Retrieving Memories from a Cognitive Architecture using Language Models for Social Robot Applications*

Thomas Sievers[1]

***Abstract*— Large Language Models (LLMs) and Vision-Language Models (VLMs) have the potential to significantly advance the development and use of cognitive architectures for social robots. Using a combined system consisting of an Adaptive Control of Thought-Rational (ACT-R) model and a humanoid social robot, I investigate how content from the declarative memory of the ACT-R model can be retrieved using real-world data obtained by the robot via an LLM or VLM, processed according to the procedural memory of the cognitive model and returned to the robot as instructions for action. In addition, real-world data captured by the robot can be stored as memory chunks in the cognitive model. This opens up possibilities for using human-like judgment and decision-making capabilities like instance-based learning and intuitive decision-making inherent in cognitive architectures with social robots and practically offers opportunities of augmenting the prompt for LLM-driven utterances with content from declarative memory thus keeping them more contextually relevant.**

I. INTRODUCTION

Large Language Models (LLMs) like ChatGPT amaze the world with their seemingly human-like abilities in some respects. They provide advanced reasoning capabilities, blending intuitive and deliberate cognitive processes [2]. LLMs also help robotic systems improve their generalization capabilities in dynamic and complex real-world environments and can significantly increase their behavior planning and execution capabilities [8]. Vision-Language Models (VLMs) are multimodal AI systems created by combining an LLM with a vision encoder that gives the LLM the ability to "see". They provide assistance with complex tasks such as creating captions and answering visual questions [4].

Cognitive architectures, on the other hand, refer both to a theory about the structure of the human mind and to a computer-based implementation of such a theory. They attempt to describe and integrate the basic mechanisms of human cognition. In doing so, they rely on empirically supported assumptions from cognitive psychology. Their formalized models can be used to react flexibly to actions in a human-like manner and – when used in a robot – to develop a situational understanding for adequate reactions. Adaptive Control of Thought-Rational (ACT-R) is a well-known and successfully used cognitive architecture [1].

Intuitive decision-making as a subjective, particularly human type of decision-making, is based on implicit knowledge that is transmitted to the conscious mind at the time of the decision through affect or unconscious cognition. Computational models of intuitive decision-making can be expressed as instance-based learning using the ACT-R cognitive architecture [12]. In instance-based learning theory (IBLT), past experiences (i.e. instances) are retrieved using cognitive mechanisms of a cognitive architecture. IBLT proposes learning mechanisms related to a decision-making process, such as instance-based knowledge and recognition-based retrieval. These learning mechanisms can be implemented in an ACT-R model [5], [6]. I think that it would be interesting to implement such behavior in a social robot as well. A combination of robot sensor technology and data processing with such an architecture offers the possibility of dealing with information from the robot's real world in cognitive models. A cognitive architecture may also be used to add a "human component" to robotic applications, as the procedural processes of a mental model behave differently – more human-like – than conventional algorithms [13].

For human-robot interaction (HRI) applications that use an LLM to generate speech and/or a VLM to recognize image content, a cognitive model can help to provide facts and context of a specific scenario for the language model. Using prompt augmentation, conclusions of a mental model can be taken into account in instruction generation for the LLM and thus reduce weaknesses such as ignorance of current individual facts and thus hallucination when relevant facts are unclear. Furthermore, language models are good at fast automatic reasoning, but less capable of high-level cognition to enable complex mental operations and "slow thinking" following the dual process theory of human cognition [9].

The cognitive architecture of ACT-R comprises a declarative and a procedural memory, whereby the declarative memory supports lexical knowledge by encoding, storing and retrieving semantic knowledge, as in humans, while the procedural memory enables the learning of habits and skills [14], [11], [3], [7]. Using the ACT-R chunk and memory system to store and process facts and impressions and utilizing knowledge from the procedural memory of the cognitive model allows an LLM to incorporate these facts, opening a path for a more reliable and evidence-based application of LLMs for individual HRI scenarios.

