

Article

Not peer-reviewed version

Retrieving Memory Content from a Cognitive Architecture by Impressions from Language Models for Use in a Social Robot

[Thomas Sievers](#) * and [Nele Russwinkel](#)

Posted Date: 22 April 2025

doi: 10.20944/preprints202504.1792.v1

Keywords: large language models; cognitive architecture; human-robot interaction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Retrieving Memory Content from a Cognitive Architecture by Impressions from Language Models for Use in a Social Robot

Thomas Sievers *  and Nele Russwinkel 

Institute of Information Systems, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck

* Correspondence: t.sievers@uni-luebeck.de

Abstract: Large Language Models (LLMs) and Vision-Language Models (VLMs) have the potential to significantly advance the development and application of cognitive architectures for human-robot interaction (HRI) to enable social robots with enhanced cognitive capabilities. An essential cognitive ability of humans is the use of memory. We investigate a way to create a social robot with a human-like memory and recollection based on cognitive processes for a better comprehensible and situational behavior of the robot. Using a combined system consisting of an Adaptive Control of Thought-Rational (ACT-R) model and a humanoid social robot, we show how recollections from the declarative memory of the ACT-R model can be retrieved using 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. Real-world data captured by the robot can be stored as memory chunks in the cognitive model and recalled, for example by means of associations. This opens up possibilities for using human-like judgment and decision-making capabilities 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. We illustrate the use of such an approach in HRI scenarios with the social robot Pepper.

Keywords: large language models; cognitive architecture; human-robot interaction

1. Introduction

The human brain, complex as it is as a result of unplanned evolution, stores memories in a rather disorganized way compared to the precise addressing system of computers [1,2]. It is organized according to context, topics are interwoven in complex associative networks and are retrieved by clues and associations, not by pure addressing in a memory as with a computer. Human memory is superior to that of computers in certain respects, for example through automatic prioritization with regard to rewards and frequently used elements [3,4]. However, memories that are unused soon decay into oblivion. And unlike the addressing system of a computer as the result of planned engineering following the von Neumann architecture, the human way of thinking is susceptible to the formation of false recollections through incorrect associations.

But perhaps the Cartesian mind-body dualism, which ignores the importance of brain-inspired embodiment for the development of human-like Artificial Cognition (ACo), leads to a dead end [5]. Personal experiences, partly mediated by social interaction and accessible to recollection, are essential for an Artificial Intelligence (AI) agent to develop cognitive abilities and learn about its environment. Such an agent should be able to imagine actions using mental representations before executing them and justify them using the cognitive ability of prospection, including planning, predicting, imagining scenarios and possible future events. A cognitive architecture seems to be crucial to equip the agent with such capabilities. In this work, we ask the question whether a social robot can be endowed with human-like memory capabilities by using a cognitive architecture for information retrieval and

processing and a Large Language Model (LLM) that is also capable of vision. A key criterion here is the ability to retrieve recollections from memory using associated expressions and to prepare the ground for further human-like cognitive abilities. We show a technical method for implementation.

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 [6]. With their sophisticated probabilistic model of language use to generate a semantically consistent text, they require a form of effective memory that has surprising similarities with central features of human memory [7]. 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, enabling robots to engage with their environment in a human-like manner [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 [9]. VLMs are capable of performing a variety of tasks after learning relationships between images and language from large data sets, such as answering questions about images, finding sentences that correspond well with images, and finding image regions that correspond to texts. These skills can be used in many ways for robotics, for example for robot movements, state recognition, object recognition, affordance recognition, relation recognition, and anomaly detection [10,11].

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 are particularly suitable for general problem solving in areas that are not clearly defined and for human cognitive modeling [12]. Cognitive architectures 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 regarding human behavior for adequate reactions. We apply the well-known and successfully implemented cognitive architecture Adaptive Control of Thought-Rational (ACT-R) for our test scenarios [13]. ACT-R is a hybrid cognitive architecture with a symbolic and sub-symbolic structure. It is based on psychological principles and offers a comprehensive model of cognitive processing. ACT-R models run in real time and are robust to errors or unexpected events.

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 [14]. 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 [15,16]. A social robot would benefit from such abilities, for example to solve problems or develop a kind of intuition for the situation. With regard to Human-Robot Interaction (HRI), 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 [17].

In HRI applications that use an LLM for speech generation and/or a VLM for image content recognition, a cognitive model can use its memory to provide facts and context of a particular scenario to the language model, bringing in personalized experiences by recalling unique memories collected by the robot. With the help of prompt augmentation, conclusions of a mental model can be taken into account in instruction generation via prompts for the LLM and thus reduce weaknesses of the

language model 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 [18]. A combination of the advantages of both approaches would result in a system that performs involuntary reactions as automatic tasks (e.g. formulating a sentence) and also incorporates and takes into account more complex human-like cognitive abilities such as memory. This would open up the possibility of significantly increasing the capabilities of a social robot in terms of incorporating and assessing human actions and intentions.

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 [19–22]. Using the ACT-R chunk and memory system to store, retrieve and process facts and impressions of a real-world scenario and utilizing knowledge from the procedural memory of the cognitive model allows an LLM to incorporate these facts into utterances, opening a path for more reliable, evidence-based and human-like application of LLMs.

For testing our approach, we 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 [23]. However, the proposed methods can also be used beyond an HRI context. As a cognitive architecture, we used the standalone application of ACT-R 7, for which we created a cognitive model in LISP [24]. The connection between the cognitive model and our robot application, which also contained the Application Programming Interface (API) to the LLM, was realized via a TCP/IP connection of the *dispatcher*, which acts as a server for client applications to ACT-R [25].

To incorporate different sensory abilities from the robot, we used the visual capabilities of the GPT model in one test method and text generation in another. The operation of the cognitive model we employed 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 worth remembering 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. In principle, it is possible to accumulate factual knowledge a priori by creating corresponding chunks in declarative memory. In the following, we show a technical implementation of such an architecture as a combination of LLM/VLM integration, humanoid social robot and cognitive architecture and give examples of applications.

2. Related Work

Robots are able to use Large Multi-modal Models (LMMs) to comprehend and execute tasks based on natural language input and environmental cues [26]. The integration of foundation models such as LLMs and VLMs can effectively improve robot intelligence [8,27]. VLMs help, for example, to equip robots with the ability for physically based task planning [28]. The use of GPT-4V for analyzing videos of humans performing tasks to obtain textual explanations of environmental and action details, combined with a GPT-4-based task planner that encodes these details into a symbolic task plan for a robot, was presented by Wake et al. [29].

Yoshida et al. investigated possibilities for the development of a “minimal self” with a sense of agency and ownership in a robot that was able to mimic human movements and emotions by using human knowledge from language models [30]. They did not implement a model of a cognitive architecture but used GPT-4’s motion generation and image recognition capabilities. VLMs as a basis for metacognitive thinking can enable robots to understand and improve their own processes, avoid hardware failures and thus increase their resilience [31].

