A Natural Language-enabled Virtual Assistant for Human-Robot Interaction in Industrial Environments

LI, Chen; Chrysostomou, Dimitrios; Yang, Hongji

Published in:
IEEE International Conference on Software Quality, Reliability and Security

DOI (link to publication from Publisher):
10.1109/QRS-C57518.2022.00107

Publication date:
2022

Document Version
Early version, also known as pre-print

A Natural Language-enabled Virtual Assistant for Human-Robot Interaction in Industrial Environments

Chen Li[1,*], Dimitrios Chrysostomou[2], and Hongji Yang[3]
[1,2]Department of Materials and Production, Aalborg University, DK-9220, Aalborg East, Denmark
[3]School of Computing and Mathematical Sciences, University of Leicester, Leicester LE1 7RH, U.K
cl@mp.aau.dk, dimi@mp.aau.dk, hongji.yang@leicester.ac.uk
*Corresponding author: Chen Li(cl@mp.aau.dk)

***Abstract*—This paper introduces a natural language-enabled virtual assistant (VA), called Max, developed to enhance human-robot interaction (HRI) with industrial robots. Regardless of the numerous natural language interfaces already available for commercial use and social robots, most VAs remain tightly bound to a specific robotic system. Besides, they lack a natural and efficient human-robot communication protocol to advance the user experience and the required robustness for use on the industrial floor. Therefore, the proposed framework is designed based on three key elements. A Client-Server style architecture that provides a centralised solution for managing and controlling various types of robots deployed on the shop floor. A communication protocol inspired by human-human conversation strategies, i.e., lexical-semantic strategy and general diversion strategy, is used to guide Max's response generation. These conversation strategies are embedded in Max's architecture to improve the engagement of the operators during the execution of industrial tasks. Finally, the state-of-the-art pre-trained model, Bidirectional Encoder Representations from Transformers (BERT), is fine-tuned to support a highly accurate prediction of requested intents from the operator and robot services. Multiple experiments were conducted for validating Max's performance in a real industrial environment.**

Keywords*—*Human-robot interaction; Natural language processing; Interactive systems; Client-server systems

I. INTRODUCTION

Human-robot interaction (HRI) has been the focus of groundbreaking research for decades [1]. Coupled with the rapid development of Artificial Intelligence (AI), various advanced technologies such as real-time object detection [2], deep reinforcement learning [3], and natural language processing (NLP), are introduced to enhance HRI for industrial robots [4]. However, as one of the latest articles of WEIRD magazine mentioned, "As Robots Fill the Workplace, They Must Learn to Get Along" [5]. The presence of advanced technologies alone, do not suffice for natural communication and interaction with multiple types of robots that are designed for different purposes and used by various operators [6]. For example, the Mobile Industrial Robot (MiR) focuses on internal logistics technologies such as navigation and obstacle avoidance. However, it only communicates with the user based on light signals. At the same time, many industrial robot manipulators are branded as collaborative, often due to marketing reasons and primarily based on their control and safety strategies. However, while they revolutionized the industry enabling human-robot collaboration (HRC) outside of safety fences, they barely incorporate ways to engage users into a natural language dialogue to enhance the communication and interaction with them [7]. Therefore, it is necessary to introduce a flexible and scalable NLP-enabled interface for HRI and HRC, which works with various industrial robots on the shop floor based on a well-designed architecture that provides a centralized way for robot management and maintenance.

There are many mature language-enabled VAs, e.g., Alexa [1], Google Assistant [2] and Siri [3] available commercially and used by millions of users worldwide. Their key feature is their impressive capacity for handling robust NLP and continuous natural dialogues. Nevertheless, most of those products are designed in the context of entertainment and remain unsuitable for direct use in robotics, especially in an industrial context.

