Personalized Interactions With a Social Robot Based on Recollections From a Cognitive Model

Thomas Sievers*, Nele Russwinkel

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

Abstract

In order to equip social robots with human-like cognitive abilities such as memory and language comprehension, cognitive architectures in combination with Large Language Models (LLMs) have the potential to act as suitable components. We demonstrate the use of an Adaptive Control of Thought-Rational (ACT-R) model in combination with an LLM to store experiences from human-robot interaction (HRI) with a humanoid social robot in the declarative memory of the cognitive architecture. These experiences can be retrieved from memory as associated recollections and used for prompt augmentation of the LLM to enable personalized interactions with references to previous encounters. This type of memory also enables the creation, storage and updating of person models from interactions with different people, so that the robot can get to know these people better during temporally unrelated interactions and react to them individually. In special application scenarios, it may also be necessary to connect data sources via interfaces to expand the robot's knowledge base. In such cases, the question of decision-making arises from a general dialog situation to clarify when such additional sources must be used for a robot response. The system for our practical study uses special chunks in ACT-R's declarative memory for such decisions. We demonstrate the use of such a system with the social robot Navel.

Keywords

social robots, human-robot interaction, cognitive architecture, ACT-R, large language model

1. Introduction

Social robots must be designed and developed to meet the needs of their social environment and be able to respond appropriately and comprehensibly [1]. Their understanding of social norms and expectations should guide their decision-making, and they are required to produce mental simulations to understand and predict the thoughts, feelings and intentions of others in a broader social context. To achieve this, the mental states of humans in the loop are modeled under the term "human-aware AI", and a human-centric perspective should improve the acceptance of social robots in Human-Robot Interaction (HRI) by increasing human trust for successful collaboration [2].

Furthermore, when it comes to developing truly human-like cognitive abilities, individual experiences mediated by social interactions between humans and machines are essential for an AI agent [3]. The gradual acquisition of knowledge and skills leads to an autonomous decision-making ability through interaction with the physical and social 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, i.e. the mental simulation of actions, including planning, predicting, imagining scenarios and possible future events. Norbert Wiener, the inventor of cybernetics, was a great advocate of a fundamental commonality between natural and artificial paradigms in terms of intelligence and cognition [4]. However, the AI community often leans in the opposite direction when considering the extent to which one should focus on a human brain-inspired approach to the development of intelligent agents exhibiting situated cognition. Situated cognition as an approach to learning supports the idea that learning takes place when an individual does something – usually in social interaction with others – in the real world [5]. Such learning takes place in a situated activity that has a social, cultural and

Perspectives on Humanities-Centered AI and Formal & Cognitive Reasoning Workshop 2025,(CHAI 2025 & FCR 2025), Joint Workshop at the 48th German Conference on Artificial Intelligence, September 16, 2025, Potsdam, Germany

*Corresponding author.
t.sievers@uni-luebeck.de (T. Sievers); nele.russwinkel@uni-luebeck.de (N. Russwinkel)
0000-0002-8675-0122 (T. Sievers); 0000-0003-2606-9690 (N. Russwinkel)

physical context and where cognition cannot be separated from this context.

In order to interact naturally with a robot, we would expect it to be able to remember and relate to what we have already experienced or discussed with it in the past within a contextual framework. This requires an episodic memory for the robot that is consistent and able to recall recollections associated with us, representing a kind of "narrative self" that includes identity and continuity over time [6]. A long-term memory that functions analogously to human memory, with storage of individual facts and experiences, recollection through associations of current experiences with stored experiences and functions such as forgetting rarely used or reinforcing frequent memories, is likely to be a beneficial cognitive ability for a robot in HRI. A reference to concrete recollections would also be useful with regard to the explanatory or black box problem of AI [7].

But how can a robot store individual experiences in interaction with humans, remember them later and learn from them? How is it possible for the robot to build up a wealth of individual experience that goes beyond short-term interactions? If this is to happen in a way that is comparable to human abilities, a cognitive architecture that has been tried and tested in cognitive psychology is likely to be essential. For this reason, we chose the Adaptive Control of Thought-Rational (ACT-R) [8] as the cognitive architecture to develop human-like memory functionalities for the robot. A Large Language Model (LLM) was used to formulate the dialog parts of the robot and the recollections to be stored. The associated recollections were in turn provided to the LLM via prompt augmentation for consideration in its utterances.

