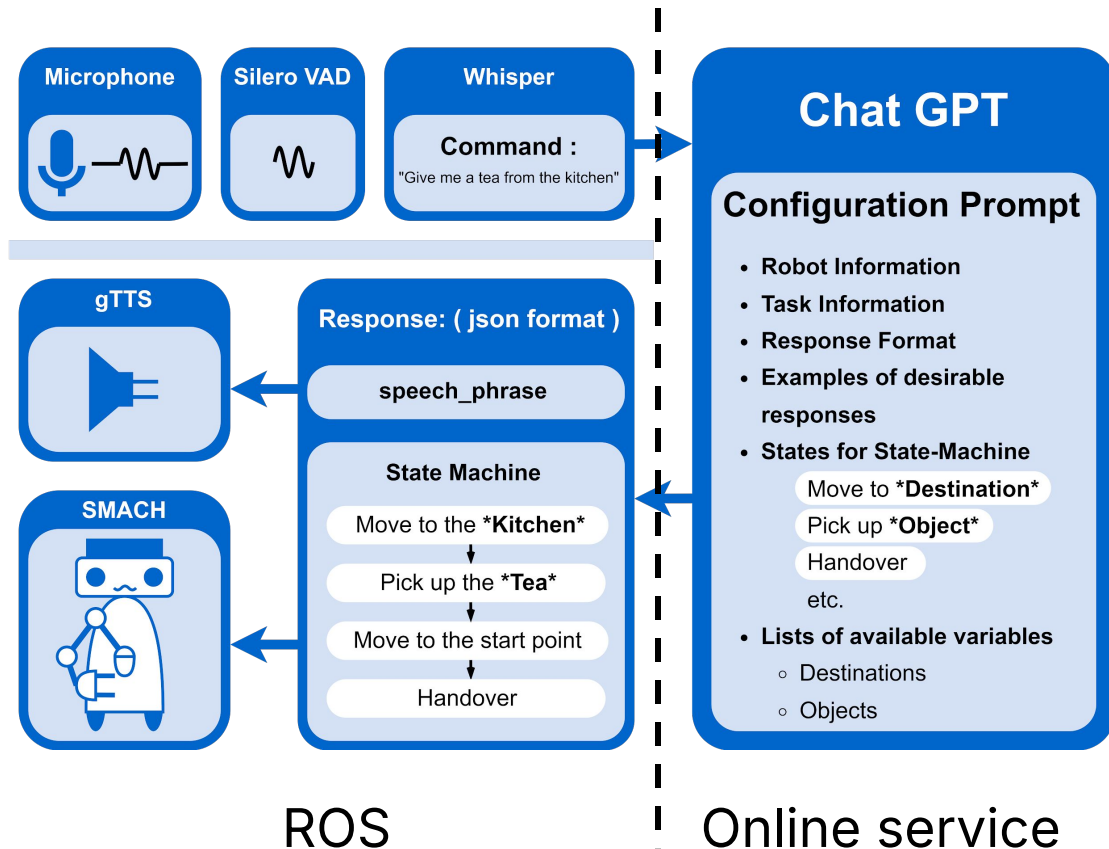
台車部 - 胴部

ChatGPTを活用した
対話や動作計画システムを
ROSノードで構築

プロンプト次第で
自然言語による対話や
動作計画が可能



競技会や展示会に向けた
デモンストレーションを開発



RoboCup@Home 2024 タスクでの ChatGPT活用

▽場所: Eindhoven(オランダ)

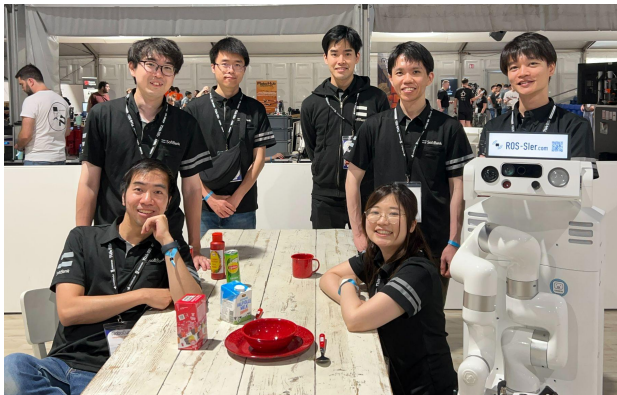
▽日程: 7/15 ~ 7/22

家庭環境を模したフィールドで

人認識・物体認識等を組み合わせた

タスクをルールの中から選択して

実施し合計スコアを競う



結果

- **4位**(19チーム中)