On the other hand, many research efforts utilize verbal cues to enhance HRI in industrial environments [8]. Specifically, Li et al. developed a language-enabled virtual assistant, Bot-X, to control a production line composed of eight Festo CP Factory and a KUKA robot for product assembly task [9]. Maksymova et al. proved that numerous models could be used for voice control of an industrial robot such as logical, semantic networks, and Petri Nets in the context of collaborative assembly [10]. Li et al. proposed an end-to-end dialogue system for HRI in four industrial tasks [11]. Additionally, Bingol and Aydogmus evaluated the capabilities of deep neural networks for the classification of a set of commands in a natural speech recognition system for the interactive control of an industrial manipulator in various industrial tasks [12]. González-Docasal et al. progressed even further and integrated a semantic interpreter able to extract semantic information from transcribed spoken content to enable an industrial robot to understand the intention of the operator and execute a collaborative task [13].

Naturally, since industrial robots mainly assist the users with manufacturing tasks [14], task-completion experience is usually set as the primary evaluation goal of robots' performance in most of the aforementioned scenarios [13]. However, an operator remains the most flexible entity on the shop floor which needs to handle a versatile range of tasks and tools with robots' cooperation where task-completion is not the only requirement, e.g., collaborative products assembly,

[1]https://www.amazon.com/b?ie=UTF8&node=17934671011
[2]https://developers.google.com/assistant
[3]https://developer.apple.com/siri/

and resource management and logistics assisted by mobile robots [15].

The proposed VA is based on our previous concept work [16] and motivated by multiple studies from social, service, and lately, industrial robotics, which have proven that creating a pleasant and symbiotic human-robot collaboration often improves the user's engagement and leads to increased productivity [17]. Key elements of such successful HRI usually are the strong sense of commitment from the operator [18] and the enhanced user experience [19].

As a result, we developed an intelligent VA to enable these elements in industrial use cases. We call it Max, and it utilizes a Client-Server (CS) style architecture and RESTful APIs to provide a scalable and flexible NLP solution for intuitive HRI. It supports various industrial robots on the shop floor by maintaining industrial robot services on the server-side alone while the robot control agent lies on the Max client. Furthermore, powered by the state-of-the-art (SOTA) model and inspired by human-human communication, Max can understand the operator's intent, track the dialogue history and enhance the user experience by generating humanized responses.

In Section II of this paper, we describe the proposed intelligent VA presenting its system architecture and core components. We present the experiments and evaluate Max's performance in Section III and Section IV respectively, and we finalize the paper with reflections and concluding remarks in Section V.

II. PROPOSED SYSTEM

The architecture of the natural language-enabled VA, Max, is depicted in Fig. 1. Following the CS-style architecture, language and robot services are hosted on the Max server side. As language services, we use spoken language understanding, dialogue state tracker, and response generator, while as robot services, we use robot skills and scripts for robot control. Max client is mainly composed of Microsoft cognitive speech service and the robot control agent. RESTful HTTP requests support the communication between Max server and client. Max client provides a voice interface (i.e., microphone and speaker) to interact with the human operator and leverages the different protocols (e.g., OPC UA, RESTful APIs) to communicate with the shop floor robots.

A. Spoken language understanding

In this module, we fine-tuned the pre-trained base BERT model, which provides contextualised sentence representation and can learn the meaning of the words in the given context to predict the operator's intent and key slots from the operator's utterance. The model is trained and validated on a dialog dataset containing the task-related conversations of two different industrial robots (i.e., MiR200, Franka Emika Robot), and it achieves intent accuracy of 97.7% and slot F1 of 96.8%.

B. Dialogue state tracker

There could be a few stages involved in finishing a manufacturing operation. As a result, a VA may need to engage in several rounds or turns of dialogue in order to acquire all of the information that is essential to complete the task. This module includes a pre-defined JSON file that specifies all of the slots for completing each job. Specifically, Max makes sure that the state is correct and then changes it depending on the slots that were requested. The conversation will go up to the point when Max has filled all of the necessary slots.

C. Response Generator

To generate the response, a simple template-based strategy is integrated to the Max. Max might start out from a semantic representation confirming that the *phone box* will be delivered to *warehouse*, where the gaps represented by *#object* and *#position* respectively e.g., "Sure, I will deliver the *#object* to *#position*".

