# How can I help you? An Intelligent Virtual Assistant for Industrial Robots

LI, Chen; Park, Jinha; Kim, Hahyeon; Chrysostomou, Dimitrios

*Published in:*
HRI 2021 - Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction

*DOI (link to publication from Publisher):*
10.1145/3434074.3447163



*Publication date:*
2021

*Document Version*
Accepted author manuscript, peer reviewed version

# How can I help you? An Intelligent Virtual Assistant for Industrial Robots

Chen Li
Aalborg University
Dept. of Materials and Production
Aalborg East, Denmark

Jinha Park
Aalborg University
Dept. of Materials and Production
Aalborg East, Denmark

Hahyeon Kim
Aalborg University
Dept. of Materials and Production
Aalborg East, Denmark

Dimitrios Chrysostomou
Aalborg University
Dept. of Materials and Production
Aalborg East, Denmark

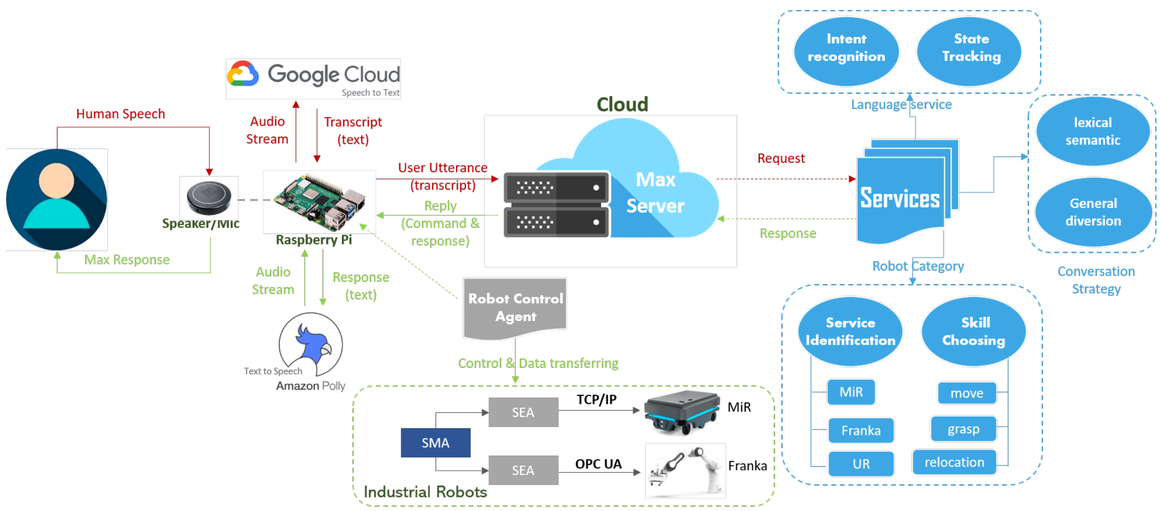


**Figure 1: Architecture overview of the proposed intelligent virtual assistant, Max.**

## ABSTRACT

In the light of recent trends toward introducing Artificial Intelligence (AI) to enhance Human-Robot Interaction (HRI), intelligent virtual assistants (VA) driven by Natural Language Processing (NLP) receives ample attention in the manufacturing domain. However, most VAs either tightly bind with a specific robotic system or lack efficient human-robot communication. In this work, we implement a layer of interaction between the robotic system and the human operator. This interaction is achieved using a novel VA, called Max, as an intelligent and robust interface. We expand the research work in three directions. Firstly, we introduce a RESTful style Client-Server architecture for Max. Secondly, inspired by studies of human-human conversations, we embed conversation strategies into human-robot dialog policy generation to create a more natural and humanized conversation environment. Finally, we evaluate Max over multiple real-world scenarios from the exploration of an unknown environment to package delivery, with the means of an industrial robot.


*HRI '21 Companion, March 8–11, 2021, Boulder, CO, USA*

ACM ISBN 978-1-4503-8290-8/21/03…$15.00
https://doi.org/10.1145/3434074.3447163

## CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; • **Computing methodologies → Discourse, dialogue and pragmatics**; • **Computer systems organization → Client-server architectures**.