Rank	Team Name	Final Score	Stage 1+2
1	NimbRo@Home	8,852.47	4,918.04
2	Tidyboy-OPL	7,495.17	4,246.52
3	SocRob@Home	6,900.60	4,212.74
4	Chief Scientist Office		3,303.41
5	LAR@Home		2,819.09
6	RoboFEI		2,130.65

- **自然言語命令からの動作計画**
 - GPSRタスク
- **対話と情報抽出**
 - Receptionistタスク
- **人や物体の特徴分析**
 - Stickler for the Rulesタスク

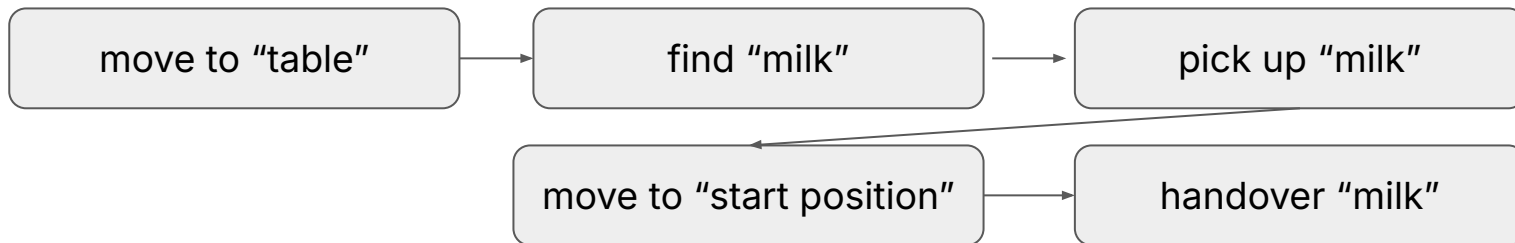


自然言語命令からの動作計画

自然言語で命令を受け取り
あらかじめ実装された機能を
組み合わせ行動を計画

命令：
「テーブルからミルクをとってきて」

ChatGPTが生成した行動計画：



ROSノード内に以下の動作を実装 ChatGPTはこれらの中から動作を 組み合わせ行動を計画

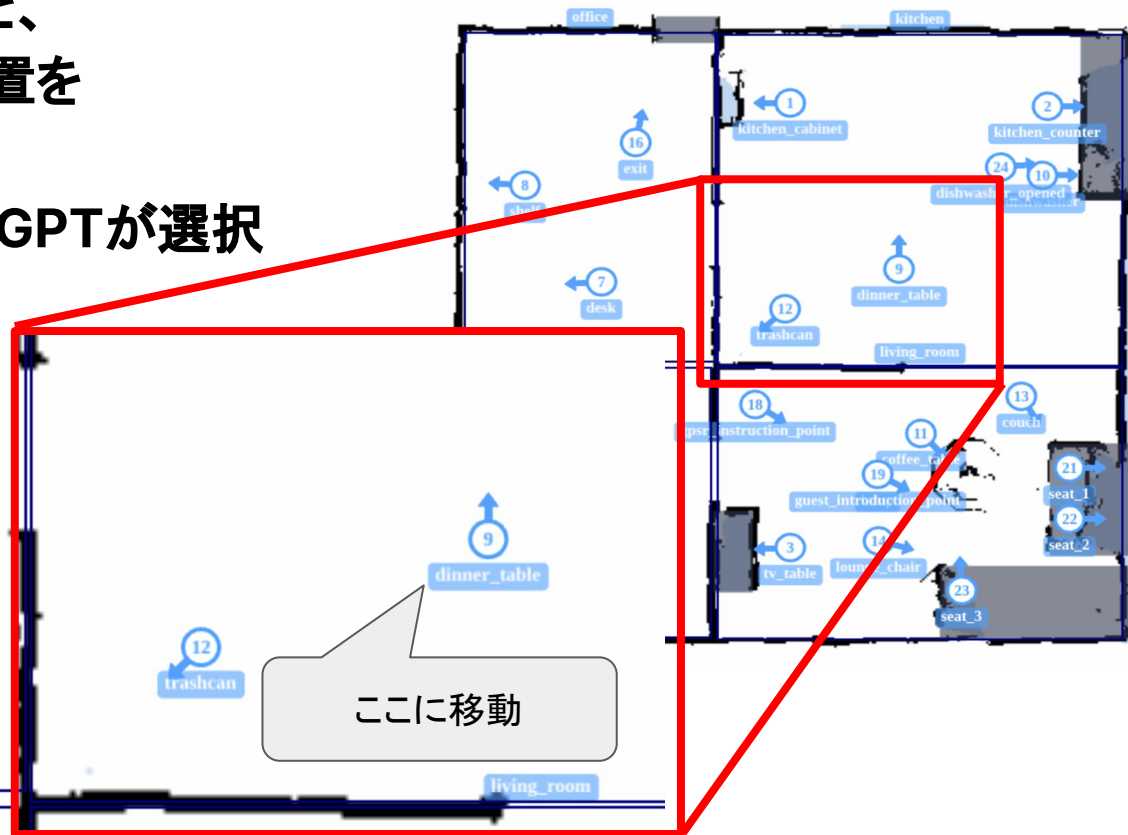
- 発話
- 人や物体の認識
- 指定した場所に移動
- 物体の把持、手渡し
- 人の追跡、先導
- 画像分析、特徴抽出

