

# A CONVERSATIONAL AI AGENT FOR INDUSTRIAL MAINTENANCE REPORTING

Industrial Design Engineering

Master's Thesis

Samuel Kernan Freire

Delft University of Technology  
Industrial design engineering  
MSc Integrated Product Design  
Graduation Thesis

July 2020

Supervising Team:

Chair: Assistant Professor Dr. Zoltán Rusák

Mentor: Assistant Professor Dr. Doris Aschenbrenner

Author:

Samuel Kernan Freire

4091787

## Acknowledgements

Firstly, I would like to thank my supervisors, Zoltán Rusák and Doris Aschenbrenner. Between our trips to Poland and Germany for DIAMOND, the COVID-19 restrictions and the explorative nature of my graduation, it has been a whirlwind from beginning to end but I could always count on Zoltán's guidance and support. He helped ensure my work stood on a solid foundation and encouraged me to "follow my heart". I would like to thank Doris for her unending enthusiasm, problem-solving orientation and encouragement to aim high.

Secondly, I would like to thank the DIAMOND partners and colleagues. With special mentions to Santiago Ruiz Arenas for our discussions, for offering a helping hand and contribution to the transcript coding and Stefan Wellsandt, the project manager, for his constructive feedback and openness for collaboration. I would also like to thank the DIAMOND representatives at Whirlpool and Stryker, Enrica Bosani and Ralf Schwanbeck.

Thirdly, I would like to thank my friends and family for their support throughout the graduation and facilitating some "time off". Special thanks to Margaret for proof-reading my thesis.

Last but not least, I would like to thank Kim, my fiancé, for being there for me through thick and thin and her eternal belief in me.

# 1 TABLE OF CONTENTS

2	Abstract.....	8
3	Introduction.....	9
3.1	Background.....	9
3.2	Objectives and high-level requirements .....	10
3.3	Stakeholders.....	11
3.4	Limitations and risks.....	11
3.5	Assumptions and exclusions.....	11
3.6	Research questions .....	12
3.7	Project outline.....	12
4	Research and analysis.....	14
4.1	Conversational AI agent.....	14
4.1.1	Types .....	15
4.1.2	Hardware.....	18
4.1.3	Market analysis.....	18
4.1.4	Competitors .....	20
4.1.5	System architecture .....	21
4.1.6	Current capabilities and challenges.....	22
4.1.7	Natural Conversation Framework.....	24
4.1.8	Standard development process.....	24
4.1.9	Interaction guidelines .....	25
4.2	Conversational agents in maintenance applications.....	26
4.2.1	Industrial maintenance .....	26
4.2.2	Industrial maintenance knowledge.....	27
4.2.3	Context analysis .....	28
4.2.4	Aircraft maintenance .....	33
4.2.5	Healthcare .....	34
4.2.6	Problems and potential use cases.....	35
4.3	Programme of requirements .....	37
4.3.1	Technology.....	37

4.3.2	User .....	37
4.3.3	Conversation design .....	37
4.3.4	Business .....	38
4.3.5	Knowledge creation.....	38
5	Development.....	39
5.1	Approach.....	39
5.2	Use case selection.....	40
5.2.1	Design and relevancy criteria .....	40
5.2.2	Selected use-case .....	40
5.3	Ideation.....	41
5.3.1	Agent persona.....	41
5.3.2	Utterance pair brainstorming.....	42
5.3.3	The current scenario .....	43
5.3.4	Envisioned scenario (1-3 years).....	44
5.4	Related work.....	46
5.4.1	Automatic medical scribes.....	47
5.4.2	Automatic meeting summarisation .....	47
5.4.3	Smart glasses for (maintenance) documentation .....	47
5.4.4	Augmented reality for (maintenance) documentation .....	48
5.5	Conversation analysis.....	48
5.5.1	Coding demonstration.....	49
5.5.2	General patterns .....	49
5.5.3	Internal state patterns.....	52
5.5.4	Relevance for the reporting use case .....	57
5.5.5	Evaluation of conversation analysis .....	57
5.6	Scientific experiment .....	58
5.6.1	Hypotheses.....	58
5.6.2	Method .....	59
5.6.3	Dependent variables.....	60
5.6.4	Independent variables.....	60
5.6.5	Participants.....	60
5.6.6	Results.....	60

5.6.7	Discussion .....	65
5.6.8	Qualitative analysis of the utterance types .....	65
5.6.9	Qualitative analysis regarding the development of the solution .....	66
5.6.10	Conclusions .....	67
5.6.11	Reflection on the experiment .....	68
5.6.12	Recommendations for future research .....	68
5.7	Solution requirements .....	68
5.7.1	Functional requirements .....	69
5.7.2	Usability requirements.....	69
5.7.3	Technical requirements .....	69
5.7.4	Scope of Development and testing.....	69
5.8	Prototyping .....	70
5.8.1	Key components .....	70
5.8.2	FormPolicy prototype .....	71
5.8.3	Free-form prototype .....	72
5.8.4	Final prototype .....	73
5.8.5	Test of the final prototype.....	76
5.9	Roadmap for future development .....	79
5.9.1	Performance (speed of the processing).....	79
5.9.2	Accuracy of intent classifier and entity extraction .....	79
5.9.3	Handling "real" speech.....	79
5.9.4	More advanced context tracking and information requests.....	80
6	Evaluation and conclusions .....	81
6.1	The Application.....	81
6.2	Scientific experiment .....	81
6.3	Prototype.....	81
6.4	Conversational Design method.....	82
6.5	Relevance beyond this project.....	82
6.6	Future work .....	82
7	Reflection .....	83
8	References .....	84
9	Appendices.....	90

9.1	Project brief.....	90
9.2	Use case selection.....	97
9.3	Utterance pair brainstorming .....	98
9.3.1	In scope and on topic.....	98
9.3.2	Out of scope and on topic.....	98
9.3.3	In scope and off topic.....	99
9.3.4	Out of scope and off topic .....	99
9.4	Experiment plan .....	100
9.4.1	Problem.....	100
9.4.2	Hypotheses.....	100
9.4.3	Variables .....	100
9.4.4	Test setup.....	101
9.4.5	Materials .....	102
9.4.6	Test plan.....	102
9.4.7	Corona measures.....	103
9.4.8	Analysis plan.....	103
9.4.9	NASA-TLX.....	104
9.4.10	Consent form.....	105
9.4.11	Preliminary questions .....	106
9.4.12	Maintenance report .....	107
9.4.13	scripts.....	108
9.4.14	Quality of reporting measure.....	110
9.4.15	Reference image for component names.....	112
9.5	DIAMOND survey for Whirlpool and Stryker .....	113
9.6	Evaluation of Mycroft .....	117
9.6.1	First impressions.....	117
9.6.2	Picroft.....	118
9.7	Rasa evaluation.....	119
9.8	Code .....	122
9.8.1	Dependencies.....	122
9.8.2	Classes for handling procedural statements and context .....	122
9.8.3	Classes for sending the report by email.....	129

9.8.4	Classes for status request feature.....	131
9.9	Experiment results .....	133
9.10	Prototype tests.....	135
9.10.1	Entity extraction .....	135
9.10.2	"IT" component Tracking and clarification .....	137
9.10.3	Intent classification.....	139





## 2 ABSTRACT

This project is an exploration of potential applications of Conversational AI Agents (CAIA) for industrial maintenance. Specifically, it involves the scientific validation and development of a CAIA for a promising application: the automatic creation of information-rich maintenance reports by conversing with a technician while they perform industrial maintenance. The choice for this application was based on literature research, (in-situ) context analyses and a review of CAIA frameworks and design guidelines. These revealed that maintenance workers rely heavily on their own experience and intuition when solving problems but mechanisms for capturing and accessing this were non-existent. This knowledge is highly valuable and can represent a significant part of a company's worth. Furthermore, maintenance technicians reported that the existing reporting mechanisms were a nuisance. Lastly, audits revealed that maintenance reports were frequently incomplete or of poor quality. In turn, CAIAs (Conversational AI Agents), have various affordances that make them well-suited to the context of industrial maintenance. They are (1) hands and gaze-free, (2) highly efficient (faster than writing or typing, facilitate multitasking and they provide faster access to specific information), (3) they can adapt to the skill level of the user and (4) impose a minimal cognitive load.

A between-subjects experiment with 24 participants, which involved changing a bicycle inner tube, was used to test three hypotheses regarding the potential value of the application. All three hypotheses compare using a CAIA for reporting whilst performing maintenance, as opposed to writing the reports on paper afterwards. They posited that using the CAIA would result in (1) reports of higher quality (more information relevant to the understanding of the task), (2) time saving, and (3) a lower perceived workload (NASA TLX). T-tests confirmed that all three hypotheses were true. These results indicate that using a CAIA for live-reporting has a clear value proposition for the industrial maintenance domain. Critically, it demonstrated that it could facilitate the capture of valuable "expert knowledge". Future research could explore integrating multi-modal information capture (e.g. through smart-glasses), identify additional uses for the captured data (e.g. for prescriptive maintenance or providing tips to maintenance technicians) and improve the functionality and usability of the existing application.

A prototype was built using the open-source frameworks, Rasa and Mycroft, to demonstrate the technical feasibility of the functional requirements. These requirements include (1) uttering "continuers" in response to the user describing their work, (2) tracking conversational context, (3) asking for clarifications when there is a lack of mutual understanding and (4) asking for status updates when the user is silent for more than a set amount of time. The main challenges for the future development of the prototype are (1) reducing the response time of the CAIA, (2) the accuracy of the intent classifier and entity extractor and (3) improving the handling of fragmented/lengthy user input. Some of the prototype's features rely on inflexible, hard-coded logic, therefore, future work should collect more conversational data and explore the use of machine learning algorithms.

### 3 INTRODUCTION

This is an explorative project regarding the potential applications of Conversational AI Agents (CAIAs) in industrial maintenance. Through literature research and a context analysis at EU factories (Section 4), it was concluded that the industrial maintenance reporting process has the potential to be significantly improved through the implementation of a CAIA. The proposed application is a CAIA that automatically generates an information-rich maintenance report by conversing with maintenance technicians while they perform their work. A scientific experiment (Section 5.6) showed that this way of reporting resulted in a higher quality report, saved time and reduced the perceived workload. The technical feasibility of the solution is demonstrated by a prototype (Section 5.8). Lastly, the generated knowledge is evaluated and a roadmap for future development is outlined (Sections 5.9 and 6).

#### 3.1 BACKGROUND

In 2019, there were 3.25 billion digital voice assistants in use worldwide and forecasts

expect this to reach 8 billion by 2023 (*Number of voice assistants in use worldwide 2019-2023*, 2019). Most of this growth is accounted for by consumer-oriented products like Amazon Alexa, Google Assistant and Baidu's Duer. However, the intense growth is fueling technological advancements that flow over to the industrial market, which is seeing renewed growth. Nowadays, much of the underlying technology is readily accessible (e.g. open-source speech-to-text services like Mozilla's Deep Speech and Baidu's Deep Speech 2, open-source development frameworks like Rasa and open-source voice assistants like Mycroft). Coupled with a better understanding of natural language and an immense amount of training data, the state-of-the-art offerings are becoming more intelligent and getting better at participating in natural conversation, otherwise known as Conversational AI Agents (CAIAs).

Lotterbach & Peissner (2005) concluded that voice-interfaces presented various benefits to industrial applications. These included: intuitive use, efficiency in inputting complex information, they are hands and eyes free, computer knowledge is not required, can



**Figure 1 Industry 4.0 visualisation** ("Hexagon Invests in New Smart Factory | Fabricating and Metalworking," n.d.)

adapt to user preference, and be used in adverse conditions. These same points are echoed by Moore & Aral (2019) and companies like Google (“Conversation design process—Is conversation the right fit?”, n.d.).

The factories of the future will be more efficient by being more adaptable and flexible. As a result, the required competencies of factory workers are expected to change significantly (Ansari, Erol, & Sihm, 2018), (Lorenz, Rießmann, Strack, Lueth, & Bolle, 2015). These researchers predict that the factory worker of the future will need to be more multi-disciplinary and more skilled in IT. Also, a dynamic factory environment can impose additional cognitive challenges. To deal with this, Industry 4.0 workers need systems that enhance their cognitive capabilities (e.g. perception, memory, reasoning, decision, motor response, etc.)(Romero, Bernus, Noran, Stahre, & Berglund, 2016). This could be achieved using an intelligent computer system, i.e. a digital assistant.

