Conversational Language Models for Human-in-the-Loop Multi-Robot Coordination

Demonstration Track

William Hunt ⬤
University of Southampton
Southampton, United Kingdom
W.Hunt@soton.ac.uk

Toby Godfrey ⬤
University of Southampton
Southampton, United Kingdom
tmag1g21@soton.ac.uk

Mohammad D. Soorati ⬤
University of Southampton
Southampton, United Kingdom
M.Soorati@soton.ac.uk

ABSTRACT

With the increasing prevalence and diversity of robots interacting in the real world, there is need for flexible, on-the-fly planning and cooperation. Large Language Models are starting to be explored in a multimodal setup for communication, coordination, and planning in robotics. Existing approaches generally use a single agent building a plan, or have multiple homogeneous agents coordinating for a simple task. We present a decentralised, dialogical approach in which a team of agents with different abilities plans solutions through peer-to-peer and human-robot discussion. We suggest that argument-style dialogues are an effective way to facilitate adaptive use of each agent's abilities within a cooperative team. Two robots discuss how to solve a cleaning problem set by a human, define roles, and agree on paths they each take. Each step can be interrupted by a human advisor and agents check their plans with the human. Agents then execute this plan in the real world, collecting rubbish from people in each room. Our implementation uses text at every step, maintaining transparency and effective human-multi-robot interaction.

KEYWORDS

Mixed Human-Robot teams; Multi-robot coordination and collaboration; Large Language Models

ACM Reference Format:
William Hunt ⬤, Toby Godfrey ⬤, and Mohammad D. Soorati ⬤. 2024. Conversational Language Models for Human-in-the-Loop Multi-Robot Coordination: Demonstration Track. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 − 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Multirobot systems, while still reasonably rare in everyday life, are becoming increasingly common in domains such as agriculture [1, 11], search and rescue [5], and construction [20]. It is projected that the household robotics market will grow considerably this decade [18], leading to large numbers of robots interacting not only with their users but also with each other. This presents increasing demand for robotic systems which can flexibly adapt to new tasks without prior training. There is also a recent trend towards

generalist robotics; the same model or overarching structure can be deployed on a wide variety of platforms and tasks [14]. To this end, it is desirable to move towards a domain-agnostic coordination structure that is common across platforms so that the group can be treated as a single entity. A popular area of recent development in the AI landscape is that of Large Language Models (LLMs) [21]; agents that operate in the domain of text through next-symbol prediction. LLMs may present an effective way to interpret task or agent descriptions, as well as calling on internalised understanding and reasoning to develop the available information. LLMs can also facilitate some degree of communication between agents that helps them organise while keeping the control flow understandable and accessible to a human operator. Multi-modal AI development is now influencing robotics and other agents-based research fields [6] by using LLMs to power end-to-end approaches which can understand and use human-written inputs to inform their actions [2]. This fits into an overarching vision towards a multimodal, generalist agent which can understand and operate on text, images, and other inputs in robotics [16]. Although conversational agents such as ChatGPT are typically used to model a human-agent conversation, some works have focused on modelling a conversation between multiple agents. This is typically done through "role-playing"; an agent is told to "imagine" that it is a person with a certain role and then enters a conversation whilst assuming that role [10]. This process can be used to model internal monologue for a single agent who "talks to themself" about what they can perceive and do [9]. Conversational approaches can also use multiple personas, each of whom brings a different specialised perspective to the collective generation of text by editing the group solution to fulfill different goals that they each have for the end result [19].

Role-playing has been used for a variety of tasks including debate [3], auctions, haggling [7, 13], and checking with a human supervisor that they are not hallucinating [17]. These approaches set the conditions for a dialogue and leave the agents to talk, assuming that an intelligent solution emerges naturally from the conversation. This has been used to create a team of software developers who each write code and pass responsibilities to the next developer [8]. LLMs have also been integrated into robotic simulators for communication to organise who performs each task, allowing robots to coordinate their workspaces and decide which of them is able to reach an object [12]. A similar approach simulates agents fetching items in a house, they communicate and incorporate short-term memory to request assistance from each other on the fly [22]. Some works include diverse skillsets where each agent has different capabilities and skills, such as a work where the agents build a team with the required skills before planning and acting [4].

We present a proof-of-concept system that leverages the knowledge of pretrained models by building language-based agents which talk to each other, and with humans, using natural language. This allows agents to discuss and debate their strategies towards a collaborative solution to a high-level mission objective with observation or assistance from a human supervisor. This forms a pipeline that allows agents to take a high-level task description and autonomously perform every step of the process, from planning to assignment.