## KEYWORDS

Human-robot interaction; Natural Language Processing; Virtual Assistant; Client-Server architecture; User Experience

**ACM Reference Format:**
Chen Li, Jinha Park, Hahyeon Kim, and Dimitrios Chrysostomou. 2021. How can I help you? An Intelligent Virtual Assistant for Industrial Robots. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3434074.3447163

# 1 INTRODUCTION

In order to leverage Artificial Intelligence (AI) to enhance Human-Robot Interaction (HRI), manufacturers need to identify the critical issues of the interaction between operators and industrial robots [8, 14]. The recent technological rise of AI technologies facilitates industry and research stakeholders to implement more efficient natural language-based methods for HRI [5, 9, 12, 13, 21, 22]. Several voice-enabled virtual assistants (VA), e.g., Alexa [1] and Siri [2], are widely available in the context of entertainment or personal service. They particularly excel in having robust natural language processing (NLP) capacities and being able to handle continuous natural dialogues. However, outside of the entertainment domain, the manufacturing environment mainly focuses on limited, fixed, and atomic actions, e.g., pick up material/tools.

In particular, to support a flexible and scalable manufacturing working environment, VAs need to have an intuitive and extendable architecture able to adapt into various situations, learning capabilities to understand human intents and the ability to control robots without human intervention.

Furthermore, the current evaluation standards of such VA are more concerned with task-completion experiences. The available research on improving the user experience during industrial HRI is limited [11, 15, 17]. As the most flexible entity in the manufacturing systems, the human operator plays an essential role in overall productivity. Therefore, it is crucial to design a user-friendly and human-aware interface to enhance the interaction between industrial robots and human operators [6, 7, 19].

In this work, we present the development of an innovative VA, named Max, for industrial settings based on a scalable and easily maintained Client-Server (CS) style architecture. The communication between Max's server and client is implemented via highly flexible RESTful API calls. To enhance the user experience, we introduce human-human conversation strategies based on neural conversational models [23, 24], and specialized dialogue generation policies [10, 20, 25]. This way, we can guide Max's response generation while leveraging the state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) [4] model for interpreting human utterance.

# 2 METHODOLOGY

## 2.1 Architecture Overview

The proposed VA, Max, consists of three parts, i.e., voice service, spoken language understanding component, and a robot control agent. The Max Client (i.e., voice service and robot control agent) is implemented on a Raspberry Pi 4. The Max Server is deployed on University Cloud hosting the spoken language understanding component. Figure 1 illustrates a high-level system architecture of Max.

CS style architecture is explicitly chosen to improve scalability and reduce maintenance cost. The communication between client and server is achieved through RESTful style API calls. The main motivation behind this is to enable a loosely coupled language interface to be able to work with various industrial robots. The main responsibilities of the client are to 1) continuously listen to the operator's speech, 2) translate the speech into a transcript, 3) send HTTP requests which wrap the transcript as a parameter to the server, 4) invoke robot control agent to control the robot according to the response from the server, and 5) provide the vocal response to the human operator. The Max server supports three services: 1) human intent recognition with dialogue state tracking, 2) robot service maintenance and 3) conversation strategy-embedded response generation.

Since the robot service identification and robot skills repository are maintained in server-side, the client does not need to bind with the specific robot. Max's client sends an update request to the server when the operator tries to access the robot, which is not recognized by the client. Thus, the robot control algorithm will be updated in the back-end. Therefore, all the maintenance are handled in the central server, and each client is independent of the connected robot. Furthermore, security and access rights are defined at the time of set-up of the server so different operators will be assigned different roles when they access the server through the client.

## 2.2 Voice Service

The Max voice service mainly includes a speech-to-text service and a text-to-speech service. To comprehend the operators' intent, Max leverages the Google speech-to-text service, Automatic Speech Recognition (ASR) API, to recognize the operators' voice signal and transcribe it. These transcripts (i.e., human utterance) are then sent to the spoken language understanding service on Max's server-side for further processing, e.g., human intent identification.

Max's response is composed of two parts, text reply for human operator and commands for controlling the robot (see section 2.4). To provide a natural and humanized response, Max supports two text-to-speech solutions to convert the text reply into an audio sequence and replay it through the speaker; an offline solution based on the Python package, Pyttsx, and an online one based on Amazon Polly service.

## 2.3 Spoken Language Understanding Service

*2.3.1 Human Intent Identifier.* Different from the open-domain conversations, the dialogue between operator and robot is mainly related to the specific manufacturing tasks. Therefore, Max is designed as a task-oriented dialogue system.