Communication between a human worker and a digital assistant requires a Human Machine Interface (HMI). Natural language interfaces have been around since 1966 (Weizenbaum, 1966) but a recent resurgence in interest and development has propelled the technology into the spotlight. New machine-learning (ML) algorithms, persistent internet connections and a better understanding of human conversation have enabled natural language interfaces that are significantly smarter than 20 years ago (Moore & Aral, 2019).

In this project, I explored the potential of CAIAs for assisting industrial maintenance personnel. I identified promising use cases based on literature research, an evaluation of state-of-the-art CAIA research and

development, in-situ context analyses and interviews with industry players. This project is partnered with DIAMOND (Digital Intelligent Assistant for Predictive Maintenance as a Response to Demanding Employee Skill Requirements), a European Institute of Innovation and Technology (EIT) project with two corporate partners, Whirlpool and Stryker who provided access to their facilities and personnel for research. For the research, I chose a specific use case: automatic creation of corrective maintenance reports. These were based on several factors including the user and client needs, feasibility, the affordances of CAIAs and potential for generating new scientific knowledge. To validate the solution and collect knowledge for further development an experiment was executed whereby the CAIA was simulated by a human. Finally, a prototype was created as a proof-of-concept for technical feasibility. The solution development was based on the Natural Conversation Framework (Moore & Aral, 2019), human-AI guidelines (e.g. Amershi et al., 2019) and conversational agents from other sources, such as Rasa and Google. The approach is described in more detail in the project outline below (Section 3.6.1).

### 3.2 OBJECTIVES AND HIGH-LEVEL REQUIREMENTS

The primary objectives of this project are to explore the potential for CAIAs in modern industrial maintenance (Industry 4.0) and develop a solution for a specific problem. This translates into multiple high-level requirements: (1) the solution should solve key challenges that the industry is facing, (2) the solution should leverage core affordances of CAIAs, (3) the solution fits the needs and interests of the end-users, and (4) the solution should be technically feasible.





**Figure 2 Whirlpool factory floor in Łódź, Poland** (“#PlacesthatMatter: Łódź, Poland. | by Whirlpool EMEA | Whirlpool Corporation | Medium,” n.d.)

### 3.3 STAKEHOLDERS

This project is partnered with an EU research project, DIAMOND, that aims to develop a Digital Intelligent Assistant (DIA) that assists maintenance personnel in the use of predictive maintenance systems. DIAMOND has two corporate partners, Whirlpool (Section 4.2.3.2) and Stryker (Section 4.2.3.1) and two academic partners, Technical University of Delft (TU Delft) and BIBA, a research institute affiliated with the University of Bremen. BIBA is leading the project and is primarily interested in learning about the technical process of developing and implementing a DIA for industrial maintenance. They plan on implementing a prototype and integrating it with real data sources from the corporate partners.

The key stakeholders for this project are the end-users: maintenance technicians in the manufacturing industry, and the DIAMOND partners.

### 3.4 LIMITATIONS AND RISKS

The planned user testing coincided with the COVID-19 pandemic. Due to the necessary social distancing and travel restrictions, also affecting the DIAMOND partners, the frequency of communication was reduced and the possibility of a second visit or user testing impossible. Instead, an experiment was devised that could be performed locally by regular citizens, performing maintenance on a bicycle (Section 5.6).

### 3.5 ASSUMPTIONS AND EXCLUSIONS

Assumptions were made regarding the technical feasibility of the solution and its potential value to the client. Some of these factors were tested during the user experiments and others will need to be tested in future work.

The following aspects are not within the scope of my graduation project: the integration of the prototype to client systems, the embodiment

of the hardware, and support for other languages besides English.

### 3.6 RESEARCH QUESTIONS

The high-level research question defined below was the starting point for this project.

*What are the potential use cases for conversational AI agents in Industry 4.0?*

The mid-level research questions presented below, aim to uncover information relevant to the development of a prototype for this use case.

- 1) *How are the different utterance types (procedural, internal state, referential, task status, inquiries) related to each other in conversations on machine maintenance?*
- 2) *What types of conversational sequences and expansions are characteristic for a conversation on machine maintenance?*
- 3) *How should state-of-the-art conversational UX design processes be adapted for the Industry 4.0 context?*

The goal of the low-level scientific research questions shown below is to validate hypotheses regarding the value of the proposed application (these are discussed and tested in Section 5.6)

- 1) *Do users provide higher quality reporting (more relevant details) through a dialogue with a Conversational AI Agent while they work, as opposed to writing reports retroactively?*
- 2) *Do users perceive a lower workload (NASA-TLX score) of the entire task (repair and reporting) when using a Conversational AI Agent for reporting during the task as opposed to writing reports afterwards?*

- 3) *Do users take less time to complete their task (repair and reporting) when using a Conversational AI Agent to report as opposed to writing reports afterwards?*

### 3.7 PROJECT OUTLINE

To make effective decisions throughout the project, I familiarised myself with various relevant domains. The primary domains are maintenance within Industry 4.0 (Section 4.2) and CAIAs (Section 4.1). However, I also researched adjacent domains like the medical and aircraft industry where digitisation and the use of CAIAs is also growing. I performed a literature review, state-of-the-art analysis, and context analysis. This included research trips to a factory in Germany and one in Poland and multiple interviews with Industry 4.0 related stakeholders. I selected a promising use case to develop a case study (Section 5.2). The use case, automatic reporting of corrective maintenance work, was selected based on its suitability for CAIAs, the user and industry needs, its potential to provide new insights and technical feasibility.

The development process is based on a design-thinking method for CAIAs outlined by Moore & Aral (2019) in their book on Conversation UX Design (Section 4.1.7). Previous research (Lotterbach & Peissner, 2005) has concluded that it is best to use a user-centred approach for voice user-interfaces in industrial applications. I chose to use their Natural Conversation Framework as a basis for my development as it strives to create conversational agents that can hold natural conversations, as opposed to simple question-answers bots that are the mainstay of current voice assistants like Alexa, Google Assistant and Siri. A key component of the process involves using conversation analysis

to identify patterns in the language used by the end-users. They argue that it is vital that designers for conversational agents command an understanding of natural language mechanics to design effective conversational agents. This is reflected in other recent research on conversation design. As part of this process, I performed a conversation analysis (Section 5.5) between two maintenance personnel in the context of machine repair. This served multiple purposes: to familiarise me with the general conversation between mechanics and the terminology used in the domain, and to answer the two mid-level research questions. Not every aspect of Natural Conversation Framework was equally relevant for this project and throughout the process, I modified and added steps to adapt to the domain and new information.

I tested three hypotheses related to improving reporting quality, capturing worker knowledge, saving time, and reducing the workload on the workers (Section 5.6). As such, the tests serve to validate the value of the use case and generate scientific knowledge that can be used by other researchers and product developers. Due to the COVID-19 restrictions, a maintenance task was chosen that many Dutch citizens were familiar with, changing a bicycle inner tube (see Figure 3 below).

A series of final prototypes were developed to demonstrate the technical feasibility of the solution featured in the experiment (Section 5.8). These were tested using excerpts from the scientific experiment.

Lastly, I concluded and reflected on my approach results and possibilities for future research and development (Sections 6 and 7).



**Figure 3** Test participant inserting the spare inner tube



## 4 RESEARCH AND ANALYSIS

### 4.1 CONVERSATIONAL AI AGENT

Conversational AI agents are a type of virtual assistants that use a natural dialogue system to improve usability and user satisfaction. Fifteen years ago, text-based virtual assistants, then referred to as “chatterbots” featured simple, un-human-like, rule-based dialogue systems (L’Abbate, Thiel, & Kamps, 2005). The same researchers proposed that virtual assistants could be made more human-like by being more proactive. Proactive virtual assistants have been widely researched (Falk, Poulakos, Kapadia, & Sumner, 2018; L’Abbate & Thiel, 2003; Ring, Barry, Totzke, & Bickmore, 2013; W. Wu et al., 2020). Proactivity is one of the many features of natural dialogue that conversational agents aim to emulate. The other features include (Berg, 2015):

- Adaptivity of the system
- Mixed initiative
- Support of implicit confirmation
- Usage of verification questions
- Possibilities to correct information that have already been given
- Over-informativeness (give more information than has been asked for)
- Support negations
- Understand references by analysing discourse and anaphora
- Natural language generation to prevent monotonous and recurring prompts
- Adaptive and situation-aware formulation
- Social behaviour (greetings, same level of formality as the user, politeness)
- Quality of speech recognition and synthesis.

In recent years, deep learning has enabled significant advancements in natural dialogue

systems (Chen, Liu, Yin, & Tang, 2017). It has reduced the reliance on manually defining conversation sequences by leveraging vast amounts of training data. Additional topics of research related to neural networks and conversational agents include making the responses of a voice-assistant less generic (Du & Black, 2019; Gao et al., 2019), improving longer duration conversation performance (Sankar, Subramanian, Pal, Chandar, & Bengio, 2019) and using multi-domain transferable context trackers (C.-S. Wu et al., 2019).

Hirschberg & Manning (2015) have noted four things that have enabled recent advances in Natural Language Processing (NLP)(Figure 4):

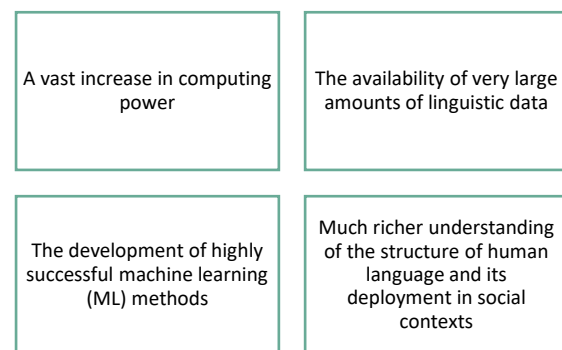


Figure 4 Enablers in Natural Language Processing

Applications of conversational AI agents are widespread and highly variable. As of 2019, 3.25 billion digital voice assistants are being used around the world (*Number of voice assistants in use worldwide 2019-2023*, 2019). Most people will be familiar with consumer-oriented smart assistants like Alexa, Google Assistant and Siri. These products have exploded in popularity in recent years and are integrated with smartphones, speakers, cars and various wearables like watches, earbuds, glasses and even rings. Alexa alone was supported by 60,000 smart home devices in 2019 (*Number of voice assistants in use worldwide 2019-2023*, 2019).

These assistants are used for productivity tasks, such as calling someone or setting an alarm, for entertainment purposes such as playing music, and for information purposes, like asking general questions or getting directions.

Some developers like Weidauer (2018), who is part of Rasa Technologies Inc, defined a roadmap of AI Assistants with five levels of intelligence. According to him, the industry is currently at level 3 of 5 “Contextual Agents”. Tracking context and user intent has been an important feature that has enabled assistants to move from simple queries like “what’s the weather today?” to more complex tasks like “Help me book a flight for my holiday” (Sun et al., 2016).

There is a lack of consensus on what a Conversational AI Agent entails and many terms are used interchangeably, e.g. conversational agents, dialogue systems, voice-user-interfaces, embodied conversational agents, smart conversational interfaces, or chatbots (Laranjo et al., 2018). However, there are some more definable aspects of CAIAs, as outlined below (Section 4.1.1).

---

#### 4.1.1 TYPES

Conversational agents have various differentiating characteristics: dialogue management, dialogue initiative, input modality, output modality and goal-orientation (Laranjo et al., 2018). See an overview in Table 1 below.

Goal-oriented agents are closed-domain, i.e. they can assist in the completion of a set of tasks. By limiting the topics of conversation, it is easier to emulate natural conversation. Goal-oriented agents typically follow a pipeline approach, i.e. the steps of natural language understanding and response selection are broken down into a sequence that best suits the task and domain. The disadvantage of this type of system is that it is a lot of work to transfer the agent to another task or domain and it is difficult to implement user feedback (Chen et al., 2017). One solution to this is to use newly developed end-to-end neural generative models. These models represent a way to implement open-domain agents. Open-domain agents attempt to address the human need for communication, affection, and social belonging (Huang, Zhu, & Gao, 2020).



Characteristic	Types	Description
Dialogue management	Finite-state	The user is let through a series of pre-determined steps.
	Frame-based	The system asks questions to elicit entities to perform a task. The dialogue is not pre-determined but depends on the user inputs and required information.
	Agent-based	Natural, dynamic dialogue that builds on preceding context and the actions and beliefs of both parties. Many variations exist, depending on the embedded intelligence of the agent.
Dialogue initiative	User	The user leads the conversation.
	System	The system leads the conversation.
	Mixed	Both the user and the system can lead the conversation.
Input modality	Spoken	The user uses spoken language to interact with the system.
	Written	The user uses written language to interact with the system.
Output modality	Spoken	The system uses spoken language to interact with the user.
	Written	The system uses written language to interact with the user.
	Visual	The system uses non-verbal communication like images, graphics, facial expression or body movements.
	Mixed	A combination of the above.
Goal-orientation	Goal-oriented	Goal-oriented (or task-oriented, closed-domain) CAIAs assist users in completing a specific task (e.g. retrieving information, booking restaurants, providing instructions etc).
	Non-goal-oriented	Non-goal-oriented (open-domain) agents interact with humans to provide reasonable responses and entertainment (Chen et al., 2017).