Since the recent successes of language models, there has been an increased interest in the interplay between LLMs and cognitive architectures. Ideas for a combination of both are emerging or the creation of a cognitive architecture based on an LLM is being considered. Niu et al. provided an overview of similarities, differences, and challenges between LLMs and cognitive science by analyzing methods for assessing potential cognitive capabilities of LLMs, discussing biases and limitations, and an integration of LLM with cognitive architectures [32]. In the novel neuro-symbolic architecture presented by Wu et al, human-centered decision making was enabled through the integration of ACT-R with LLMs by using knowledge from the decision process of the cognitive model as neural representations in trainable layers of the LLM [33]. This improved the ability for grounded decision making. González-Santamarta et. al. provided an example of integrating an LLM with a cognitive architecture to enable planning and reasoning in autonomous robots [34]. They showed how to use the reasoning capabilities of LLMs in the MERLIN2 cognitive architecture integrated in ROS 2.

The use of a cognitive architecture connected to a social robot to store and process memory chunks from a language game between the robot and a human was demonstrated by Sievers et al. [35]. An ACT-R model received words from the robot, searched for matching associations already present in memory, and provided feedback to an LLM on these associations when appropriate. This way of grounding abstract words and ideas should help to constrain possible meanings.

Knowles et al. proposed a system architecture that combined LLMs and cognitive architectures with an analogy to “fast” and “slow” thinking in human cognition [18,36]. Leivada et al. explored whether the current generation of LLMs is able to develop grounded cognition that incorporates prior expectations and prior world experiences to perceive the big picture [37]. Insights from human cognition and psychology anchored in cognitive architectures could contribute to the development of systems that are more powerful, reliable and human-like [38]. This dual-process architecture and the hybrid neuro-symbolic approach to overcoming the limitations of current LLMs is seen as particularly important.

He et al. investigated the Long-Term Memory (LTM) capabilities of AI systems such as LLM-based agents by creating a mapping between human LTM and AI LTM mechanisms and proposing an extension of current cognitive architectures for the creation of next-generation AI systems with LTM [39]. The significance of LTM for the foundation of AI self-evolution – allowing LLMs to evolve during inference – based on limited data or interactions versus large-scale training on increasingly large datasets was emphasized by Jiang et al. [40].

A human-like memory architecture to improve the cognitive abilities of LLM-based dialog agents was proposed by Hou et al. [41]. The proposed architecture enabled agents to autonomously retrieve recollections required to generate responses, thereby addressing limitations of temporal awareness of LLMs. Memories from the interaction history with the user were stored in a database containing the content and temporal context of each reminder. This allowed the agent to understand and utilize specific memory content relevant to the user in a temporal context.

In their Cognitive Architectures for Language Agents (CoALA), Sumers et al. proposed a language agent with modular memory components, a structured action space for interacting with internal memory and external environments, and a generalized decision-making process for selecting actions [42]. Their framework draws from the rich history of symbolic artificial intelligence and cognitive science and combines some longstanding findings with current research on LLMs for tasks that require basic knowledge or reasoning. Furthermore, Retrieval-Augmented Generation (RAG) can

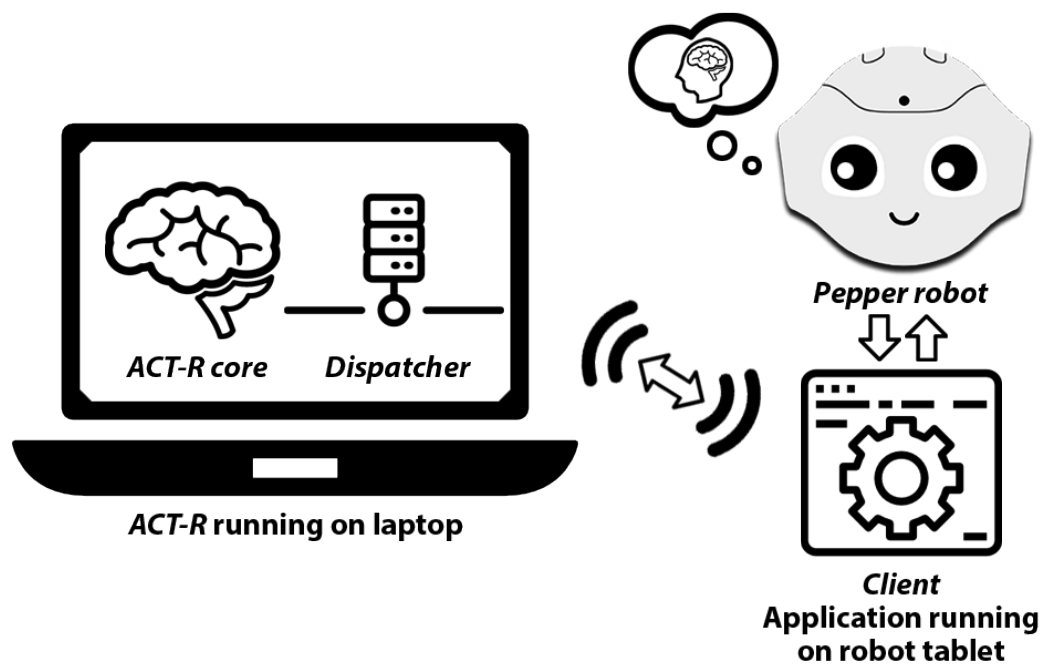


Figure 1. Connection between ACT-R / Dispatcher and the robot application.

offer a different approach to challenges such as hallucinations, outdated knowledge and untraceable reasoning processes in LLMs [43,44]. RAG does not retrieve knowledge from unknown scenarios, but provides an approach to combine external sources in LLMs with many optimizations to incorporate the appropriate knowledge.

The production of semantically consistent text requires a form of effective memory, especially in individual scenarios. Examination of the memory properties of LLMs revealed similarities with important features of human memory, but in the case of an LLM these are learned from the statistics of the training text data and not from the architecture of the LLM [45]. For a more human-like way of retrieving recollections from a memory – which is moreover explainable in terms of the accessible memory contents and thus does not pose explainability problems – we propose to use the declarative memory of a cognitive architecture in conjunction with the procedural capabilities of a cognitive model, as described below.

3. Methods

Our system consists of the humanoid social robot Pepper with a software application that connects 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 model with bidirectional communication into an application of choice. To establish a remote connection from the robot application to ACT-R, we used its remote interface – the *dispatcher* [25]. Fig. 1 shows the setup of the bidirectional connection between the dispatcher, which acts as a kind of server, and the client application of the robot.

3.1. Humanoid Social Robot Pepper

The humanoid social robot Pepper, shown in Figure 2, was developed by Aldebaran and first released in 2015 [46]. The robot is 120 centimeters tall and optimized for human interaction. It is able to interact with people through conversation, gestures and its touch screen and is therefore well suited for studies on human-robot interaction.



