Towards Human Awareness in Robot Task Planning with Large Language Models

Yuchen Liu[1,2], Luigi Palmieri[1], Sebastian Koch[1], Ilche Georgievski[2] and Marco Aiello[2]

Abstract— **The recent breakthroughs in the research on Large Language Models (LLMs) have triggered a transformation across several research domains. Notably, the integration of LLMs has greatly enhanced performance in robot Task And Motion Planning (TAMP). However, previous approaches often neglect the consideration of dynamic environments, i.e., the presence of dynamic objects such as humans. In this paper, we propose a novel approach to address this gap by incorporating human awareness into LLM-based robot task planning. To obtain an effective representation of the dynamic environment, our approach integrates humans' information into a hierarchical scene graph. To ensure the plan's executability, we leverage LLMs to ground the environmental topology and actionable knowledge into formal planning language. Most importantly, we use LLMs to predict future human activities and plan tasks for the robot considering the predictions. Our contribution facilitates the development of integrating human awareness into LLM-driven robot task planning, and paves the way for proactive robot decision-making in dynamic environments.**

I. INTRODUCTION

In recent years, with robots playing more important roles in collaborations with humans in industrial, transportation, and household environments, planning robot tasks and motions while considering motions of cohabitating humans also becomes a crucial topic [1], [2]; e.g., a robot performing household tasks while avoiding disturbing the human in the vicinity, as shown in Fig. 1. Meanwhile, a multitude of Large Language Models (LLMs) have emerged thanks to significant advancements in the research of Natural Language Processing (NLP), demonstrating the ability to generate human-like text, programming code, and service compositions with high proficiency [3]–[7]. Witnessing the capabilities of the LLMs, many researchers utilize them to tackle robot Task And Motion Planning (TAMP) problems [8]–[14]. A popular strategy is to extract common-sense knowledge from the LLMs and use it as constraints of the classical automated task planning algorithms to improve the correctness and executability of the generated plans [8], [11], [15]. Other approaches also use LLMs to generate task specifications formulated in formal planning language, e.g., Planning Domain Definition Language (PDDL) [16], such that the problem can be solved by the off-the-shelf automated task planners, as did in previous work [10], [14].

For robots operating in large and complex environments, interpreting the underlying semantic information in the

[1]Corporate Sector Research and Advance Engineering, Robert Bosch GmbH, Robert-Bosch-Campus 1, 71272 Renningen, Germany {yuchen.liu2, luigi.palmieri, sebastian.koch2}@de.bosch.com

[2]Service Computing Department, Institute of Architecture of Application Systems, University of Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany {yuchen.liu, ilche.georgievski, marco.aiello}@iaas.uni-stuttgart.de
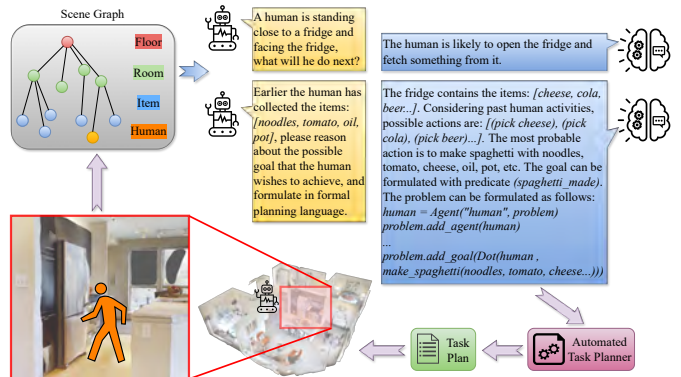
Fig. 1: An example of the human-aware robot task planning upon the *Allensville* environment [19]. A scene graph is pre-built from the environment with the floor, room, and item layers, where humans are modeled as nodes in the item layer. The edges refer to the semantic relationships. Using the scene graph as the environment representation, the human user queries a Large Language Model (LLM) to predict the next possible human activities and the goal states of the humans. An automated task planner generates a task plan considering humans as additional planning agents.

environment is an important factor leading to successful navigation [2]. Recently, many studies have investigated the potential to improve the efficiency of navigation by encoding environment topology and semantic relations in high-level representations, e.g., scene graphs [17], [18]. Particularly in high-level task planning, utilizing scene graphs can effectively enhance the planning performance by reducing the search space and the planning time [13], [14].