**Table 1 Types of conversational agents**



Figure 5 Smartphone ("Useful artificial intelligence apps for your mobile phones," n.d.)



Figure 6 Automotive assistants ("Automotive AI Assistants : BMW AI assistant," n.d.)



Figure 7 Smart glasses ("Explore Focals - North," n.d.)

---

## 4.1.2 HARDWARE

Virtual assistants have come to prominence in recent years through their mass adoption by consumers in their homes, on smart speakers, and smartphones.

Most of voice assistants systems are consumer products and come pre-installed on smartphones(*Voice Assistant Consumer Adoption Report*, 2018)(Figure 5). In recent years, the smart speaker market has been rapidly growing (USD 1.57 billion in 2017 to USD 11.79 billion by 2023 (“Smart Speaker Market by IVA, Component,” 2018). Smart watch usage has been doubling every year for the past four years and “hearables” are gaining attraction with the advancement and acceptance of smart assistants (“Smart Wearable Market | Growth, Trends, Forecast (2020-2025),” n.d.). In both cases, the healthcare segment makes up a significant portion of new sales; however, despite the industrial market being non-existent in 2018, it is expected to equal the healthcare market by 2026 (“Hearables Market Size, Share & Growth | Trends & Analysis - 2026,” 2019).

For consumers, the voice assistants tend to be cloud-based and use a smartphone as an intermediary relay. The hardware market for smart assistants consist of:

- Smart speakers (e.g. Echo (Amazon), Google Home, Mycroft V2)
- Wearables
  - Smartwatches (e.g. Samsung Galaxy Watch, Apple watch)
  - Hearables (e.g. Apple Airpods, Google Pixel Buds, Echo buds (Amazon), Microsoft Earbuds, Nuheara)
  - Glasses (e.g. North focal glasses, Echo frames (Amazon), Glass (Google)(Figure 7)

- Rings (e.g. Echo ring)
- Smartphone
- Automobile-integrated (Figure 6)

For enterprises, the system tends to be run by local servers so that the company’s data is well protected. The following hardware solutions are common:

- Hearables (e.g. wireless headsets) connected to a smartphone or other mobile computer
- Glasses (e.g. Glass Enterprise Edition (Google))

State-of-the-art “hearables” features include:

- Smart audio filtering (“Smart Earbuds for Hearing Enhancement | Nuheara,” n.d.)
  - Focusing the sound based on the user’s gaze
  - Focusing on nearby human conversation and filtering out background noise (“AirPods Pro Active Noise Cancellation and Transparency mode - Apple Support,” n.d.)
- Optimisation for an individual’s hearing response
- Audio augmented reality: cue audio based on the user’s location and gaze (“Wireless & Bluetooth Headphones | Bose,” n.d.)
- Biometric tracking
  - Heartrate
  - Blood pressure (“Valencell - The most accurate PPG Sensors for wearable technology,” n.d.)

---

## 4.1.3 MARKET ANALYSIS

---

### 4.1.3.1 GENERAL PURPOSE

The digital voice assistant market was valued at USD 2.4 billion in 2018 and forecasted to reach USD 8.9 billion by 2023. Most market

reports focus on general-purpose personal assistants, such as Apple's Siri, Amazon's Alexa, Google's Google Assistant, Samsung's Bixby and Mycroft, an open-source platform. If a third-party wants to add domain-specific services, they must develop their own "skill". Companies like Voiceflow offer services to fulfil this need. They were seed-funded by Amazon, who set aside USD 200 million for startups that support the Alexa infrastructure ("Alexa Fund," n.d.).

Alexa is HIPAA (Health Insurance Portability and Accountability Act) compliant, which makes it attractive to American medical providers as they can offer services that contain private medical information ("Amazon Announces 6 New HIPAA Compliant Alexa Skills," n.d.).

The big players like Amazon and Google can make money indirectly through the sale of additional services, marketing, collecting user data and by charging for their web services (e.g. Amazon web services) ("How Alexa, Siri, and Google Assistant Will Make Money Off You | MIT Technology Review," n.d.). Mycroft has managed to keep their costs relatively low by building on existing frameworks, like Mozilla's speech-to-text service, which in turn uses Baidu's deep speech architecture and Google's TensorFlow framework ("Initiatives - Mycroft," n.d.).

---

#### 4.1.3.2 TAILORED

Some assistants specifically focus on business customers. Companies like SoundHound and Nuance offer tailored solutions depending on customer needs. SoundHound's Houndify platform promises the ability to "Add voice AI to anything". Their customers include various major automotive manufacturers, online music platforms and

cloud health services. They boast 125+ domains of understanding and customize development for wake-words and sound processing, such as noise reduction.

---

#### 4.1.3.3 DOMAIN-SPECIFIC

As the market grows, so too do domain-specific applications. At present, the two biggest industries are healthcare (e.g. Pria and Lifepod, among others) and automotive (e.g. Cerence, among others). In 2019, Grand View Research noted that "[...] automation technologies such as Robotic Process Automation (RPA) and AI are proliferating in the healthcare sector. Thus, intelligent virtual assistants are expected to witness high demand over the next five years ("Intelligent Virtual Assistant Market Size | Industry Report, 2027," n.d.).

As doctors have to deal with a large amount of paperwork, a domain-specific voice assistant capable of capturing the relevant information could save a lot of time ("Does Your Doctor Need a Voice Assistant? | WIRED," n.d.). This is just one of the many parallels these industries share with the manufacturing industry. Other similarities include the necessity to monitor processes, health and trends, to create reports and diagnose problems.

---

#### 4.1.3.4 CONCLUSIONS

As a new player in the industry, DIAMOND cannot compete with companies that have experience of over 125+ domains like Houndify, or rely on the indirect monetization of existing products like Google or Amazon. However, as demonstrated by the domain-specific players in the healthcare and automotive industry, it is possible to operate in a niche.

---

## 4.1.4 COMPETITORS

---

### 4.1.4.1 SPIX BY SIMSOFT INDUSTRY

Released in 2019, Simsoft Industry has launched a DIA based products for Industry 4.0 applications focused primarily on the French market. Simsoft Industry offers four product packages to their client:

- VOGOF: Voice guide operations in the field for technicians
- eXtend: Smart voice assistant skills for business applications
- K-ontact: Transmission of critical technical knowledge
- SmartKit: Experiment with a smart voice assistant for technicians

Key features of their products include: the ability to work offline, machine data retrieval, problem-solving, task guidance, tutorials, instruction difficulty adapted to the user, integration into existing management tools (ERP, MES, CMMS, FMS, SCADA), local servers,

the use of images and videos, a windows tablet and Bluetooth headset (see Figure 8) and proactive assistance. As all information available on SPIX was published by Simsoft Industry, it is difficult to determine how mature the product is.

Based on information on their site and other publications, we can infer that they are currently running pilots at client facilities. They state that 200 technicians at 15 different facilities have been involved in the development of their VOGOF system. They have clients in the automotive, construction, energy, aerospace and manufacturing industries.

---

### 4.1.4.2 ATHENA BY ITSPEEX

Athena is an intelligent assistant developed specifically for machine operators. Once integrated with the machine's computer, operators can use their voice to issue commands to the machine. Athena runs locally on a server.



**Figure 8 Maintenance technician wearing an Athena headset** ("Athena, Transform Manufacturing: Voice-Activated Machining Technology in Multiple Demos : Modern Machine Shop," n.d.)



---

#### 4.1.4.3 LEX<sup>x</sup>

Lex<sup>x</sup> is an intelligent assistant purpose-built for maintenance technicians in the aircraft, logistics and utility industry. It offers troubleshooting support during problem diagnosis by learning from existing technical documentation as well as from reports. Furthermore, it allows technicians to access and create reports with their voice, thereby enabling the capture and retention of knowledge. At the time of writing, no publicly available demonstration of Lex<sup>x</sup> could be found.

---

#### 4.1.4.4 CONCLUSIONS

SPIX by Simsoft Industry is marketed as a jack-of-all-trades for Industry 4.0 applications. This allows them to try out lots of things and discover the most lucrative niches. Conversely, Athena and Lex<sup>x</sup> are already focused on smaller subdomains. Their existence provides some evidence that there is a market for CAIAs in industrial maintenance. However, I do not have any insight into their financial or developmental situation.

---

#### 4.1.5 SYSTEM ARCHITECTURE

The architecture discussed in the following section is that of a goal-oriented, closed-domain agent. This represents the state-of-

the-art of conversational AI agents in widespread use.

---

#### 4.1.5.1 NATURAL LANGUAGE UNDERSTANDING (NLU)

An NLU pipeline for goal-oriented agents typically contains three main components: tokenisers, featurisers and entity/intent recognition. It resides between the incoming user utterance and the dialogue manager (Figure 9 Basic CAIA components).

*Tokenisers* convert text input into tokens using whitespaces as separators.

*Featurisers* create features for each token and return a matrix (number of tokens x number of features). For English, Rasa supports using pre-trained word embeddings, i.e. it recognises that 'dog' is a similar word to 'cat' and creates a contextual vector representation of each sentence as opposed to treating words individually. Using non-pre-trained word embeddings (e.g. for non-English use-cases) requires more training data.

*Entity recognition* is when pieces of information are extracted from a sentence, e.g. to extract the colour from the following sentence: "the car was blue". *Intent classification* is when the intention of a user's statement is predicted, e.g. when the user says, "tell me about the weather", the intent is "weather request".

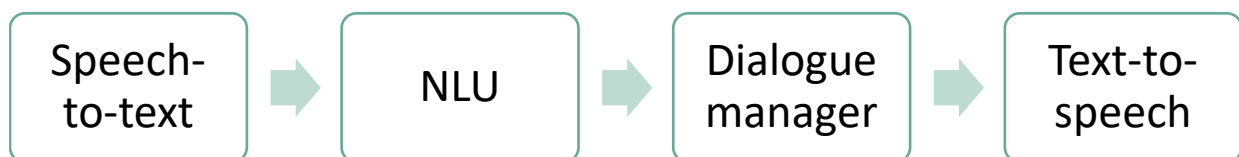


Figure 9 Basic CAIA components

Additional (custom) components can be added to the pipeline, such as sentiment analysis. State-of-the-art NLU, like Rasa, relies heavily on being able to classify intents of the user. For basic, narrow scenarios in which the user requests information and the agent responds, an intent-based system can suffice. However, as conversations get longer, less predictable, and more natural, the intent-based system can become a hindrance (Nichol, 2019). In natural conversation, people frequently have multiple intents in their statements: e.g. “Okay and add pears to my shopping list”. This utterance contains the affirmative intent, “okay”, and the intent to add an item to their shopping list. One way to deal with this is by using multi-intent classifiers.

---

#### 4.1.5.2 DIALOGUE MANAGER

The dialogue manager determines how the assistant should respond. Depending on how the agent is implemented, it can take numerous factors into account: intents, entities, user sentiment, context, the task at hand and any other logic implemented by the developer.