Figure 2. Pepper showing a picture of what it sees in dialog with a human

Pepper is equipped with four directional microphones in his head and speakers for voice output. The robot features an open and fully programmable platform so that developers can program their own applications to run on Pepper using software development kits (SDKs) for programming languages like C++, Python, Java or Kotlin. In a normal dialog mode with humans, our robot application forwarded utterances of the human dialog partner as input to the OpenAI API, which returned a dictionary with the status and response of the API. With each API call, the entire dialog was transferred to the GPT model. This allows the model to constantly ‘remember’ what was previously said and refer to it as the dialog progresses, enabling a comprehensive exchange between human and robot.

3.2. Cognitive Model

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 programmed in LISP for our test scenarios should receive chunks with keywords from the robot application and checking whether chunks for these keywords are already stored in the declarative memory of the model in a suitable combination. We assumed the transfer of three keywords. The programmed 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 or for all three matching keywords. In this case, the associated memory content was called up and the content of a special slot of the memory content chunk from this recollection was returned to the robot application.

Our LISP code 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 memory content we wanted to retrieve. For example, this phrase could be a remembered fact used to constrain a system prompt for the LLM that the robot used to talk to humans. Each chunk stored in the declarative memory also has an arbitrary name. Figure 3 shows the process of searching for a matching memory chunk. Having found a corresponding memory

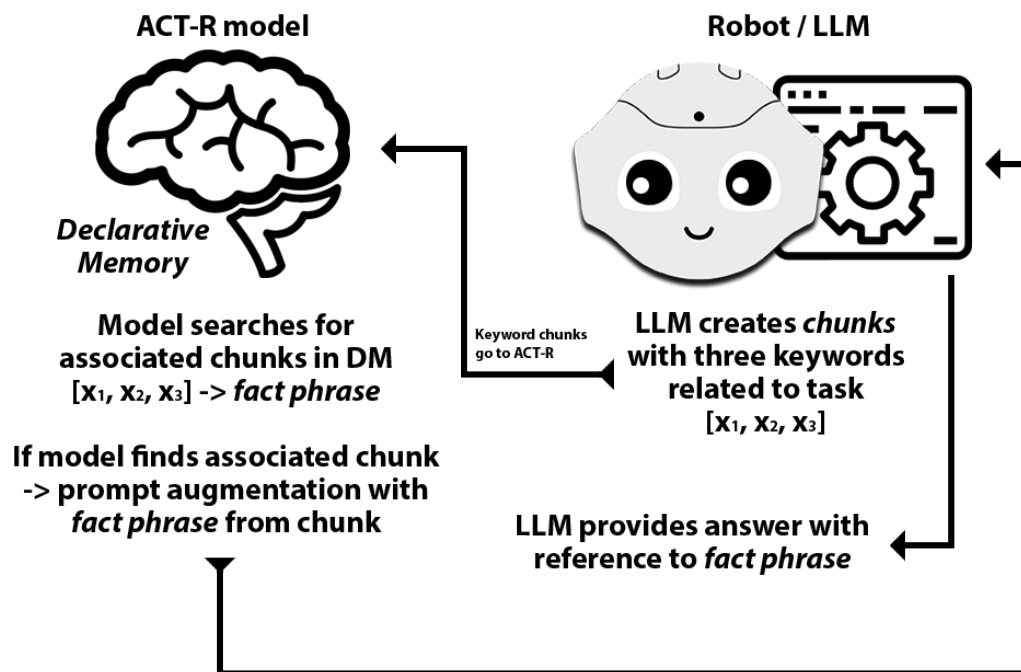


Figure 3. Transfer of keywords to ACT-R and return of retrieved memory facts to the LLM for prompt augmentation.

chunk, the cognitive model returned the sentence stored in the 'phrase' slot of this chunk as a *fact phrase*.

The following is an example of possible content contained in such a memory chunk for a train station scenario assumed in one of our tests: asso-one *platform-1* asso-two *location* asso-three *direction-assistance* phrase *Platform 1 is the neighboring platform right here.*

3.3. Basic Settings of the LLM

We chose GPT-4o to create the conversational parts of the robot. The OpenAI API provides various hyperparameters that can be used to control the model behavior during an API call. The value for *temperature* was set to 0 in order to obtain consistent responses and exclude any randomness as far as possible. Positive values of *presence penalty* would penalize new tokens, depending on whether they already appear in the text, increase the probability that the model will talk about a new topic. We set this value to zero. Positive values of *frequency penalty* would penalize new tokens based on their previous frequency in the text, reducing the likelihood that the model will repeat the same line verbatim. This value was also set to zero. For the instruction of the GPT model, we used prompts with zero-shot prompting [47] for the system role to have the LLM perform the desired tasks as a completion task. These settings applied equally to the use of the model's vision capabilities and to text generation.

4. Content Retrieval from the Declarative Memory of ACT-R

We demonstrated 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, we assumed an utterance of a human in a dialog with a social robot, where eventually the robot is supposed to search 'its' (ACT-R) memory for matching recollections based on this utterance. In the second example, the goal was to assign the visual impression of the robot to existing recollections. In both cases, we wanted to retrieve a suitable recollection – if available – and have the LLM use its content through prompt augmentation.

4.1. Example Application 1: Recollection Based on Text

The assumed scenario for this example application was as follows: A humanoid social robot is standing at a train station on one of the tracks and offers its services to travelers looking for information. Understanding the content of human questions – especially in relation to certain scenarios defined via the system prompt – and answering them with the help of an LLM generally works well. However, the LLM probably does not know all details of the current scenario and has no individual recollection of recent encounters between humans and the robot and therefore no experience of the particular situation. It is therefore advantageous if such factual knowledge can be provided to the LLM for questions about specific local conditions, for example, so that it can be incorporated into an answer that is otherwise perhaps freely formulated by the language model.

The LLM was instructed via system prompts to understand the basic setting and everything we needed it to output. A special feature of this application was the instruction to first create a summary of the question in the form of three keywords or phrases consisting of a few words instead of a complete answer and put them in square brackets for better programmatic handling. These keywords or phrases were passed as chunks to the cognitive model to search for a matching recollection. We stored memory chunks including suitable *fact phrases* for this test application as a priori knowledge in the declarative memory. For each API call that was used to generate a response to the human and not keywords or phrases, the entire dialog part was transferred to the GPT model without the previously generated keywords or phrases so that the model could ‘remember’ what was previously said in the dialog and refer to it as it progressed. The dialog text returned by the API was forwarded to the robot’s voice and tablet output.

4.1.1. Prompting the LLM

The system prompt for the LLM consisted of explanations of the scenario at hand in general and the instruction to output three keywords or short phrases as chunks that well represent the essential content of the human’s question, separated by commas in square brackets, for example for the question ‘Where can I find platform 1?’ corresponding keywords could be [Platform 1, Location, Direction Assistance].

For the scenario we used a system prompt like ‘You are a robot, your name is Pepper. You offer help and information to travelers in a train station by asking them where they want to go and how you can help.’