States
Movement • Move to Destination: Move robot to a specified location. ◦ Userdata Needed: ▪ goal_name: Location to move to. This can be set to a room name or destination.
Object Handling • Pick: Robot picks up an object. ◦ Userdata Needed: ▪ pick_object: Name of the graspable object or object category • Handover: Robot hands over an object. No userdata needed. ◦ Userdata is not needed. • Place: Robot places an object. ◦ Userdata is not needed.
Perception • Infer Object: Robot finds an object that fits the given abstract description. ◦ Userdata Needed: ▪ will_inform_result: Will inform the result to the other person in after this state or not. ◦ Do not execute this state twice in a row. • Find Person: Robot finds a specified person or people around the current location. ◦ Userdata Needed: ▪ person_num: Number of people to find. If set to 0, the robot will find all people matching the other criteria. ▪ person_name: Name of the person. If set to an empty string, the robot will find people without considering the name.

実際のプロンプト(抜粋)

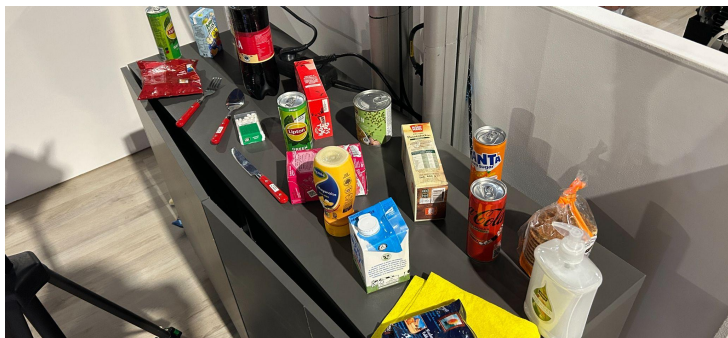
フィールドのそれぞれの部屋名と、
その部屋にある家具の名称・位置を
ROSノードに事前に登録
→ 既知の目的地の中から ChatGPTが選択

例: ディナーテーブルに移動



ルールで規定されているカテゴリと
物体名、および SOAR が把持可能かを
記載

把持する物体の認識には
RGBDカメラ+YOLOv8を使用



Graspable Objects

You can pick objects only from the following list. If you are asked to pick an object not on this list, inform the user and ask if another object is acceptable.

- **drinks:** cola, ice_tea, water, milk, big_coke, fanta, dubbelfris
- **food:** cornflakes, pea_soup, curry, pancake_mix, hagelslag, sausages, mayonaise
- **fruits:** pear, plum, lemon, orange, banana, apple
- **dishes:** bowl, plate, fork, spoon, knife, cup
- **cleaning_supplies:** sponges, soap, dishwasher_tab
- **decorations:** candle
- **snacks:** candy, pringles, tictac

You know the following objects too, but you cannot pick them because their shape is not suitable for your gripper.

- **fruits:** peach, strawberry
- **cleaning_supplies:** washcloth
- **snacks:** stroopwafel, liquorice, crisps

実際のプロンプト (抜粋)

命令に対して正しい実行順序を出力できるよう具体例を記載

例1 User: Tell me what is the lightest object on the kitchen cabinet.

```
"state_sequence": ["move_to_destination", "infer_object", "move_to_destination", "inform"],  
"userdata_sequence": [  
  {"goal_name": "kitchen_cabinet"},  
  {"will_inform_result": true},  
  {"goal_name": "gpsr_instruction_point"}]
```

結果を伝える前に move_to をいれて人に移動してから報告

例2 User: Look for a snack in the living then fetch it and bring it to Angel in the living.