---

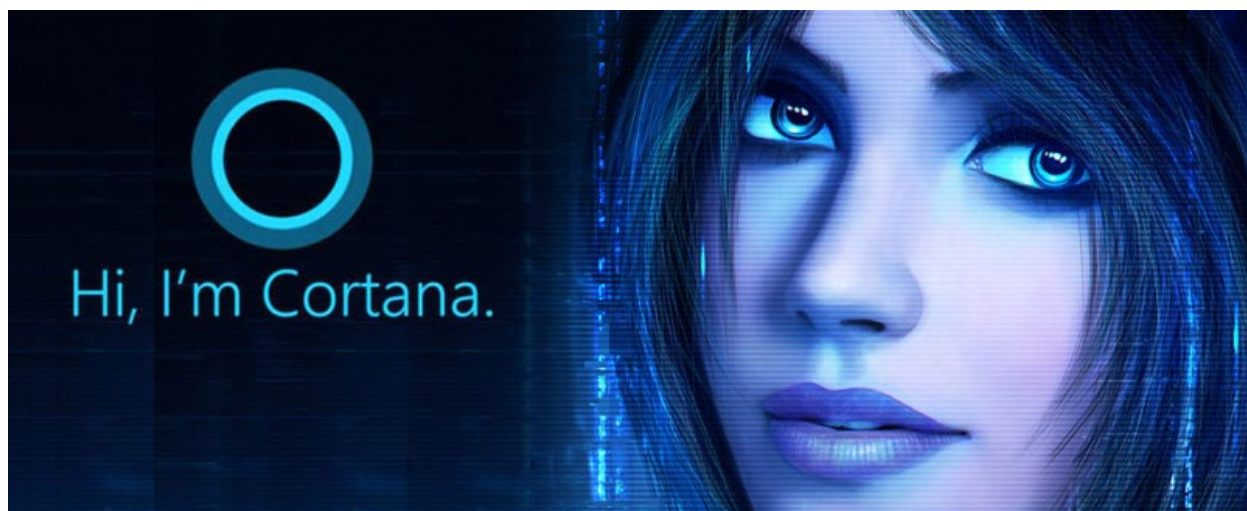
#### 4.1.6 CURRENT CAPABILITIES AND CHALLENGES

Porcheron, Fischer, Reeves, & Sharples (2018) have noted that user interactions with so-called conversational agents like Alexa, cannot be considered analogous to human conversation. They point out that interactions are preconfigured by adjacency pairs, whereas natural conversation is often unstructured. One way to achieve truer natural conversation is by employing an open-domain conversation agent. However, further research and development are needed to improve semantics, consistency, and interactiveness (Huang et al., 2020). In the meantime, in including natural conversation in this project, it will be necessary to (partially) rely on preprogrammed sequences.

---

##### 4.1.6.1 GENDER BIAS

A recent UNESCO report (West, Kraut, & Chew, 2019) has pointed out clear gender bias in voice assistants, namely that they are almost all (young) females (see Figure 10 below). Whereas earlier research (Mitchell, Ho, Patel, & MacDorman, 2011) has shown that users have a preference for female voice assistants, many



**Figure 10 Microsoft Cortana's face** (“How to activate ‘Hey, Cortana!’ in Windows 10 Laptop / PC,” n.d.)

others have raised concerns about the potential negative implications of this bias (Feine, Gnewuch, Morana, & Maedche, 2020; Fessler, 2017; West et al., 2019), specifically, many of the assistants are presented as young, servile woman and may contribute to the tolerance of sexual harassment and verbal abuse (West et al., 2019). A similar discussion is also being held regarding robots (Alesich & Rigby, 2017).

The UNESCO report (West et al., 2019) provides a list of recommendations on four themes to help address the bias and associated problems: (1) document and build evidence, (2) Create new tools, rules and processes, (3) Apply gender-responsive approaches to digital skills development and (4) Ensure oversight and incentives. Beyond calling for further research, awareness, better digital education and more gender-balanced development teams, a few of their recommendations are implemented in the prototype developed for this project, namely (Figure 11):

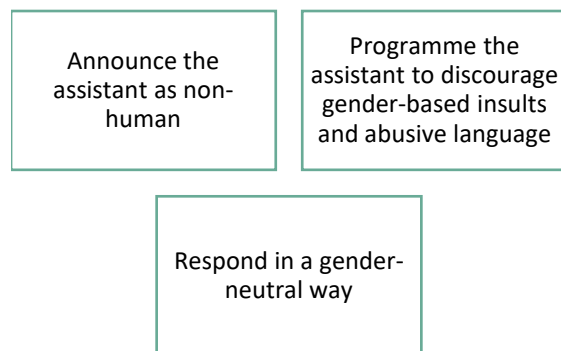


Figure 11 Anti-gender-bias features

#### 4.1.6.2 USER ACCEPTANCE

Early research on the acceptance of new technologies within the field of AI and IOT showed that data security, compatibility with other devices and the relationship to the

device have a major impact, whereas performance, hedonistic motivation and price were less significant (Kessler & Martin, 2017). Other research on the anthropomorphism of VAs showed that overall quality and ease-of-use were far more important factors (Kääriä, 2017). However, this may be because VAs are still unreliable and basic functionality must come first (Burbach et al., 2019). The same researchers also concluded that privacy was the most important factor for user acceptance, not price. Other research has shown that the perceived privacy risk is an important factor for how males interact with voice-user-interfaces, whereas trust and mobile self-efficacy are more important factors for females (Nguyen, Ta, & Prybutok, 2019).

For in-home VAs, three main motivating factors were identified: utilitarian benefits, symbolic benefits and social benefits provided by voice assistants. Like other studies, they also noted that privacy issues had a moderating affect (McLean & Osei-Frimpong, 2019).

A recent study outlined four guidelines for improving VA usage in the context of smart homes:

- 1) Authentication and authorisation: from a privacy and security perspective it would be useful if the VA could identify the user.
- 2) Activity-based interaction: users need a more natural, efficient, customizable, adaptable interaction experience.
- 3) Situated dialogue: the VA should remember dialogue history and other information sources to understand the context of the user.
- 4) Explainability and transparency: users would like more information on the



system's capabilities and available features.

---

#### 4.1.7 NATURAL CONVERSATION FRAMEWORK

The Natural Conversation Framework (Moore & Aral, 2019) consists of four parts as depicted below (Figure 12).

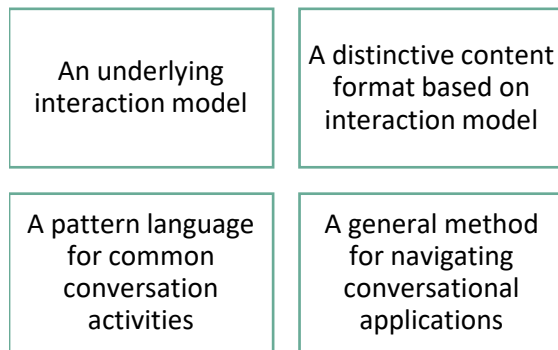


Figure 12 Natural Conversation Framework parts

During the conversational analysis process later in this report (Section 5.5), I attempt to identify how the above components are represented in the context of Industry 4.0.

---

#### 4.1.8 STANDARD DEVELOPMENT PROCESS

In academia and industry alike, a user-centred design approach is standard (Ghosh & Pherwani, 2015; "How do I get started? - Conversation design process - Conversation design," n.d.; Moore & Aral, 2019). These are all inspired by the user-centred design method defined by Norman (1988) and the five-stage design thinking model by Plattner (2016).

I will outline the process as defined by Moore & Aral (2019) as it is representative of other state-of-the-art conversational design methods.

- 1) Empathise: Developing an understanding of the users' needs, motivations and behaviours. Ideally, this involves observing and engaging with users. It is good practice to record conversations with the user for later consultation and consult existing conversation analysis literature on similar scenarios.
- 2) Define: Use the results of the emphasize stage to define a user persona and outcome statements. Choose a suitable use case based on the users' needs and fit conversation design.
- 3) Ideation: Three conversational UX specific exercises are suggested for ideation
  - a) The shape of the conversation: using conversation analysis to uncover structure and patterns in the conversations.
  - b) Create an agent persona
  - c) Utterance pair brainstorming: synthesizing the phrases and answers that the agent should be able to handle. Once collected, they can be sorted and prioritized depending if they are in-scope and/or on-topic.
- 4) Prototype: Use mock-ups and pseudo-code to evaluate early designs before building the functional prototype.
- 5) Test: Iterate frequently through the process of mock-up, design, build and test. Conversation UX design requires extensive beta testing. Early rounds of testing will likely be focused on collecting user and agent actions hadn't been identified up to this point. Next, the focus will be on fine-tuning the agent by adding new intent variations and synonyms of entities. Being flexible, reflective and opportunistic is important.

Besides the five phases, it is recommended to familiarise oneself with conversation design

and analytics (“How do I get started? - Conversation design process - Conversation design,” n.d.; Moore & Aral, 2019). (Moore & Aral, 2019)

This method is implemented in Section 5.3 and evaluated in Section 6.4.

---

#### 4.1.9 INTERACTION GUIDELINES

---

##### 4.1.9.2 COGNITIVE LOAD

Research has shown that when driving a car, the cognitive load of talking to a conversational AI Agent is comparable to talking hands-free on a mobile phone (Large,

---

##### 4.1.9.1 USABILITY

The guidelines outlined below (Table 2) were extracted from various papers related to human interaction with conversational agents, AIs or voice-user-interfaces (VUIs). I will use these guidelines in conjunction with those defined in the Natural Conversation Framework to guide the development process.

Burnett, Anyasodo, & Skrypchuk, 2016). A lot of research has been performed on phoning whilst driving a car (e.g. Liu & Ou, 2011; Nunes & Recarte, 2002) and the consensus is that the effect on driving is highly dependent on the cognitive load of the conversation, i.e. the

Source	Guidelines for the CAIA
Ghosh & Pherwani (2015)	<ul style="list-style-type: none"> <li>• Moderate level of expression</li> <li>• Moderately reactive to user’s mood space</li> <li>• Proactive: interrupt when a matter is more urgent than the current task</li> <li>• Avoid all external updates until requested (e.g. between tasks)</li> <li>• Personality should mirror the behaviour of human colleagues</li> <li>• Do not be over sympathetic</li> <li>• Inventiveness is desirable</li> <li>• Informal language is preferred</li> <li>• Pauses are expected when giving suggestions, or before enquiries</li> <li>• Speech should be faster for common statements</li> <li>• Interjections should precede new contexts, following a user’s inquiry and spoken quickly</li> </ul>
Porcheron, Fischer, Reeves, & Sharples (2018)	<ul style="list-style-type: none"> <li>• The CAIA should consider the social environment when inquiring something of the user to avoid awkward, strange or improper situations</li> <li>• The CAIA should be more transparent regarding misunderstandings and its intent</li> </ul>
Amershi et al. (2019)	<ul style="list-style-type: none"> <li>• Support efficient correction: Make it easy to edit, refine, or recover when the AI system is wrong</li> <li>• Support efficient invocation: Make it easy to invoke or request the AI system’s services when needed</li> <li>• Match relevant social norms: Ensure the experience is delivered in a way that users would expect, given their social and cultural context</li> <li>• Time services based on context: Time when to act or interrupt based on the user’s current task and environment</li> </ul>

Table 2 Design guidelines for CAIAs and AIs

content. However, natural chit-chat does not appear to have a significant effect on drivers' ability to drive safely. This insight might indicate that the cognitive load of CAIAs can be minimised by conversing in a natural, casual manner.

## 4.2 CONVERSATIONAL AGENTS IN MAINTENANCE APPLICATIONS

In the 21<sup>st</sup> Century, various technologies such as industrial internet of things (IIOT) (Xu, He, & Li, 2014), artificial intelligence (AI) (Dopico et al., 2016), big data analytics (Lee, Lapira, Bagheri, & Kao, 2013), cloud computing (Zhang et al., 2014), cyber-physical-systems (CPS) (Leitão et al., 2016) and new ways of human-machine-interaction (HCI) (Krupitzer et al., 2020) have reached a maturity level that allows factories, humans and machines to reduce the cost of communication, and increase flexibility for manufacturing, mass customization capabilities, production speed and quality (Ansari et al., 2018). This development is commonly referred to as Industry 4.0, a term coined in Germany and commonly used throughout Europe. It is

synonymous with Smart Manufacturing in the USA and South Korea (Kang et al., 2016).

### 4.2.1 INDUSTRIAL MAINTENANCE

Industrial maintenance, otherwise known as plant maintenance, refers to the pursuit of maximising machine reliability and uptime in a manufacturing environment.

Reason (1997) defines maintenance activities as:

- 1) Unscheduled operations, including corrective maintenance, and disturbance- and failure preventive operations (opportunity-based maintenance)
- 2) Scheduled disturbance- and failure-preventive operations
- 3) Inspections
- 4) Calibration and testing

There are four types of preventative (scheduled) maintenance strategies implemented in industry: preventative, condition-based, predictive and prescriptive maintenance.



**Figure 13 A Maintenance technician inspecting a machine** ("Highest Paying Diesel Mechanic Jobs - Diesel Mechanic Guide," n.d.)

---

#### 4.2.1.1 PREVENTATIVE

The most basic form of preventative maintenance involves following a use-based or time-based schedule. A common example is car maintenance. The car manual might specify that oil must be changed every 20,000 km or once a year. The goal is that the oil is changed before it deteriorates to the point that it may cause a defect. The main problem with this strategy is that it doesn't account for non-typical use scenarios or external factors. For example, someone who drives long distances in a warm climate or has a small oil-leak may need to change the oil more frequently. Conversely, someone who is very easy on their car would be fine changing the oil every three years. In the first example, the car may unexpectedly break down, whereas in the second case, the owner is unnecessarily spending time and money on replacing the oil.