2. Related Work

With the recent successes of language models in various fields, interest in the interplay between LLMs and cognitive architectures has also increased, and ways of integrating LLMs with cognitive architectures have been discussed [9, 10, 11, 12]. Insights from human cognition and psychology that underlie cognitive architectures can contribute to the development of systems that are more efficient, reliable and human-like [13]. The storage of episodic recollections has long been studied in simulation models of human memory [14]. Paplu et al. investigated the use of long-term memory to generate context for customized interactions and the connection with the interaction partner on an emotional level in a personalized HRI [15]. They employed a MySQL database to store the memory content.

We demonstrated in previous work the ability of an ACT-R cognitive architecture connected to a social robot to store and process recollections with the help of an LLM [16]. An integration of ACT-R with LLMs was applied for human-centered decision making by using knowledge from the decision process of the cognitive model as neural representations in trainable layers of the LLM [17]. Sumers et al. proposed a language agent with a framework for modular memory components, a structured action space for interaction with internal memory and the external environment, and a generalized decision making process that combines insights from symbolic artificial intelligence and cognitive science [18]. A conceptual framework for a combination of sequential cognitive situation modeling and continuous motion control of a robot was proposed by Hao et al. to overcome the separation between the cognitive and physical levels in HRI [19]. In order to equip the robot with a good anticipatory model of the individual that can adapt to different individual situation representations, a wide range of sensory input data from the human was required during the interaction. Knowles et al. proposed a system architecture that combined LLMs and cognitive architectures with an analogy to "fast" and "slow" thinking in human cognition [20, 11].

3. Methodology

Our approach to generate and utilize individual experiences for a robot in social interaction with humans combines the memory capabilities of the ACT-R cognitive architecture with the linguistic capabilities of an LLM to both formulate recollections and refer to those recollections in utterances

Figure 1: Social robot Navel

during a subsequent interaction. In addition, individual characteristics of the interaction partner (e.g. preferences, interests, etc.) are stored in a person model so that they can be referred to later.

Special application scenarios (for example in a care context) may make it necessary to connect external sources or internal data of an organization via interfaces in order to expand the robot's knowledge base for the provision of information. Our proposed system uses special chunks in the declarative memory of the cognitive architecture, to decide from a general dialog situation on a variety of topics between humans and robots when such an additional source must be used for a robot response.

The humanoid social robot Navel as shown in Fig. 1 serves as a platform for the implementation of our system.

3.1. Social Robot Navel

Designed as a care robot by navel robotics GmbH, Navel's task is to autonomously increase the well-being of people in need of care by additional emotional and cognitive activation and to relieve the burden on caregivers [21]. The robot has been used in several nursing homes since October 2023. Based on a Linux OS the robot features a Python SDK to program custom behavior with direct access to all functions like face detection, emotion recognition, sound processing, and sound source localization. It has a height of 72 cm.

3.2. Cognitive Model

Cognitive architectures like ACT-R 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 human cognitive modeling [22]. 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 to develop a situational understanding regarding human behavior for adequate reactions. 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 [23, 24, 25, 26]. To create and run ACT-R cognitive models directly on the robot as part of our application, we used the Python package “pyactr” [27].

3.2.1. Creating Recollections

In ACT-R, declarative knowledge is represented in the form of chunks, i.e. representations of individual properties as strings, each of which can be accessed via a labeled slot. The cognitive model we developed to test our hypothesis should receive chunks with the name of the person speaking and keywords for the memory association as well as the actual memory fact as a phrase from the robot application. Then it had to check whether there were already any memory chunks with this same name indicating an existing recollection for this person. Given the name, the model also searched for matching chunks already stored in the declarative memory for the specified keywords. For simplicity reasons we assumed the transfer of not more than three keywords. The model’s productions of the procedural memory checked all combinations of the sequence of keywords for a match with memory content and generated a hit if at least one of the keywords were matching. In this case, the associated recollection was called up and returned to the robot application for LLM prompt augmentation. Generally, the new chunk was stored in the declarative memory supplemented by the time at which this recollection was created in order to be able to make a temporal classification if necessary.

