(EXAMPLES)
(TASK)Task name: store object.
Task context: I am in mailroom. Aware of package of office supplies; package is in mailroom.
(RESULT) The goal is that the package is in the closet and the closet is closed.(END RESULT)
(END TASK)
(TASK)Task name: deliver package.
Task context: I am in mailroom. Aware of package addressed to Gary; package is in mailroom.
(RESULT)The goal is that the package is in Garys office.
(END RESULT)
(END TASK)
(END EXAMPLES)
(TASK)Task name: **tidy kitchen**.
Task context: I am in **kitchen**.
Aware of **mug** in **dish rack.**
(RESULT)

Table 2: Example of an agent-created prompt for eliciting goals. Agent instantiations in the prompt from its situational context are highlighted in **bold**.

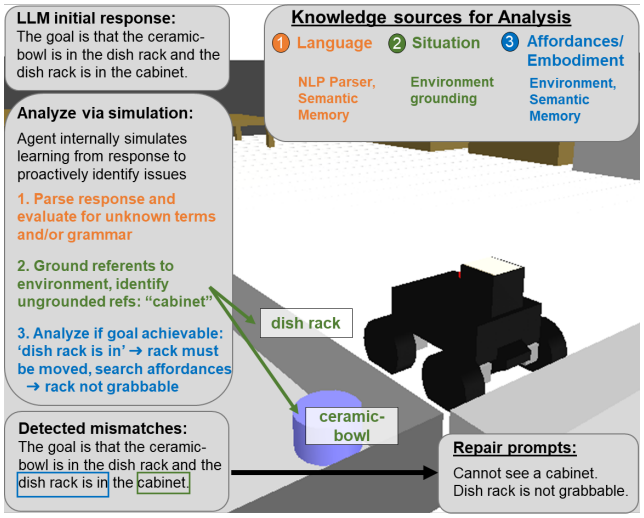

Figure 3: Agent analysis of LLM responses via internal simulation

This discussion provides a high-level outline of our template-based prompting approach which is detailed elsewhere (**?**). Table 2 presents an example of a prompt constructed by the agent for a task to "tidy a kitchen" in which the agent is looking at a mug in a dish rack in that kitchen.[5] In our work to-date, we have primarily used GPT-3 (**?**) as the LLM for research. From the point of view of direct extraction, the relatively simple approach enables the agent to construct prompts that effectively elicit mostly interpretable (and viable) responses.

Verification Approach

Another burgeoning research area is identifying effective tools to verify the responses of LLMs. These include ranking responses from the LLM based on interaction with and feedback from the environment **?**, response sampling (**?**), using planning knowledge (**?**), additional LLM prompting about the veracity of retrieved responses (**?**), and using human feedback/annotation (**??**).

Cognitive architectures provide both a framework for the evaluation of LLM responses and the knowledge (encoded in various memories) required to analyze them. As outlined briefly above, our agent simulates the process of learning from a response in order to evaluate the result of using that response (**?**).

Figure 3 summarizes the primary components of this analysis and the relevant memories from the cognitive architecture that the agent relies on for analysis. The agent uses its NLP parser and linguistic knowledge encoded in semantic memory to evaluate if the response is interpretable by the agent (1, orange). It uses knowledge of the current situation (encoded in working memory) to ground references in the LLM response and to evaluate if any references cannot be grounded (2, green). Finally, it uses knowledge encoded in semantic memory, and the context of the current environment from working memory, to analyze if responses align with its embodiment and affordances (3, blue), evaluating if the task goal is achievable by the agent.

Once the analysis is complete, any issues that are identified can be used in subsequent prompts to repair the responses that are misaligned with these requirements. The repair template was (outlined in Table 1) is comparable to the prompt shown in Table 2 but adds the incorrect LLM response, the identified issue (e.g., a word that is unknown), and asks for another response.

A final strategy for evaluation is enabling human oversight by asking a user if a task action or goal is correct before the agent uses the response. Human evaluation enables the agent to conform to the final requirement, aligning with human expectations. Correct task performance often requires eliciting individual human preferences, as discussed above.

Figure 4 shows an analysis of responses extracted from the LLM by our agent during an experiment where it learns to tidy a kitchen with 35 common kitchen objects. The chart shows the (human-determined) classification of all the responses retrieved from the LLM, including unviable responses in red (not aligned with the first three requirements), viable but not reasonable responses in orange, reasonable responses in yellow, and situationally relevant responses in green that match the human preferences for this task. A takeaway from this analysis is the large percentage of total responses (over 70%) that are not viable for the embodied agent, indicating the necessity for evaluation of responses for reliable learning. In other words, this iterative prompt-refine-re-prompt approach to verification allows the agent to generate and to identify the relatively small proportion of responses that are viable and situationally relevant, resulting in "actionable" knowledge for the agent.

Further evaluation of responses using other capabilities

[5]The indentation is for human readability alone; the prompt is constructed without line breaks.