II. METHODS

My system consisted of the humanoid social robot Pepper with a software application that connected to OpenAI's GPT-4o model via an API, and a TCP/IP connection over a wireless network to the current version of the ACT-R 7 cognitive architecture software running on a remote computer. ACT-R offers the technical possibility of integrating a cognitive

*This work was not supported by any organization

[1]Thomas Sievers is with Institute of Information Systems, University of Lübeck, 23562 Lübeck, Germany t.sievers@uni-luebeck.de

model with bidirectional communication into an application of choice using it's remote interface.

I applied OpenAI's Generative Pretrained Transformer (GPT) language model in a setting with a social robot in dialog with a human, where the robot used the GPT model to generate speech, processed the content of the dialog and recognized image content from its front camera [10]. However, the proposed methods can also be used beyond an HRI context.

For testing, I used the vision capabilities of the GPT model in one method and text generation in another. The cognitive model remained the same in both cases. In the vision method, the language model was instructed to analyze the content of images from the front camera on the robot's head and describe the main content of each image in three keywords or key phrases. In the other case, the utterances of a person in dialog with the robot were processed by an LLM so that the core content of the human question or problem was also expressed in three keywords. These keywords or phrases were passed as chunks to an ACT-R cognitive model, where they were processed with productions from procedural memory to search the declarative memory for existing memory content indexed by chunks in the same or similar form. The memory content contained an additional chunk that represented the actual recollection and could, for example, represent a fact in the form of a short sentence. If there was a positive correlation between keyword chunks from the LLM and memory chunks, this recollection was passed to the robot application and thus to the LLM for prompt augmentation, which generated a response based on this knowledge. An a-priori factual knowledge could be accumulated by creating corresponding chunks in the declarative memory.

In ACT-R, declarative knowledge is represented in the form of chunks, i.e. representations of individual properties, each of which can be accessed via a labeled slot. The cognitive model should receive chunks with keywords from the robot application and checking whether corresponding chunks for these keywords were already stored in the declarative memory of the model. I assumed the transfer of three keywords. Productions of the procedural memory checked all combinations of the sequence of keywords for a match with memory content and generated a hit for two out of three matching keywords. In this case, the associated memory content was called up and the corresponding recollection was returned to the robot application.

The cognitive model defined a chunk type for memory content in declarative memory as follows: *(chunk-type keyword asso-one asso-two asso-three phrase)*

This chunk type 'keyword' featured three labeled slots 'asso-one', 'asso-two' and 'asso-three' as well as a 'phrase' slot, which stored the recollection. For example, this could be a remembered fact used to complete a system prompt for an LLM that the robot used to talk to humans. Figure 1 shows the process of searching for a matching memory chunk. Having found a corresponding memory chunk, the cognitive model returned the sentence stored in the 'phrase' slot of this chunk as a *fact phrase*.

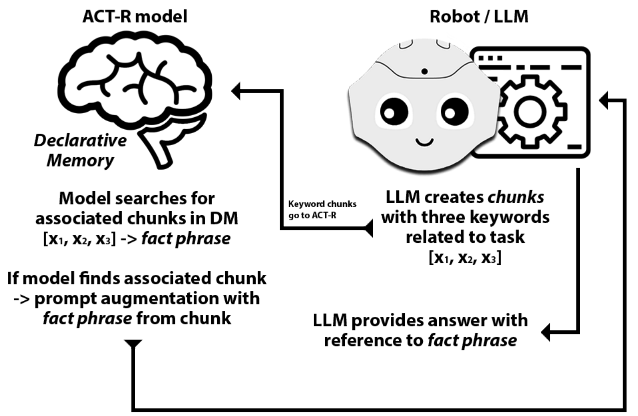

Fig. 1. Transfer of keywords to ACT-R and return of retrieved memory facts to the LLM for prompt augmentation.

I tested the use of the GPT-4o model to retrieve specific recollections from the declarative memory of an ACT-R model using two examples. In the first example, I assumed a dialog between a human and a social robot, where eventually the robot is supposed to search 'its' (ACT-R) memory for matching recollections based on this utterance. The second example deals with the assignment of a visual impression to recollections. The robot's front camera took a picture every few seconds to capture an up-to-date impression of what the robot saw, and the LLM was instructed to create a description of the content in the form of three keywords.