3.2.2. Person Model

According to Person Model Theory (PMT), we generally understand others based on specific background knowledge that we accumulate over the course of our lives and use to develop “person models” of ourselves, other people and groups [28, 29]. These person models are the basis on which we recognize and evaluate people with both mental and physical characteristics. To enable the robot to remember individual characteristics of the people it has dealt with, we stored them in a person model linked to the person’s name in ACT-R’s declarative memory. A memory chunk was to be created for each person with slots such as interest, hobby, task in the organization and special sensitivities. One of the robot’s goals during interaction was to gradually fill these gaps with relevant content by asking specific questions about the slots that were still open. To do this, the system prompts for the LLM were expanded to include a question about the next open slot in the person model. The LLM was then supposed to provide a keyword, which was stored in the person model chunk.

3.2.3. Cognitive Control for Task Selection

We also defined a chunk type for task selection in ACT-R, which essentially contained a keyword slot and a slot for the assigned task. Our test scenario included tasks such as querying a lunch menu via an interface, consulting or completing the person model or engaging in conversation using recollections from previous interactions with the conversation partner. Data from external sources was transferred to the LLM system prompt, as were recollections or other remembered data. The declarative memory was equipped with several chunks of this type in order to be able to use results from the model in the robot application for a selection of predefined tasks. The existing assignment between keyword and task could be dynamically extended to include the remaining keywords from the keyword triple. Each utterance of the human in dialog with the robot started a search via the productions of the cognitive model for all three keywords created by the LLM to determine whether there was a stored memory chunk with a task in the declarative memory for one of these keywords.

Since Retrieval Augmented Generation (RAG) offers an approach to combine external sources in LLMs with many optimizations to include relevant knowledge, it could provide a different approach to make external data available [30, 31]. But here too, an instance would be needed to make the decision on the appropriate action selection.

3.3. Prompting the LLM

We applied OpenAI's Generative Pretrained Transformer (GPT) language model to generate speech and process the dialog content of an interaction for finding the keywords and fact phrase to store in the declarative memory of the ACT-R model [32]. The system prompts for the LLM differed depending on the task. In the event that a person model needed to be completed, the LLM was instructed to specifically ask for missing characteristics such as the interests of the human interlocutor and to generate a term from the answer, which was then saved in the corresponding slot of the person model. Otherwise, in addition to formulating an answer in the dialog, the LLM had the task of creating the fact phrase to be stored as a recollection and three suitable keywords from the facts just discussed.

We chose the model gpt-4o-mini to create all conversational parts of the robot. For the instruction of the GPT model, we used prompts with zero-shot prompting for the system role to have the LLM perform the desired tasks as a completion task [33].

4. Interaction Process Comprising Individual Experiences

We tested our proposed system with the social robot Navel. Fig. 2 schematically shows the course of an interaction between the robot and a human whereby the robot can remember past encounters and characteristics of the conversation partner. At the very beginning, the application loads all previous contents of the declarative memory from a text file saved for this purpose so that they are available to the cognitive model even after a restart. The robot waits until a person appears in its field of vision and looks at it. If it perceives a person, it greets them. As soon as the human speaks to the robot, an attempt is made to derive a task for the robot from the content of what is said, as described in Chapter 3.2.3, and to recognize whether data from external sources is required, for example.

Now it is important whether the robot can determine the identity of the person. Since it is currently not possible to access Navel's camera recordings for a visual detection, we use speaker recognition, i.e. we check whether the robot recognizes the speaker's voice by comparing it with audio samples of already known speakers. If the recognition fails, the robot asks for the name of the speaker, creates voice samples for recognition and generates a person model for later completion. Once a person has been recognized, the declarative memory of the ACT-R model is searched for recollections concerning this person and information from the associated person model. If the person model is incomplete, an attempt is made to complete it.

Content from recollections, the person model or external sources is fed to the LLM via the prompt for consideration and reference in the conversation. As described in Chapter 3.3, the LLM generates a response as well as keywords and fact phrases for storage in memory, whereby the name of the interaction partner and the date and time of the interaction are also stored for possible later reference. If the person responds to the robot's answer, a new turn begins, otherwise the interaction ends.