One key contribution of our approach is the inclusion of human-human conversation strategies to the interaction with the industrial robots. inspired by [20], we borrowed two dialogue strategies, lexical-semantic strategy and general diversion strategy, to Max to provide a highly dynamic and humanised conversation environment. The video [4] shows how Max may prompt the operator with a request and get right into manufacturing operations, provide alternatives if the desired activity can't be completed, and redirect the conversation when it realizes it's not connected to the task at hand. Since this is the case, Max can carry out the desired actions and facilitate dynamic conversations to enhance the user experience.

D. Industrial Robot Service

As discussed before, Max is not tightly bound with specific industrial robots as it is implemented with a CS-style architecture. This module supports the available robot services and skills registered on the server-side and the corresponding robot control scripts. Therefore, operators only need to focus on maintaining those registrations and scripts on the server-side without investing much time and effort in the configuration of the shop floor robots.

E. Voice Interface

For Max's client, a voice-enabled interface is being developed in order to provide effective communication. To enable Max's client voice interface, the cognitive speech services offered by Microsoft are integrated. These services are known as Speech-to-Text and Text-to-Speech. After being wrapped into an HTTP request formatted in a RESTful style, the operator's utterance transcripts are then transmitted to Max's server where the text response (which is a JSON string) is generated. Therefore, the text-to-speech service is leveraged to produce a natural human sounds based on the generated text response. A voice interface of this kind frees up the operator's hands while also allowing for natural conversation on the shop floor, which is essential for effective HRI.

[4]http://y2u.be/lYnh2cOeeE0

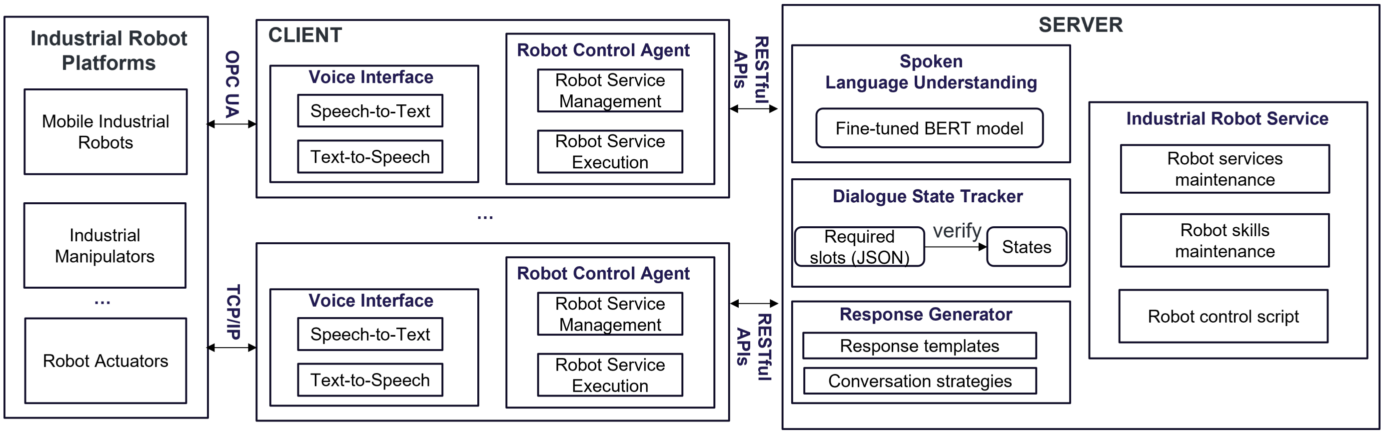

Figure 1. The architecture of the proposed natural language-enabled VA.

F. Robot Control Agent

in order to be agnostic of robotics platforms, two components, robot service management (RSM) and the robot service execution (RSE), are implemented on the Max client. RSM regularly synchronizes the local robot services and skills with the Max server on a monthly basis by default. This allows RSM to update the robot services and skills that have been registered on the local client-side as well as the robot controlling scripts. RSE calls the associated robot control script, which provides robot control functions (such as mark position) and communication protocols (such as TCP/IP, OPC-UA) for each robot type.

