Comparing Rule-based and LLM-based Methods to Enable Active Robot Assistant Conversations

Jan Leusmann
LMU Munich
Munich, Germany
jan.leusmann@ifi.lmu.de

Chao Wang
Honda Research Institute EU
Offenbach am Main, Germany

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

ABSTRACT

In the future, robots will be a ubiquitous part of our daily interactions. In contrast to home assistants, robots will assist us by responding to questions and requests and interacting with the environment and humans directly. For this and with the help of recent advancements in AI, we propose shifting from passive to active robots, which can communicate with humans to understand how and when they should perform supportive tasks, which will be especially supportive in collaborative settings. In this paper, we envision two different approaches to how this can be implemented. (1) Rule-based approaches where the dialogue is implemented as a state machine and the conversation procedure is static and predictable. (2) LLM-based approaches, which can dynamically adapt to any situation and ambiguous human input. We compare these two ideas of how robots can first detect the state in which a human is and how they can engage in conversations, and then discuss the advantages and disadvantages of these approaches regarding trust, usability, performance, user agency, and perception of the robot.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

human computer interaction, robot, human-robot interaction, voice user interface

1 INTRODUCTION

Robots have undergone major technological advancements. With this, robots are developing from autonomous agents toward collaborative companions. As such, robots can now perform various tasks to support humans daily. However, a key component in collaborative settings is fluent communication. In human-human scenarios, we often need to ask questions about how a task should be performed or whether it is a good time to help. While this is often intuitive in human-human scenarios depending on how close the relationship between the users is, in human-robot interaction (HRI), it is currently unclear how to best enable this kind of communication. Current technical devices often react predictably to human input as we have learned their interaction modalities. Today, most voice assistants are passive agents and only react to user requests through a wake word. Here, most communication models are rule-based, making the interaction flow static and repeatable. This requires users to learn to communicate with these systems to get their desired results correctly. On the one hand, this makes learning how to interact with these systems simple. On the other hand, this might lead to frustration when the system can not dynamically react to user requests. Large-language models (LLMs) have recently drastically changed the possibilities of interacting with technical systems. Due to their potential to "understand" every human request to a certain extent, they enable more fluid interaction. Thus, LLMs offer to enhance the HRI and overcome the narrow rule-based responses of traditional voice user interfaces in robots.

When envisioning future agents to support users in their daily lives, we see a shift from passive agents to active agents interacting directly with the environment and users. Active engagement requires more communication so these agents do not disturb users but help users. Furthermore, users would also prefer proactive voice agents [32]. In contrast to traditional voice agents, robots also offer the possibility of expressions or non-verbal communication in dialogues. User agency plays an important role here. We do not envision agents just to be servants, which do all tasks for humans, but a collaborative workflow, leveraging users' desire to still perform certain tasks but getting support when needed. This requires the agents to ask questions and start conversations in certain situations actively [16]. However, we do not want agents to guide us in every step, as this would take away user agency. Thus, it is important to perform this active engagement gracefully, e.g., through the use of humor [36]. The current advancements in the use of LLMs enable robots to grasp the state of the environment in an accurate and detailed way [1, 4, 29]. Wang et al. [35] created a system that can intake information from various sources and output different sequences of atomic actions to fulfill user requests. This enables agents to react to any situation and not only situations that can be classified based on training data.

Trust plays a vital role in HRI. Factors like reliability, competence, transparency, and predictability all foster trust in HRI [3, 13, 24], and transparent communication of capabilities and limitations play a significant role in building trust. Understanding and enhancing trust in HRI is essential for improving communication effectiveness and user acceptance in various contexts and increases the use of robots [23]. User agency is another crucial aspect of HRI. Users should be given a sense of control and autonomy in their interactions with robots [10]. Robots that facilitate user agency by providing options for input, decision-making, and task customization can enhance user satisfaction and collaboration.

In this work, we compare traditional rule-based approaches against LLM-based approaches. We discuss their advantages and disadvantages and how they affect user agency and trust. Then, we propose a study design on how to investigate the effect of both

CUI@CHI 2024: Building Trust in CUIs – From Design to Deployment , May 11, 2024, Honolulu, HI, USA

approaches in terms of trust in the agent, usability, fluidity, performance, and perception.

2 COMPARISSON

These two approaches, described in the following, reflect the two outer points of a continuum, as LLM-based approaches can be fine-tuned, which leads to a hybrid approach.

2.1 Rule-Based Approach

The main part of a voice user interface circle is dialog management. Here, the detected intents in a user query get processed, and the agent generates a fitting response based on the response and state of the dialogue [7, 16, 27]. Often, interaction with voice assistants is task-oriented [6], and we envision goal-orientated interaction.