---

#### 4.2.1.2 CONDITION-BASED

Condition-based maintenance involves identifying parameters that indicate machine health, such as temperature, and setting a threshold to trigger maintenance work. To continue with the car example, a car engine is equipped with various temperature sensors and if one of them reaches a certain threshold, the engine warning light on the dashboard will light up. This warning indicates that the driver should stop the car and seek maintenance help. For manufacturing machines, relevant parameters include vibration, pressure, frequency or fluid levels.

---

#### 4.2.1.3 PREDICTIVE

Predictive maintenance involves identifying patterns in sensor data that might indicate a problem. Because these patterns can be complex, involving the interplay of numerous sensors over time, a machine is more suited to

the task than a human. Machines can be taught to identify faults in the machine through a process called machine learning (ML). An ML algorithm is presented with data labelled as healthy and unhealthy data. It uses this data to learn patterns such that it can identify a problem when it arises, this intelligence is captured in a mathematical model. Compared to condition-based maintenance, preventative maintenance does not require that a human analyse and set thresholds for each parameter, as the algorithm does the work. However, the process of setting up the sensors, digital infrastructure and analysis can be difficult and expensive.

---

#### 4.2.1.4 PRESCRIPTIVE

Prescriptive maintenance takes the predictive maintenance a step further and prescribes which maintenance tasks should be performed to alleviate the problem. In the case of a car, it might detect that a specific sparkplug is not functioning properly due to the characteristics of the measured vibrations and suggest that the mechanic replace it. This has the potential to greatly reduce time spent diagnosing problems, however, it puts a lot of responsibility on the AI to make the correct decision.

---

### 4.2.2 INDUSTRIAL MAINTENANCE KNOWLEDGE

As mentioned in the report introduction, the factories of the future will be more efficient by being more adaptable and flexible. As a result, the required competencies of factory workers are expected to change significantly (Ansari et al., 2018), (Lorenz et al., 2015). These researchers predict that the factory worker of the future will need to be more multi-disciplinary and more skilled in IT. Also, working in a dynamic factory results in

increased cognitive load. To deal with this, Industry 4.0 workers need systems that enhance their cognitive capabilities (e.g. perception, memory, reasoning, decision, motor response, etc.)(Romero et al., 2016). This could be achieved using an intelligent computer system, i.e. a digital assistant.

Human intellectual capital represents up to 80% of the total resources in a modern organisation (Kans, 2019). There are two types of knowledge, explicit and tacit. Explicit knowledge can be articulated and easily communicated between individuals, whereas tacit knowledge is learnt from experience and cannot be explicitly defined (Polanyi, 1966).

Solutions for capturing and sharing tacit knowledge have been explored in many ways, especially through novel technologies like augmented reality (Aromaa, Heimonen, Väättänen, & Aaltonen, 2015). However, the researchers mention various challenges to its successful implementation: (1) the solutions are focused on the visualisation of maintenance as opposed to transferring tacit knowledge, (2) co-location and co-presence are important for the transfer of tacit knowledge, (3) professionals find it difficult to communicate over a digital medium, and (4) it may be difficult to learn from videos if the expert doesn't verbalise the cues they are responding to.

Research has shown that workers would be more willing to store their tacit knowledge in a digital format if the process was enjoyable (Desouza, 2003).

Aromaa et al. (2015) suggest that wearable technologies that are connected to the company's information and reporting system could facilitate the sharing and utilisation of tacit knowledge.

---

## 4.2.3 CONTEXT ANALYSIS

The in-situ context analysis was performed at two manufacturing facilities, a Whirlpool factory in Poland, and a Stryker factory in Germany. The main outcome of the analysis was the identification of possible use cases for a digital intelligent assistant. To validate the findings and collect additional information, questionnaires were sent out to the maintenance personnel at both factories. The results of the context analysis and questionnaire are outlined in the following section.

---

### 4.2.3.1 STRYKER

Stryker is a medical devices and equipment manufacturer based in Michigan, USA. Their facility in Kiel, Germany, is a world leader in the manufacturing of titanium intramedullary (inserted into bones) nails (see Figure 14).



Figure 14 Manufacturing stages of Stryker's titanium intramedullary nails

---

#### 4.2.3.1.1 MACHINES

Their plant features approximately 200 manufacturing machines in total. These include 110 CNC (computer numerical control) machines, such as turning, gun drilling, milling, bending, grinding, sand/glass blasting, and electropolishing machines.



They have recently purchased new machines, but the infrastructure and personnel are not ready to support all the available digital features, such as remote control and monitoring. One of these new machines is depicted in Figure 15 Traub TNL20 automatic lathe below. It is already fitted with many sensors, features a robotic arm and digital interfaces for external access.

Unfortunately, Stryker Kiel does not have inhouse capabilities for developing Industry 4.0 features, like using sensor data for predictive maintenance or remotely monitoring the machines.



Figure 15 Traub TNL20 automatic lathe

#### 4.2.3.1.2 MAINTENANCE

Over the past few years, Stryker Kiel has been making its maintenance procedures more efficient. The number of inhouse maintenance technicians has been drastically reduced in favour of flexible external service providers. One of whom checks the fluid levels of all the machines every day. They use handheld computers to scan fluid measurement points on machines, and manually enter the levels. Their cart and handheld computer is shown in Figure 16 below.



Figure 16 Fluid cart used by external service providers

#### 4.2.3.1.3 REPORTING

One of the maintenance coordinators we spoke to said that the reporting process was his biggest “pain”. This was attributed to the number of documents that had to be filled in and be approved by line managers and quality control engineers. Audits had also revealed that these reports were frequently incomplete or missing.

#### 4.2.3.1.4 STAKEHOLDERS

**Machine operators:** Operate the machines and perform basic cleaning tasks. The cleaning tasks are performed on a schedule, e.g. every day or once a week. They usually operate multiple machines. If an error occurs that isn’t easily solvable, they will alert the line manager who will find a maintenance technician to help.

**Maintenance technicians:** the inhouse (Stryker) maintenance technicians perform the more complex corrective maintenance tasks. Technicians from external service providers perform the bulk of the preventative maintenance tasks.

**Maintenance coordinators:** Senior maintenance technicians that advise and coordinate maintenance operations. If the technicians need to contact the machine manufacturer for advice this will likely go through the maintenance coordinators.

**Line managers:** Monitor and approve maintenance work on the machines.

**Maintenance planner:** Plans preventative and autonomous maintenance tasks.

---

#### 4.2.3.2 WHIRLPOOL

Whirlpool Corporation is the leading major appliance manufacturer in the world, with approximately \$20 billion in annual sales, 77,000 employees and 59 manufacturing and technology research centres in 2019. The company markets Whirlpool, KitchenAid, Maytag, Consul, Brastemp, Amana, Bauknecht, JennAir, Indesit and other major brand names in nearly every country throughout the world.

Whirlpool EMEA, the European regional division of Whirlpool Corporation, has a manufacturing footprint of 13 industrial sites in 6 countries, producing the home appliances delivered and sold in 35 markets.

In 2014, Whirlpool EMEA acquired Indesit Company, a direct competitor based in Italy. The factory that is participating in DIAMOND is a former Indesit factory in Lodz, Poland. They currently produce dryers (as seen in Figure 17 below).

Many of the systems, such as spare-parts inventories, reporting, maintenance scheduling or quality control are partially digitised but are not all remotely accessible and not integrated.

---

#### 4.2.3.2.1 MACHINES

DIAMOND is focused on the semi-automated dryer drum assembly line (see Figure 18 below). This line is also the focus of an existing EU research project on PdM, Uptime.



**Figure 17** A defective drum been removed after quality control

Like Stryker, Whirlpool uses modern machinery, but the existing digital infrastructure for accessing the information remotely covers only some of the available



**Figure 18** Whirlpool dryers at Lodz, Poland

data for maintenance purposes. Some data (e.g. Overall Equipment Effectiveness) are loaded into vertical digital systems, so are remotely available for digital access but not connected horizontally to other legacy modules. On the other side, to read some values, e.g. to accessing error codes or machine states, from the programmable logic controllers (the central processing unit of the machines), still require physically plugging a laptop in. Currently, a graphical user interface is used to access machine information and change machine parameters. The production line has three sections, each with a dedicated computer interface. A paper manual can be referenced to help find the desired information or setting. Figure 15 below depicts the paper manual and one of the computer interfaces.



Figure 19 Main computer interface

#### 4.2.3.2.2 MAINTENANCE

At Whirlpool, machine operators are given the autonomy to fix small problems and perform preventive maintenance tasks within the Autonomous Maintenance plan. However, if the problem is complex, they will involve the machine setter or maintenance technicians, which normally operate within Professional Maintenance plan or perform corrective interventions on machine breakages producing digital documents called EWO (Emergency Worker Orders). Despite the availability of manuals (Figure 19) and digitised maintenance reports, the workers prefer to use their own experience and intuition when solving problems. In other cases, they are forced to rely on their intuition. For example, when the machine returns multiple errors but does not indicate which error caused the stoppage. Unless they have recently encountered a similar scenario, the technicians may resort to an intuition-directed trial-and-error process to uncover the root problem.



Figure 20 Maintenance information markers



Many of the machines featuring information markers, as depicted in Figure 20 above, to help workers follow maintenance procedures are still not digitalized

#### 4.2.3.2.3 REPORTING

Although the quality control documents (see Figure 21) are hand-written, they are later digitised. The same is true for maintenance reports (EWO cards), which are used whenever a corrective maintenance job takes more than 10 minutes. Concerning reporting, the results of the questionnaire showed that: the situation had improved with the introduction of the EWO cards (which include a sketch area). However, tracking changes to machine parameters is frequently a problem. and the process of digitising reports can be time-consuming. The maintenance technicians would appreciate getting tips on how to solve problems Furthermore, maintenance management point out that that pen and paper is the best way to report because of the average age of the maintenance technicians.

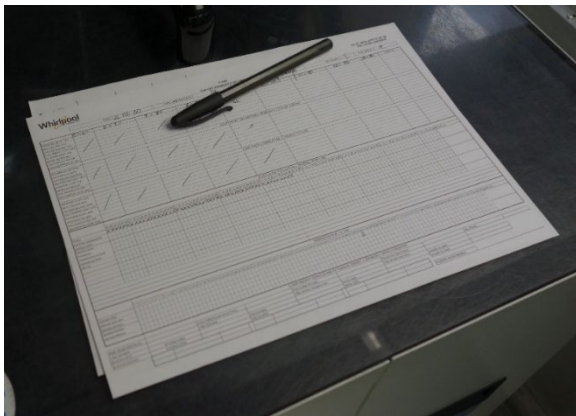


Figure 21 Document for tracking production quality

#### 4.2.3.2.4 STAKEHOLDERS

**Maintenance Technician:** Executes lengthier or complex maintenance tasks. Maintenance technicians have a specialisation, e.g. electrician, programming and mechanics.

**Maintenance Coordinator:** Coordinates and assigns maintenance technicians to unplanned maintenance jobs.

**Machine Operator:** Responsible for the operation and monitoring one of the production line stations. They perform simple autonomous maintenance tasks like inspection, cleaning and lubrication. They also can perform simple corrective maintenance tasks.

**Setter:** Responsible for monitoring the production line and changing machine parameters. If the production line has a problem, they will help the machine operator fix the problem. Otherwise, they will alert the maintenance coordinator that help is needed.

**Maintenance planner:** Plans preventative and autonomous maintenance tasks.

#### 4.2.3.3 USER PERSONA

The conclusions from the context analysis from the perspective of a maintenance technician are outlined in the following user persona:

#### 4.2.3.4 CONCLUSIONS

The main findings of the context analysis are the DIAMOND partner factories are the following:

- Twelve potential user scenarios for a CAIA were identified. These ranged from checking spare-part availability to creating a report by voice (see Appendix 9.2). These scenarios fall into five themes: reporting, monitoring processes, maintenance planning, problem-solving, task execution.

- Both factories still rely heavily on paper-based administration and reporting.
- Neither factory has experience using a PDM system.
- Most computer systems (e.g. machine computers, spare-parts database, quality control database, planning
- The technicians, setters, operators and maintenance coordinators have a lot of knowledge about how to fix the machine but most of this is not recorded elsewhere (on paper or digitally).
- Tasks related to reporting are considered the biggest pain for maintenance technicians.

software) are not interconnected or remotely accessible.

- DIAMOND will also need to develop the interface with the existing information systems.
- Maintenance personnel rely heavily on their intuition when diagnosing machine problems.

estimate the health of aircraft components. Upon detecting an anomaly, it sends a report to the maintenance coordinator centre. The maintenance coordinators check the report and then assign the task to a maintenance technician.

