A Conversational User Interface for Instructional Maintenance Reports

Kernan Freire, Samuel; Niforatos, Evangelos; Rusak, Zoltan; Aschenbrenner, Doris; Bozzon, Alessandro

DOI
10.1145/3543829.3544516

Publication date
2022

Document Version
Final published version

Published in
Proceedings of the 4th International Conference on Conversational User Interfaces, CUI 2022

A Conversational User Interface for Instructional Maintenance Reports

S. Kernan Freire
Delft University of Technology
Delft, The Netherlands
s.kernanfreire@tudelft.nl

E. Niforatos
Delft University of Technology
Delft, Netherlands
e.niforatos@tudelft.nl

Z. Rusak
Royal HaskoningDHV
Amersfoort, Netherlands
zoltan_rusak@msn.com

D. Aschenbrenner
Hochschule Aalen
Aalen, Germany
doris.aschenbrenner@hs-aalen.de

A. Bozzon
Delft University of Technology
Delft, Netherlands
a.bozzon@tudelft.nl

ABSTRACT

Maintaining a complex system, such as a modern production line, is a knowledge-intensive task. Many firms use maintenance reports as a decision support tool. However, reports are often poor quality and tedious to compile. A Conversational User Interface (CUI) could streamline the reporting process by validating the user's input, eliciting more valuable information, and reducing the time needed. In this paper, we use a Technology Probe to explore the potential of a CUI to create instructional maintenance reports. We conducted a between-groups study ($N = 24$) in which participants had to replace the inner tube of a bicycle tire. One group documented the procedure using a CUI while replacing the inner tube, whereas the other group compiled a paper report afterward. The CUI was enacted by a researcher according to a set of rules. Our results indicate that using a CUI for maintenance reports saves a significant amount of time, is no more cognitively demanding than writing a report, and results in maintenance reports of higher quality.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Interactive systems and tools**; **Natural language interfaces**.

KEYWORDS

conversational user interface, voice assistant, maintenance reporting, technology probe, knowledge sharing, knowledge management

ACM Reference Format:
S. Kernan Freire, E. Niforatos, Z. Rusak, D. Aschenbrenner, and A. Bozzon. 2022. A Conversational User Interface for Instructional Maintenance Reports. In *4th Conference on Conversational User Interfaces (CUI 2022), July 26–28, 2022, Glasgow, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3543829.3544516

CUI 2022, July 26–28, 2022, Glasgow, United Kingdom

ACM ISBN 978-1-4503-9739-1/22/07…$15.00
https://doi.org/10.1145/3543829.3544516

1 INTRODUCTION AND BACKGROUND

Maintaining a complex system, such as a modern production line, is a knowledge-intensive task. One third of maintenance costs are incurred by inappropriate or unnecessary maintenance practices [18]. An expert technician typically trains a new technician 1-on-1 by explaining what they are doing, but this does not scale very well, as the expert is not always available. Technicians use maintenance reports to share knowledge and learn from colleagues, but they are often of poor quality and, therefore, not useful [8]. Moreover, reports frequently have inconsistent terminology and missing information.

While there has been a strong push towards manufacturing automation in recent decades, we are now entering a new phase where intelligent systems will fully merge with the physical world in cooperation with human intelligence [20]. Voice Assistants, with a Conversational User Interface (CUI), are beginning to emerge in manufacturing and focus on providing instructions and decision support [1, 22]. Using a voice-based CUI allows technicians to keep their hands and eyes free to work on the machines. Therefore, the technician can describe their actions and thought process while working. Furthermore, Wizard of Oz studies in the automotive context have shown positive effects of the use of CUIs on aspects such as cognitive demands / work load and environmental participation [13].

AI-powered decision support systems are revolutionizing the way knowledge is shared between technicians (e.g., [4]). However, these systems are mainly based on predefined knowledge bases that require a lot of resources to create and maintain [7]. Previous research has explored natural language processing (NLP) on existing maintenance reports to automatically discover knowledge, but numerous data quality issues were found [8]. Others concluded that technicians often describe problems informally, leading to inconsistencies and inaccuracies in the data; certain maintenance data, such as the actual root cause of a problem, are not always collected; and once the data is collected, it is often not used for future diagnosis [23]. Clearly, the poor quality of reports inhibits (AI-facilitated) knowledge sharing among technicians.