In the case that a chunk generation was required, we augmented this system prompt with ‘You don’t answer the question, but always create three keywords in square brackets that describe the traveler’s problem.’

And in the situation where the ACT-R model could not find any suitable recollection in the declarative memory, the system prompt was augmented with ‘You tell the traveler that you don’t know the answer to the current question and don’t make up an answer under any circumstances.’ This should prevent the language model from inventing a seemingly suitable answer without a factual basis.

4.1.2. Results

If a matching chunk was found in the declarative memory, the *fact phrase* contained in this chunk was transferred back to the robot application and used to augment the LLM’s system prompt, e.g. with ‘Platform 1 is the neighboring platform right here.’ as shown in Figure 3. This *fact phrase* was used by GPT-4o to respond with, for example ‘Platform 1 is right here, neighboring this information desk. You can easily access it by walking straight ahead.’.

The LLM therefore used the fact correctly, albeit in its own wording and with some embellishment from the scenario conditions. Details of a dialog run can be seen in Figure 4 showing an output in the graphical tool of the QiSDK robot emulator [48]. The output of the keywords in square brackets was only for testing purposes and would not be spoken by the robot in a production application. Part of a corresponding trace output of the running ACT-R model is shown in Figure 5.

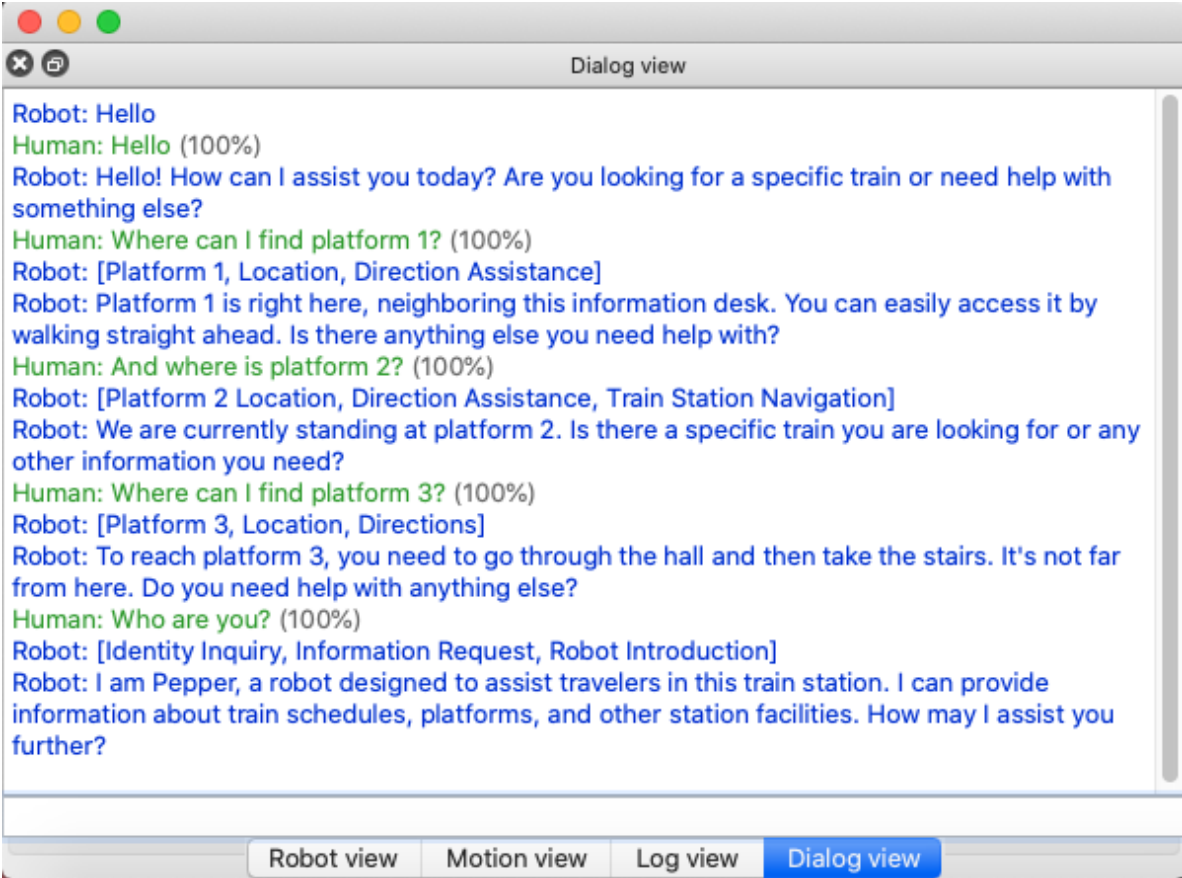
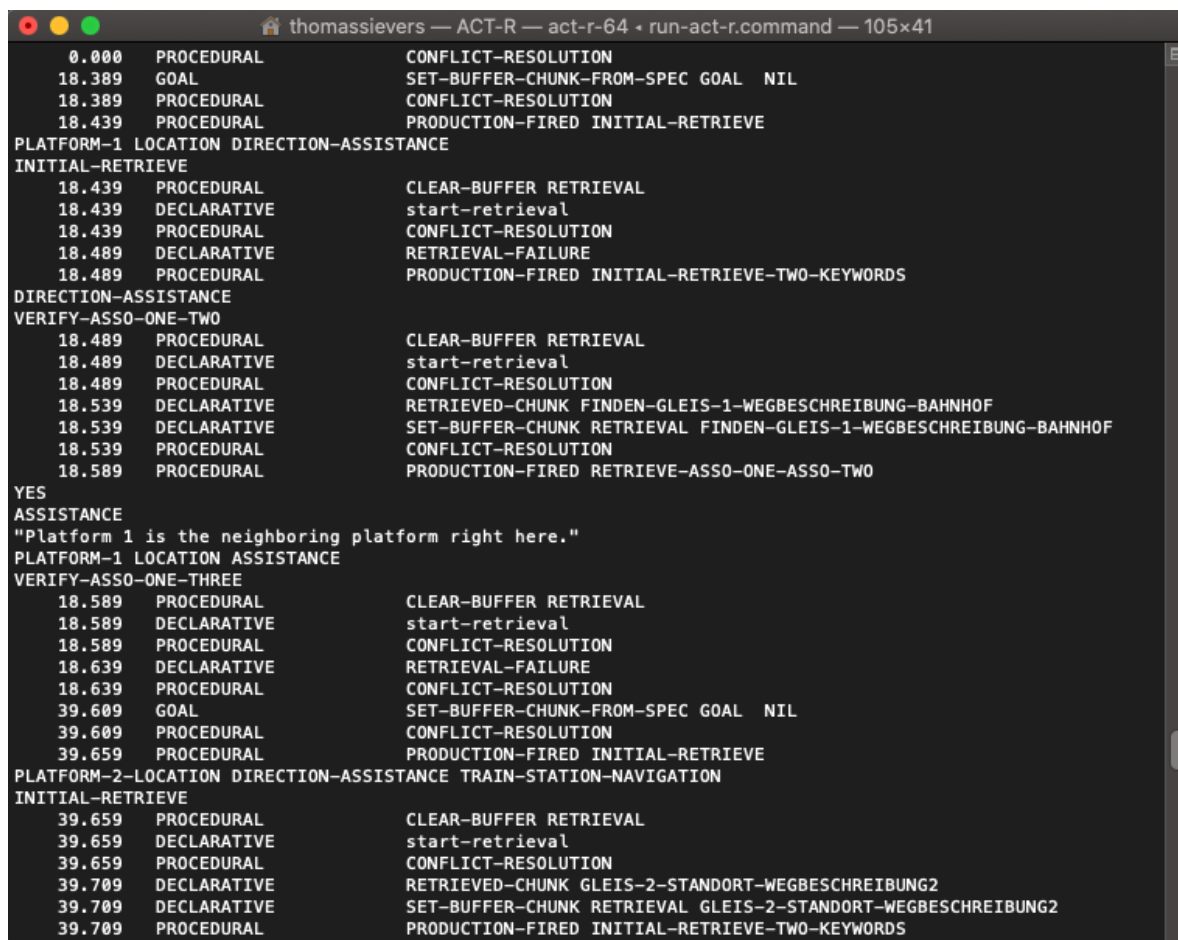