III. Experimental Setup

We demonstrate the effectiveness of our approach with a set of experiments performed in the AAU learning factory [21]. As the main robot, we use one of our autonomous industrial mobile manipulators, namely Little Helper (LH) [22]. In this iteration, LH combines a MiR 200 with a Franka Emika Robot, as shown on the top side of Fig. 2. The 2D map of the AAU learning factory is generated by LH as shown on the bottom side of Fig. 2.

We tested Max in three different industrial scenarios i.e, *environment exploration*, *package delivery* and *embedded conversation strategy*. These scenarios explore and test Max's capabilities to handle typical, everyday tasks from an industrial shop floor.

In the *environment exploration* scenario, the operator navigates LH inside the AAU learning factory to explore the working environment. Updating and checking the position of the robot on the map are identified as two major tasks (see Tasks 1 and 2 in Table I).

In the *package delivery* scenario, LH needs to deliver a box from one place to another according to the operator's verbal instruction (Task 3 in Table I). The box pickup place *workshop* and destination *warehouse* are marked on the map (see Fig. 2) and are generated already in Task 1. The intent *DELIVERY* is tested here, and slot values *recipient*, *to be delivered object*, *object size*, *object color* and *destination*.

In the last scenario, we evaluate Max's performance in utilising embedded conversation strategies (Tasks 4, 5, and 6). These tasks test specifically whether the embedded human-human conversation strategies can bootstrap the user experience by improving the user's engagement. In Task 4, Max can initialise activities (e.g., report the currently scheduled tasks) to detect operators' greetings. For Task 5, Max provides task-related options to the operator when it cannot perform the requested tasks. Task 6 tests if Max can provide different responses for the same request.

The minimum, maximum, and average ambient noise levels during our experiments were 44.6 dB(A), 80.5 dB(A), and 69.0 dB(A), respectively. Six users, a mix of developers and everyday users of the system, were tasked to interact with Max and control LH based on verbal commands. The performed tasks and respective intents are presented in Table I.

TABLE I
SELECTED TASKS AND INTENTS FOR THE THREE SCENARIOS

Task ID	Task description	Intent
1	Update the location	*POSITIONUPDATE*
2	Check the location	*POSITIONCHECK*
3	Deliver a package	*DELIVERY*
4	Initial activities	*GREETING*
5	Switch a topic	*ASKHELP*
6	Don't repeat	*MISSIONCHECK*

Each task was reproduced 30 times to verify Max's accuracy and validate its overall performance. The evaluation involves two main levels; the language level, where we evaluate how well Max can recognise the operator's intent and the slot value during the dialogue [23], and task level, where the task success rate and dialogue time costs are evaluated [24]. Furthermore, the ability to handle multiple tasks is also considered to evaluate the scalability of the proposed architecture. Overall, six metrics are used to evaluate the performance of Max in these scenarios.

- **Intent Error Rate (IER)**: refers to the rate where the model cannot predict the requested intent of the user correctly.