5. Discussion and Initial Findings

Our investigations into the optimal use of the ACT-R cognitive architecture to generate a long-term memory for a social robot are still in their infancy. In particular, meaningful comparative studies on the perception of the described memory capabilities of the robot in interaction with humans are lacking.

However, as a proof-of-concept for technical feasibility, we can confirm both the generation and storage of recollections of events during a dialog interaction as well as retrieval and reference during the conversation. Individual recollections can be combined into a chain of episodic memory and used for prompt augmentation. Task selection and the inclusion of the person model works in principle in the way described, but with an unsatisfactory level of reliability to date. A detailed quantitative evaluation is still pending. An increase in accuracy in the differentiation of possible tasks from the dialog, e.g. to include data from an external source via interface, could possibly be increased by including the

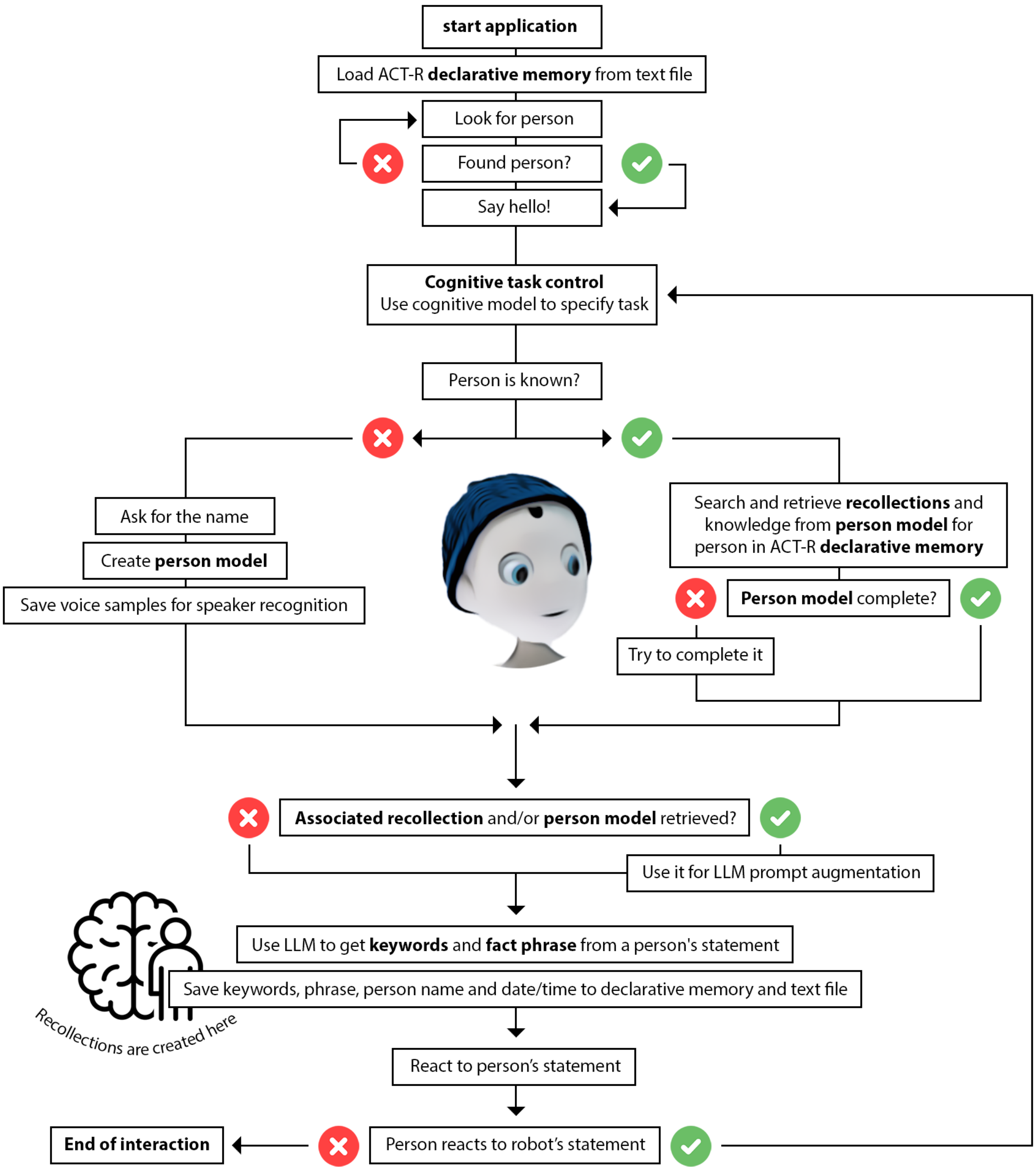

Figure 2: Interaction process with access to individual previous experiences (supplemented graphic from [34])

reasoning capabilities of the LLM in this decision-making process in addition to stored memory content and procedural memory of the ACT-R model. This would need to be examined in the future.

Nonetheless, the basic possibility of accessing recollections available in plain text is an advantage with regard to the explanatory or black box problem of AI or LLMs. Actual retrieval is influenced by the productions of procedural memory in the ACT-R model as well as by settings such as noise, utility for the subsymbolic processes and the like.

Eventually, the effectiveness of the process depends very much on the exact formulation of the prompts for the LLM, especially if the robot's task changes during the ongoing interaction (e.g., completing the person model vs. conversation considering remembered knowledge).

6. Conclusion

Our idea that individual recollections of a robot retrieved by a model of the ACT-R cognitive architecture create an enriched, more human-like personalized interaction experience between a robot and a human has yet to be proven.