Surprisingly, the research on incorporating human awareness into robot task planning has not seen meaningful progress recently. Early approaches mainly consider the human-aware task planning problems as scheduling problems, where humans have their own agenda, and the robot should plan its tasks by avoiding conflicts with humans' agenda [20]–[22]. However, most of these approaches rely on pre-provided humans' agendas. Despite the enormous advancement in LLMs and their wealth of common-sense knowledge, many previous LLM-based robot task planning approaches do not consider the presence of humans. Studies that leverage LLMs in human-aware task planning are yet rarely seen [23].

Observing such a research gap, in this paper, we propose a novel approach to incorporate human awareness in robot task planning with LLMs. Our approach presents the following key contributions:

i) We introduce a novel combination of scene graphs and LLMs. To enable human awareness, we encode humans with their semantic relationships to other static objects into scene graphs, and use LLMs to predict future human activities based

Fig. 2: The architecture of the approach.

Fig. 3: Illustration of the *Allensville* scene and the corresponding 3D scene graph [19]. The scene graph has three layers: floor (red), room (green), and item (blue) layers. The humans and part of the item nodes are not visualized.
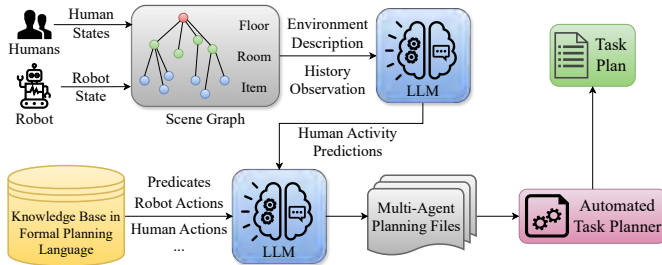
on past observations of the relationships.

ii) Considering humans as additional planning agents with the predicted activities being their goals, our approach transforms the human-aware single-robot task planning problem into a multi-agent task planning problem and can achieve better avoidance of humans during task planning. Utilizing LLMs to further ground the problem specifications into formal planning language, our approach can effectively solve the multi-agent problem jointly with the robot and humans while ensuring the plan's executability.

iii) Our contribution facilitates the development of integrating human awareness into LLM-driven robot task planning, and paves the way for proactive robot decision-making in dynamic environments.

II. RELATED WORK

A. 3D Scene Graphs

3D scene graphs are graph structures used to model large-scale physical environments. They offer an efficient hierarchical abstraction for both spatial and semantic domains. 3D scene graphs were first introduced in [19] as a structure to connect buildings, rooms, objects, and cameras on multiple layers. Following this introduction, Rosinol *et al.* [24], [25] and Hughes *et al.* [26] investigated the construction of 3D scene graphs from sensor data in dynamic environments with humans. Wald *et al.* [27] introduce semantic 3D scene graphs that represent object relationships as edges between nodes representing the object instances in the scene. Given their success, 3D scene graphs have started to be integrated into robotic systems for applications such as navigation [28]–[30] or task and motion planning [13], [14], [18].