CUIs can be a viable alternative to creating maintenance reports on paper or graphical user interfaces. Voice-based CUIs are increasingly prevalent in the healthcare domain to support clinical workflows [17, 25]. CUIs in healthcare relieve physicians of the burden of documentation by using a digital scribe [5, 24]. In fact, CUIs have a positive impact on the accuracy and productivity of documentation [10]. Giving a verbal description of your actions and thoughts is intuitive, and designers use CUIs to better understand the user's thought process [6, 15]. We believe that using a CUI in the manufacturing domain, which uses documentation in a similar way to healthcare, could result in improved maintenance report quality, while reducing the technician's workload.

Here, we evaluate the use of a CUI to create maintenance reports using a Technology Probe [3]. We compare two groups that change the inner tube of a bicycle and report their work. Participants in one condition report their work during the task by voice, while the other group writes a report on paper afterward. We measure the duration of tasks, perceived workload, and the quality of reports.

2 USER STUDY

Figure 1: Experiment setup for changing a bicycle inner tube

We conducted a between-subjects Technology Probe study with 24 participants to compare using a CUI to create an instructional maintenance report while changing a bicycle inner tube versus writing the report on paper afterwards. The probe consisted of a researcher enacting the voice-enabled CUI. Therefore, we avoided the influence of additional factors such as the accuracy of the speech-to-text model and other NLP pitfalls. We purposely selected the task of changing an inner tube (the inflatable tube within a bicycle tire) as the study was carried out in the Netherlands and many people are proficient in it. Additionally, it contains several challenging steps where knowledge sharing could be relevant. We designed the user study to answer the following research questions:

RQ1: How does the quality of reporting compare between using a CUI during a maintenance task versus filling out a report on paper afterward?

RQ2: How does interacting with a CUI to file a report while performing a maintenance task influence the overall duration of the task as opposed to filling out a report on paper after the task?

RQ3: How is the perceived workload affected when comparing the use of a CUI during a maintenance task versus writing the report on paper afterward?

2.1 Methodology

Before starting the task, the participants provided their informed consent and completed a demographic questionnaire. Then we presented them with an inverted bicycle (the handlebars and seat resting on the ground), tire levers, a bicycle pump, and the spare inner tube (see Fig. 1). We then informed them that the front tire was flat and asked them to change the inner tube, reattach the wheel to the bicycle, and pump it up. During this task, we asked them to report on their work so that a novice technician could understand what steps they took, how and why. The "paper" group was asked to write the report on paper after changing the inner tube, while the "CUI" group was asked to report the information verbally using the (simulated) CUI while changing the inner tube. We instructed them to perform the tasks in an efficient manner. They were asked to indicate when they finished and then fill in a NASA-TLX questionnaire.

To create the CUI condition for the experiment, we used a Technology Probe. A researcher enacted the CUI by standing 1.5 meters behind the participant and conversing directly with them according to a set of rules. No attempt was made to fool the participants into thinking they were talking to a computer. The researcher was facing away from the participant to minimize the feeling of being watched and discourage them from turning around to talk or observe facial expressions. The simulated CUI features were based on the capabilities of a prototype built with a state-of-the-art CUI framework and CUIs from the recent literature [2, 12, 16].

Reporting a maintenance task is largely a one-sided conversation that closely resembles the story-telling style defined by Moore [19]. Therefore, the CUI responds with interjections and "continuers" to simulate a voice assistant that is listening and prompts elaboration, the same as a human conversation partner might. Using these cues has been shown to positively affect user motivation to continue the conversation [14]. Furthermore, a voice assistant can improve the quality of the report by attempting to extract named entity's and ask for clarification if ambiguous (e.g., when the user uses a pronoun). The full set of rules used for the experiment is described below.

- `Continuers`: Respond with a "continuer" whenever the participant finishes an utterance (e.g., "Okay," "nice," "mhm," "uh-huh")
- `Interjections`: Respond with interjections or acknowledgments when the participant indicates that they have completed a subtask (e.g., "Okay," "cool," "nice," "great")
- `Checkups`: If the participant has been silent for 15–20 seconds, ask them: "how are you doing?" or "how's it going?" If they are silent again within the next minute, wait 25–30 seconds before asking again
- `Pronouns`: If the participant uses a pronoun or general noun (e.g., it, that, tool), but it is not obvious which object they are referring to (e.g., from the previous utterance), ask which object they are referring to.