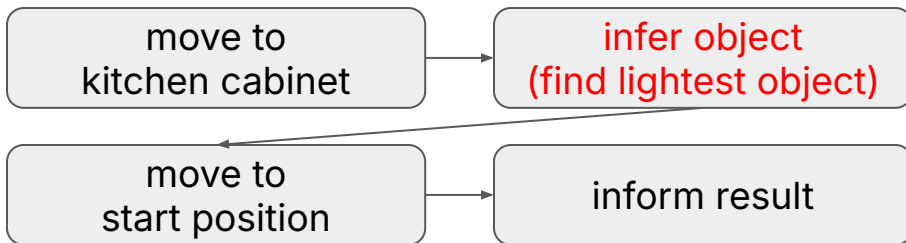
```
"state_sequence": ["move_to_destination", "pick", "find_person", "speak", "handover"],  
"userdata_sequence": [  
  {"goal_name": "living"},  
  {"pick_object": "snack"},  
  {"person_num": 1, "person_name": "Angel", "feature_explanation": "", "infer_from_person": ""},  
  {"ask_name": false, "find_location": "living", "will_inform_result": false},  
  {"speech_phrase": "Here is the snack for you."}]
```

find_person を使用して人物の位置を把握

GPSR (General Purpose Service Robot) タスク

命令: 「キッチンキャビネットにある
最も軽いものは何ですか？」

生成された動作:



動作計画には成功したが
最も軽いものの判定を誤った

定義済みの場所

“*What’s the meaning of life, for the people, and part of the solution*”

Available Destinations

You can move to the following destinations:

- "washing_machine"
- "kitchen_table"
- "coffee_table"
- "kitchen_cabinet"
- "bed"
- "shelf"
- "sofa"
- "gym_instruction_point"
- "hallway"

[illegible][illegible][illegible]

Task Planning

You will plan a sequence of states to accomplish a given command. Some states require *userdata* to execute, so you will need to provide it as well.

Response Format

All your responses should be in JSON format with the following keys:

- state_sequence**: List of states to execute to accomplish the tasks.
- userdata_sequence**: List of [key: value] required by the states. The number of *userdata* lists must match the number of state sequences. Use an empty `[]` for any state that does not need *userdata*.

対話と情報抽出

ゲストの名前や好きな飲み物を
尋ね人物情報を記憶

ChatGPTは話者の発言や画像
から以下の人物情報を抽出

- 性別
- 年代
- 髪・服の色
- 名前
- 好きな飲み物

これらの情報を集めるための
発話内容を出力



田中です

コーラが
好きです

ようこそ！
あなたの名前を
教えてください

田中さん、
好きな飲み物は
なんですか？

ChatGPTの抽出結果

性別: 男性

年代: 20代

髪の色: 黒色

服の色: グレー

名前: 田中

好きな飲み物: コーラ



Receptionistタスク

1. ゲストの名前や好きな飲み物を尋ね人物情報を記憶
2. リビングまで先導
3. ゲストをホストに紹介
4. ゲストに対し空席への着席を指示して案内を完了

▶▶4

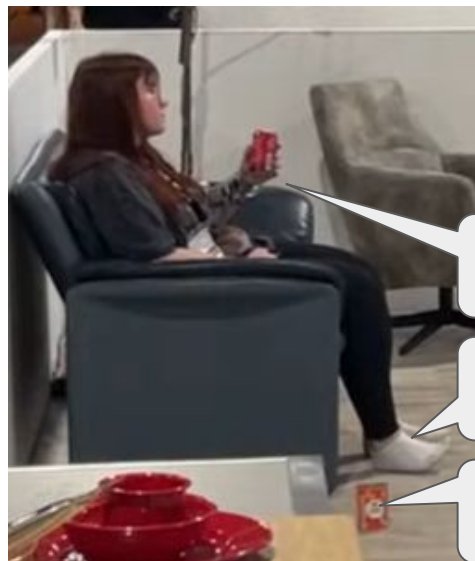


人や物体の特徴分析

画像をChatGPTに与え
以下のルールを守っているか
確認させる

- 靴を脱いでいるか
- 手に飲み物を持っているか
- 床に物を捨てていないか

ルールに違反している場合は
ルールに従うよう指示



飲み物を持っている

靴を履いていない

床に物を捨てている

Stickler for the Rulesタスク

1. 部屋を巡回し人を探す
2. ルールに違反している場合は
ルールに従うよう指示



様々なタスクに挑戦(4倍速)