B. LLM-based Robot Task and Motion Planning

In the past years, the development of incorporating LLMs into robot TAMP methodologies has achieved significant progress. A well-known approach involves leveraging LLMs to capture and employ their embedded rich semantic and common-sense knowledge to proficiently interpret the description of the environment written in Natural Language (NL) [9], [11]. However, many researchers have already proven that the high-level task plans directly generated by LLMs are often incorrect and not executable since the actionable knowledge is not grounded specifically in the environment [31]. Thus, various approaches are proposed for grounding LLMs' output into affordable actions [8], [12], [15], or into formal planning language such as PDDL for obtaining high-quality plans [10], [32] or even efficiently tackling long-term tasks [14].
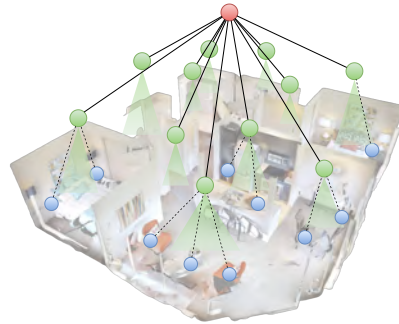
Understanding the semantic relations embedded in large environments is a crucial factor for the robots navigating in such scenarios. Scene graphs work as efficient environment representations thanks to the hierarchical structure and the compactly encoded semantic relationships [33]. Utilizing LLMs to reason over scene graphs allows the comprehension of large environments, and enables a significant performance boost in robot task planning approaches in terms of the planning success rate [13], [14].

C. Human-aware Robot Task Planning

A vast amount of studies in human-aware robot task planning have been introduced in the past decades [1], [20]–[23], [34]. Classical approaches treat it as a scheduling problem with pre-provided or predicted human schedule, and plan the tasks for the robots while trying to avoid disturbing the humans [21], [22]. However, they only consider the presence of one human in the environment. To enable crowd awareness, several recent approaches consider the crowd as a whole entity and model its own long-term behaviors. Palmieri *et al.* [35], Surma *et al.* [34] and Liu *et al.* [1] use Map of Dynamics (MoDs) to model the human motion patterns from past observations, and then plan tasks or motions for the robot considering the MoDs to achieve lower cost of traversing through human flow.

Nevertheless, to the best of our knowledge, despite the power of the LLMs, only one work contributes to LLM-driven human-aware robot task planning. Graule *et al.* [23] used a pre-trained LLM to predict the next probable human activity in terms of which object in the environment the human is likely to interact with, and plan the next actions for the robot to avoid disturbance to the human. However, they refer to the disturbance to the same-room occupation of the robot and the human, which is less realistic. Moreover, the applied task planning strategies are solely greedy policies. The vision of long-term planning is missing. Furthermore, they also consider one single human. To summarize, integrating LLMs into human-aware robot task planning with multiple cohabitating humans still remains an open research topic.

III. APPROACH

A. Problem Statement

We focus on solving human-aware robot task planning problems with LLMs. We consider a mobile robot that

moves in household environments and performs different tasks. In the same space there are also multiple humans moving around. The robot should complete its own task while avoiding disturbing the humans. Given a 3D scene graph as the environment representation, and the knowledge base of the robot's and humans' actions, the LLM will predict future human activities and generate the task specifications, such that an automated task planner can solve the problem.

In order to distinguish the *object* keyword in formal planning language (e.g., PDDL) from objects in scene graphs, in the following, we refer to the scene graphs objects as *items*. Additionally, we define *agent* as an object type in the domain specification, while *robot* and *humans* as instances of *agent* in the problem instances. We assume that humans have rational actions and do not interact with each other. We also assume the full observability of all nodes in the scene graphs.

B. Architecture

Fig. 2 illustrates the system architecture. The input consists of the states of the robot and the humans, i.e., the position, the currently performing action, etc., a pre-built scene graph as the environment representation, and a knowledge base containing the domain knowledge, such as the object types and the affordable actions of the robot and the humans with preconditions and effects. Eventually, the approach returns a high-level task plan consisting of multiple symbolic actions. A simple example of a task plan looks like: *(goto room A), (pick item X), (goto room B), (drop item X)*.

The scene graph has a hierarchical structure with three layers of nodes: floor, room, and item layers. Fig. 3 shows an example of the scene graph's layout with the static items. The room nodes are annotated with their neighboring rooms, and the item nodes contain several attributes, e.g., accessibility, states, and affordable actions. The robot and humans are integrated into the scene graph and modeled as nodes in the item layer, as illustrated in Fig. 1. An essential prerequisite for predicting future human activities is to predict the semantic relationships between the human node and the item nodes in the vicinity of the human. This can be done by utilizing existing methods such as [27] that predict such relationships and encode them as edges in scene graphs.