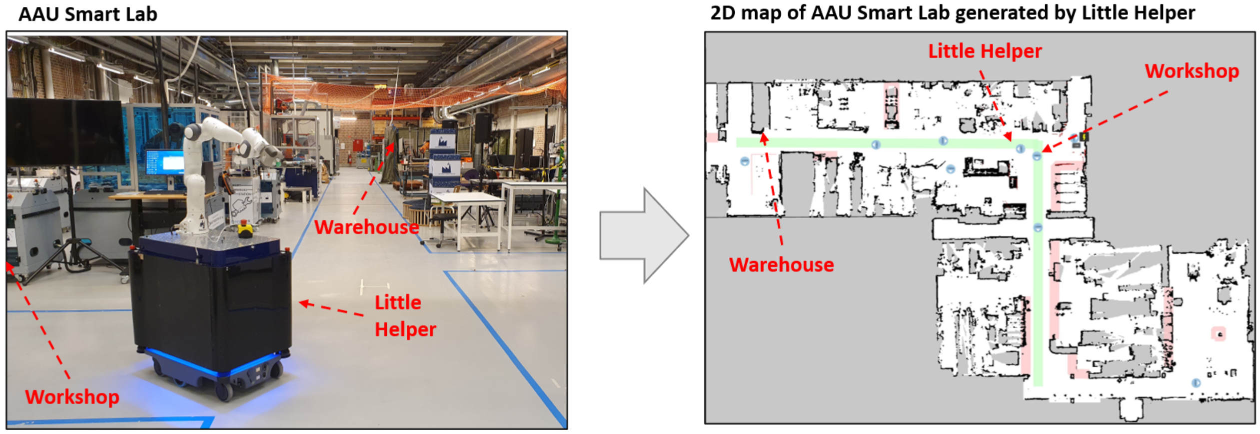

Figure 2. AAU learning factory and the corresponding 2D map generated by the robot (top). The locations of the robot (Little Helper), workshop and warehouse, are marked on the map (bottom).

- **Slot Error Rate (SER)**: is the rate of incorrect slot values, i.e., the model selects an incorrect value for the slot, and slot detection fails, i.e., the model cannot recognise the slot.
- **Task-Success Rate (TSR)**: corresponds to the rate of Max's ability to retrieve all the required information to complete the task without encountering any intent or slot error problem.
- **Average Communication Time (ACT)**: the required time for a task to be completed on average. The ACT is equal to the meantime from the beginning of the conversation to its end for 30 experiments of each task. In general, high intent/slot error rate requires more communication time.
- **Parallel Requests Handling (PRH)**: CS-style architecture brings flexibility and scalability to Max, but it also suffers from traffic congestion. PRH is used to measure Max's parallel processing capacity.
- **Average Service Updating Time (ASUT)**: measures the average time cost for updating the local robot control scripts. ASUT is equal to the meantime from sending a synchronisation request to completing an update of the local robot control scripts.

IV. System Evaluation

In the first scenario, Max failed to recognise the operator's intent and slot values in eight experiments and five experiments, respectively, for Task 1. For Task 2, the corresponding occurrences numbers are six and two separately. Max completed the first two tasks with a TSR of 0.60 and 0.76, respectively, while requiring 24.70 sec and 15.28 sec on average to perform the communication. Table II reports the corresponding IER, SER, TSR, and ACT for all tasks.

In the second scenario, Max failed to recognise the operator's intent three times and the slot value 13 times. Due to the complexity of the task and the ambient noise, Max achieved a completion rate of only 0.50 and required the highest communication time of 31.08 seconds on average (see Table II).

In the last scenario, Max's client remains in standby mode until it receives confirmation from the operator. The intent error rate varies depending on the complexity of the dialogue, and the task completion rate is 0.80, 0.70, and 0.83 per task. Similarly, the average communication time varies noticeably from 27.07 to 25.04 and 11.12 seconds, respectively (see Table II).

TABLE II
Performance of the VA in terms of IER, SER, TSR and ACT metrics

Task ID	IER	SER	TSR	ACT (in secs)
1	0.26	0.16	0.60	24.70
2	0.2	0.06	0.76	15.28
3	0.1	0.43	0.50	31.08
4	0.2	None	0.80	27.07
5	0.3	None	0.70	25.04
6	0.17	None	0.83	11.12

Additionally, to evaluate the overall parallel processing capacity, we run stress tests on Max's local server and the AAU Cloud using Siege[5], an HTTP load testing and benchmarking utility. The tests focus on the total elapsed time for the given number of transactions, the transaction rate, the actual maximum concurrent number of the connections, and the throughput. Table III shows the test results of the local server and the AAU Cloud.

TABLE III
PRH of Max's local server and AAU Cloud.

Server type	Transactions	Elapsed time	Transaction rate	Concurrency	Throughput
Local Server	2,000	10.09 sec	198.22 trans/sec	2.01	0.16 MB/sec
AAU Cloud	100,000	7.04 sec	14204.55 trans/sec	289.79	11.65 MB/sec

Finally, to evaluate the ASUT, another 30 experiments were conducted on MiR robot. The tests focus on calling the

[5]https://www.joedog.org/siege-home/

functions which are not defined in the local robot control script. For the 30 experiments, the minimum, maximum, and average service updating times were 5.6, 6.3, and 5.7 seconds respectively, measured on the local server.