2.2 Measures

We collected participants' demographics, including experience with changing inner tubes of bicycles, thinking aloud, writing reports, wordiness, and English proficiency. We measured the duration of the task in seconds. For both conditions, the timer started when participants were instructed to start. For the "Paper" condition, the task to change the inner tube and create the report was measured separately, as they were performed consecutively. For the CUI condition, the inner tube change and reporting were performed in parallel, so only one time was measured. For both conditions, the timer stopped when the participant indicated that they had completed their tasks. The participants evaluated the perceived workload (NASA-TLX [9]) of the task as a whole upon completion. We followed a coding protocol to quantitatively measure the quality of the report provided by the participants (Table 1). The paper reports were digitized and an audio recording of the "CUI" group was transcribed to facilitate coding. We define the quality of the report as the frequency of useful and unique pieces of information that can be used to determine how the task was performed and why it was performed in this way. The coding protocol is based on one previously used to code conversations between an expert and a novice working on a bicycle repair task [11]. Only unique information is counted towards the report quality score to avoid favoring the CUI, where participants may be more inclined to repeat information.

Table 1: Report quality coding scheme

Information type	Definition
Procedural	Instructions furthering task completion (e.g., "next, remove the cover")
Task State	State of the task or objects within the task (e.g., "the tire is flat", "the quick release is very tight").
Referential	Utterances pertaining to the identification or location of task objects. (e.g., "the big disc in the middle")
Internal State	Intentions, knowledge, emotions, and so forth (e.g., "because, it might be slippy")

Although the original coding protocol classified entire user utterances, we coded at a lower level, as a single user utterance often contained multiple types of information. For example, a user could provide a description of a task step and an explanation in one utterance: "you have to turn the end of the pump to make it fit properly to the valve." We awarded points based on the following rules: (1) the procedural, internal state, and task state information must be unique at a report level, (2) the referential information must be unique at the utterance level, (3) the referential information only gets a point when it appears in conjunction with another point scoring information (procedural, internal state, task state), and (4) the information is relevant to the understanding of the task. By including requirements for the uniqueness and relevance of the information, we aim to reduce the potential bias that would favor the CUI condition as participants may include irrelevant details or repeat information when describing their actions and thoughts aloud.

2.3 Participants

We recruited 24 participants (20 male, 4 female) with ages ranging from 18 to 64 years. Most of the participants ($N = 15$) were 25–34 years old, followed by the 18–24 age bracket ($N = 6$). All participants must have changed a bicycle inner tube at least once, be confident that they could do so again without assistance, and be able to communicate clearly in spoken and written English.

3 RESULTS

To decide on our statistical methods, we first performed all the necessary pre-tests, such as Shapiro-Wilk tests of normality and Levene's tests of homogeneity of variance. We omit the pre-tests for brevity. Depending on the statistical test at hand, we report averages and standard deviations (parametric), or median values (non-parametric). We found no significant differences in the relevant experience (English language) and skills (inner tube replacement) of the participants between the two conditions.

3.1 Report quality

Next, we investigated the quality of the reports collected in the "Paper" and the "CUI" conditions based on the protocol we described above. A series of Shapiro-Wilk tests showed that procedural, referential, and total report quality scores were normally distributed between the two conditions, whereas task state and internal state scores were not. On the one hand, a series of independent-samples t-tests unveiled a significant difference in the average procedural knowledge scores between the "Paper" ($M = 23.92, SD = 6.708$) and the "CUI" ($M = 42.5, SD = 11.35$) conditions ($t(22) = -4.883, p < .001$), a significant difference in the average referential information scores between the "Paper" ($M = 37.67, SD = 12.908$) and the "CUI" ($M = 65.92, SD = 23.446$) conditions ($t(22) = -3.3656, p < .001$), and a significant difference in the average total report quality scores between the "Paper" ($M = 63.33, SD = 19.42$) and the "CUI" ($M = 126.33, SD = 42.059$) conditions ($t(22) = -4.711, p < .001$) (see Fig. 2). On the other hand, two Mann-Whitney U tests displayed a significant difference in the median task state scores between the "Paper" ($Mdn = .0$) and the "CUI" ($Mdn = 6$) conditions ($U = 15.5, p < .001$), and a significant difference in the median internal state scores between the "Paper" ($Mdn = 1$) and the "CUI" ($Mdn = 13.5$) conditions ($U = 16.5, p < .01$) (see Fig. 3). These findings show that participants who performed a maintenance task, such as changing a bicycle inner tube, produce knowledge of significantly higher quality when using the "CUI," compared to when filling in a report at the end (RQ1).