Advantages: *Transparency:* Rule-based systems are usually transparent, as their functionality is based on predefined rules. Transparency issues only occur for voice assistants of large companies, which might be interested in personal user data [11, 19]. *Consistency and Control:* As the responses are created based on rules, they are predictable and internally consistent, and both users and developers have a lot of control over them.

Disadvantages: *Limited flexibility:* Rule-based systems are inherently limited by exactly those rules. When a user query is outside of scope, these systems can not positively react to those queries. *Lack of adaptability:* Rule-based systems are usually implemented with one certain personality and can usually not adapt. *Hard to scale:* Creating a rule-based system that scales up to react to every single user intent is not feasible, as the option space for those rules is way too high. Maintaining a working rule set in a changing environment adds to this problem.

2.2 Dynamic LLM-based Approach

LLM-based voice assistants represent a significant advancement over their rule-based counterparts, particularly in the domain of dialog management. Like rule-based systems, LLM-based assistants analyze user queries to discern intents. However, they employ language understanding through large general pre-trained word embeddings to generate contextually appropriate responses. Furthermore, as the trend shifts towards goal-oriented interaction between robots and humans, LLM-based assistants are pivotal in facilitating seamless outcome-driven communication.

Advantages: *Natural language understanding:* Due to the largely trained models, LLMs perform great at being able to respond to any user query, even if outside of the original scope, giving LLM-powered voice assistants richer interaction patterns [18]. *Adaptability:* LLMs are still constantly improved and trained and can also be trained locally to a specific scenario, allowing them to adapt to user preferences [26]. *Broad coverage:* LLMs enable agents to respond to user queries that are completely out of the original context of the user agent.

Disadvantages: *Unpredictable:* LLMs are often a black box, making it hard to understand how certain decisions are made. *Unexpected responses:* Due to their nature of just predicting the most probable next word in each sentence, the output of LLMs can sometimes be unexpected [26]. *Data dependency:* This is more of an issue for LLMs, which should work in specific contexts. LLMs need a large amount of data to generate good embeddings. *Bias:* If there were bias in the training data, there would be bias in the output [8].

2.3 Context Considerations

The effectiveness of rule-based and LLM-based methods depends on task type, user group, and environment. Rule-based approaches excel in structured tasks like commands, while LLM-based methods thrive in dynamic conversational interactions [17]. While novice user groups might prefer the predictability of rule-based systems, the dynamic nature of LLM-based systems might also be advantageous for not experienced users, as LLMs are great at extrapolating the context from only partly given information. While more technical affine user groups might prefer the dynamic behavior of LLM-based systems at first, oftentimes, expert users want shortcuts and interactions to be efficient [9]. LLM-based systems are great in complex environments, as they can leverage context to generate a fitting response [28, 35].

2.4 Future Work

We introduced the need for active assistant robot conversational agents and compared rule-based and LLM-based approaches by reviewing their advantages and disadvantages. While LLM-based methods are way more capable, especially in terms of responding to unexpected user queries, it is currently unclear how this affects trust and user agency. To solve this, we propose to conduct a user study with a robotic system comparing a rule-based system and an LLM-based system for the conversational agent. We propose to measure trust [25], user preference, user agency [20], usability [5], interaction fluency [22], and general perception of the robot [2].

ABOUT THE AUTHORS

Jan Leusmann. (https://leusmann.io/) is a 2nd year PhD student at the LMU Munich (Germany). His research is situated in the domain of Human-Robot Interaction and focuses on fostering understandable and intuitive communication between robots and humans, e.g., [15]. Recently, he investigated how we can enable, improve, and understand communication between robots and humans, e.g., [14, 16].

Chao Wang (https://wallacewangchao.github.io/) is a Senior Research Scientist at Honda Research Institute Europe. His work focuses on human-AI interaction applied to autonomous driving vehicles and robots, e.g., leveraging LLM to enable multi-modal interactions between humans and physical robots [12, 35], human-vehicle cooperation on prediction-level [34], and utilizing Augmented Reality to improve explainability for human-robot training and cooperation [33].

Sven Mayer (https://sven-mayer.com/) is an assistant professor of computer science at LMU Munich (Germany). His research sits at the intersection between HCI and AI. He uses AI to design, build, and evaluate future human-centered interfaces. In particular, he envisions enabling humans to outperform their performance in collaboration with the machine. A major breach of his research focuses on human-robot interaction, e.g., [14–16, 21, 30, 31].

REFERENCES

[1] Michael Ahn, Anthony Brohan, Noah Brown, and et al. 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. arXiv:2204.01691

[2] Christoph Bartneck, Elizabeth Croft, and Dana Kulic. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[3] Deborah R. Billings, Kristin E. Schaefer, Jessie Y.C. Chen, and Peter A. Hancock. 2012. Human-Robot Interaction: Developing Trust in Robots. In *Proc. of HRI.* ACM, 109–110. https://doi.org/10.1145/2157689.2157709

[4] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, and et al. 2023. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. https://doi.org/10.48550/arXiv.2307.15818 arXiv:2307.15818 [cs]

[5] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (1995).