V. Discussion & Conclusions

The proposed natural language-enabled VA, Max, benefits from the CS-style architecture, RESTful style APIs, and centralised management, enabling high efficiency in multiple HRI scenarios in industrial robots. The addition of the industrial robot service and the robot control agent can also interact efficiently with various industrial robots.

Our proposed VA still has room for improvement as it requires significant processing power for handling many parallel requests and a well-designed security strategy to maintain the privacy of the interaction environment. Furthermore, other factors, e.g., the operator's accent and voice volume, are also influential aspects of the intent/slot error rate. Future work will focus on ways to suppress the ambient noise and enhance speech so as the VA and, consequently the robot system, can communicate with the workers with fewer interruptions and errors.

The encouraging results based on the two embedded conversation strategies prove the Max can support an active interaction during various manufacturing tasks. It provides task-related suggestions, attracts the operator's attention successfully, and forms a diverse and thoughtful dialogue to improve user engagement in HRI for industrial robots. An extensive HRI usability study was postponed due to COVID-19 restrictions; however, it remains a central part of our future work to collect feedback on the naturalness and coherence of Max's generated dialogue and responses.

Acknowledgment

The authors would like to acknowledge support by EU's SMART EUREKA programme under grant agreement S0218-chARmER, Innovation Fund Denmark (Grant no. 9118-00001B), and the H2020-WIDESPREAD project no. 857061 "Networking for Research and Development of Human Interactive and Sensitive Robotics Taking Advantage of Additive Manufacturing – R2P2"

References

[1] C. Jost, B. Le Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge, and N. Mirnig, Eds., *Human-robot interaction : evaluation methods and their standardization*. Springer, 2020, vol. 12. [Online]. Available: https://doi.org/10.1007/978-3-030-42307-0

[2] J. F. Buhl, R. Grønhøj, J. K. Jørgensen, G. Mateus, D. Pinto, J. K. Sørensen, S. Bøgh, and D. Chrysostomou, "A dual-arm collaborative robot system for the smart factories of the future," *Procedia manufacturing*, vol. 38, pp. 333–340, 2019. [Online]. Available: https://doi.org/10.1016/j.promfg.2020.01.043

[3] N. Arana-Arexolaleiba, N. Urrestilla-Anguiozar, D. Chrysostomou, and S. Bøgh, "Transferring human manipulation knowledge to industrial robots using reinforcement learning," *Procedia Manufacturing*, vol. 38, pp. 1508–1515, 2019. [Online]. Available: https://doi.org/10.1016/j.promfg.2020.01.136

[4] E. Lithoxoidou, S. Doumpoulakis, A. Tsakiris, C. Ziogou, S. Krinidis, I. Paliokas, D. Ioannidis, K. Votis, S. Voutetakis, E. Elmasllari *et al.*, "A novel social gamified collaboration platform enriched with shop-floor data and feedback for the improvement of the productivity, safety and engagement in factories," *Computers & Industrial Engineering*, vol. 139, p. 105691, 2020. [Online]. Available: https://doi.org/10.1016/j.cie.2019.02.005

[5] W. Knight. (2021, February) As robots fill the workplace, they must learn to get along. [accessed 02/02/2021]. [Online]. Available: https://www.wired.com/story/robots-fill-workplace-must-learn-get-along/

[6] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018. [Online]. Available: https://doi.org/10.1016/j.mechatronics.2018.02.009

[7] S. Hjorth and D. Chrysostomou, "Human–robot collaboration in industrial environments: A literature review on non-destructive disassembly," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102208, 2022. [Online]. Available: https://doi.org/10.1016/j.rcim.2021.102208

[8] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015. [Online]. Available: https://doi.org/10.1016/j.robot.2014.09.031

[9] C. Li and H. J. Yang, "Bot-x: An ai-based virtual assistant for intelligent manufacturing," *Multiagent and Grid Systems*, vol. 17, no. 1, pp. 1–14, 2021. [Online]. Available: https://doi.org/10.3233/MGS-210340

[10] S. Maksymova, R. Matarneh, V. Lyashenko, and N. Belova, "Voice control for an industrial robot as a combination of various robotic assembly process models," *Journal of Computer and Communications*, 2017. [Online]. Available: https://doi.org/10.4236/jcc.2017.511001

[11] C. Li, X. Zhang, D. Chrysostomou, and H. Yang, "Tod4ir: A humanised task-oriented dialogue system for industrial robots," *IEEE Access*, vol. 10, pp. 91 631–91 649, 2022.