3.2 Task Completion Times

An independent-samples t-test revealed that the average task completion time (change inner tube and create the report) was significantly shorter ($t(14.532) = 3.889, p < .01$) for the "CUI" ($M = 574s, SD = 171.446$) as opposed to the "Paper" ($M = 1085.58s, SD = 422.215$) condition. Surprisingly, an independent-sampled t-test did not display a significant difference ($t(22) = .209, p = .418$) between the time to only change the inner tube for the "CUI" ($M = 539, SD = 172.888$) and for the "Paper" ($M = 554.17, SD = 183.304$) condition (see Fig. 4). These findings indicate that participants who used the CUI to report the inner tube change process were significantly faster to complete the entire task (inner tube change and report) than those who had written a report after the inner tube change (RQ2). Interestingly, using the CUI, while changing the inner tube, did not appear to significantly slow the participants.

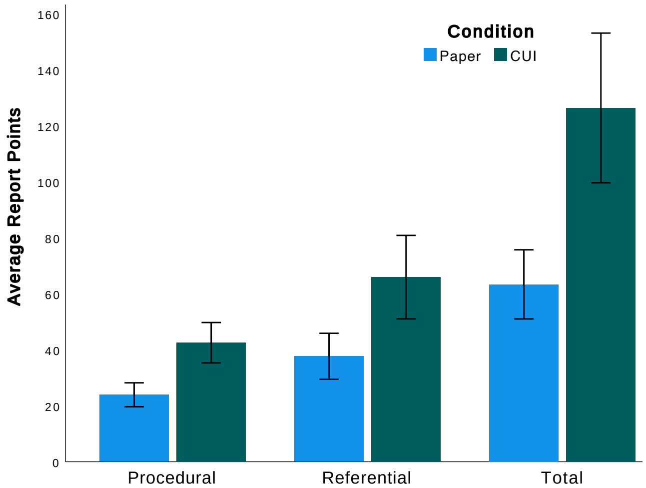

Figure 2: Average procedural, referential and total report quality per condition. All differences are significant ($p < .05$ or lower).

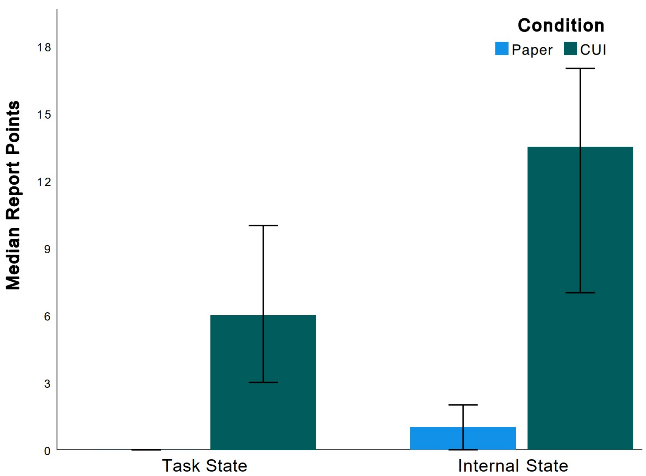

Figure 3: Median task state and internal state instances per condition. All differences are significant ($p < .05$ or lower).