2 DEMONSTRATION

An LLM is used to create a conversation where agents build paths to be executed on hardware, which in turn can detect problems and prompt the LLM to re-plan. The pipeline uses text at every step of the process to retain deep meaning from end to end. A Python program takes human input, and calls the GPT API ("gpt-4-vision-preview") [15] for language generation. The conversation produces plans which are passed directly to the robots. In brief, the steps are: **(1) Agent Ego**: The "system message" (identity) for each agent is set to a description of each agent, plus some general guidance on how to debate; **(2) Environment Description**: We provide a flowchart-style environment model that shows agents which rooms are connected with arrows (see Fig.2c); **(3) Human Supervisor**: The supervisor presents a task to the agents; **(4) Discussion**: Agents discuss the task and plan their approach; **(5) Calling the supervisor**: Agents call the supervisor when done, or if they need help. The supervisor can ask for alterations if desired or approve the plan. Agents are added to the chat and a human supervisor is given a chat box to speak with them. When agents discuss, they start each message with their name and it is added to the log. From each agent's perspective, they perceive every other agent's message as one from the human, but the name tags allow them to understand the conversation properly. The system prompt, which the LLM considers with every message, encourages agents to negotiate with each other. This is important because the default GPT configuration is polite and rarely contradicts its interlocutor, however for collective planning agents should point out mistakes. When they reach a decision, a path of rooms for each agent is extracted and passed to hardware for execution. Two TurtleBot3 robots (see Fig.2a) are controlled with ROS2 Humble on a Raspberry Pi 4, they are each equipped with a LIDAR and an optical camera. ROS gathers data from the LIDAR and controls the differential-drive motors. The LIDAR is used for collision avoidance to protect the hardware as it moves around the environment. The optical camera is used to detect ArUco markers which indicate rooms.

We demonstrate a conversational planning system deployed on real robots to simulate autonomous waste collection. A bin is mounted on top of each robot using a 3D-printed structure (see Fig.2b), turning the robots into mobile bins (see Fig.2a). Participants can engage by typing a task into the supervisor PC, and then watching the agents converse. The participants can also offer advice to the robots or point out issues before approving the plan. The robots then move around the arena to solve the task in real time. Participants are encouraged to provide vague and unusual instructions or interrupt the execution with novel information to test the system and explore the challenges of the approach as well as its potential. A cut-down example of this process is shown in Fig.1, where the
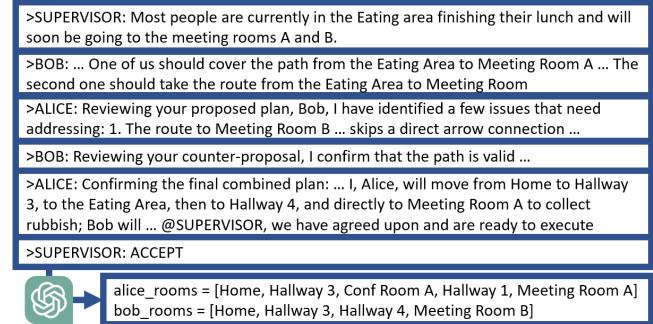

>SUPERVISOR: Most people are currently in the Eating area finishing their lunch and will soon be going to the meeting rooms A and B.

>BOB: … One of us should cover the path from the Eating Area to Meeting Room A … The second one should take the route from the Eating Area to Meeting Room

>ALICE: Reviewing your proposed plan, Bob, I have identified a few issues that need addressing: 1. The route to Meeting Room B … skips a direct arrow connection …

>BOB: Reviewing your counter-proposal, I confirm that the path is valid …

>ALICE: Confirming the final combined plan: … I, Alice, will move from Home to Hallway 3, to the Eating Area, then to Hallway 4, and directly to Meeting Room A to collect rubbish; Bob will … @SUPERVISOR, we have agreed upon and are ready to execute

>SUPERVISOR: ACCEPT

alice_rooms = [Home, Hallway 3, Conf Room A, Hallway 1, Meeting Room A]
bob_rooms = [Home, Hallway 3, Hallway 4, Meeting Room B]

Figure 1: An example dialogue where the rooms are extracted.

(a) Cleaning Robots

(b) 3D printed connection

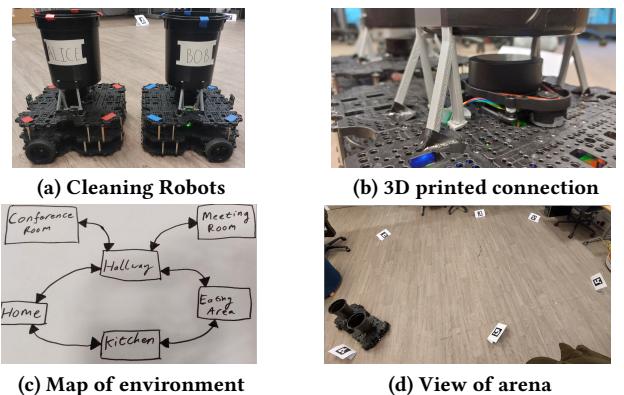**(c) Map of environment**

(d) View of arena

Figure 2: Cleaning Robots and the demonstration setup.