Figure 4. Dialog output in the graphical tool of the QiSDK robot emulator.



```

thomassievers — ACT-R — act-r-64 * run-act-r.command — 105x41
0.000 PROCEDURAL CONFLICT-RESOLUTION
18.389 GOAL SET-BUFFER-CHUNK-FROM-SPEC GOAL NIL
18.389 PROCEDURAL CONFLICT-RESOLUTION
18.439 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE
PLATFORM-1 LOCATION DIRECTION-ASSISTANCE
INITIAL-RETRIEVE
18.439 PROCEDURAL CLEAR-BUFFER RETRIEVAL
18.439 DECLARATIVE start-retrieval
18.439 PROCEDURAL CONFLICT-RESOLUTION
18.489 DECLARATIVE RETRIEVAL-FAILURE
18.489 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE-TWO-KEYWORDS
DIRECTION-ASSISTANCE
VERIFY-ASSO-ONE-TWO
18.489 PROCEDURAL CLEAR-BUFFER RETRIEVAL
18.489 DECLARATIVE start-retrieval
18.489 PROCEDURAL CONFLICT-RESOLUTION
18.539 DECLARATIVE RETRIEVED-CHUNK FINDEN-GLEIS-1-WEGBESCHREIBUNG-BAHNHOF
18.539 DECLARATIVE SET-BUFFER-CHUNK RETRIEVAL FINDEN-GLEIS-1-WEGBESCHREIBUNG-BAHNHOF
18.539 PROCEDURAL CONFLICT-RESOLUTION
18.589 PROCEDURAL PRODUCTION-FIRED RETRIEVE-ASSO-ONE-ASSO-TWO
YES
ASSISTANCE
"Platform 1 is the neighboring platform right here."
PLATFORM-1 LOCATION ASSISTANCE
VERIFY-ASSO-ONE-THREE
18.589 PROCEDURAL CLEAR-BUFFER RETRIEVAL
18.589 DECLARATIVE start-retrieval
18.589 PROCEDURAL CONFLICT-RESOLUTION
18.639 DECLARATIVE RETRIEVAL-FAILURE
18.639 PROCEDURAL CONFLICT-RESOLUTION
39.609 GOAL SET-BUFFER-CHUNK-FROM-SPEC GOAL NIL
39.609 PROCEDURAL CONFLICT-RESOLUTION
39.659 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE
PLATFORM-2-LOCATION DIRECTION-ASSISTANCE TRAIN-STATION-NAVIGATION
INITIAL-RETRIEVE
39.659 PROCEDURAL CLEAR-BUFFER RETRIEVAL
39.659 DECLARATIVE start-retrieval
39.659 PROCEDURAL CONFLICT-RESOLUTION
39.709 DECLARATIVE RETRIEVED-CHUNK GLEIS-2-STANDORT-WEGBESCHREIBUNG2
39.709 DECLARATIVE SET-BUFFER-CHUNK RETRIEVAL GLEIS-2-STANDORT-WEGBESCHREIBUNG2
39.709 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE-TWO-KEYWORDS

```

Figure 5. ACT-R trace output of the running model.

4.2. Example Application 2: Recollection Based on Vision

No specific scenario was assumed for this example. The system's ability to retrieve recollections from declarative memory by recognizing visual impressions from the robot's perspective should be demonstrated. The procedure was basically similar to that in example 1. Again, the LLM was instructed to create a description of the content in the form of three keywords, but this time it was about the content of an image taken with the robot's camera equipment. The robot's front camera took a picture every few seconds to get an actual impression of what the robot was seeing. To prepare a response to a human utterance based on visual impressions, the robot application sent the currently captured image to the GPT-4o model as a completion task, using the type *'image_url'* instead of *'text'*.

The three image-related keywords from the model's response were transmitted as chunks to the cognitive model to search for comparable chunks in declarative memory. Here, too, we have previously stored suitable memory chunks for our test scenario in the declarative memory. In the event of a match, the content of the *'phrase'* chunk slot was transferred to the robot application for use in prompt augmentation.

In our test applications we focused on usage of recollections for prompt augmentation of a GPT model. But the other way around, it was also possible with this system to write memory chunks and thus visual impressions, which are provided with keywords, into the declarative memory in order to retrieve them later in the outlined manner. Below we give an example of a system prompt that not only generates the three keywords, but also a *'phrase'* to describe the image content.

4.2.1. Prompting the LLM

For chunk generation, including a phrase that describes what is contained in this image and is to be stored in the *'phrase'* slot, we provided a system prompt like the following *'I am robot, my name is Pepper. I don't answer but create three keywords followed by a sentence with a short description describing what is in this image. The three keywords are enclosed by square brackets. The short sentence is enclosed by round brackets, for example [person, indoor, computer] (A sentence with a short description describing what is in this image).'*'.

For augmenting the prompt with a recollection we used for example *'I am robot, my name is Pepper. I can see what is described in the following:'* followed by the *phrase* recalled from declarative memory. In this way, the LLM could receive and process information about visual recollections stored in the past if similar sensory impressions and thus similar keywords are currently present.

4.2.2. Results

Figure 6 shows the generation of individual chunks from the keywords *'keyboard'*, *'indoor'*, *'christmas tree'* and in a frame the phrase *'A person is sitting indoors in front of a christmas tree appearing to engage in conversation or video call.'* supplied by the LLM in the running ACT-R model. These chunks were transferred to the declarative memory of the cognitive model, not only separately, but also as a composite chunk consisting of keywords and phrase. To be able to save keywords and phrases as chunks, spaces had to be replaced by a *'-'*. The procedure was reversed when the phrase was used for prompt augmentation. A retrieval of a recollection by the ACT-R model is shown in Figure 7 in framed areas. The model searches the stored memory chunks using the supplied keywords to find a matching recollection, the content of which is passed from the *'phrase'* slot to the robot application for use in prompt augmentation.