As the scene graph is formulated in programming code, after obtaining the extended scene graph with humans, it can be easily integrated into the prompt for the LLM for predicting future human activities with probabilities. For more accurate prediction, we take the history of the observations into account, i.e., the historical states of humans as well as the items that humans have interacted with in the past.

The core idea of our approach is to transform the human-aware single-robot task planning problem into a multi-agent task planning problem in such a way that humans are considered as additional planning agents, with the predicted activities as their goals, as did previously in many studies that consider human motion predictions as planning problems [36]. Using the available predicates and the action knowledge from the knowledge base, the LLM can transform the predicted human activities into the goal states of humans and formulate them in formal planning language. As a result, we obtain the necessary multi-agent planning specifications.

Algorithm 1: Transforming Human Awareness into Multi-Agent Task Planning Problem with LLM

Data: *domain*, *problem* Π, *K*, *SG*, *LLM*
Result: π

1   $object\_types, predicates, actions \leftarrow$
    $LLM(K.get\_knowledge(domain))$
2   Add *object_types*, *predicates*, and *actions* to *domain*
3   $problem.add\_agent(SG.get\_robot\_node())$
4   **for** $n_i \in SG$ **do**
5      $problem.set\_init\_state(n_i)$
6   **end**
7   **for** $n_h \in SG^t,\ t \in [1, t_n], n \in \mathbb{Z}^+$ **do**
8      $problem.add\_agent(n_h)$
9      $E_h^t, I_h^t \leftarrow SG^t.get\_edges\_and\_neighbors(n_h)$
10     $\boldsymbol{H}_h^{1:t_n} \leftarrow \{E_h^{1:t_n} \cup I_h^{1:t_n}\}$
11     $\{g_{h,m}^{t_{n+1}} \sim p_m \mid \sum_{m=1}^{M} p_m(g_{h,m}^{t_{n+1}} \mid \boldsymbol{H}_h^{1:t_n}) = 1, M \in \mathbb{Z}^+\} \leftarrow$
      $LLM(n_h, \boldsymbol{H}_h^{1:t_n})$
12   **if** *no predicates or actions correspond to* $g_{h,1:M}^{t_{n+1}}$ **then**
13       $predicates_h', actions_h' \leftarrow LLM(g_{h,1:M}^{t_{n+1}})$
14   Add $predicates_h'$ and $actions_h'$ to *domain*
15   **end**
16     $problem.add\_goal(n_h, \arg\max_m p_m(g_{h,m}^{t_{n+1}}))$
17   **end**
18   $\pi \leftarrow \Pi(domain, problem)$

Querying an off-the-shelf automated task planner, the multi-agent task planning problem can be solved by generating a symbolic task plan.

C. Transforming Human Awareness into Multi-Agent Task Planning Problem

While automated planning techniques deliver optimal solutions, tackling task planning problems with dynamic objects such as humans still remains a challenging topic. Therefore, we propose a method to convert dynamic objects into planning agents. Knowing the action knowledge, the initial states, and the desired goal states of these agents, it is achievable to avoid conflicts between the robot and the other agents, i.e., humans, when planning jointly, thus enabling human awareness.