[6] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proc. of CHI.* 1–12. https://doi.org/10.1145/3290605.3300705

[7] D. A. Dhinagaran, L. Martinengo, M. R. Ho, S. Joty, T. Kowatsch, R. Atun, and L. T. Car. 2022. Designing, Developing, Evaluating, and Implementing a Smartphone-Delivered, Rule-Based Conversational Agent (DISCOVER): Development of a Conceptual Framework. *mHealth and uHealth* (2022).

[8] Satyam Dwivedi, Sanjukta Ghosh, and Shivam Dwivedi. 2023. Breaking the Bias: Gender Fairness in LLMs Using Prompt Engineering and In-Context Learning. *Rupkatha Journal on Interdisciplinary Studies in Humanities* 15, 4 (2023). https://doi.org/10.21659/rupkatha.v15n4.10

[9] Gabriel Haas, Michael Rietzler, Matt Jones, and Enrico Rukzio. 2022. Keep it Short: A Comparison of Voice Assistants' Response Behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22).* Association for Computing Machinery, New York, NY, USA, Article 321, 12 pages. https://doi.org/10.1145/3491102.3517684

[10] Guy Hoffman and Wendy Ju. 2014. Designing Robots with Movement in Mind. *Journal of Human-Robot Interaction* 3, 1 (2014), 91–122.

[11] Trung Dong Huynh, William Seymour, Luc Moreau, and Jose Such. 2023. Why Are Conversational Assistants Still Black Boxes? The Case For Transparency. In *Proc. of CUI.* ACM, 1–5. https://doi.org/10.1145/3571884.3604319

[12] Frank Joublin, Antonello Ceravola, Pavel Smirnov, Felix Ocker, Joerg Deigmoeller, Anna Belardinelli, Chao Wang, Stephan Hasler, Daniel Tanneberg, and Michael Gienger. 2023. CoPAL: Corrective Planning of Robot Actions with Large Language Models. *arXiv preprint arXiv:2310.07263* (2023).

[13] Kornelia Lazanyi and Beata Hajdu. 2017. Trust in Human-Robot Interactions. In *2017 IEEE 14th International Scientific Conference on Informatics.* IEEE, 216–220. https://doi.org/10.1109/INFORMATICS.2017.8327249

[14] Jan Leusmann, Carl Oechsner, Johanna Prinz, Robin Welsch, and Sven Mayer. 2023. A Database for Kitchen Objects: Investigating Danger Perception in the Context of Human-Robot Interaction. In *Extended Abstracts of CHI.* ACM, Article 6, 9 pages. https://doi.org/10.1145/3544549.3585884

[15] Jan Leusmann, Chao Wang, Michael Gienger, Albrecht Schmidt, and Sven Mayer. 2023. Understanding the Uncertainty Loop of Human-Robot Interaction. https://doi.org/10.48550/arXiv.2303.07889 arXiv:2303.07889 [cs.HC]

[16] Jan Leusmann, Jannik Wiese, Moritz Ziarko, and Sven Mayer. 2023. Investigating Opportunities for Active Smart Assistants to Initiate Interactions With Users. In *Proc. of MUM.* ACM, 495–498. https://doi.org/10.1145/3626705.3631787

[17] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang. 2023. AgentBench: Evaluating LLMs as Agents. arXiv:2308.03688 [cs.AI]

[18] Amama Mahmood, Junxiang Wang, Bingsheng Yao, Dakuo Wang, and Chien-Ming Huang. 2023. *LLM-Powered Conversational Voice Assistants: Interaction Patterns, Opportunities, Challenges, and Design Guidelines.* arXiv:2309.13879 [cs] http://arxiv.org/abs/2309.13879

[19] Marco Manca, Parvaneh Parvin, Fabio Paternò, and Carmen Santoro. [n. d.]. Integrating Alexa in a Rule-based Personalization Platform. In *Proc. of the 6th EAI International Conference on Smart Objects and Technologies for Social Good.* ACM, 108–113. https://doi.org/10.1145/3411170.3411228

[20] Emily Mower, David J. Feil-Seifer, Maja J. Mataric, and Shrikanth Narayanan. [n. d.]. Investigating Implicit Cues for User State Estimation in Human-Robot Interaction Using Physiological Measurements. In *RO-MAN.* IEEE, 1125–1130. https://doi.org/10.1109/ROMAN.2007.4415249

