AutoGPT+P: Affordance-based Task Planning using Large Language Models

Timo Birr, Christoph Pohl, Abdelrahman Younes and Tamim Asfour

***Abstract*— Robots need to understand their environment and plan actions to complete tasks effectively. While recent approaches combine Large Language Models (LLMs) with traditional planning algorithms to improve reasoning capabilities, they face several limitations: they can't easily dynamically adapt to changes in the environment, may generate unreliable plans due to LLM hallucinations, and are constrained by the closed-world assumption of classical planners. We propose *AutoGPT+P*, which combines an affordance-based scene representation with a planning system. By deriving planning domains based on affordances - action possibilities of objects offered to an agent in the environment - *AutoGPT+P* enables symbolic planning with arbitrary objects. Given a task description specified by the user in natural language, it generates and executes plans that handle incomplete information by exploring the scene or suggesting alternatives. *AutoGPT+P* achieves a 98% success rate on the SayCan instruction set and 79% on a new dataset of 150 complex scenarios, including tasks with missing objects. The code and dataset are available at `https://git.h2t.iar.kit.edu/sw/autogpt-p`.**

I. Introduction

Instructing robots in natural language is a very intuitive way to interact with an assistive robot. However, mapping these often ambiguous commands into precise, physically feasible actions that are grounded in the real world and constrained by physical limitations remains challenging. While recent approaches use LLMs to directly generate plans from the user-specified task in natural language, they are limited by the constrained reasoning capabilities of current LLMs and struggle especially with tasks that require a long sequence of steps to complete [4]. To mitigate the problem of limited logical reasoning capabilities of LLMs, approaches such as LLM+P [3], have combined the natural language understanding capabilities of LLMs with classical planning methods. However, these systems are limited by the closed-world assumption, which requires that all objects in the scene needed for the task are already known. *AutoGPT+P* addresses these issues by combining dynamic exploration of the environment and the ability to suggest alternatives when needed objects are not immediately available with an LLM+P-inspired planning system under the closed-world assumption. In addition, it improves LLM+P by automatically detecting semantic and syntactic errors in the LLM-generated goal and prompting the LLM to correct its errors.

II. Method

AutoGPT+P, described in [2] is a novel system that integrates an affordance-based scene representation with a task planning framework to allow a robot to plan and execute a sequence of actions to solve a task specified by the user in natural language. By exploiting affordances – the action possibilities offered by objects in a scene – *AutoGPT+P* dynamically generates planning domains that allow flexible task planning independent of concrete object classes.

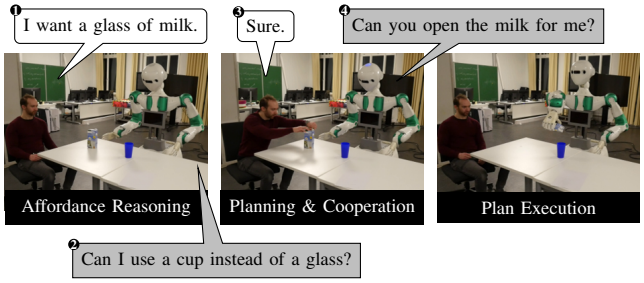

Fig. 1. ARMAR-DE solves the task given in natural language by detecting the objects within the scene, reasoning about their affordances, generating a plan to solve the task including asking for help, and finally executing the plan.

The system relies on a two-step process to symbolically represent scenes. First, objects in the scene are detected using standard object detection methods. Then, an Object Affordance Mapping (OAM) is conducted to associate objects with predefined affordances, such as a knife affordance for cutting. This object-affordance mapping is generated automatically by ChatGPT, using its common sense reasoning capabilities. Several strategies, such as including the logical combination of binary yes/no prompts, are employed to improve precision and recall during OAM generation.

Building on this affordance-based representation, *AutoGPT+P* uses a feedback loop, as seen in Figure 2, that iteratively executes three main tools until a plan is found: `Plan`, `Explore`, and `Suggest Alternative`. The `Plan Tool` (see Figure 3) generates optimal plans

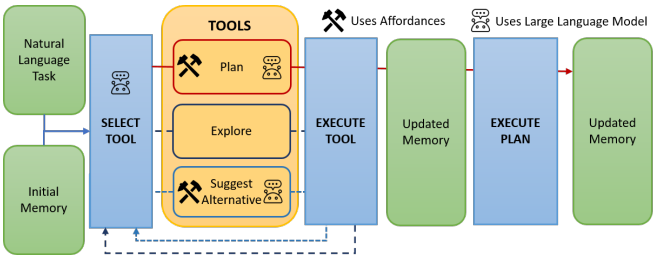

Fig. 2. Overview of the *AutoGPT+P* feedback loop. Green boxes symbolize inputs and outputs, while blue boxes symbolize discrete steps of the process. The tool selection process chooses one of the tools in the yellow *Tools* box.