The main findings of my interviews with KLM personnel are:

- They are positive about the financial savings that predictive maintenance is providing and continue to expand the number of components that they monitor.
- Whenever a new component is added to Prognos, they provide the maintenance coordinators with a short training explaining how the model works and a proof-of-concept.

#### 4.2.4 AIRCRAFT MAINTENANCE

KLM is a Dutch commercial airline and one of the ten largest companies by employee numbers in The Netherlands. A year and a half ago, KLM introduced a predictive maintenance system, Prognos. The system uses sensor data and mathematical models to



Figure 22 Aircraft maintenance ("Predictive Maintenance Template | Azure AI Gallery," n.d.)

- Maintenance coordinators call the data engineers for reassurance when the model suggests an action that they are uncertain of.
- Some maintenance coordinators embrace the systems, while others do not use it.
- In the beginning, the developers of Prognos had to call the maintenance coordinators to remind them to act on the predictions.
- KLM maintenance employees are happy to use new technology if the value to them is clear
- The average age of maintenance coordinators is 55 years.
- Following predictive maintenance advice can go against the intuition of workers when they do not perceive a problem.
- KLM selects components to add to Prognos based on the potential cost benefits and ease of modelling.

#### 4.2.5 HEALTHCARE

The use of virtual assistants in the healthcare industry is widespread, e.g. clinical decision support (Ahamed, Sharmin, Ahmed, Hague, & Khan, 2006), providing emotional support (Provoost, Lau, Ruwaard, & Riper, 2017), decision support for medical research (Belciug & Gorunescu, 2020), guiding users through family health history (Ponathil, Ozkan, Welch, Bertrand, & Chalil Madathil, 2020), as a health coach (Gupta et al., 2018), or as a medical scribe (Quiroz et al., 2019)(Figure 24).

Many of the above tasks have parallels in the maintenance industry. In both cases, professionals must perform complex decision making related to diagnosis of problems, perform complex tasks, track the “health” of systems (i.e. humans or machines) over time and maintain detailed documentation of their work.



Figure 23 A medical scribe with a laptop, voice recorder and paper

Research has shown that clinical documentation is associated with increased cognitive load (Wachter & Goldsmith, 2018), distractions (Campbell, Sittig, Ash, Guappone, & Dykstra, 2006), information loss (Shachak, Hadas-Dayagi, Ziv, & Reis, 2009), and clinician burnout (Friedberg et al., 2014). Considering the parallels to the maintenance industry, some of these problems may also be prevalent there.

Automatic medical scribes use many of the same technologies in CAIAs, e.g. NLP (natural language processing) to make sense of conversations between a doctor and patient. This poses many challenges to the system: 1. Poor audio, 2. Conversations are unstructured, feature overlap and incomplete statements, 3. Information extract is challenging as existing systems are based on more standardised text inputs, 4. Summaries are challenging due to the high variability in each set of notes, 5. Lack of clinical data (Quiroz et al., 2019). The researchers state that the main hurdle to further research and development is the lack of conversation transcripts for analysis in the domain. This is likely a similar issue for the maintenance industry, where companies are eager to protect their intellectual property.

---

#### 4.2.6 PROBLEMS AND POTENTIAL USE CASES

The following list represents the six highest priority problems:

- 1) The amount of reporting and administration. All users we spoke to dislike having to do the reporting and administration but recognise its benefits.
- 2) The quality of the reporting (incomplete or unclear). This is more of an issue for corrective maintenance as the

troubleshooting processes can be lengthy and complex.

- 3) Not every step of a preventative maintenance task was performed. Audits are both Whirlpool and Stryker revealed that many tasks were skipped during preventative maintenance, however, the reason behind this is unclear.
- 4) Changes to machine parameters are poorly tracked. This is a common issue at Whirlpool as the machine parameters are frequently changed without properly documenting the reason.
- 5) Users get stuck during corrective maintenance work. By the nature of the variability and complexity of modern moderns, troubleshooting corrective maintenance is non-trivial. The users expressed an interest in receiving tips if they get stuck.
- 6) Error code messages do not provide enough information. Sometimes the machine computers don't provide any information, are missing context or the sheer number of error codes hinder a root-cause analysis.

The use cases outlined below (Table 3) are outlined by user type and complexity. The basic use cases are technically feasible within the timeframe of DIAMOND and the advanced use cases give perspective for future developments.



User	Maintenance tasks	Basic use cases	Advanced use cases
Machine operator	Autonomous maintenance: cleaning, lubrication, inspection. Perform basic corrective fixes (at WHL)	Create maintenance reports Check spare part availability Check machine status Retrieve error code meaning Call/message/email colleagues Guide user through preventative maintenance procedures Check changes to machine parameters	Provide personal guidance through maintenance troubleshooting Advise operator to perform extra tasks based on PdM/RxM Capture explicit/implicit user knowledge
Maintenance technician	Preventative maintenance Corrective maintenance Advises operators on easy fixes (at WHL)	[Same as above]	[Same as above] Analyse machine or product parameter trends over time
Maintenance coordinator	Advices maintenance techs Oversees maintenance work Assigns unplanned tasks	[Same as above] Assign tasks to mechanics Digital authorization/approval of maintenance tasks	Provide decision support using AI/PdM
Maintenance planner	Creates preventative maintenance plan	Retrieve availability of technicians Call/message/email	Provide decision support using AI/PdM

Table 3 Use case overview

## 4.3 PROGRAMME OF REQUIREMENTS

The following list of requirements is based on the results of the Research and analysis phase (Section 4). This includes a context analysis of a Whirlpool (Section 4.2.3.2) and Stryker factory (Section 4.2.3.1), interviews with KLM E&M (Section 4.2.4) and other companies, and a general analysis of CAIAs, including an exploration of methods and best practices for conversational AI design (Section 4.1).

This programme of requirements to steer the choice for the use case in Section 5.2. Furthermore, future work can use it as a guide for the development of CAIAs for Industry 4.0.

Requirements are designated by an R followed by the category and a number: R.[Y].[X]. Wishes follow the same pattern but start with a W.

---

### 4.3.1 TECHNOLOGY

- R.T.1 The solution should operate reliably in a noisy (shop floor) environment
- R.T.2 The solution should be possible to develop within the timeframe of six months by a small team of developers (2-3). [DIAMOND]
- R.T.3 The desired solution must be technically feasible using existing technologies and frameworks
- R.T.4 The underlying technology should support other languages for the later stages of development [DIAMOND]
- R.T.5 The architecture of the AI agent must allow for the addition of new features and intelligence
- R.T.6 The AI agent should be able to access details about the user from a database
- R.T.7 The AI agent should be able to communicate with client systems through APIs

- R.T.8 The AI agent should be compatible with voice and text interaction
- R.T.9 The underlying framework should support interactive learning
- W.T.1 Where possible, open-source frameworks and technologies should be used

---

### 4.3.2 USER

- R.U.1 The solution should address something that the user perceives as a problem
- R.U.2 The user should be able to opt-out of any personal data collection
- R.U.3 The user should perceive the solution as an enabler, not as a threat that might displace themselves or their colleagues.
- R.U.4 The solution should be safe to use (i.e. not distract the user from dangers)
- R.U.5 The solution should enable the user to make maintenance decisions efficiently and effectively

---

### 4.3.3 CONVERSATION DESIGN

- R.C.1 The AI agent should be able to explain its capabilities, data usage and limitations to the user
- R.C.2 The user should be able to converse with the AI agent in a natural way
- R.C.3 The AI agent should be able to adjust its response complexity based on the user's experience with the system
- R.C.4 The AI agent should be able to track the context of its conversation with the user

- R.C.5 The AI agent should be able to handle regular conversation navigation
- R.C.6 The AI agent should communicate in a consistent style throughout the interaction
- R.C.7 The AI agent should be accessible to users who are inexperienced using digital technologies
- R.C.8 The AI agent should not overwhelm the user with information
- R.C.9 The AI agent should address a problem that the users identify with
- R.C.10 The conversation style of the agent should match that of the user (Qiu, Gadiraju, & Bozzon, 2020)

---

#### 4.3.4 BUSINESS

- R.B.1 The solution should allow the clients to maintain complete control over their data
- R.B.2 The solution should provide value to the clients upon which a business case could be built

- R.B.3 The value of the solution should be explainable to client leadership
- R.B.4 The solution is relevant for manufacturing companies besides the DIAMOND partners, Whirlpool and Stryker

---

#### 4.3.5 KNOWLEDGE CREATION

- R.K.1 The solution should support capturing usage data for further development and analysis
- W.K.1 The solution should support the exploration of topics related to tacit knowledge transfer [DIAMOND 2.0]
- W.K.2 The solution should support the exploration of topics related to prescriptive maintenance [DIAMOND]
- W.K.3 The solution can be used to facilitate the transfer of knowledge from retiring maintenance technicians to the newer generations.

## 5 DEVELOPMENT

The development phase consists of choosing a specific use case for the CAIA, testing hypothesis regarding the value of the use case, analysing similar solutions and developing a prototype.

Through this process, I was able to evaluate the potential value, hurdles and future research within the scope of CAIAs and Industrial Maintenance.

### 5.1 APPROACH

The use case for which to develop a solution was selected using a weight-criteria method to reduce bias, as outlined below (Section 5.2). The approach for the development is based on

the Natural Language Framework by Moore & Aral (2019), as outlined in Section 4.1.7 and 4.1.8, a state-of-the-art framework specifically designed for conversational AI agents. According to them, an understanding of existing conversations between the target users is one of the key requirements for the successful development of a CAIA. Therefore, I performed a conversational analysis of existing transcripts in an industrial maintenance setting (Section 5.5). Furthermore, I performed a literature analysis of related solutions (Section 5.4) to determine how others have approached similar problems and what the state-of-the-art techniques are. I supplemented their approach with a scientific experiment (Section 5.6) to validate the

ID	Criteria description	Weight
D1	Plays to the strengths of conversation design according to the guidelines of Google, IBM and Rasa: <ul style="list-style-type: none"><li>- Users can do this task while their hands and eyes are busy</li><li>- Users are used to having human-to-human conversations on the topic</li><li>- The current way of accomplishing the task would take longer</li><li>- Users can do this task while multi-tasking</li><li>- Users feel comfortable talking about the subject</li><li>- There are users of varying levels of expertise</li><li>- Capturing personalised user information would be beneficial</li><li>- Accessing information over documents would be beneficial</li></ul>	0.4
D2	Frequency of occurrence	0.1
D3	The positive impact of each usage (short and long-term value)	0.1
D4	Solves something that the users perceive as a problem	0.2
D5	How easily can the required technology and infrastructure be implemented	0.2

Table 4 Design criteria and weighting



potential value of the solution and to create knowledge for future work. Then I used the new knowledge from the previous steps to develop a series of prototypes (Section 5.8). The final prototype was tested according to the requirements defined in Section 5). Finally, I outlined a roadmap for future development (Section 5.9).

## 5.2 USE CASE SELECTION

To inform the choice of a use case objectively, I scored them with weighted-criteria. The criteria fall into two equally weighted categories, design and relevancy. The weights of the criteria were determined based on my perception of their contribution to the success, academic and industrial relevance of the project.

### 5.2.1 DESIGN AND RELEVANCY CRITERIA

The design criteria are based on Google's and IBM's respective conversation design guidelines.

The list of criteria above (Table 4) is slightly modified compared to Google's conversation fit checklist as their list assumes that the new feature is replacing a graphical user-interface. Many processes in the maintenance

procedures at Whirlpool and Stryker still rely on pen and paper or making phone calls. It should be noted that other companies, who are more modernized will have already adopted digital technologies like using digital reporting on tablets. Google's guidelines stipulate that interactions should be brief, with minimal back-and-forth dialogue, however, this is at odds with natural human conversation as discussed in the book on Conversation UX design by Moore and Arar.

Furthermore, I considered the relevancy of the use-case to the project and further research (Table 5):

### 5.2.2 SELECTED USE-CASE

The five use-cases presented in Table 6 scored highest in the weight-criteria matrix (see Appendix 9.2 for the complete scoring). As shown in the table, creating maintenance reports scored highest. Reflecting, on this outcome, I believe it affords interesting opportunities related to collecting data on the terminology and conversation style of maintenance workers, offering insights on how they reason, and a platform for developing the usability of conversational AI agents beyond simple question-answers or basic information retrieval.