[21] Carl Oechsner, Sven Mayer, and Andreas Butz. 2022. Challenges and Opportunities of Cooperative Robots as Cooking Appliances. In *Proc. of the 2022 Workshop on Engaging with Automation (AutomationXP22).* https://sven-mayer.com/wp-content/uploads/2022/05/oechsner2022challenges.pdf

[22] Matthew K.X.J. Pan, Elizabeth A. Croft, and Günter Niemeyer. 2018. Evaluating Social Perception of Human-to-Robot Handovers Using the Robot Social Attributes Scale (RoSAS). In *Proc. of HRI.* ACM, 443–451. https://doi.org/10.1145/3171221.3171257

[23] Tracy Sanders, Alexandra Kaplan, Ryan Koch, Michael Schwartz, and P. A. Hancock. 2019. The Relationship Between Trust and Use Choice in Human-Robot Interaction. *Human Factors* 61, 4 (2019), 614–626. https://doi.org/10.1177/0018720818816838

[24] Tracy L. Sanders, Tarita Wixon, K. Elizabeth Schafer, Jessie Y. C. Chen, and P. A. Hancock. 2014. The Influence of Modality and Transparency on Trust in Human-Robot Interaction. In *CogSIMA.* IEEE, 156–159. https://doi.org/10.1109/CogSIMA.2014.6816556

[25] Kristin E. Schaefer. [n. d.]. Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI". In *Robust Intelligence and Trust in Autonomous Systems.* Springer US, 191–218. https://doi.org/10.1007/978-1-4899-7668-0_10

[26] Kotaro Shukuri, Ryoma Ishigaki, Jundai Suzuki, Tsubasa Naganuma, Takuma Fujimoto, Daisuke Kawakubo, Masaki Shuzo, and Eisaku Maeda. 2023. Meta-Control of Dialogue Systems Using Large Language Models. (2023). https://doi.org/10.48550/ARXIV.2312.13715

[27] Jagdish Singh, Minnu Helen Joesph, and Khurshid Begum Abdul Jabbar. 2019. Rule-Based Chabot for Student Enquiries. *Journal of Physics: Conference Series* 1228, 1 (2019), 012060. https://doi.org/10.1088/1742-6596/1228/1/012060

[28] Daniel Tanneberg, Felix Ocker, Stephan Hasler, Joerg Deigmoeller, Anna Belardinelli, Chao Wang, Heiko Wersing, Bernhard Sendhoff, and Michael Gienger. 2024. To Help or Not to Help: LLM-based Attentive Support for Human-Robot Group Interactions. arXiv:2403.12533 [cs.RO]

[29] Sai Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. 2023. ChatGPT for Robotics: Design Principles and Model Abilities. https://doi.org/10.48550/arXiv.2306.17582 arXiv:2306.17582 [cs]

[30] Steeven Villa and Sven Mayer. 2022. Cobity: A Plug-And-Play Toolbox to Deliver Haptics in Virtual Reality. In *Proc. of Mensch Und Computer.* ACM, 78–84. https://doi.org/10.1145/3543758.3543775

[31] Steeven Villa, Sven Mayer, Jess Hartcher-O'Brien, Albrecht Schmidt, and Tonja-Katrin Machulla. 2022. Extended Mid-air Ultrasound Haptics for Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 578 (2022), 25 pages. https://doi.org/10.1145/3567731

[32] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In *Proc. of CHI.* ACM, Article 254, 15 pages. https://doi.org/10.1145/3411764.3445536

[33] Chao Wang, Anna Belardinelli, Stephan Hasler, Theodoros Stouraitis, Daniel Tanneberg, and Michael Gienger. 2023. Explainable Human-Robot Training and Cooperation with Augmented Reality. In *Extended Abstracts of CHI.* ACM, Article 449, 5 pages. https://doi.org/10.1145/3544549.3583889

[34] Chao Wang, Derck Chu, Marieke Martens, Matti KrüGer, and T. H. Weisswange. 2022. Hybrid Eyes: Design and Evaluation of the Prediction-Level Cooperative Driving with a Real-World Automated Driving System. In *Proc. of AutomotiveUI.* ACM, 274–284. https://doi.org/10.1145/3543174.3546832

[35] Chao Wang, Stephan Hasler, Daniel Tanneberg, Felix Ocker, Frank Joublin, Antonello Ceravola, Joerg Deigmoeller, and Michael Gienger. 2024. Large Language Models for Multi-Modal Human-Robot Interaction. arXiv:2401.15174 [cs.RO]

[36] Nima Zargham, Leon Reicherts, Vino Avanesi, Yvonne Rogers, and Rainer Malaka. 2023. Tickling Proactivity: Exploring the Use of Humor in Proactive Voice Assistants. In *Proc. of MUM.* ACM, 294–320. https://doi.org/10.1145/3626705.3627777