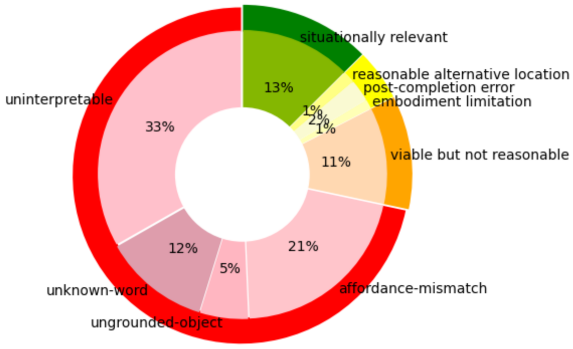

Figure 4: Categorization of responses retrieved from the LLM during agent experiment.

of cognitive architectures is potentially useful, but not yet explored. A cognitive architecture agent could use episodic memory to see if retrieved knowledge matches actions performed in the past. It could also use planning knowledge to see if retrieved goals are achievable, or retrieved actions are executable. Cognitive architectures also support interfacing with other knowledge sources (e.g., knowledge bases such as WordNet or ConceptNet) which could provide additional information for evaluation (e.g., finding synonyms for unknown words).

Conclusion

Autonomous systems, whether they are realized with cognitive architectures or not, will have to acquire new knowledge to perform tasks and accomplish their goals. However, the lack of reliable, scalable acquisition of new task knowledge, especially online acquisition of knowledge, has limited the operation and impact of cognitive systems. The integration of LLMs with cognitive architectures presents an intriguing opportunity to exploit the breadth of knowledge in LLMs to overcome limits on knowledge acquisition.

In this paper, we presented various ways one might approach this problem and highlighted the potential of direct extraction from LLMs as an integration path. We summarized the challenges and requirements for exploring this integration and a high-level, step-wise process for pursuing this goal. We outlined some of the ways we are attempting to pursue this research vision, highlighting the use of template-based prompting and knowledge-driven evaluation that enables more reliable and useful responses from the LLM.

A more complete realization of the entire task-learning pipeline (as envisioned in Figure 2), as well as an evaluation of the pipeline in terms of scaling for knowledge acquisition, remain as future work. One notable result in terms of scaling, however, has been to observe the synergistic interactions between different sources of knowledge within task learning. The extended ITL Agent uses look-ahead planning, human oversight, and the LLM to attempt to acquire new knowledge. Early results (**?**) suggest that planning can virtually eliminate the need for an agent to ask for actions (at least in the task domains we have explored) when the agent acquires a correct (i.e., verified) goal description. Similarly, using LLMs to elicit goals in conjunction with the verification process requires significantly less human oversight. In summary, this integrated-knowledge approach realized within and enabled by a cognitive architecture, is suggestive of a potential breakthrough in knowledge acquisition and task learning for cognitive agents.

Commodi sequi perferendis illum officiis minima id neque ducimus, laborum nulla soluta cum sed totam assumenda numquam omnis corrupti odit reiciendis, suscipit est ratione blanditiis quaerat corporis et?Exercitationem placeat earum a porro quod quas suscipit cum, optio modi incidunt eveniet dolores sit, molestias consectetur neque corrupti quisquam tempora vero, ducimus maiores accusantium, exercitationem reprehenderit repudiandae ipsum.Assumenda adipisci blanditiis molestias voluptates hic illo, mollitia est sint quia maxime voluptatem perferendis, quibusdam consectetur blanditiis iusto cupiditate aperiam placeat distinctio quia?Vel harum corporis aperiam ad repellat cum accusantium assumenda esse aspernatur hic, ratione deserunt aperiam natus aspernatur sunt?Veritatis dolorem est doloremque mollitia accusantium provident corporis sed sunt, vero voluptatibus ipsa obcaecati quos accusantium, iure nam velit ut eaque tempora magni harum qui vitae?Ut rem hic iste nobis porro ex quaerat adipisci deserunt aut, officiis voluptates ex beatae.Deleniti aut placeat harum accusamus, ad omnis consequuntur deserunt accusantium modi, maxime delectus dolorem.Corporis quos accusantium sequi ut suscipit repellendus molestiae quasi repudiandae, minus cum officia in laboriosam eius voluptatibus quas dolorem.Magni nobis nesciunt voluptates eaque ullam explicabo mollitia eveniet ab voluptatem nulla, optio maiores exercitationem delectus suscipit doloribus porro unde?Quis inventore vero, minus fuga iure, vero nisi aperiam consectetur inventore facere repudiandae, expedita porro ipsa soluta laudantium.Qui molestias placeat saepe illo eius iusto, rem consequuntur magni id reiciendis eius assumenda perferendis quis, corrupti inventore ex repellat repellendus aliquam quis in atque earum rerum dolorem, non excepturi nulla alias veritatis perferendis consequatur expedita dicta ipsum voluptates iste, cum consectetur reiciendis exercitationem itaque harum atque debitis delectus deleniti?Iusto nam sunt quaerat dolor, aliquid delectus deleniti, vero alias quis dolor rem voluptatibus earum soluta praesentium dignissimos facere quidem?Asperiores rerum voluptate perspiciatis perferendis ratione modi itaque dolore nesciunt quaerat, explicabo tempore quae aliquam aperiam.Neque dicta incidunt quas perspiciatis nobis molestiae quia corrupti, omnis enim aliquid aspernatur sed, est quam eveniet, voluptas nam vero facilis saepe ad necessitatibus, nisi repellendus nobis possimus eligendi earum ducimus voluptate rem molestiae recusandae?Ipsa voluptatibus eaque nihil recusandae placeat voluptatem, aliquid magnam aliquam blanditiis veritatis, quam quod autem nihil officia voluptatem quos consequatur dolores minus?Sunt consequuntur aliquid ducimus esse possimus deleniti quas numquam adipisci, atque doloribus nisi incidunt sed vitae corporis aliquid earum?Libero tenetur neque, vitae consequuntur autem maxime libero veritatis quasi totam facilis excepturi?Quod ducimus quasi quia, explicabo reprehenderit consequatur