ID	Criteria description	Weight
R1	The goal of my project: evaluate the potential for a conversational AI agent for maintenance tasks in the manufacturing industry	0.6
R2	The DIAMOND project: currently focused on getting a working MVP to test basic features at Whirlpool and Stryker	0.3
R3	The proposed follow-up DIAMOND project on tacit knowledge capture	0.1

Table 5 Relevancy criteria and weighting

Use-case	Score	Comments
Creating maintenance reports	9.25	High scores on all criteria and is especially interesting for its potential for capturing rich data on how maintenance technicians approach their work.
Retrieve error code meaning	8.15	A good candidate but loses points due to occurring less frequently and aligning worse with my project goals and potential for future work.
Call / message/ email colleagues	7.6	Less interesting from a conversational design perspective as interactions would be basic and does not solve something the users consider to be a problem.
Guide the user through preventative maintenance procedures	7.5	An interesting topic that could build on existing research, however, the users did not see this as a problem.
Check spare part availability	7.45	A useful feature but does not take advantage of many conversation UX affordances or potential for future research

Table 6 Top five use cases

Furthermore, it has the added benefit of being tested and implemented without having to work with company-specific APIs as the created maintenance reports could be sent by email. Lastly, I would expect the accuracy and completeness of the reporting to improve if it performed in real-time. It is widely accepted that the accuracy of recall diminishes over time (Oberauer, K., & Lewandowsky, S. (2008)). Interviews and surveys with the users also revealed that they find the process of writing down their work to be time-consuming and an additional burden, which might affect their willingness to do so in detail.

**5.3 IDEATION**

The ideation process as defined in the Natural Conversation Framework (Section 4.1.7), involves defining the agent persona and

utterance pair brainstorming. Together these are used to define the initial features for development.

---

**5.3.1 AGENT PERSONA**

To be successful in eliciting details and explanation from the maintenance technicians, the agent will need to pose questions without being annoying. If the agent presents itself as all-knowing, yet asks many questions, the user may perceive it as arrogant. Therefore, the agent’s persona can play an important role in the acceptance of the system. Moore & Aral (2019) suggest that the personality of the agent should reflect the company’s brand and be consistent across use-cases. They define three aspects to consider: the agent’s job description, personality and self-knowledge.

---

### 5.3.1.1 JOB DESCRIPTION

A dedicated assistant to support technicians during corrective maintenance jobs.

Responsibilities and duties	Qualifications
<ul style="list-style-type: none"><li>- Create maintenance reports</li><li>- Retrieve error code descriptions and solutions</li><li>- Search the database of previous maintenance reports</li><li>- Check spare part availability</li><li>- Call/message/email colleagues</li><li>- Retrieve machine parameters</li></ul>	<ul style="list-style-type: none"><li>- The ability to interact with machine information systems</li><li>- The ability to communicate with company mail servers, phones and messaging.</li><li>- The ability to create and submit maintenance reports</li><li>- The ability to access databases on spare part availability, error code descriptions and solutions</li></ul>

---

### 5.3.1.2 PERSONALITY

According to Moore & Aral (2019), adjusting the level of detail to suit the user is vitally important for a conversational AI agent. In the case of an assistant, I expect this to also entail the frequency of contact, the level of insights it provides, providing rationale versus empathy and leaving the decision up to the user or not. Therefore, the agent will need to adjust its extraversion to introversion level based on the experience and needs of the user. If the user is experienced and requires little assistance, the assistant should be more passive.

Based on the requirements defined in the analysis section on CAIA guidelines. The assistant should be helpful, have a sense of humour that aligns with that of the maintenance technicians i.e. mostly casual but serious or playful when the situation calls for it.

Considering that the agent is in early development and likely to make errors, its personality should reflect an apprentice in learning.

Lastly, in accordance with the guidelines on gender bias discussed in Section 4.1.6.1, the agent should respond in a gender neutral manner and not tolerate insults (e.g. gender related insults).

---

### 5.3.1.3 SELF-KNOWLEDGE

To be accepted by the users, I believe that the agent should be open and transparent about its origins, purpose and capabilities. If asked, it should be able to explain these topics.

In accordance to the gender bias guidelines (Section 4.1.6.1) it should be able to explain that it is not human and does not have a gender.

---

### 5.3.2 UTTERANCE PAIR BRAINSTORMING

According to Moore & Aral (2019), the utterance pairs should be based on the use case, outcome statement and agent persona. For this project, I will also consider the research questions that I would like to answer through user testing. The pairs depicted below were selected as a high priority. The remain pairs were deemed either off-topic or out of scope for this project. The full list can be found in appendix 9.2.

<p>U: I don't understand what you need from me?  A: I need you to explain your actions and thoughts step-by-step.</p>	<p>U: [Procedural statement]/[Internal statement]/[Task status].  A: okay / I see / Go on / Interesting</p>
<p>A: That was it?  U: Yes, I'm done. / No, I'm still fastening the bolts.</p>	<p>A: Please keep me updated on your progress.  U: Sorry, I was just cleaning up the oil that I split.</p>
<p>U: Okay, that was everything.  A: Thank you, I'll create the report for you now.</p>	<p>A: Can you take me through the steps you took?  U: First, I checked if the oil levels were high enough.</p>

Figure 24 High priority utterance pairs

### 5.3.3 The current scenario

Performing maintenance and documenting are currently two distinct steps. The current scenario is described first (Figure 26).

1. The machine operator notices an issue and if s/he is unable to solve it themselves, they will call in the maintenance department.

2. The operator fills in the maintenance technician on the situation and the steps they've already taken.
3. They set to work, trying to uncover the root cause of the problem and how to fix it. If the maintenance technician is unfamiliar with the problem, s/he may consult the documentation of previous repairs, go through the machine manual or search online.
4. If the problem is urgent or they remain stuck, they may involve more people.
5. Upon the completion of a corrective maintenance task that took longer than 10 minutes, an EWO card must be filled in on paper (at Whirlpool). Although the exact make-up of a corrective maintenance report may differ per company, they generally contain a description of the work carried out and information about the involved machine and personnel.
6. This involves describing the steps taken to diagnose and solve the problem. If applicable, there is space for a sketch.

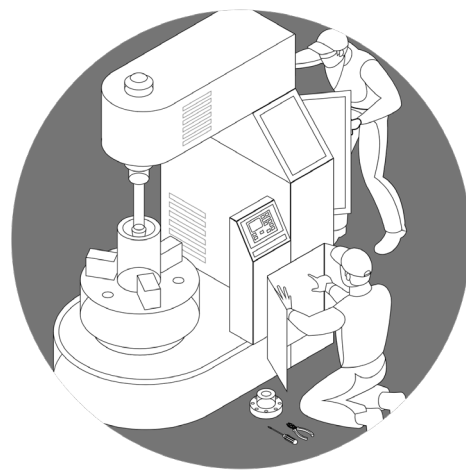


Figure 25 Current maintenance reporting scenario

7. Every EWO is then digitised by the maintenance technicians so that it can be easily accessed later.

The current scenario for corrective maintenance reporting at Whirlpool and Stryker is tried and tested. For many years, documentation was of lesser importance because machines were simpler, and experience technicians were always available to answer questions. As seen at Stryker and Whirlpool, the number of experienced/skilled technicians is dwindling, there is pressure for maintenance to become more efficient and flexible, and an increase in the complexity of problems with modern machinery. The last two points are discussed in a paper by (Lotti, Villani, Battilani, & Fantuzzi, 2018), who conclude that machine operators and technicians require more support tools to enable a smooth and sustainable interaction. They note that this is especially true for machine setup, maintenance and troubleshooting. The notion that maintenance personnel need better support tools is not new but has long been hampered by the underestimation of the importance of maintenance in the manufacturing industry (Setchi & White, 2003).

#### 5.3.4 ENVISIONED SCENARIO (1-3 YEARS)

I envision a scenario in which the maintenance technician and operator chat with a conversational AI agent during a corrective maintenance process (Figure 27).

This will enable richer, real-time, more accurate and efficient documentation of their work. Additionally, this presents an opportunity to move beyond the basics in reporting and attempt to capture expert tacit knowledge. The data collected from each repair can be analysed and in time, the AI can begin providing advice if other users get stuck.

The AI will also be able to retrieve information or perform other tasks as described in the other use-cases discussed in the previous sections. A sample dialogue for an inexperienced user is shown below. Note that it includes some out-of-scope features, like spare-part checks, to demonstrate handling user extra user requests during reporting.

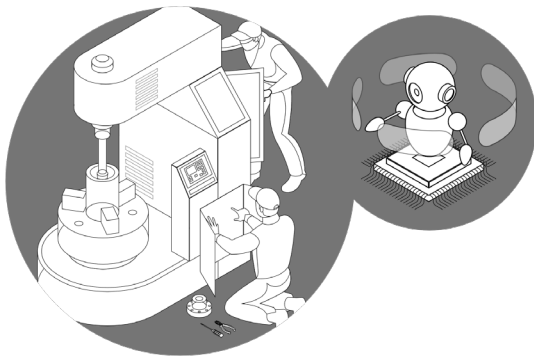


Figure 26 Envisioned maintenance reporting scenario

1	U:	((Activates conversational AI agent))	31	A:	Okay.
2	U:	Hi, I'm starting corrective maintenance	32		(420.0)
3		for machine X892 at 10:52.	33	U:	((Activates conversational AI agent))
3	A:	Hi John, I've created a new	34	U:	The sensor was full of dust. It's useless.
4		maintenance report for machine X892	35		Do we have one in stock?
5		with a start time of 10:52.	36	A:	Yes, you can find one in row 9, section 2.
6	U:	Thanks.	37	U:	Thanks.
7	A:	No problem, have you had a chance to	38	A:	No problem.
8		assess the situation yet?	39		(590.0)
9	U:	Yes, the operator reported error code	40	U:	((Activates conversational AI agent))
10		898 and the machine computer	41	U:	I replaced the sensor, and everything
11		indicates that something is wrong with	42		works again.
12		the airflow sensor A234.	43	A:	Great! I've noted 11:15 as the end time of
13	A:	Ok, thanks.	44		the repair.
14		(1.0)	45		(1.0)
15	A:	Let me know when you've decided on a	46	A:	Anything else to add to your report?
16		plan-of-action.	47	U:	Sorry, what?
17		(4.0)	48	A:	I asked if you had anything else you
18	U:	Has this error code been solved before?	49		wanted to add to the report?
19	A:	Yes, would you like me to send you the	50	U:	Ah okay, no.
20		EWO card?	51	A:	Okay, I've sent you a draft of the report
21	U:	Yes, please.	52		to your tablet for you to check and add a
22		(35.0)	53		sketch.
23	U:	((Activates conversational AI agent))	54	U:	Thanks.
24	U:	Ok, I'm going to start by checking the	55	A:	Anything else I can help you with?
25		fuses.	56	U:	Nope, thank you.
26	A:	Ok, let me know how you get along.	57	A:	Okay, bye.
27		(130.0)	58	U:	Bye.
28	U:	((Activates conversational AI agent))			
29	U:	The fuses look ok. I'm going to remove			
30		the sensor and test it now.			

Figure 27 Example dialogue of the envisioned scenario (U=User, A=Agent)



The resulting report might look like the one depicted in Table 7.

The reporting aspects of the envisioned scenario requires that the CAIA can track conversational context, proactively ask questions, deal with complex user utterances, extract and structure relevant information and send a report. To evaluate if these features are feasible with current technologies, I analysed related work (Section 5.4), and tested various features with a prototype built in Mycroft and Rasa, open-source CAIA frameworks (4.8 and 4.9).

To evaluate my hypothesis regarding the proposed value of this solution (saved time, reduced perceived workload and higher quality reports), I conducted a scientific experiment (Section 5.6).

#### 5.4 RELATED WORK

The purpose of examining related work at this stage is to understand how others have approached and solved similar problems and to evaluate the technical feasibility of what I am proposing.

Type of maintenance	Corrective
Machine	X892
Maintenance technician	John Smith
Start-time	10:52
End-time	11:15
Trigger	Machine reported error code 898: malfunctioning airflow sensor A234.
Troubleshooting steps	Talked to the operator about error code and malfunctioning airflow sensor Checked an EWO card for the same error code Checked fuses: they were ok Checked sensor: The sensor was full of dust. It's useless. Checked sensor availability in the warehouse: it was available at Row 9, Section 2. Replaced the airflow sensor A234 Reported problem was solved
Critical action	Replacing the airflow sensor A234
Materials used	1 x Airflow sensor A234 (Row 9, Sec 2)
Tags	Error code 898, airflow sensor A234, machine X892, sensor full of dust

Table 7 Envisioned maintenance report

The use of augmented reality and in some cases, CAIAs, have been extensively researched for aiding maintenance technicians perform tasks (Aschenbrenner et al., 2018; De Crescenzo et al., 2011; Henderson & Feiner, 2011). However, an analysis of existing literature and solutions did not yield any uses of voice assistants for maintenance reporting. Insights from adjacent research and applications are discussed below.