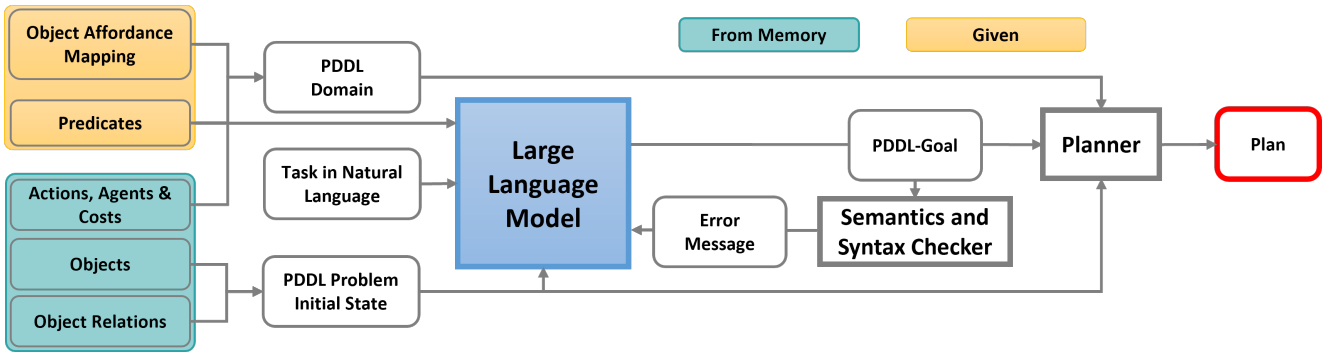

Fig. 3. Overview of the `Plan Tool`. Rounded boxes represent the input and the output of the components that are represented as rectangles.

under the closed-world assumption by combining LLMs with symbolic planners. First, it converts the user-specified task in natural language into a goal specified in the Planning Domain Definition Language (PDDL). As context, the LLM receives the initial scene state derived from the affordance-based scene representation. To further increase reliability, the planning tool incorporates an automatic self-correction mechanism that detects errors, such as semantically contradictory or syntactically invalid goal states, e.g., the same object such as an apple being simultaneously in the robot's and the user's hand, and prompts the LLM to iteratively revise its output. Finally, a symbolic planner generates a plan using the goal state, the initial state, and the planning domain. This domain is dynamically derived based on the available robot and human capabilities, allowing for easy adaptation to changing robot capabilities or humans with constrained motor skills.

If not all needed objects are in memory, the `Explore Tool` can be called to update the scene representation by searching for missing objects at the specified location. If a desired object is not in the scene at all, the `Suggest Alternative Tool` can identify substitutes for unavailable objects by reasoning about shared affordances and guiding the LLM through a structured decision process. The feedback loop is designed to allow *AutoGPT+P* to plan without being restricted by the closed-world assumption of symbolic planners and thus the `Plan Tool`.

III. Evaluation

We evaluated our `Plan Tool` as well as the full feedback loop in simulation. Our `Plan Tool` achieved a 98% success rate on the SayCan [1] set of instructions. Here our automatic error correction had a significant impact on the success rate, improving it by 20%. The full feedback loop was evaluated on a newly created dataset, which included missing object tasks requiring alternative suggestions and more implicit user commands, and we observed a reduced success rate of 79%. *AutoGPT+P* had the most problems interpreting implicit user commands, which require reasoning about the scene context and using the tools in the optimal order to solve the user task.

In experiments on the robot, we validated that the generated plans were not only logically feasible, but could actually be successfully executed on the real robot. However, challenges such as false object detection and limited robustness during execution were observed.

IV. Conclusion

AutoGPT+P combines the natural language processing capabilities of LLMs with the logical problem solving capabilities of symbolic planners. By enabling dynamic adaptation to incomplete knowledge about the environment and missing objects, the system demonstrates its potential for practical use in complex tasks. In addition, dynamic domain generation from agent capabilities allows for flexible deployment on robots with different capabilities. However, the system shows weaknesses in uncertainty assessment: if a task given by the user is ambiguous, the system should ask for clarification instead of starting the task immediately. Furthermore, for real-world deployment, real-time feedback is urgently needed to retry failed actions or dynamically adjust plans.

Acknowledgments

The research leading to these results has received funding from the European Union's Horizon Europe programme under grant agreement No. 101070292 (HARIA) and No. 101070596 (euROBIN) anfrom the Carl Zeiss Foundation through the JuBot project and was (partially) supported by the German Federal Ministry of Education and Research (BMBF) under the Robotics Institute Germany (RIG).

References

[1] Michael Ahn et al. *Do As I Can, Not As I Say: Grounding Language in Robotic Affordances*. 2022. arXiv: `2204.01691 [cs.RO]`.

[2] Timo Birr et al. "AutoGPT+P: Affordance-based Task Planning using Large Language Models". In: *Proceedings of Robotics Science and Systems. Delft, Netherlands*. 2024.

[3] Bo Liu et al. "LLM+P: Empowering Large Language Models with Optimal Planning Proficiency". In: (2023). arXiv: `2304.11477 [cs.AI]`.

[4] Karthik Valmeekam et al. *Large Language Models Still Can't Plan (A Benchmark for LLMs on Planning and Reasoning about Change)*. 2023. arXiv: `2206.10498 [cs.CL]`.