5. Discussion

On the one hand, cognitive models can be used within the framework of their cognitive architectures to anticipate human behavior and thus make human cognitive processes more understandable – even for a robot. 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

```

0.000 PROCEDURAL CONFLICT-RESOLUTION
#|Warning: Creating chunk A-PERSON-IS-SITTING-INDOORS-IN-FRONT-OF-A-CHRISTMAS-TREE-APPEARING-TO-ENGAGE-IN-CONVERSATION-OR-VIDEO-CALL.
ts |#
#|Warning: Creating chunk CHRISTMAS-TREE with no slots |#
#|Warning: Creating chunk INDOOR with no slots |#
#|Warning: Creating chunk KEYBOARD with no slots |#
8.889 GOAL SET-BUFFER-CHUNK-FROM-SPEC GOAL NIL
8.889 PROCEDURAL CONFLICT-RESOLUTION
8.939 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE
KEYBOARD INDOOR CHRISTMAS-TREE
INITIAL-RETRIEVE
8.939 PROCEDURAL CLEAR-BUFFER RETRIEVAL
8.939 DECLARATIVE start-retrieval
8.939 PROCEDURAL CONFLICT-RESOLUTION
8.989 DECLARATIVE RETRIEVED-CHUNK KEYBOARD-INDOOR-CHRISTMAS-TREE
8.989 DECLARATIVE SET-BUFFER-CHUNK RETRIEVAL KEYBOARD-INDOOR-CHRISTMAS-TREE
8.989 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE-TWO-KEYWORDS
CHRISTMAS-TREE
VERIFY-ASSO-ONE-TWO
8.989 PROCEDURAL CLEAR-BUFFER RETRIEVAL
8.989 DECLARATIVE start-retrieval
8.989 PROCEDURAL CONFLICT-RESOLUTION
9.039 DECLARATIVE RETRIEVAL-FAILURE
9.039 PROCEDURAL CONFLICT-RESOLUTION
9.089 PROCEDURAL PRODUCTION-FIRED VERIFY-ASSO-ONE-ASSO-TWO
CHRISTMAS-TREE

```

Figure 6. ACT-R model creates chunks for keywords and phrase

```

"Start checking for all three keywords"
9.589 PROCEDURAL CLEAR-BUFFER RETRIEVAL
9.589 DECLARATIVE start-retrieval
9.589 PROCEDURAL CONFLICT-RESOLUTION
9.639 DECLARATIVE RETRIEVED-CHUNK KEYBOARD-INDOOR-CHRISTMAS-TREE
9.639 DECLARATIVE SET-BUFFER-CHUNK RETRIEVAL KEYBOARD-INDOOR-CHRISTMAS-TREE
9.639 PROCEDURAL CONFLICT-RESOLUTION
9.689 PROCEDURAL PRODUCTION-FIRED RETRIEVE-ASSO-ONE-ASSO-TWO-ASSO-THREE
YES
A-PERSON-IS-SITTING-INDOORS-IN-FRONT-OF-A-CHRISTMAS-TREE-APPEARING-TO-ENGAGE-IN-CONVERSATION-OR-VIDEO-CALL.
KEYBOARD INDOOR CHRISTMAS-TREE
FINISHED
9.689 PROCEDURAL CLEAR-BUFFER RETRIEVAL
9.689 PROCEDURAL CONFLICT-RESOLUTION
21.388 GOAL SET-BUFFER-CHUNK-FROM-SPEC GOAL NIL
21.388 PROCEDURAL CONFLICT-RESOLUTION
21.438 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE
KEYBOARD INDOOR CHRISTMAS-TREE
INITIAL-RETRIEVE
21.438 PROCEDURAL CLEAR-BUFFER RETRIEVAL
21.438 DECLARATIVE start-retrieval
21.438 PROCEDURAL CONFLICT-RESOLUTION
21.488 DECLARATIVE RETRIEVED-CHUNK GOAL-CHUNK0-0
21.488 DECLARATIVE SET-BUFFER-CHUNK RETRIEVAL GOAL-CHUNK0-0
21.488 PROCEDURAL PRODUCTION-FIRED INITIAL-RETRIEVE-TWO-KEYWORDS
CHRISTMAS-TREE
VERIFY-ASSO-ONE-TWO

```

Figure 7. ACT-R model retrieves chunk with keywords and phrase from declarative memory

more humane way. Such processes can therefore be used to make human actions understandable for the machine, but also to make the robot's actions more comprehensible for the human in HRI. Kambhampati introduced the term *human-aware AI* to refer to AI systems that are explainable to humans in the loop and capable of working synergistically with humans [49].

The possibility of using cognitive architectures to unlock human-like judgment and decision-making capabilities such as instance-based learning and intuitive decision making for social robots can be an opportunity for greater acceptance and trust through more common ground and similarities in the way we – humans and robots – think. With the help of LLMs and VLMs and the corresponding sensors, robots can access and interpret the same information that is available to humans. The use of declarative and procedural memory in cognitive models makes it possible to process the data provided by robot sensors and language models in a human-centered way. We only outline the possibilities here with our example applications for storing visual impressions or facts together with keywords and using them for prompt augmentation of an LLM. In general, there are far more options available for a programmatic implementation of cognitive models with ACT-R or similar frameworks in combination with social robots. Harnessing the memory and decision-making capabilities of cognitive models can even be a step towards Continual Learning (CL) that can adapt to the ever-changing needs, preferences and environments of users [50].

The proposed system and procedure is not without shortcomings and open problems. It is well known that LLMs and VLMs are sometimes prone to hallucinations. One possible way to reduce the risk of such hallucinations and the reproduction of made-up statements by an LLM such as OpenAI's GPT would be to provide the language model with additional information tailored to individual scenarios via the system prompt, as we suggest in our application example based on remembered facts. However, even such a method of constraining the LLM does not offer absolute certainty for the exclusion of hallucinations.

So far, we have only tested the method with a few, mostly predefined memory contents, but we consider it scalable with regard to more complex cognitive models and the use of concepts such as *forgetting*, *learning new facts* and *reinforcement of recollections* or further possibilities made possible by the cognitive architecture of ACT-R.

The time required to retrieve relevant recollections could play a limiting role depending on the amount of memory chunks available. In addition, for each interaction with the human user, our system requires two consecutive API calls to the GPT model, both of which have a certain latency. Together, these time aspects may cause a noticeable and unnatural delay in interaction, even if there was no significant difference to conventional use with just one call to the GPT API in our tests.