In our work, we fine-tune the base BERT model, which has larger feed-forward networks. We train it on our human-labeled training dataset. The dataset mainly provides dialogues of the manufacturing tasks of using a Mobile Industrial Robot (MiR200) (e.g., *please delivery this box to the warehouse*). BERT encodes the user utterance (including intents, slots annotated with inside–outside–beginning (IOB) tags and slots values) predicts the requested intent (i.e., intent requested by the operator for a given robot service) and requested slots (i.e., requested by the operator in the current utterance).

*2.3.2 Conversation Strategies.* Comparing with the open-domain dialogue systems, task-oriented dialogue systems are easier to maintain due to specific task domains and pre-built knowledge while they suffer from lower flexibility and user experience.

In our work, we study the generic conversational strategies which have been proposed in open-domain conversations [3, 16, 18]. Two conversational strategies, lexical-semantic strategy and general diversion strategy [25], are selected to increase the task completion rate and enhance the user experience with a high dynamic and humanized conversation environment.

**Lexical semantic strategy**. Different from [25], we apply the *don't repeat yourself* strategy to our VA instead of the human user. Max can respond differently but remains in the same context when the operator asks the same things. For example, Max may say: "My battery is fine at this moment." or "I am fully charged and ready to work." when the operator queries the battery level of the robot. **General diversion strategy**. We introduce two general diversion strategies: *i) initial activities* and *ii) switch a topic*, to provide options to the operators and attract their attention when the current task is impossible to continue. Initial activities mean that Max should be able to initiate a request to start the manufacturing tasks at the appropriate time. For example, Max may say: "There are two scheduled tasks today. Would you like me to do them now?". The robot should be able to switch to a task-related topic, if the current task is impossible to continue, by responding "Sorry, the location is not registered in the system. Do you want to mark it on the map now?".

*2.3.3 Robot Skill Identifier.* Max is designed to be robot-agnostic and, therefore, can support various kinds of industrial robots such as mobile and manipulators. To enable such extended support of robot services, we define a unified JSON format schema to maintain the robot control service on the server's side. Thus, we allow easy extension and integration of new robot services and APIs.

## 2.4 Robot Controller Agent

The robot controller agent, as a core part of Max's client, assists with the control of the robot according to the operator's instructions. There are two types of robot controller agents implemented for Max, i.e., service maintenance agent (SMA) and service execution agent (SEA). SMA chooses the right SEA for manufacturing tasks based on the operator's voice commands. The maintenance of SEAs is also performed through SMA, e.g., updating SEA in the back-end if there is a new version available on Max's server-side. SEA is the low-level robot control algorithm which communicates directly with the robot. In general, the communication between Max's client and robots may vary depending on the supported protocols from the robot, e.g., TCP/IP, OPC-UA. (see Fig. 1).

# 3 EXPERIMENTAL RESULTS

The experiments conducted for this work are based on MiR 200, a safe, cost-effective industrial mobile robot that quickly automates shop floor internal transportation and logistics. We consider the following three scenarios to evaluate Max's performance: *i) collaborative environment exploration*, *ii) package delivery* and *iii) conversation strategy-embedded response generation* (see Table 1).

## 3.1 Collaborative Environment Exploration

As an initial test for evaluating Max's performance, we chose the collaborative exploration of a shop floor environment. Building a 2D digital shop floor map of a factory hall is essential for the planning of autonomous internal transportation tasks and the calculation of the robot's operational capacity. Two simple tasks are identified in this scenario (see Tasks #1 and #2 in Table 1). The tested intentions here are *Check_location* and *Update_location*.

**Table 1: Tasks for intent verification**

| Task id | Task description | Intent |
|---|---|---|
| 1 | Remove the position from the digital map | *Update_location* |
| 2 | Check the current location on the digital map | *Check_location* |
| 3 | Deliver the package to the operator in the predefined destination | *Deliver_package* |
| 4 | Initial activities | *Greeting* |
| 5 | Switch a topic | *Ask_help* |
| 6 | Don't repeat yourself | *Check_mission* |

In this scenario, Max's server returns the predicted requested intent and slot values (e.g., *Shelf A*) from operator's utterance to Max's client (see figure 2). The SMA calls MiR's SEA, which controls the MiR 200 through REST API calls, according to the requested service. The SEA sends the *HTTP Delete* and *Get* requests to remove *Shelf A* and obtain the *storage room's* position from digital map respectively.