Algorithm 1 is about how the single-robot task planning problem is transformed into a multi-agent task planning problem. It takes the specifications of *domain* and *problem*, the automated task planner Π, the knowledge base *K*, the scene graph *SG*, and the *LLM* as inputs. Firstly, the *LLM* extracts the domain knowledge about object types and actions from the knowledge base *K*, formulates them in formal planning language, and adds to the domain specification (L. 1-2). Subsequently, the algorithm loops through the item nodes n_i and the human nodes n_h from the scene graph *SG* considering all time steps $t \in [1, t_n]$. In the loop, the algorithm adds the initial states of each item (L. 5), as well as the historical observations of the humans (L. 7-10), including the item nodes I_h^t which the humans have interacted in the past, and the semantic relationships (i.e., edges E_h^t) between humans and the items. Based on past observations of humans $\boldsymbol{H}_h^{1:t_n}$, the *LLM* can predict *M* numbers of possible future activities in terms of the desired goal states $g_h^{t_{n+1}}$ for the humans with corresponding probabilities $p(g_h^{t_{n+1}} \mid \boldsymbol{H}_h^{1:t_n})$ (L. 11). If there exist no predicates or actions in the domain that

correspond to the predicted goal states, we can again leverage the rich common-sense knowledge from the *LLM* to generate new ones and append them to the domain (L. 12-15), as we did in our previous work [14]. Finally, we assign the goal with the highest probability to the corresponding human (L. 16). As a result, the task plan π can be obtained by solving the problem with the planner Π (L. 18).

IV. CONCLUSION

We propose a novel approach for enabling human awareness in LLM-based robot task planning, where we integrate humans as nodes in a hierarchical scene graph for a unified environment representation, and use LLMs to predict future human activities based on the semantic relationships between humans and other static objects from the scene graph. A key contribution is the transformation of the single-robot task planning problem into a multi-agent problem, where humans are considered as additional planning agents with predicted activities as their goal states. Human awareness can be achieved by jointly planning tasks for the robot with humans to avoid disturbances.

We leave for future work to experiment with the proposed architecture and algorithm. In particular, we plan to implement replanning functionalities when encountering failures or false predictions, evaluate the approach against other competitive baselines with various tasks and scenarios, and prepare experiments in photo-realistic simulations and real-world applications.

ACKNOWLEDGMENTS

REFERENCES

[1] Y. Liu, L. Palmieri, I. Georgievski, and M. Aiello, "Human-flow-aware long-term mobile robot task planning based on hierarchical reinforcement learning," *IEEE Robotics and Automation Letters*, 2023.

[2] Y. Goel, N. Vaskevicius, L. Palmieri, N. Chebrolu, K. O. Arras, and C. Stachniss, "Semantically informed mpc for context-aware robot exploration," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 11 218–11 225.

[3] J. Achiam *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[4] A. Chowdhery *et al.*, "Palm: Scaling language modeling with pathways," *Journal of Machine Learning Research*, vol. 24, no. 240, pp. 1–113, 2023.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[6] H. Touvron *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[7] M. Aiello and I. Georgievski, "Service composition in the chatgpt era," *Service Oriented Computing and Applications*, pp. 1–6, 2023.

[8] W. Huang, P. Abbeel, D. Pathak, and I. Mordatch, "Language models as zero-shot planners: Extracting actionable knowledge for embodied agents," in *International Conference on Machine Learning*. PMLR, 2022, pp. 9118–9147.

[9] C. H. Song, J. Wu, C. Washington, B. M. Sadler, W.-L. Chao, and Y. Su, "Llm-planner: Few-shot grounded planning for embodied agents with large language models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2998–3009.

[10] B. Liu *et al.*, "Llm+ p: Empowering large language models with optimal planning proficiency," *arXiv preprint arXiv:2304.11477*, 2023.

[11] Y. Ding, X. Zhang, C. Paxton, and S. Zhang, "Task and motion planning with large language models for object rearrangement," *arXiv preprint arXiv:2303.06247*, 2023.

[12] M. Ahn *et al.*, "Do as i can, not as i say: Grounding language in robotic affordances," *arXiv preprint arXiv:2204.01691*, 2022.

[13] K. Rana, J. Haviland, S. Garg, J. Abou-Chakra, I. Reid, and N. Suenderhauf, "Sayplan: Grounding large language models using 3d scene graphs for scalable robot task planning," in *Conference on Robot Learning*. PMLR, 2023, pp. 23–72.