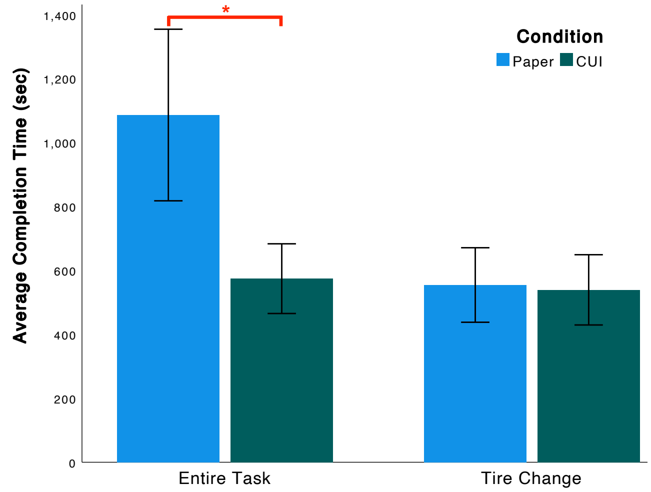

Figure 4: Average completion times for the inner tube task and inner tube-changing task per condition. Only marked differences are significant ($p < .05$ or lower).

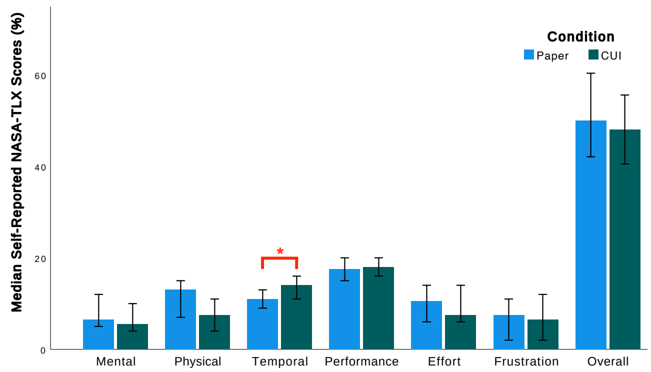

Figure 5: Median self-reported NASA-TLX values per condition. Only marked differences are significant ($p < .05$ or lower).

3.3 Perceived Workload

Finally, we had a look at the participants NASA-TLX responses for both the "Paper" and the "CUI" conditions. A series of Shapiro-Wilk tests showed that self-reported effort, frustration, and total workload were normally distributed across the two conditions, as opposed to self-reported mental, physical, and temporal demand, and performance that were not. On the one hand, a series of independent-samples t-tests displayed no significant difference in average self-reported effort ($t(22) = .179, p = .860$), frustration ($t(22) = -.344, p = .734$), and total workload ($t(22) = .102, p = .920$) between the "Paper" and the "CUI" conditions (see Fig. 5). On the other hand, a series of Mann-Whitney U tests displayed no significant difference in median self-reported mental ($U = 59.5, p = .467$) and physical demand ($U = 43.5, p = .098$), as well as performance ($U = 70.5, p = .931$), but a significant difference in the median self-reported temporal demand between the "Paper" ($Mdn = 11$) and the "CUI" ($Mdn = 14$) conditions ($U = 37.5, p < .05$) (see Fig. 5). These findings show that there were no significant differences in the perceived workload values between the "Paper" and the "CUI" condition (RQ3). However, self-reported temporal demand was found to be significantly higher for the "CUI" as opposed to the "Paper" condition. This indicates that the participants perceived the interaction with the "CUI," while changing the inner tube, as more time-pressed than when changing an inner tube and filling in a report afterwards (RQ3).

4 DISCUSSION

Overall, report quality was significantly in favor of the "CUI" group / condition. This result can be explained by several factors: (1) by thinking aloud as participants performed the task, the "CUI" group was less likely to forget details, (2) the participants were inclined to provide more details because talking is much more efficient than writing on paper, (3) participants who took longer, or had challenges performing the task, inevitably talked more, including describing unproductive steps. On the contrary, participants in the "Paper" group usually only reported the steps to complete the task, thus excluding any exploratory steps or explanations.

Interestingly, CUI-based reporting was not found to affect maintenance task completion time. In fact, the time to change the inner tube did not differ significantly between the two conditions ("CUI" and "Paper"). When reporting time is factored in, using a CUI can save technicians and their firm a significant amount of time. Clearly, the ability to report by voice during the maintenance task, rather than afterward, was the reason. Surprisingly, the overall perceived workload did not differ significantly between the "CUI" and "Paper" conditions. However, the perceived temporal demand for the "CUI' group was reported as significantly higher than that for the "Paper" group. We attribute this finding to the fact that participants who interacted with the CUI were multitasking and perhaps felt pressed to maintain the conversation flow. The significance of this finding should be noted, as it suggests that designers of conversational systems for on-the-job knowledge acquisition should adjust conversational flow to task progress.