Progress along this path could lead to a memory system that supports the accumulation of knowledge at ever higher levels of abstraction, and would possibly also be capable of prospection, i.e. the mental simulation of actions. This would require a significantly expanded ACT-R model. However, the collection of personal information when dealing with humans requires responsible handling of this data. Although the recollections are stored directly on the robot, the use of an external LLM could represent a flaw in terms of data protection.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL in order to: Grammar and spelling check. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] H. Mahdi, S. A. Akgun, S. Saleh, K. Dautenhahn, A survey on the design and evolution of social robots — past, present and future, Robotics and Autonomous Systems 156 (2022). doi:10.1016/j.robot.2022.104193.

[2] S. Sreedharan, A. Kulkarni, S. Kambhampati, Explainable human-ai interaction: A planning perspective, 2024. URL: https://arxiv.org/abs/2405.15804. arXiv:2405.15804.

[3] G. Sandini, A. Sciutti, P. Morasso, Artificial cognition vs. artificial intelligence for next-generation autonomous robotic agents, Frontiers in Computational Neuroscience 18 (2024). doi:10.3389/fncom.2024.1349408.

[4] N. Wiener, Cybernetics: Or Control and Communication in the Animal and the Machine, MIT Press, Cambridge, MA, 1948. doi:10.7551/mitpress/11810.001.0001.

[5] B. G. Wilson, K. M. Myers, Situated cognition in theoretical and practical context [book chapter], in: Theoretical foundations of learning environments, Lawrence Erlbaum Associates Publishers, 2000, pp. 57–88.

[6] S. Gallagher, Philosophical conceptions of the self: implications for cognitive science, Trends in Cognitive Sciences 4 (2000) 14–21. doi:10.1016/S1364-6613(99)01417-5.

[7] J. E. T. Taylor, G. W. Taylor, Artificial cognition: How experimental psychology can help generate explainable artificial intelligence, Psychonomic Bulletin & Review 28 (2021) 454–475. doi:10.3758/s13423-020-01825-5.

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

[9] M. Binz, E. Schulz, Turning large language models into cognitive models, 2023. URL: https://arxiv.org/abs/2306.03917. arXiv:2306.03917.

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

[11] K. Knowles, M. Witbrock, G. Dobbie, V. Yogarajan, A proposal for a language model based cognitive architecture, Proceedings of the AAAI Symposium Series (2024). URL: https://api.semanticscholar.org/CorpusID:267206594.

[12] Q. Niu, J. Liu, Z. Bi, P. Feng, B. Peng, K. Chen, M. Li, L. K. Yan, Y. Zhang, C. H. Yin, C. Fei, T. Wang, Y. Wang, S. Chen, Large language models and cognitive science: A comprehensive review of similarities, differences, and challenges, 2024. URL: https://arxiv.org/abs/2409.02387.