[12] M. C. Bingol and O. Aydogmus, "Performing predefined tasks using the human–robot interaction on speech recognition for an industrial robot," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103903, 2020. [Online]. Available: https://doi.org/10.1016/j.engappai.2020.103903

[13] A. González-Docasal, C. Aceta, H. Arzelus, A. Álvarez, I. Fernández, and J. Kildal, "Towards a natural human-robot interaction in an industrial environment," in *Conversational Dialogue Systems for the Next Decade*. Springer, 2020, pp. 243–255. [Online]. Available: https://doi.org/10.1007/978-981-15-8395-7_18

[14] S. Kumar, C. Savur, and F. Sahin, "Survey of human–robot collaboration in industrial settings: Awareness, intelligence, and compliance," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 280–297, 2021. [Online]. Available: https://doi.org/10.1109/TSMC.2020.3041231

[15] G. Aceto, V. Persico, and A. Pescapé, "A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3467–3501, 2019. [Online]. Available: https://doi.org/10.1109/COMST.2019.2938259

[16] C. Li, J. Park, H. Kim, and D. Chrysostomou, "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*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 220–224. [Online]. Available: https://doi.org/10.1145/3434074.3447163

[17] L. Pérez, S. Rodríguez-Jiménez, N. Rodríguez, R. Usamentiaga, D. F. García, and L. Wang, "Symbiotic human–robot collaborative approach for increased productivity and enhanced safety in the aerospace manufacturing industry," *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 3, pp. 851–863, 2020. [Online]. Available: https://doi.org/10.1007/s00170-019-04638-6

[18] M. Székely, H. Powell, F. Vannucci, F. Rea, A. Sciutti, and J. Michael, "The perception of a robot partner's effort elicits a sense of commitment to human-robot interaction," *Interaction Studies*, vol. 20, no. 2, pp. 234–255, 2019. [Online]. Available: https://doi.org/10.1075/is.18001.sze

[19] E. Prati, M. Peruzzini, M. Pellicciari, and R. Raffaeli, "How to include user experience in the design of human-robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102072, 2021. [Online]. Available: https://doi.org/10.1016/j.rcim.2020.102072

[20] Z. Yu, Z. Xu, A. W. Black, and A. Rudnicky, "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, 2016, pp. 404–412. [Online]. Available: https://www.aclweb.org/anthology/W16-3649

[21] M. Nardello, O. Madsen, and C. Møller, "The smart production laboratory: A learning factory for industry 4.0 concepts," in *CEUR Workshop Proceedings*, vol. 1898. CEUR Workshop Proceedings, 2017.

[22] C. Schou, R. S. Andersen, D. Chrysostomou, S. Bøgh, and O. Madsen, "Skill-based instruction of collaborative robots in industrial settings," *Robotics and Computer-Integrated Manufacturing*, vol. 53, no. June 2016, pp. 72–80, 2018. [Online]. Available: https://doi.org/10.1016/j.rcim.2018.03.008

[23] J. Schatzmann, B. Thomson, and S. Young, "Error simulation for training statistical dialogue systems," in *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*. IEEE, 2007, pp. 526–531. [Online]. Available: https://doi.org/10.1109/ASRU.2007.4430167

[24] J. Deriu, A. Rodrigo, A. Otegi, G. Echegoyen, S. Rosset, E. Agirre, and M. Cieliebak, "Survey on evaluation methods for dialogue systems," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 755–810, 2021. [Online]. Available: https://doi.org/10.1007/s10462-020-09866-x