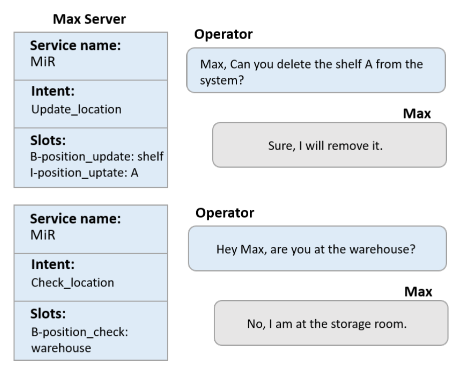


**Figure 2: The predicted dialogue service, requested intents and requested slot values (shown with the blue rectangle) for the operator utterance.**

## 3.2 Package Delivery

Package delivery is the second scenario for the evaluation of Max's performance. As seen in Table 1 and Task #3, Max has to handle the request to deliver a package to a human operator in a target location according to oral instruction.

The intent tested here is *Deliver_package*. Similarly to the previous scenario, Max's server returns the predicted intent and requested slot values (e.g, warehouse, box, small) from operator's utterance to Max's client, as Fig. 3 shows. The extracted slot values are then set as parameters for a HTTP Post request (i.e., package delivery request) which will be sent to MiR 200's internal web server by SEA.

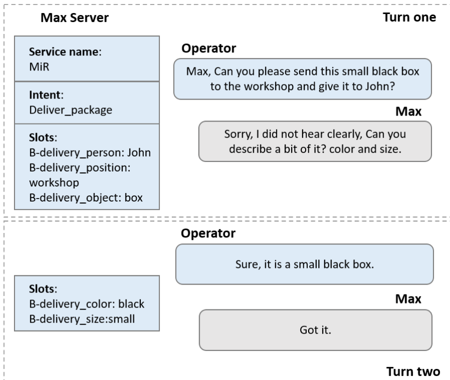


**Figure 3: Max uses two turns to obtain all the requested slot values from operator's utterance.**

## 3.3 Testing Conversation Strategy

The last scenario for the evaluation of Max's performance is focused on exploring how embedded conversation strategies can improve the task-completion rate and bootstrap the user experience. Tasks #4, #5 and #6 are implemented for this scenario as listed in Table 1. The intents tested in this scenario are *i) Greeting*, *ii) Ask_help* and *iii) Check_mission*. Figure 4 illustrates the embedded conversation strategies between Max and an operator. In this case, Max's client remains in standby mode until it received a confirmation from the operator.

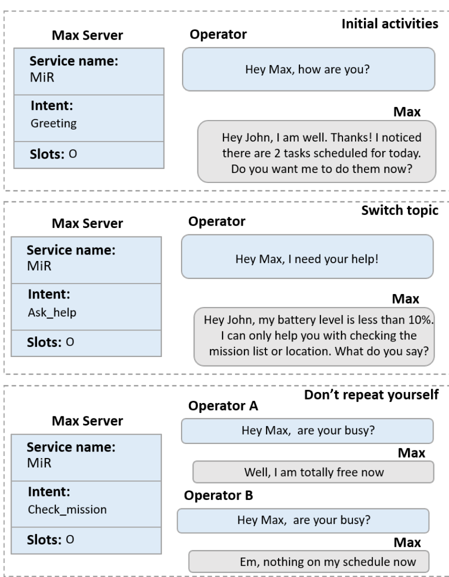


**Figure 4: The embedded conversation strategies.**

# 4 DISCUSSION & CONCLUSION

Taking advantage of the CS style architecture and RESTful style API design, Max provides a more flexible and scalable range of services for HRI in industrial settings. One shortcoming of the CS style paradigm is the traffic congestion problem, i.e., the response time of the server may become longer when a high number of simultaneous requests is sent from the clients. Therefore, the deployment (e.g., number of servers, load balance strategy) of the server-side needs to be carefully designed and tested. In our case, Max's server is deployed on a local server and a Cloud server cluster so as the client requests are forwarded to the cloud in case the number of local requests reaches the allowed upper limit.

Such dynamic load balance adjustment is achieved by using Nginx[1] (i.e., HTTP reverse proxy) and Gunicorn[2] (WSGI HTTP Server). The results from the stress testing received by Siege[3] indicate that the actual maximum concurrent number is 289.79/second for 100,000 transactions (with transaction rate 14204.55 trans/sec) in 7.04 seconds. A future improvement will be to ensure security by verifying the operator's authority, to avoid the unintentional or malicious REST API calls for robot control.