[14] Y. Liu, L. Palmieri, S. Koch, I. Georgievski, and M. Aiello, "Delta: Decomposed efficient long-term robot task planning using large language models," *arXiv preprint arXiv:2404.03275*, 2024.

[15] Y. Chen, J. Arkin, Y. Zhang, N. Roy, and C. Fan, "Autotamp: Autoregressive task and motion planning with llms as translators and checkers," *arXiv preprint arXiv:2306.06531*, 2023.

[16] D. McDermott *et al.*, "Pddl-the planning domain definition language," 1998.

[17] Z. Ravichandran, L. Peng, N. Hughes, J. D. Griffith, and L. Carlone, "Hierarchical representations and explicit memory: Learning effective navigation policies on 3d scene graphs using graph neural networks," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9272–9279.

[18] C. Agia *et al.*, "Taskography: Evaluating robot task planning over large 3d scene graphs," in *Conference on Robot Learning*. PMLR, 2022.

[19] I. Armeni *et al.*, "3d scene graph: A structure for unified semantics, 3d space, and camera," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[20] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and R. Chatila, "Toward human-aware robot task planning," in *AAAI spring symposium: to boldly go where no human-robot team has gone before*, 2006.

[21] M. Cirillo, L. Karlsson, and A. Saffiotti, "Human-aware task planning: An application to mobile robots," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 1, no. 2, pp. 1–26, 2010.

[22] U. Köckemann, "Constraint-based methods for human-aware planning," Ph.D. dissertation, Örebro university, 2016.

[23] M. A. Graule and V. Isler, "Gg-llm: Geometrically grounding large language models for zero-shot human activity forecasting in human-aware task planning," *arXiv preprint arXiv:2310.20034*, 2023.

[24] A. Rosinol, A. Gupta, M. Abate, J. Shi, and L. Carlone, "3d dynamic scene graphs: Actionable spatial perception with places, objects, and humans," in *Robotics: Science and Systems (RSS)*, 2020.

[25] A. Rosinol *et al.*, "Kimera: From slam to spatial perception with 3d dynamic scene graphs," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1510–1546, 2021.

[26] N. Hughes, Y. Chang, and L. Carlone, "Hydra: A real-time spatial perception system for 3D scene graph construction and optimization," 2022.

[27] J. Wald, H. Dhamo, N. Navab, and F. Tombari, "Learning 3d semantic scene graphs from 3d indoor reconstructions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[28] Z. Seymour, N. C. Mithun, H.-P. Chiu, S. Samarasekera, and R. Kumar, "Graphmapper: Efficient visual navigation by scene graph generation," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 4146–4153.

[29] Q. Gu *et al.*, "Conceptgraphs: Open-vocabulary 3d scene graphs for perception and planning," *arXiv preprint arXiv:2309.16650*, 2023.

[30] H. Chang *et al.*, "Context-aware entity grounding with open-vocabulary 3d scene graphs," in *7th Annual Conference on Robot Learning*, 2023.

[31] V. Pallagani *et al.*, "On the prospects of incorporating large language models (llms) in automated planning and scheduling (aps)," *arXiv preprint arXiv:2401.02500*, 2024.

[32] T. Silver, V. Hariprasad, R. S. Shuttleworth, N. Kumar, T. Lozano-Pérez, and L. P. Kaelbling, "Pddl planning with pretrained large language models," in *NeurIPS 2022 Foundation Models for Decision Making Workshop*, 2022.

[33] B. Chen *et al.*, "Open-vocabulary queryable scene representations for real world planning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 11 509–11 522.

[34] F. Surma, T. P. Kucner, and M. Mansouri, "Multiple robots avoid humans to get the jobs done: An approach to human-aware task allocation," in *2021 European Conference on Mobile Robots (ECMR)*. IEEE, 2021, pp. 1–6.

[35] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras, "Kinodynamic motion planning on gaussian mixture fields," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 6176–6181.

[36] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.