Faculty Of Engineering and Natural Sciences
Department of Electrical and Electronic Engineering
April 17, 2025

EE422/ CS421 Introduction to Robotics

Progress Report: Week 2

Almira Demirkıran 200202150
E-mail: almira.demirkiran@std.antalya.edu.tr

Mehmet Tosun 190202009
E-mail: mehmet.tosun@std.antalya.edu.tr

Abdallah Rasthy 200201125
E-mail: ali.rashty@std.antalya.edu.tr

Rustam Akhmedov 210201113
E-mail: rustam.akhmedov@std.antalya.edu.tr

This paper distils a week's investigation into robotic task planning with Large Language Models (LLMs), highlighting their transformer architecture, natural language capabilities, and robot operating system compatibility. Drawing from existing literature, it explores the capacity of LLMs to enhance robotic autonomy, highlights some challenges, and maps out possible future lines of research.

Study Context and Outcomes

The use of Large Language Models (LLMs) in robotics is a revolutionary transformation where robots can understand and execute complex tasks using natural language instructions. Last week, research involved examining the technical basis of LLMs, their role in task planning, and their implementation in robotic systems. LLMs, with their training on huge text corpora, are especially good at generating text, reasoning, and planning and are, therefore, very appropriate for transforming high-level human commands into concrete robotic actions. This report summarises these conclusions, highlighting the transformer architecture, task decomposition as a process, and ROS acting as a conduit between LLMs and robot hardware. By citing recent studies, it provides an academic summary of the potential and constraints of LLMs in robotics.

The research during the week concerned itself with comprehending how LLMs can complement robotic capabilities by processing natural language. One of the key areas of emphasis was defining and functioning of LLMs, which are AI models trained on large amounts of data to perform tasks like answering questions, generating text, and planning to act. Models such as GPT-4 and Google's Gemini were studied for their ability to process natural language instructions with human-like competence, enabling natural human-to-robot interaction. This capability is particularly valuable to robotics, in which robots must interpret vague or unclear directions, such as "make a meal," and translate them into doable steps.

One of the most important research elements was transformer architecture, which LLMs are based upon. Introduced by Vaswani et al. (2017), transformers utilize self-attention mechanisms to prioritize words in a sentence to retain contextual knowledge. By training over billions of pieces of text examples, LLMs acquire strong sense of logic, structure of language, and common sense. The research highlighted how such a type of architecture makes it possible for LLMs to interpret directions and generate reasonable plans, multimodal models like GPT-4V building on capabilities through the incorporation of visual and textual inputs. However, computational expense and transparency within transformers were points that still had to be looked into.

LLM task planning in robotics allows autonomous high-level command decomposition into structured sequences of actions, beyond the limit of rule-based methods. High-level commands like 'clean the kitchen' are mapped into hierarchical sub-tasks via natural language comprehension, planning modules, and action execution layers. The study demonstrated the applicability of LLMs in dynamic environments, outperforming the conventional static methodologies.

The integration of LLMs and the Robot Operating System (ROS) was also explored as an efficient mechanism for plan transformation to action. ROS, as a middleware software framework, ensures end-to-end transmission of plans' outputs by LLMs with robotic systems smoothly. Newer studies, e.g., Vemprala et al. (2024), show that LLMs can generate ROS executable code employing APIs like OpenCV in order to guide through the surroundings. These models, including LLM-Planner and EmbodiedGPT, were tested to see if they could combine vision and language processing for real-time task execution. The outcomes observed the ability of LLMs to enhance robotic generalization, while limitations in API-based systems were reported.

Opportunities and Challenges

The research this week revealed huge potential for LLMs in robotics. Through enabling natural human-robot interaction, they enable non-experts to issue complex commands, widening the applicability of robotic systems. Pretrained LLMs also have zero-shot generalization, carrying over to new tasks without large-scale retraining, which accelerates development. The integration of vision-language models also enhances robots' environmental data processing capabilities, allowing embodied tasks in real-world settings.

In spite of these benefits, a number of challenges were recognized. The black-box nature of LLMs is concerning in terms of transparency, especially for safety-critical systems where it is necessary to understand the model's rationale. The gap between simulation and real-world deployment is still a challenge since robots have to generalize to unpredictable environments. Additionally, the computational requirement of transformers makes their deployment on resource-limited robotic platforms limited, and there is a need to research lightweight models. These challenges emphasize the need for more robust architectures, standardized assessment protocols, and efficient fine-tuning techniques

Conclusion

The week's research illuminates the revolutionary potential of LLMs for robotic task planning and execution, driven by their transformer architecture, natural language understanding, and ROS integration. Through enabling robots to execute and respond to complex instructions, LLMs promote autonomy and responsiveness in dynamic environments. However, overcoming transparency issues, real-world robustness, and computational efficiency is necessary to unlock their full potential.

References

Vaswani, A., Shazeer, N., Parmar, N., Uszoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 30.

Vemprala, S., & Kapoor, R. (2024). ChatGPT in Robotics: Leveraging NLP for Task Planning. Mewburn Ellis.

Liu, Y., & Zhang, H. (2023). EmbodiedGPT: A Model for Embodied AI. MDPI.

Macenski, S., Foote, T., Gerkey, B., Lalancette, C., & Woodall, W. (2022). Robot Operating System 2: Design, Architecture, and Uses in the Wild. Science Robotics, 7(66).

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F., … & McGrew, B. (2023). GPT-4 Technical Report. arXiv preprint arXiv:2303.08774.