The performed experiments took place in our workshop where the background noise is high, resembling an industrial environment. We observed that the intent error rate, i.e., misunderstanding operator's intent, and slot error rate, i.e., incorrect prediction of slot value reach 20% to 30% respectively in the workshop while they remain less than 10% and 5% respectively in a quiet, office environment.

The accuracy of the prediction of requested slot values depends on the ambient noise, operators' voice volume and the physical distance between the operator and Max's client, as well as the length of the sentence. In our package delivery experiment, Max took two turns to predict the entire requested intent and slot values. Further work will investigate the noise suppression methods to filter out the steady-state noise of the environment, such as the sound of ventilation.

Task-completion experiences are usually considered as the primary criteria for industrial HRI evaluation. However, it is also important to mention that the overall effectiveness of the HRI profoundly relies on human productivity and user experience. Based on our experiments, we observed that the two proposed human-human conversation strategies, attract the attention of the operators and create a pleasant interaction.


# ACKNOWLEDGMENTS

The authors would like to acknowledge support by the H2020-WIDESPREAD project no. 857061 "Networking for Research and Development of Human Interactive and Sensitive Robotics Taking Advantage of Additive Manufacturing – R2P2". This research was also supported by MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program related to Robotics(P0008749) supervised by the Korea Institute for Advancement of Technology (KIAT). Authors would also like to thank the research assistants Martin Bieber and Galadrielle Humblot-Renaux for their significant help with setting up the robot demos.


[1] https://www.nginx.com/
[2] https://gunicorn.org/
[3] https://www.joedog.org/siege-home/

## REFERENCES


[1] Amazon. 2020. *Developer Documentation.* Retrieved Oct 2, 2020 from https://developer.amazon.com/documentation

[2] Apple. 2020. *SiriKit.* Retrieved Oct 5, 2020 from https://developer.apple.com/documentation/sirikit/

[3] Rafael E. Banchs and Haizhou Li. 2012. IRIS: a Chat-oriented Dialogue System based on the Vector Space Model. In *Proceedings of the ACL 2012 System Demonstrations.* Association for Computational Linguistics, Jeju Island, Korea, 37–42. https://www.aclweb.org/anthology/P12-3007

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[5] L. El Hafi, S. Isobe, Y. Tabuchi, Y. Katsumata, H. Nakamura, T. Fukui, T. Matsuo, G. A. Garcia Ricardez, M. Yamamoto, A. Taniguchi, Y. Hagiwara, and T. Taniguchi. 2020. System for augmented human–robot interaction through mixed reality and robot training by non-experts in customer service environments. *Advanced Robotics* 34, 3-4 (2020), 157–172. https://doi.org/10.1080/01691864.2019.1694068

[6] Abdelfetah Hentout, Mustapha Aouache, Abderraouf Maoudj, and Isma Akli. 2019. Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017. *Advanced Robotics* 33, 15-16 (2019), 764–799. https://doi.org/10.1080/01691864.2019.1636714

[7] Jhih-Yuan Huang, Wei-Po Lee, Chen-Chia Chen, and Bu-Wei Dong. 2020. Developing Emotion-Aware Human–Robot Dialogues for Domain-Specific and Goal-Oriented Tasks. *Robotics* 9, 2 (2020), 31. https://doi.org/10.3390/robotics9020031

[8] Andreas Huber and Astrid Weiss. 2017. Developing Human-Robot Interaction for an Industry 4.0 Robot: How Industry Workers Helped to Improve Remote-HRI to Physical-HRI. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 137?138. https://doi.org/10.1145/3029798.3038346

[9] Hiroshi Ishiguro, Tatsuya Kawahara, and Yutaka Nakamura. 2020. Autonomous Dialogue Technologies in Symbiotic Human-Robot Interaction. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 650?651. https://doi.org/10.1145/3371382.3374855

[10] Ziming Li, Julia Kiseleva, and Maarten de Rijke. 2019. Dialogue generation: From imitation learning to inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6722–6729. https://doi.org/10.1609/aaai.v33i01.33016722

[11] Jessica Lindblom, Beatrice Alenljung, and Erik Billing. 2020. *Evaluating the User Experience of Human–Robot Interaction.* Number 1 in Springer Series on Bio- and Neurosystems. Springer, 231–256. https://doi.org/10.1007/978-3-030-42307-0_9