And it also seems to be a disadvantage and increases the complexity that the implementation of the cognitive architecture in our test applications runs as a standalone version on an extra computer and not on the robot itself. However, this problem could be solved by using other robot models and thus other ways of implementing the cognitive architecture such as *pyactr* as an ACT-R implementation for the Python-based robot *navel* [51,52].

6. Conclusion

We propose a design and development approach for a combined system consisting of an ACT-R cognitive model and a humanoid social robot to endow the robot with human-like memory capabilities. Recollections from the declarative memory of the ACT-R model could be retrieved using real-world data obtained by the robot via an LLM or VLM. The procedural memory, which consists of the productions of the cognitive model, was used to retrieve these recollections and return them to the robot as instructions for action. In addition, real-world data captured by the robot could be stored as memory chunks in the cognitive model's declarative memory.

In an example application, such a system was used to improve the correctness and accuracy of GPT-4o's reasoning capabilities by using recollections for prompt augmentation. Another example

application delivered keyword labeled visual impressions of the robot to the declarative memory or retrieved recollections based on these impressions. We tested this system with the social robot Pepper. In principle, however, this method can also be used independently of HRI scenarios.

Regarding explainability aspects, our system provides approaches for a possible constraint of LLMs to generate robot utterances in HRI by comprehensible memory contents. The type of possible connection or integration of the cognitive model depends on the robot's operating system and on whether or which ACT-R implementations or implementations of a comparable cognitive architecture are available for it. For example, there are direct implementations for Python, which means that the standalone version of ACT-R on an external computer and thus also the TCP/IP connection between the cognitive model and the robot application could be avoided.

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 memory skills and 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. We are 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.

Author Contributions: Conceptualization, Thomas Sievers and Nele Russwinkel; methodology, Thomas Sievers and Nele Russwinkel; writing—original draft preparation, Thomas Sievers.; writing—review and editing, Thomas Sievers and Nele Russwinkel; supervision, Nele Russwinkel; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACT-R	Adaptive Control of Thought-Rational
API	Application Programming Interface
AI	Artificial Intelligence
ACo	Artificial Cognition
CL	Continual Learning
CoALA	Cognitive Architectures for Language Agents
GPT	Generative Pretrained Transformer
HRI	Human-Robot Interaction
IBL	Instance-based Learning Theory
LLM	Large Language Model
LMM	Large Multi-modal Models
LTM	Long-Term Memory
RAG	Retrieval-Augmented Generation
TCP/IP	Transmission Control Protocol/Internet Protocol
VLM	Vision-Language Model

References

1. Bisaz, R.; Travaglia, A.; Alberini, C. The Neurobiological Bases of Memory Formation: From Physiological Conditions to Psychopathology. *Psychopathology* **2014**, *47*. <https://doi.org/10.1159/000363702>.

2. Zhang, J. How memories are stored in the brain: the declarative memory model, 2024, [arXiv:q-bio.NC/2403.17985].
3. Myers, N.E.; Stokes, M.G.; Nobre, A.C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences* **2017**, *21*, 449–461. <https://doi.org/10.1016/j.tics.2017.03.010>.
4. Braun, E.K.; Wimmer, G.E.; Shohamy, D. Retroactive and graded prioritization of memory by reward. *Nature Communications* **2018**, *9*, 2041–1723. <https://doi.org/10.1038/s41467-018-07280-0>.
5. Sandini, G.; Sciutti, A.; Morasso, P. Artificial cognition vs. artificial intelligence for next-generation autonomous robotic agents. *Frontiers in Computational Neuroscience* **2024**, *18*. <https://doi.org/10.3389/fncom.2024.1349408>.
6. 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* **2024**, *12*. <https://doi.org/10.18103/mra.v12i6.5383>.
7. Janik, R.A. Aspects of human memory and Large Language Models, 2024, [arXiv:cs.CL/2311.03839].
8. Jeong, H.; Lee, H.; Kim, C.; Shin, S. A Survey of Robot Intelligence with Large Language Models. *Applied Sciences* **2024**, *14*. <https://doi.org/10.3390/app14198868>.
9. 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, [arXiv:cs.CV/2404.07214].
10. Kawaharazuka, K.; Obinata, Y.; Kanazawa, N.; Okada, K.; Inaba, M. Robotic Applications of Pre-Trained Vision-Language Models to Various Recognition Behaviors. In Proceedings of the 2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids). IEEE, 2023, p. 1–8. <https://doi.org/10.1109/humanoids57100.2023.10375211>.
11. Tan, R.; Yang, H.; Jiao, S.; Shan, L.; Tao, C.; Jiao, R. Smart Robot Manipulation using GPT-4o Vision. In Proceedings of the 7th European Industrial Engineering and Operations Management Conference, Augsburg, Germany. IEOM Society International, 2024. <https://doi.org/10.46254/EU07.20240272>.
12. Kurup, U.; Lebiere, C. What can cognitive architectures do for robotics? *Biologically Inspired Cognitive Architectures* **2012**, *2*, 88–99. <https://doi.org/10.1016/j.bica.2012.07.004>.
13. Anderson, J.R.; Bothell, D.; Byrne, M.D.; Douglass, S.; Lebiere, C.; Qin, Y. An integrated theory of the mind **2004**. *111*, 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>.
14. 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* **2015**, *4*, 180–190. Modeling and Aiding Intuition in Organizational Decision Making, <https://doi.org/https://doi.org/10.1016/j.jarmac.2014.06.002>.
15. Gonzalez, C.; Dutt, V.; Lebiere, C. Validating instance-based learning mechanisms outside of ACT-R. *Journal of Computational Science* **2013**, *4*, 262–268. PEDISWESA 2011 and Sc. computing for Cog. Sciences, <https://doi.org/https://doi.org/10.1016/j.jocs.2011.12.001>.
16. Gonzalez, C.; Lerch, J.F.; Lebiere, C. Instance-based learning in dynamic decision making. *Cognitive Science* **2003**, *27*, 591–635. [https://doi.org/https://doi.org/10.1016/S0364-0213\(03\)00031-4](https://doi.org/https://doi.org/10.1016/S0364-0213(03)00031-4).
17. 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**.
18. Kahneman, D. *Thinking, fast and slow*; Farrar, Straus and Giroux: New York, 2011.
19. Whitehill, J. Understanding ACT-R - an Outsider's Perspective, 2013.
20. 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* **2018**, *9*. <https://doi.org/10.3389/fpsyg.2018.01828>.
21. Eichenbaum, H. Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory. *Neuron* **2004**, *44*, 109–120. <https://doi.org/https://doi.org/10.1016/j.neuron.2004.08.028>.
22. Hayne, H.; Boniface, J.; Barr, R. The Development of Declarative Memory in Human Infants: Age-Related Changes in Deferred Imitation. *Behavioral neuroscience* **2000**, *114*, 77–83. <https://doi.org/10.1037/0735-7044.114.1.77>.
23. OpenAI. The most powerful platform for building AI products. Technical report, 2025.
24. Carnegie Mellon University. ACT-R Software. Technical report, 2024.
25. Sievers, T.; Russwinkel, N. How to use a cognitive architecture for a dynamic person model with a social robot in human collaboration. *CEUR Workshop Proceedings* **2024**.