agents must clean up after lunch at a conference. They define two roles, correct a mistake, and decide on the division of labour. [1]

3 CONCLUSION

Robotic systems would benefit greatly from being able to understand, interpret, and utilise otherwise ignored text data. Inter-agent conversation may be a useful tool for decentralised mission planning with a human in the loop. We demonstrate conversational multi-agent coordination that allows agents to be represented with few-word names and calls on the deep knowledge of Large Language Models. The language-based approach can leverage expert opinion across many scenarios as the entire system is understandable and can be interfaced directly by a human user as much as is required. The proposed demonstration uses a language model to allow two robots to plan and execute a garbage collection task with a human supervisor in the loop and a large screen will display the conversations. Participants can interact and interrupt the system with different textual inputs to learn more about the capabilities and limitations of using language models for multi-robot coordination.

ACKNOWLEDGMENTS

---

[1]Our demo video can be found here: https://www.youtube.com/watch?v=cVCwG8aLIvI

REFERENCES

[1] Daniel Albiero, Angel Pontin Garcia, Claudio Kiyoshi Umezu, and Rodrigo Leme de Paulo. 2022. Swarm robots in mechanized agricultural operations: A review about challenges for research. *Computers and Electronics in Agriculture* 193 (2022), 106608.

[2] Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, et al. 2023. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on Robot Learning*. PMLR, 287–318.

[3] Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2023. Chateval: Towards better llm-based evaluators through multi-agent debate. *arXiv preprint arXiv:2308.07201* (2023).

[4] Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chen Qian, Chi-Min Chan, Yujia Qin, Yaxi Lu, Ruobing Xie, et al. 2023. AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors in Agents. *arXiv preprint arXiv:2308.10848* (2023).

[5] Jediah R. Clark, Mohammad Naiseh, Joel Fischer, Marise Galvez Trigo, Katie Parnell, Mario Brito, Adrian Bodenmann, Sarvapali D. Ramchurn, and Mohammad Divband Soorati. 2022. Industry Led Use-Case Development for Human-Swarm Operations. arXiv:2207.09543 [cs.RO]

[6] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. 2023. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378* (2023).

[7] Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142* (2023).

[8] Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, et al. 2023. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352* (2023).

[9] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. 2023. Inner Monologue: Embodied Reasoning through Planning with Language Models. In *Conference on Robot Learning*. PMLR, 1769–1782.

[10] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for" mind" exploration of large scale language model society. *arXiv preprint arXiv:2303.17760* (2023).

[11] Chris Lytridis, Vassilis G Kaburlasos, Theodore Pachidis, Michalis Manios, Eleni Vrochidou, Theofanis Kalampokas, and Stamatis Chatzistamatis. 2021. An overview of cooperative robotics in agriculture. *Agronomy* 11, 9 (2021), 1818.

[12] Zhao Mandi, Shreeya Jain, and Shuran Song. 2023. Roco: Dialectic multi-robot collaboration with large language models. *arXiv preprint arXiv:2307.04738* (2023).

[13] Nathalia Nascimento, Paulo Alencar, and Donald Cowan. 2023. Self-adaptive large language model (llm)-based multiagent systems. In *2023 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*. IEEE, 104–109.

[14] Open X-Embodiment Collaboration. 2023. Open X-Embodiment: Robotic Learning Datasets and RT-X Models. https://robotics-transformer-x.github.io.

[15] OpenAI. 2023. Introducing ChatGPT and Whisper APIs. https://openai.com/blog/introducing-chatgpt-and-whisper-apis

[16] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, et al. 2022. A generalist agent. *arXiv preprint arXiv:2205.06175* (2022).

[17] Allen Z Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu, Leila Takayama, Fei Xia, Jake Varley, et al. 2023. Robots that ask for help: Uncertainty alignment for large language model planners. *arXiv preprint arXiv:2307.01928* (2023).

[18] Straits Research. 2023. *Household Robotics Market: Information by Application (Robotic Vacuum Mopping, Lawn Mowing), Offering (Products, Services), and Region - Forecast till 2030.*

[19] Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei, and Heng Ji. 2023. Unleashing Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration. *arXiv preprint arXiv:2307.05300* (2023).

[20] Justin K Werfel, Kirsten Petersen, and Radhika Nagpal. 2011. Distributed multi-robot algorithms for the TERMES 3D collective construction system. In *Proceedings of Robotics: Science and Systems*. Institute of Electrical and Electronics Engineers.

[21] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2023. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *arXiv preprint arXiv:2304.13712* (2023).

[22] Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and Chuang Gan. 2023. Building Cooperative Embodied Agents Modularly with Large Language Models. *arXiv preprint arXiv:2307.02485* (2023).