4.1 Limitations

We used a Technology Probe in which a researcher simulated the CUI instead of implementing one. In doing so, we minimized the effects of the quality of the CUI at the cost of realism. However, we believe that we maintained sufficient realism, since the simulated CUI followed a set of rules based on existing CUI capabilities. We also assume that a coherent report can be automatically generated from the information provided in the CUI condition. We believe that this is a valid assumption given the state-of-the-art capabilities of digital scribes in healthcare [25]. Another limitation is that the participants knew that they were conversing with a human and not a computer. Although this may bias the results, we believe the effects are insignificant, as previous research on automated interviewing has shown that humans disclose the same information to conversational agents compared to a human [21]. Furthermore, the researcher was not watching the participant's progress, nor could the participant see the researcher's face. Next, our participants were not factory technicians. However, we believe that our findings can still be generalized, as our participants were technically qualified enough to know how to change the inner tube of a bicycle, and we explained the objective of the report to them. The coding protocol we used to measure the quality of the report is based on the frequency of several types of information. However, it does not consider its accuracy or whether important information was omitted. We recognize that our protocol can only approximate the amount of useful information collected. Finally, a researcher who was already involved in the study performed the coding. Therefore, the reliability and accuracy of the coding was not independently validated.

5 CONCLUSION AND FUTURE WORK

Using a CUI for maintenance reports saves a significant amount of time, is no more cognitively demanding than writing a report, and produces maintenance reports of higher quality—reports contained more utterances that were relevant to understand or to perform the assigned task. In the next step, we will recruit an additional coder to independently rate the reports collected for the two conditions and validate our coding scheme. Future work will explore how to improve the design of the CUI to enhance the user experience and knowledge acquisition. For example, we will investigate how conversation flow can be best adapted to task progress by asking for task updates or employing context awareness. To increase realism, we will develop a prototype capable of uttering continuers, interjections, clarifying pronouns, and exploring how to process the collected information into a useful format for other technicians or AI assistants.

ACKNOWLEDGMENTS

This work is funded by the European Union's Horizon 2020 research and innovation program via the project COALA "COgnitive Assisted agile manufacturing for a LAbor force supported by trustworthy Artificial Intelligence" (Grant agreement No 957296).

REFERENCES

[1] André da Silva Barbosa, Felipe Pinheiro Silva, Lucas Rafael dos Santos Crestani, and Rodrigo Bueno Otto. 2018. Virtual assistant to real time training on industrial environment. In *Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0*. IOS Press, 33–42.

[2] Dennis Benner, Edona Elshan, Sofia Schöbel, and Andreas Janson. 2021. What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents. In *International Conference on Information Systems (ICIS)*.

[3] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI Interprets the Probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1077–1086. https://doi.org/10.1145/1240624.1240789

[4] Mario Casillo, Francesco Colace, Loretta Fabbri, Marco Lombardi, Alessandra Romano, and Domenico Santaniello. 2020. Chatbot in Industry 4.0: An Approach for Training New Employees. In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. 371–376. https://doi.org/10.1109/TALE48869.2020.9368339

[5] Enrico Coiera, Baki Kocaballi, John Halamka, and Liliana Laranjo. 2018. The digital scribe. *NPJ digital medicine* 1, 1 (2018), 1–5. https://doi.org/10.1038/s41746-018-0066-9

[6] Lynne Cooke. 2010. Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach. *IEEE Transactions on Professional Communication* 53, 3 (2010), 202–215. https://doi.org/10.1109/TPC.2010.2052859

[7] Jet Cullen and Alan Bryman. 1988. The knowledge acquisition bottleneck: time for reassessment? *Expert Systems* 5, 3 (1988), 216–225. https://doi.org/10.1111/j.1468-0394.1988.tb00065.x

[8] Brett Edwards, Michael Zatorsky, and Richi Nayak. 2008. Clustering and classification of maintenance logs using text data mining. *Volume 87-Data Mining and Analytics 2008* (2008), 193–199.