III. DISCUSSION

With the help of LLMs and VLMs and the corresponding sensors, robots can access and interpret the same information that is available to humans. I only outline the possibilities here with my example applications for storing visual impressions or facts together with keywords and using them for prompt augmentation. There are far more options available for a programmatic implementation of cognitive models with ACT-R in combination with social robots. On the one hand, cognitive models can be used within the framework of their cognitive architectures to anticipate human behavior and thus make it more comprehensible. On the other hand, cognitive processes can use declarative and procedural memory to produce decisions which, when used to control the actions of a robot, allow the robot to act in a more humane way.

The use of a cognitive architecture such as ACT-R enables the inclusion and investigation of further cognitive principles and processes in interaction with a social robot and LLMs independent of declarative memory retrieval. Future work should consider these aspects in more detail. Further experiments are needed to optimize the ACT-R models and system prompts for the LLM, as well as ongoing evaluation in studies with different people and different robot systems for various tasks. I am confident that this will open up a wide range of possibilities for future research into how cognitive architectures and their models can add a human touch to a social robot in HRI.

References

[1] Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind **111**(4), 1036–1060 (2004). https://doi.org/10.1037/0033-295X.111.4.1036

[2] Dubey, S., Ghosh, R., Dubey, M., Chatterjee, S., Das, S., Benito-León, J.: Redefining cognitive domains in the era of chatgpt: A comprehensive analysis of artificial intelligence's influence and future implications. Medical Research Archives **12** (06 2024). https://doi.org/10.18103/mra.v12i6.5383

[3] Eichenbaum, H.: Hippocampus: Cognitive processes and neural representations that underlie declarative memory. Neuron **44**(1), 109–120 (2004). https://doi.org/https://doi.org/10.1016/j.neuron.2004.08.028

[4] Ghosh, A., Acharya, A., Saha, S., Jain, V., Chadha, A.: Exploring the frontier of vision-language models: A survey of current methodologies and future directions (2024), https://arxiv.org/abs/2404.07214

[5] Gonzalez, C., Dutt, V., Lebiere, C.: Validating instance-based learning mechanisms outside of act-r. Journal of Computational Science **4**(4), 262–268 (2013). https://doi.org/https://doi.org/10.1016/j.jocs.2011.12.001, pEDISWESA 2011 and Sc. computing for Cog. Sciences

[6] Gonzalez, C., Lerch, J.F., Lebiere, C.: Instance-based learning in dynamic decision making. Cognitive Science **27**(4), 591–635 (2003). https://doi.org/https://doi.org/10.1016/S0364-0213(03)00031-4

[7] Hayne, H., Boniface, J., Barr, R.: The development of declarative memory in human infants: Age-related changes in deferred imitation. Behavioral neuroscience **114**, 77–83 (02 2000). https://doi.org/10.1037/0735-7044.114.1.77

[8] Jeong, H., Lee, H., Kim, C., Shin, S.: A survey of robot intelligence with large language models. Applied Sciences **14**(19) (2024). https://doi.org/10.3390/app14198868

[9] Kahneman, D.: Thinking, fast and slow. Farrar, Straus and Giroux, New York (2011)

[10] OpenAI: The most powerful platform for building ai products. Tech. rep. (2025), https://openai.com/api/

[11] Quam, C., Wang, A., Maddox, W.T., Golisch, K., Lotto, A.: Procedural-memory, working-memory, and declarative-memory skills are each associated with dimensional integration in sound-category learning. Frontiers in Psychology **9** (2018). https://doi.org/10.3389/fpsyg.2018.01828

[12] Thomson, R., Lebiere, C., Anderson, J.R., Staszewski, J.: A general instance-based learning framework for studying intuitive decision-making in a cognitive architecture. Journal of Applied Research in Memory and Cognition **4**(3), 180–190 (2015). https://doi.org/https://doi.org/10.1016/j.jarmac.2014.06.002, modeling and Aiding Intuition in Organizational Decision Making

[13] Werk, A., Scholz, S., Sievers, T., Russwinkel, N.: How to provide a dynamic cognitive person model of a human collaboration partner to a pepper robot. ICCM (2024), https://mathpsych.org/presentation/1452

[14] Whitehill, J.: Understanding act-r - an outsider's perspective (2013), https://arxiv.org/abs/1306.0125