[12] Keting Lu, Shiqi Zhang, Peter Stone, and Xiaoping Chen. 2020. Learning and Reasoning for Robot Dialog and Navigation Tasks. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue.* Association for Computational Linguistics, 1st virtual meeting, 107–117. https://www.aclweb.org/anthology/2020.sigdial-1.14

[13] Cynthia Matuszek. 2018. Grounded Language Learning: Where Robotics and NLP Meet. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence.* International Joint Conferences on Artificial Intelligence Organization, 5687–5691. https://doi.org/10.24963/ijcai.2018/810

[14] Sonja K Ötting, Lisa Masjutin, Jochen J Steil, and Günter W Maier. 2020. Let's Work Together: A Meta-Analysis on Robot Design Features That Enable Successful Human–Robot Interaction at Work. *Human Factors* (2020), 0018720820966433. https://doi.org/10.1177/0018720820966433

[15] Elisa Prati, Margherita Peruzzini, Marcello Pellicciari, and Roberto Raffaeli. 2021. How to include User eXperience in the design of Human-Robot Interaction. *Robotics and Computer-Integrated Manufacturing* 68 (2021), 102072. https://doi.org/10.1016/j.rcim.2020.102072

[16] Alan Ritter, Colin Cherry, and William B. Dolan. 2011. Data-Driven Response Generation in Social Media. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Edinburgh, Scotland, UK., 583–593. https://www.aclweb.org/anthology/D11-1054

[17] Matthew Rueben, Shirley A. Elprama, Dimitrios Chrysostomou, and An Jacobs. 2020. *Introduction to (Re)Using Questionnaires in Human-Robot Interaction Research.* Number 1 in Springer Series on Bio- and Neurosystems. Springer, 125–144. https://doi.org/10.1007/978-3-030-42307-0_5

[18] Maria Schmidt, Jan Niehues, and Alex Waibel. 2017. *Towards an Open-Domain Social Dialog System.* Springer Singapore, Singapore, 271–278. https://doi.org/10.1007/978-981-10-2585-3_21

[19] Minija Tamosiunaite, Mohamad Javad Aein, Jan Matthias Braun, Tomas Kulvicius, Irena Markievicz, Jurgita Kapociute-Dzikiene, Rita Valteryte, Andrei Haidu, Dimitrios Chrysostomou, Barry Ridge, Tomas Krilavicius, Daiva Vitkute-Adzgauskiene, Michael Beetz, Ole Madsen, Ales Ude, Norbert Krüger, and Florentin Wörgötter. 2019. Cut & recombine reuse of robot action components based on simple language instructions. *The International Journal of Robotics Research* 38, 10-11 (1 Sept. 2019), 1179–1207. https://doi.org/10.1177/0278364919865594

[20] Chongyang Tao, Wei Wu, Can Xu, Wenpeng Hu, Dongyan Zhao, and Rui Yan. 2019. One Time of Interaction May Not Be Enough: Go Deep with an Interaction-over-Interaction Network for Response Selection in Dialogues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, Florence, Italy, 1–11. https://doi.org/10.18653/v1/P19-1001

[21] Jesse Thomason, Aishwarya Padmakumar, Jivko Sinapov, Nick Walker, Yuqian Jiang, Harel Yedidsion, Justin Hart, Peter Stone, and Raymond Mooney. 2020. Jointly improving parsing and perception for natural language commands through human-robot dialog. *Journal of Artificial Intelligence Research* 67 (2020), 327–374. https://doi.org/10.1613/jair.1.11485

[22] J. Thomason, A. Padmakumar, J. Sinapov, N. Walker, Y. Jiang, H. Yedidsion, J. Hart, P. Stone, and R. J. Mooney. 2019. Improving Grounded Natural Language Understanding through Human-Robot Dialog. In *2019 International Conference on Robotics and Automation (ICRA)*. 6934–6941. https://doi.org/10.1109/ICRA.2019.8794287

[23] Zhiliang Tian, Rui Yan, Lili Mou, Yiping Song, Yansong Feng, and Dongyan Zhao. 2017. How to Make Context More Useful? An Empirical Study on Context-Aware Neural Conversational Models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Vancouver, Canada, 231–236. https://doi.org/10.18653/v1/P17-2036

[24] Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869* (2015).

[25] Zhou Yu, Ziyu Xu, Alan W Black, and Alexander Rudnicky. 2016. Strategy and Policy Learning for Non-Task-Oriented Conversational Systems. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue.* Association for Computational Linguistics, Los Angeles, 404–412. https://doi.org/10.18653/v1/W16-3649