[9] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[10] Joseph Joseph, Zena EH Moore, Declan Patton, Tom O'Connor, and Linda Elizabeth Nugent. 2020. The impact of implementing speech recognition technology on the accuracy and efficiency (time to complete) clinical documentation by nurses: A systematic review. *Journal of clinical nursing* 29, 13-14 (2020), 2125–2137. https://doi.org/10.1111/jocn.15261

[11] Robert E. Kraut, Susan R. Fussell, and Jane Siegel. 2003. Visual Information as a Conversational Resource in Collaborative Physical Tasks. *Human–Computer Interaction* 18, 1-2 (June 2003), 13–49. https://doi.org/10.1207/S15327051HCI1812_2 Publisher: Taylor & Francis.

[12] Aman Kumar and Binil Starly. 2021. "FabNER": information extraction from manufacturing process science domain literature using named entity recognition. *Journal of Intelligent Manufacturing* (2021), 1–15. https://doi.org/10.1007/s10845-021-01807-x

[13] David R. Large, Gary Burnett, and Leigh Clark. 2019. Lessons from Oz: Design Guidelines for Automotive Conversational User Interfaces. In *Proceedings of*

the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 335–340. https://doi.org/10.1145/3349263.3351314

[14] Naeun Lee and Sangwon Lee. 2018. The Effect of Back-Channeling Cues on Motivation to Continue Human-Machine Textual Interaction. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication* (Langkawi, Malaysia) *(IMCOM '18)*. Association for Computing Machinery, New York, NY, USA, Article 49, 4 pages. https://doi.org/10.1145/3164541.3164626

[15] Clayton Lewis. 1982. *Using the" thinking-aloud" method in cognitive interface design.* IBM TJ Watson Research Center Yorktown Heights.

[16] Ondrej Miksik, I Munasinghe, J Asensio-Cubero, S Reddy Bethi, ST Huang, S Zylfo, X Liu, T Nica, A Mitrocsak, S Mezza, et al. 2020. Building proactive voice assistants: When and how (not) to interact. *arXiv preprint arXiv:2005.01322* (2020).

[17] Izidor Mlakar, Valentino Šafran, Daniel Hari, Matej Rojc, Gazihan Alankuş, Rafael Pérez Luna, and Umut Ariöz. 2021. Multilingual Conversational Systems to Drive the Collection of Patient-Reported Outcomes and Integration into Clinical Workflows. *Symmetry* 13, 7 (2021), 1187. https://doi.org/10.3390/sym13071187

[18] R Keith Mobley. 2002. *An introduction to predictive maintenance.* Elsevier. https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3

[19] Robert J Moore and Raphael Arar. 2019. *Conversational UX design: A practitioner's guide to the natural conversation framework.* Morgan & Claypool.

[20] Saeid Nahavandi. 2019. Industry 5.0—A human-centric solution. *Sustainability* 11, 16 (2019), 4371. https://doi.org/10.3390/su11164371

[21] Matthew D Pickard and Catherine A Roster. 2020. Using computer automated systems to conduct personal interviews: Does the mere presence of a human face inhibit disclosure? *Computers in Human Behavior* 105 (2020), 106197. https://doi.org/10.1016/j.chb.2019.106197

[22] Manex Serras, Laura García-Sardiña, Bruno Simões, Hugo Álvarez, and Jon Arambarri. 2020. Dialogue enhanced extended reality: Interactive system for the operator 4.0. *Applied Sciences* 10, 11 (2020), 3960. https://doi.org/10.3390/app10113960

[23] Thurston Sexton, Michael P. Brundage, Michael Hoffman, and K C Morris. 2017. Hybrid datafication of maintenance logs from AI-assisted human tags. In *2017 IEEE International Conference on Big Data (Big Data)*. 1769–1777. https://doi.org/10.1109/BigData.2017.8258120

[24] Cameron G Shultz and Heather L Holmstrom. 2015. The use of medical scribes in health care settings: a systematic review and future directions. *The Journal of the American Board of Family Medicine* 28, 3 (2015), 371–381. https://doi.org/10.3122/jabfm.2015.03.140224

[25] Marieke M van Buchem, Hileen Boosman, Martijn P Bauer, Ilse MJ Kant, Simone A Cammel, and Ewout W Steyerberg. 2021. The digital scribe in clinical practice: a scoping review and research agenda. *NPJ digital medicine* 4, 1 (2021), 1–8. https://doi.org/10.1038/s41746-021-00432-5