[13] R. Sun, Can a cognitive architecture fundamentally enhance llms? or vice versa?, 2024. URL: https://arxiv.org/abs/2401.10444.
[14] D. L. Hintzman, Minerva 2: A simulation model of human memory, Behavior Research Methods, Instruments, & Computers 16 (1984) 96–101. doi:10.3758/BF03202365.
[15] S. Paplu, R. F. Navarro, K. Berns, Harnessing long-term memory for personalized human-robot interactions, in: 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids), 2022, pp. 377–382. doi:10.1109/Humanoids53995.2022.10000213.
[16] T. Sievers, N. Russwinkel, Retrieving memory content from a cognitive architecture by impressions from language models for use in a social robot, Applied Sciences 15 (2025). doi:10.3390/app15105778.
[17] S. Wu, A. Oltramari, J. Francis, C. L. Giles, F. E. Ritter, Cognitive llms: Towards integrating cognitive architectures and large language models for manufacturing decision-making, 2024. URL: https://arxiv.org/abs/2408.09176.
[18] T. R. Sumers, S. Yao, K. Narasimhan, T. L. Griffiths, Cognitive architectures for language agents, 2024. URL: https://arxiv.org/abs/2309.02427. arXiv:2309.02427.
[19] C. Hao, N. Russwinkel, D. F. Haeufle, P. Beckerle, Conceptualizing the integration of cognitive and physical models to enable continuous human-robot interaction (2025 forthcoming).
[20] D. Kahneman, Thinking, fast and slow, Farrar, Straus and Giroux, New York, 2011.
[21] navel robotics GmbH, navel, Technical Report, 2025. URL: https://navelrobotics.com.
[22] U. Kurup, C. Lebiere, What can cognitive architectures do for robotics?, Biologically Inspired Cognitive Architectures 2 (2012) 88–99. doi:10.1016/j.bica.2012.07.004.
[23] J. Whitehill, Understanding act-r - an outsider's perspective, 2013. URL: https://arxiv.org/abs/1306.0125.
[24] C. Quam, A. Wang, W. T. Maddox, K. Golisch, A. Lotto, Procedural-memory, working-memory, and declarative-memory skills are each associated with dimensional integration in sound-category learning, Frontiers in Psychology 9 (2018). doi:10.3389/fpsyg.2018.01828.
[25] H. Eichenbaum, Hippocampus: Cognitive processes and neural representations that underlie declarative memory, Neuron 44 (2004) 109–120. doi:10.1016/j.neuron.2004.08.028.
[26] H. Hayne, J. Boniface, R. Barr, The development of declarative memory in human infants: Age-related changes in deferred imitation, Behavioral neuroscience 114 (2000) 77–83. doi:10.1037/0735-7044.114.1.77.
[27] jakdot, pyactr, Technical Report, 2025. URL: https://github.com/jakdot/pyactr.
[28] A. Newen, Understanding others: The person model theory, in: T. K. Metzinger, J. M. Windt (Eds.), Open MIND, MIND Group, Frankfurt am Main, 2015. doi:10.15502/9783958570320.
[29] A. Newen, The person model theory and the question of situatedness of social understanding, in: The Oxford Handbook of 4E Cognition, Oxford University Press, 2018. doi:10.1093/oxfordhb/9780198735410.013.25.
[30] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, 2024. URL: https://arxiv.org/abs/2312.10997. arXiv:2312.10997.
[31] D. Sanmartin, Kg-rag: Bridging the gap between knowledge and creativity, 2024. URL: https://arxiv.org/abs/2405.12035. arXiv:2405.12035.
[32] OpenAI, The most powerful platform for building AI products, Technical Report, 2025. URL: https://openai.com/api/.
[33] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, D. Amodei, Language models are few-shot learners, 2020. doi:10.48550/arXiv.2005.14165.
[34] T. Sievers, N. Russwinkel, A path to gradual individual experience and recollection for social robots based on a cognitive architecture, in: 17th International Conference on Social Robotics + AI, 2025 forthcoming.