# Developing Self Awareness in Robots with Inner Speech

## Abstract

Large-scale Language Models (LLMs) have revolutionized the digital paradigm. Their progress spans across text-based pipelines, exceptional reasoning capabilities, and multi model frameworks, emerging as a leading paradigm in the field of Artificial Intelligence (AI). With rapid increase in leveraging LLM-based pipelines in Robotics domain over the past few years, these models have shown great success in task decomposition, guiding the agent for the next action, keeping a track of success ratio of previous moves, and history-retention to track the progress of given task. Despite their remarkable progress, LLMs are still subject to hallucination and struggle if the task requires commonsense knowledge, needs long sequences of reasoning, or requires a situated and embodied understanding of the context. Recent work by Google has tried to tackle some of these shortcomings by synergizing reasoning and action in LLMs so they have access to external knowledge to verify or modify their reasoning traces. To the best of our knowledge, no work has considered infusing the LLM reasoning with commonsense to improve its situated understanding of the task context. The unavailability of commonsense knowledge is another primary reason that makes the systems hallucinate in long-horizon tasks, due to the inability to make the effective decisions when internal-knowledge of language models fail. Inspired by human cognition, this research study implements the inner-speech workflow in robotic streamlines, while leveraging an external semantic network *ConceptNet* to implement the commonsense-based reasoning capabilities in LLMs. The reasoning and acting is synergized to make the system efficient for any given task of embodied environments. The system is tested in an embodied environment *ALFWorld*, and is capable to synergize reasoning and acting with commonsense knowledge. The early tests show that LLMs can become more explain able and improve their reasoning compared with the baseline when they are infused with commonsense knowledge. The system enhances the user trust by demonstrating the inner-speech and agent trajectories, to increase deployment flexibility in real-world settings.