人に付いていく



朝食を並べる



机の上を片付ける



注文を聞いて持ってくる



ルール違反を見つける



挙手している人の指示を聞く



今後の課題・まとめ

ChatGPTを使用する際の課題

- **会話の応答が遅い**
 - 音声区間検出+Speech2Text+ChatGPTで数秒かかる
 - 音声区間検出で終話を検知できないと不要な処理で長時間かけてしまう
- **物体・場所の指示が具体的にないと計画できない**
 - 「赤いものを持ってきて」など抽象的な命令も実行できるよう開発中
 - 指定された物体が既知のリストにない場合はChatGPTで画像認識
- **デバッグが難しい**
 - ChatGPTが画像からどの程度の情報まで理解できているか不明
 - 正しい動作を実行できているのかわかりにくい
- **動作に失敗した場合でも復帰動作をしてタスクを続けたい**
 - 失敗原因をもとに対話的に再計画するよう設計を変更中

その他の課題

- 「はい」「YES」など短文が聞き取りづらい
 - 特に環境音が多い場合に顕著で、音声区間検出や音声認識に課題
- 移動しながらの画像認識が難しい
 - 時刻の同期や移動量の考慮が必要
- navigationの課題
 - テーブルに接近する動作などで、障害物回避の機能と干渉し近づけない
 - 「アームが届かない場合に、届く位置まで台車で移動する」動作ができない
- 動作計画はあっても動作に失敗することがある
 - 例: 事前に登録したゴール地点が障害物に近い、把持対象がアーム可動範囲外、テーブル全体がカメラの視野に収まらない場合など

- **RoboCup@Home2024参加**

- 4位(17チーム中)
- それぞれのタスクについて比較的安定した動作・完成度

- **今後の課題**

- ハード
 - 電子回路など課題解決
 - モジュール化によるカスタマイズ性向上・低コスト化
- ソフト
 - ChatGPTを活用し機能拡張が容易なシステムを作る
 - アームの軌道計画改善(キャリブレーション・障害物回避など)

- **RoboCup2026年の韓国大会に向け改良**

- 改良・販売(目標:売値1000万円、10台)
- オープンソース・オープンハード

YouTube





[SPECIAL MOVIES | ソフトバンク新卒採用](#)

APPENDIX

競技動画

1. 人が指さしたバッグを認識把持
2. 障害物をよけながら
人に追従
3. 目的地に着いたらバッグを渡す
4. 元の位置に戻る

▶▶4



1. ドアから入ってきたゲストの名前や好きな飲み物を聞く
2. リビングまで案内
3. ゲストを紹介
4. ホストなどすでにリビングにいる人を紹介
5. 空いてる席に座らせる



1. 以下の物体をキッチンからテーブルに持ってくる

- a. ボウル
- b. スプーン
- c. シリアル
- d. ミルク

2. シリアルとミルクをボウルにそそぐ

▶▶4



1. テーブルに置かれている物体を 棚にしまう



段ごとに同じカテゴリの物体が置かれているので、正しい段にしまう



1. 指定地点に移動
2. 音声による指示に従って行動

▶▶4



1. 事前にマップを作っていない環境でスタート
2. 手を挙げている客に近づく
3. 注文を聞く
4. 初期位置で注文された商品を受け取る
5. 同じ客のところに帰り
注文された商品を渡す

▶▶4



1. フィールド内にいる人に対して 以下のルールを確認する

- a. 禁止された部屋に入っていないこと
- b. 靴を脱いでいること
- c. 手に飲み物を
持っていること
- d. 床にごみを
捨てていないこと

2. ルールを破っていたら 守るためにすべき行動を促す



1. テーブルに並べられた以下の食器を食洗器に入れる

- a. 皿
- b. ボウル
- c. コップ
- d. フォーク
- e. スプーン

▶▶4



1. 部屋の各所に置かれている物体を正しい位置に片づける
2. 部屋の中で手を挙げている人がいれば近づきGPSR (ランダムな命令を聞いてその通りに動くタスク) を行う



End of File