Your Project Title

Your Name

***Abstract*—Write your abstract here.**

***Index Terms*—Write up to three keywords about your work.**

I. Introduction

Large Language Models (LLMs) have been shown to exhibit impressive abilities in reasoning, language understanding, and task execution. This project explores how an LLM can serve as the core of a cognitive architecture embodied in a simulated 3D robot. By assigning the agent physical tasks in a realistic environment, we aim to investigate how such systems can perceive, plan, act, and adapt in grounded, interactive settings.

A. Motivation

To fully explore the cognitive capabilities of large language models, it is essential to situate them in environments that demand embodied, goal-directed interaction. This project investigates how LLMs function as the core of a cognitive architecture within a simulated 3D robotic setting, where the agent must physically manipulate objects and execute tasks in the real world.

Unlike static text-based benchmarks, a robotic context enables examination of embodied reasoning, memory-guided action, and the capacity for grounded, anticipatory planning. By focusing on a single agent operating in a realistic environment – such as a kitchen – we can probe how LLMs reason about spatial layouts, interpret user instructions, and adapt plans based on feedback.

Importantly, the robot's (and by extension, the LLM agent's) planning and decision-making capabilities take on a more grounded role, reasoning in natural language about the physical consequences of movements and manipulative actions. This opens new directions for exploring embodied cognition and could inform future developments in cognitive robotics, assistive AI, and multi-modal agent systems.

B. Problem Statement

Describe the problem you are addressing in the work.

C. Proposed Approach

Write a short summary of your proposed approach.

II. Related Work

This project draws on multiple domains:

A. Cognitive Architectures

Cognitive architectures such as SOAR, which integrates symbolic reasoning with perceptual and motor systems for cognitive robotics [1], and ACT-R, which offers a hybrid symbolic-subsymbolic framework for modeling human cognition [2].

B. Embodied AI and Reinforcement Learning

Embodied-AI and reinforcement-learning agents have been used within simulation frameworks like OpenAI Gym, DeepMind Lab, and MineDojo. Generative models have been used to power social agents that produce believable simulacra of human behaviour, as exemplified by Park et al. [3]. Complementary work on open-ended embodied agents, such as Voyager, a GPT-4-powered agent that incrementally builds a skill library while autonomously exploring Minecraft [4], demonstrates lifelong learning in complex environments.

C. Vision-Language-Action Policies

Generalist vision-language-action policies like Octo map multimodal observations to diverse robot actions and tasks [5]. Our project extends these directions by introducing an explicit cognitive architecture that equips an embodied robot with hierarchical memory, anticipatory reasoning, and explainable decision-making.

D. LLM-Driven Robot Control

Recent research has also explored LLM-driven control of simulated or real robots capable of translating natural-language instructions into sensorimotor actions. Notable examples include:

- RT-2, a vision-language-action model that transfers web-scale knowledge to real-world robotic control [6]
- PaLM-E, an embodied multimodal language model that fuses a large-scale LLM with visual and state inputs [7]
- Code-as-Policies, which prompts code-generating LLMs to produce executable robot control policies [8]
- Instruct2Act, a framework that converts multimodal instructions into perception-planning-action code, outperforming specialist policies on tabletop manipulation tasks [9]

III. Background

This is an optional section in which you can introduce concepts, terms, or methods that are important for understanding your approach and that would not directly fit in Sec. IV. If you do not need this section, comment out the respective line in *report.tex*.

*Submitted to the Department of Computer Science at Hochschule Bonn-Rhein-Sieg in partial fulfilment of the requirements for the degree of Master of Science in Autonomous Systems

†Supervised by Supervisor 1 (Affiliation) and Supervisor 2 (Affiliation)

‡Submitted in Month 20XX

IV. Methodology

Describe all conceptual details about your approach in this section. Add any necessary subsections to improve the presentation.

Feel free to rename this section to better reflect the concrete topic you are discussing.

V. Evaluation

If your work involved experiments, describe the experimental setup and the results in this section.

VI. Conclusions

A. Summary

B. Contributions

C. Future Work

References

[1] J. E. Laird, "Cognitive robotics using the soar cognitive architecture," *AAAI Workshop - Technical Report*, p. 01, 2012.

[2] C. Ritter, J. F. Neves, M. Bagerman, and N. A. Taatgen, "Act-r: A cognitive architecture for modeling cognition," *WIREs Cognitive Science*, vol. 10, no. 3, 2018.

[3] J. S. Park, E. Kay, J. Zou, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '23. Association for Computing Machinery, 2023.

[4] G. Wang, Y. Xie, L. Dai, T. Zhao, Z. Wang, H. Zhou, H. Chen, R. Li, S. Ma, W. Zhang *et al.*, "Voyager: An open-ended embodied agent with large language models," *Transactions on Machine Learning Research*, 2024.

[5] O. Mees, Y. Li, Y. He, C. Zhang, L. Tai, and W. Burgard, "Octo: An open-source generalist robot policy," in *First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024*, 2024. [Online]. Available: https://openreview.net/forum?id=jGrtIvJBpS

[6] A. Brohan, N. Brown, I. Chebotar, O. Cortes, B. Duet, C. Finn, K. Gopalakrishnan, A. Higuera, A. Herzog *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," *arXiv preprint arXiv:2307.15818*, 2023.

[7] D. Driess, F. Xia, M. Wahid, C. Allen-Blanchette, P. Joshi, A. Bengio, B. Faenza, A. Singletary, Y. Wu, A. Gurkaynak *et al.*, "Palm-e: An embodied multimodal language model," *arXiv preprint arXiv:2303.03378*, 2023.

[8] M. Liang, B. Ichter, T. Zhang, S. Xiao, A. Zhang, F. Xia, and S. Levine, "Code as policies: Language model programs for embodied control," *arXiv preprint arXiv:2209.07753*, 2022.

[9] Y. Huang, X. Liang, B. Chen, S. Liu, D. Huang, Z. Zhu, H. Ma, J. Zhu, Y. Zhang, H. Yu *et al.*, "Instruct2act: Mapping multi-modality instructions to robotic actions with large language model," *arXiv preprint arXiv:2305.11176*, 2023.

ACKNOWLEDGMENT

Write your acknowledgments here.

STATEMENT OF ORIGINALITY

[If AI assistants have not been used, use this sentence] I, the undersigned below, declare that this work has not previously been submitted to this or any other university and that it is, unless otherwise stated, entirely my own work.

[If an AI assistant has been used, use this sentence] I, the undersigned below, declare that this work has not previously been submitted to this or any other university and that it is, unless otherwise stated, entirely my own work. The report was, in part, written with the help of the AI assistant [AI assistant name] as described in the appendix. I am aware that content generated by AI systems is no substitute for careful scientific work, which is why all AI-generated content has been critically reviewed by me, and I take full responsibility for it.

Date Signature

Appendix

Please limit the main part of the report to 20 pages (not including the references, the statement of originality, and the appendix).

In your appendix, you can add any additional details about your work, such as:

- extra results that do not necessarily belong in Sec. V
- more detailed justifications of certain algorithm design decisions
- algorithm proofs

Additionally, in the case of using AI assistants, describe in detail what content was generated using an AI assistant. In particular, name the AI assistant(s) that you used and how they were used (e.g. which prompts were used, and for which parts of the project).