26. Lykov, A.; Litvinov, M.; Konenkov, M.; Prochii, R.; Burtsev, N.; Abdulkarim, A.A.; Bazhenov, A.; Berman, V.; Tsetserukou, D. CognitiveDog: Large Multimodal Model Based System to Translate Vision and Language into Action of Quadruped Robot, New York, NY, USA, 2024; HRI '24, p. 712–716. <https://doi.org/10.1145/3610978.3641080>.
27. Wang, J.; Shi, E.; Hu, H.; Ma, C.; Liu, Y.; Wang, X.; Yao, Y.; Liu, X.; Ge, B.; Zhang, S. Large language models for robotics: Opportunities, challenges, and perspectives. *Journal of Automation and Intelligence* **2024**. <https://doi.org/https://doi.org/10.1016/j.jai.2024.12.003>.
28. Hu, Y.; Lin, F.; Zhang, T.; Yi, L.; Gao, Y. Look Before You Leap: Unveiling the Power of GPT-4V in Robotic Vision-Language Planning, 2023, [\[arXiv:cs.RO/2311.17842\]](https://arxiv.org/abs/2311.17842).
29. Wake, N.; Kanehira, A.; Sasabuchi, K.; Takamatsu, J.; Ikeuchi, K. GPT-4V(ision) for Robotics: Multimodal Task Planning from Human Demonstration, 2024, [\[arXiv:cs.RO/2311.12015\]](https://arxiv.org/abs/2311.12015). <https://doi.org/https://doi.org/10.1109/LRA.2024.3477090>.
30. Yoshida, T.; Baba, S.; Masumori, A.; Ikegami, T. Minimal Self in Humanoid Robot "Alter3" Driven by Large Language Model, 2024, [\[arXiv:cs.RO/2406.11420\]](https://arxiv.org/abs/2406.11420).
31. Leidner, D. Toward Robotic Metacognition: Redefining Self-Awareness in an Era of Vision-Language Models. *40th Anniversary of the IEEE Conference on Robotics and Automation* **2024**.
32. Niu, Q.; Liu, J.; Bi, Z.; Feng, P.; Peng, B.; Chen, K.; Li, M.; Yan, L.K.; Zhang, Y.; Yin, C.H.; et al. Large Language Models and Cognitive Science: A Comprehensive Review of Similarities, Differences, and Challenges, 2024.
33. Wu, S.; Oltramari, A.; Francis, J.; Giles, C.L.; Ritter, F.E. Cognitive LLMs: Towards Integrating Cognitive Architectures and Large Language Models for Manufacturing Decision-making, 2024.
34. González-Santamarta, M.Á.; Rodríguez-Lera, F.J.; Guerrero-Higuera, Á.M.; Matellán-Olivera, V. Integration of Large Language Models within Cognitive Architectures for Autonomous Robots, 2024, [\[arXiv:cs.RO/2309.14945\]](https://arxiv.org/abs/2309.14945).
35. Sievers, T.; Russwinkel, N.; Möller, R. Grounding a Social Robot's Understanding of Words with Associations in a Cognitive Architecture. In Proceedings of the Proceedings of the 17th International Conference on Agents and Artificial Intelligence - Volume 3: ICAART. INSTICC, SciTePress, 2025, pp. 406–410. <https://doi.org/10.5220/0013144200003890>.
36. Knowles, K.; Witbrock, M.; Dobbie, G.; Yogarajan, V. A Proposal for a Language Model Based Cognitive Architecture. *Proceedings of the AAAI Symposium Series* **2024**.
37. Leivada, E.; Marcus, G.; Günther, F.; Murphy, E. A Sentence is Worth a Thousand Pictures: Can Large Language Models Understand Human Language and the World behind Words?, 2024.
38. Sun, R. Can A Cognitive Architecture Fundamentally Enhance LLMs? Or Vice Versa?, 2024.
39. He, Z.; Lin, W.; Zheng, H.; Zhang, F.; Jones, M.W.; Aitchison, L.; Xu, X.; Liu, M.; Kristensson, P.O.; Shen, J. Human-inspired Perspectives: A Survey on AI Long-term Memory, 2025, [\[arXiv:cs.AI/2411.00489\]](https://arxiv.org/abs/2411.00489).
40. Jiang, X.; Li, F.; Zhao, H.; Wang, J.; Shao, J.; Xu, S.; Zhang, S.; Chen, W.; Tang, X.; Chen, Y.; et al. Long Term Memory: The Foundation of AI Self-Evolution, 2024, [\[arXiv:cs.AI/2410.15665\]](https://arxiv.org/abs/2410.15665).
41. Hou, Y.; Tamoto, H.; Miyashita, H. "My agent understands me better": Integrating Dynamic Human-like Memory Recall and Consolidation in LLM-Based Agents. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2024; CHI EA '24. <https://doi.org/10.1145/3613905.3650839>.
42. Summers, T.R.; Yao, S.; Narasimhan, K.; Griffiths, T.L. Cognitive Architectures for Language Agents, 2024, [\[arXiv:cs.AI/2309.02427\]](https://arxiv.org/abs/2309.02427).
43. Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, M.; Wang, H. Retrieval-Augmented Generation for Large Language Models: A Survey, 2024, [\[arXiv:cs.CL/2312.10997\]](https://arxiv.org/abs/2312.10997).
44. Sanmartin, D. KG-RAG: Bridging the Gap Between Knowledge and Creativity, 2024, [\[arXiv:cs.AI/2405.12035\]](https://arxiv.org/abs/2405.12035).
45. Janik, R.A. Aspects of human memory and Large Language Models, 2024.
46. Aldebaran, United Robotics Group and Softbank Robotics. Pepper. Technical report, 2025.
47. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners, 2020. <https://doi.org/10.48550/arXiv.2005.14165>.
48. Aldebaran, United Robotics Group and Softbank Robotics. Pepper SDK for Android. Technical report, 2025.
49. Sreedharan, S.; Kulkarni, A.; Kambhampati, S. Explainable Human-AI Interaction: A Planning Perspective, 2024, [\[arXiv:cs.AI/2405.15804\]](https://arxiv.org/abs/2405.15804).

50. Ayub, A.; De Francesco, Z.; Mehta, J.; Yaakoub Agha, K.; Holthaus, P.; Nehaniv, C.L.; Dautenhahn, K. A Human-Centered View of Continual Learning: Understanding Interactions, Teaching Patterns, and Perceptions of Human Users Toward a Continual Learning Robot in Repeated Interactions. *J. Hum.-Robot Interact.* **2024**, *13*. <https://doi.org/10.1145/3659110>.
51. jakdot. pyactr. Technical report, 2025.
52. navel robotics GmbH. navel. Technical report, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.