---

#### 5.4.1 AUTOMATIC MEDICAL SCRIBES

Electronic care documentation poses two main challenges: (1) the need to pause care to perform documentation activities or (2) perform documentation afterwards (Fabbri & Ehrenfeld, 2016). This can result in physician frustration, documentation errors and clinical errors. This is why hands-free documentation has been used (e.g. by employing medical scribe). However, medical scribes require training, are expensive to employ and there is a high throughput as they are usually medical students.

Automatic medical scribes fulfil a similar role to the one being proposed in this project. However, the key difference is that automatic medical scribes only listen and do not actively participate in the conversation. Finley et al. (2018) claim to have developed the first such system. It features many of the components that are also found in a CAIA, like speech-to-text, automatic speech recognition, natural language processing (e.g. entity extraction) and natural language generation. Because it must determine if the doctor or the patient is speaking, it uses a speaker diarisation module. This might also be useful for a CAIA when multiple maintenance personnel are working together on a problem. Another interesting feature is how they turn the doctor's speech

into a report: instead of copying it directly, they first extract information from the text and using Natural Language Generation create an understandable, well-structured report.

---

#### 5.4.2 AUTOMATIC MEETING SUMMARISATION

Making notes for a meeting is a useful but time-consuming task. In the early 2000s, automatic meeting summarisation was the subject of various large-scale EU and US research projects (Buist, Kraaij, & Raaijmakers, 2004). In recent years, with advances in natural language processing, new approaches are possible, for example using semantic analysis (Shabbir Moiyadi et al., 2016). It remains a highly researched topic, in terms of summarising written texts (Allahyari et al., 2017) and spoken language (Nihei & Nakano, 2019; Zhu, Xu, Zeng, & Huang, 2020)(Nihei & Nakano, 2019), and deserves further exploration.

---

#### 5.4.3 SMART GLASSES FOR (MAINTENANCE) DOCUMENTATION

Quint & Loch (2015) explored the possibility of using smart glasses for maintenance reporting (Figure 29) and concluded that it was a suitable device for knowledge conservation. However, its implementation was hindered by a lack of battery life, processing power and limited interaction possibilities. For future work, they suggest including the capabilities to add comments to the video, verbal and gesture commands. Other studies have been performed to test the suitability of Smart Glasses for documentation of forensic medicine (Albrecht et al., 2014). They concluded that the image quality was insufficient for clinical use but sufficient for education. They mentioned similar concerns



**Figure 28 A technician wearing smart glasses** (“Smart glasses in aid of maintenance,” n.d.)

regarding battery life. Both studies suggest further research into privacy and data-protection are needed.

#### 5.4.4 AUGMENTED REALITY FOR (MAINTENANCE) DOCUMENTATION

Researchers have proposed a multi-modal system for reading, manipulating and editing maintenance documentation. It consists of an AR headset, digital pen and tablet or smartphone (Rateau, Clay, & Bottecchia, 2018). It should be noted that the system design was based on user observation and has not been tested with the end-user group.

### 5.5 CONVERSATION ANALYSIS

For this project, the conversation analysis has multiple purposes: (1) to familiarise me with the conversations between maintenance personnel, (2) to compare the conversational patterns to existing (general) ones and if

necessary, define new ones, (3) explore situations in which the participants reveal explicit and implicit knowledge.

To provide a structure for the analysis, I will code the utterances using a system that featured in Kraut, Fussell & Siegel's (2003) paper on remote assistance for bicycle repair. The authors noted that a weakness of their coding system was that they lacked insight into the syntax of individual utterance. I observed that in natural conversation many utterances contain multiple types of information. Therefore, I plan on applying the coding to individual utterances.

From a maintenance reporting standpoint, it would be beneficial to learn about rich task descriptions (a description that includes additional details and reasoning). For example, what triggers them and how are they structured. Therefore, I am especially interested in tracking the relationship

(sub-)Utterance type	Definition
PROCEDURAL	Descriptions of tasks (e.g. "next, remove the cover")
TASK STATUS	State of task or objects within the task (e.g. "The bolts are very rusty", the cable housing is almost out")
INQUIRY	Requests for information (e.g. "Is the plug screwed?", "what are you doing now?")
REFERENTIAL	Identifying or localising task objects (e.g. "the big disc in the middle")
INTERNAL STATE or INTERNAL STATE	Intentions, reasoning, knowledge, emotions, etc. (e.g. "because it might be slippery"*)
ACKNOWLEDGEMENT	Feedback that message is heard/understood (e.g. "okay", "uh-huh")
OTHER	Non-task and un-codable communication

Table 8 Coding scheme

\*In some cases, an utterance may be both a task status and an internal state

between procedural tasks and internal states (e.g. *reasoning*). Furthermore, it is interesting to identify patterns in how task objects are identified and later referred to (*referential* utterance type). To function well, a conversational AI will need to be able to identify objects, track context and identify when the user is offering *internal states* (see coding scheme in Table 8).

The material that will be analysed was sourced from previous research on different augmented reality support applications for cooperative machine repair (Aschenbrenner et al., 2018). The researchers collected transcripts of conversations between a remote expert and technician. I analysed the eight "telephone" conversations that did not feature any augmented reality.

### 5.5.1 CODING DEMONSTRATION

The following demonstration (Figure 30) is from Aschenbrenner et al. (2018) transcript "Versuch 33 telefon" translated from German (E = expert, T = technician)

- 1 E: Wonderful, now we can take care of the Green part.
- 2 T: Yes.
- 3 E: And then please put all the cables where they belong. Is best from top to bottom all cables through.
- 4
- 5 T: Yes.
- 6 E: and please pay attention to the position of the cables, because if they are not in their correct position there may be a short circuit.
- 7

Figure 29 Coding demonstration

### 5.5.2 GENERAL PATTERNS

When instructing the user to perform a task, the conversation matches the "A3 Extended Telling". The expert utters short pieces of instructions and waits for continuers, repair initiations or inquiries from the user.

Below is an example of A3 Extended telling (Figure 31). Sometimes the technician responds with “yes” to indicate they’ve understood the message and sometimes to indicate they’re ready for the next task. The expert can infer which one it is based on how long he expects the task to take and if he is not sure, he can ask.

Below are examples (Figure 32 and Figure 33)

E: so, then you can loosen the port at the top left. You can then also hang it.  
 T: Yes.  
 T: all clear  
 E: Good. To the left and right of the orange plug you have a canbus connection.  
 T: Yes.  
 E: Who please also release and hang.  
 T: Yes.

Figure 30 A3 Extended Telling (Agent) (Attempt 36, row 18)

of the expert checking if the technician is ready for the next task. Evidently, in the first example, the expert could not discern if the technician’s “ok” indicated he was ready. In the second example, the expert acknowledges he needs to give the technician more time and says he will give him 1 minute but implicitly it is understood that the technician will let him know when he is done.

E: and make sure that the cables are also plugged in at the important points, otherwise, a short circuit may occur.  
 T: yes  
 T: o. k.  
 E: ready? ((completion check))  
 T: yes

Figure 31 A3 Extended Telling (Agent) Completion check

E: nice, next you plug the motor plug back in and then screw it back to the control cabinet  
 E: how far are you now? ((status request))  
 T: I'm just screwing tightly  
 E: OK, then I give you 1 minute  
 T: everything is clear, is fixed.  
 E: OK, then next you can tighten the posture screws at the top again.  
 T: All clear, are fixed.

Figure 32 A3 Extended Telling (Agent) Status request

Below (Figure 34) is an example of the technician expanding on the expert’s instructions by inquiring if he has understood them correctly. After answering, the expert offers additional information.

E: Beautiful. Then I would ask you to loosen the XR connection, which you will find at the top left.  
 T: top left, so the metal plug? ((confirming understanding))  
 E: Yes.  
 T: O. k.  
 E: He is also screwed.  
 T: O. k., plug is off  
 E: Next, you can disconnect the two Canbus connections. They are located to the left and right of the orange connector.

Figure 33 A3 Extended Telling (Agent) Understanding Confirmation request (User)

Here (Figure 35) the expert provides additional information to communicate the importance of the current task. Generally, the conversations between expert and technician follow the pattern of extended telling. One participant leads the conversation and pauses for a reaction after each utterance.

E: **Make sure that you insert the cables too short again too long** ((additional details 1))  
 T: Yes.  
 E: **And that the cables are stuck in the right place.** ((additional details 2))  
 T: Yes.  
 E: **otherwise there may be a short circuit.** ((reasoning))  
 T: All clear.  
 T: **OK, the lines are back on**  
 E: **well, next you can reattach the two Canbus connectors.**

Figure 34 A3 Extended Telling (Agent) Additional Details and Reasoning (Attempt 37, row 202)

During the extended telling sequence, a variety of expansions can occur, these included: completion confirmation (technician), reasoning (technician and expert), status requests (expert), validation request (technician), inquiry (technician and expert), paraphrase request, additional details, repeat request, observation, detail request, help request (technician), correction (technician and expert).

Many of the above expansions for extended storytelling are not listed in the Natural Conversation Framework. Defining the patterns of these expansions will be beneficial for building a conversation agent that provides extended instructions and when a user provides extended explanations, e.g. in the reporting use case. One of the new expansion types is especially interesting for the reporting use case, namely, status requests. If the user is performing a task and the agent doesn't have a way of determining if they've completed the first task or not, they will have to ask. Therefore, by analysing how the expert checks on the technician's progress could help steer the design for the reporting agent.

Extended storytelling with a status request has multiple variations, the most basic case, in which the user says he is ready for the next step, is outlined below (Figure 36).

- 1 U: STEP 1
- 2 A: CONTINUER/PAUSE
- 3 A: STATUS REQUEST
- 4 U: TASK STATUS
- 5 A: ACKNOWLEDGEMENT
- 6 U: STEP 2

Figure 35 Pattern Extended telling with a status request (U = User, A = Agent)

Other types of status requests expansions are listed below (Table 9) (the NCF already includes two variations of Extended telling expansions, Extended Telling with Repair (A3.0) and Extended Telling with Abort (A3.1). I've designated Extended telling with a Status request as A3.2.

- A3.2a The user indicates they are ready
- A3.2b The user indicates that they need more time
- A3.2c The user provides the task status (other than complete)
- A3.2d The user says he is stuck
- A3.2e The user inquires about something (e.g. for additional details or confirmation)

Table 9 Status request expansions

If the user is doing the storytelling about an action he is performing, it might be prudent for the agent to ask how long he expects to take. This information can be used to avoid asking for updates too frequently. Furthermore, this indicates that the agent expects the user to inform him when he is done.



The conversations between technician and expert showed that asking for a task status can trigger the technician to ask for help or reveal that he was doing the wrong thing, as shown in the examples below (Figure 37 to Figure 40).

E: nice, then I would ask you to lift out the defective servo amplifier. ((Step 1))  
E: Does it work? ((Status request))  
T: I don't look at it through the blue cable that hangs on it. Do I have to get over it somehow? ((Task status, Inquiry))  
E: best to give it up ((Additional instructions to step 1))  
T: o. k.  
T: now

Figure 36 A3 Extended Telling (Agent) Status request and Inquiry (User) (Attempt 37, row 202)

E: now you can insert the spare part in the same place. ((Step 1))  
E: are you through? ((Status request))  
T: yes, I tighten the screws straight at the top. ((Task status))  
E: please do not do it yet ((Task correction))  
T: o. k.  
E: because first, you need the plug that is at the bottom,... ((Reasoning))  
T: yes  
E: ... Plugin again. ((Step 2))  
T: o. k.

Figure 37 A3 Extended Telling (Agent) Status request, Task Correction (Agent) and Reasoning (Agent) (Attempt 33, row 104)

E: They are also screwed. Please loosen and then pull them out.  
T: o. k.  
T: yes  
E: beautiful. All of them already solved?  
T: yes  
E: next we take care of the green bar. There are cables on every side. And please solve them all now.

Figure 38 A3 Extended Telling (Agent) Status request

E: next you put this plug underneath again in the servo amplifier purely in its place.  
T: he doesn't go quite in. Do I have to pay attention to anything?  
E: he should go in quite a day.  
T: o. k.  
E: it works?  
T: yes  
E: good. Then you can screw the same plug back to the control cabinet

Figure 39 A3 Extended Telling (Agent) Status request

### 5.5.3 INTERNAL STATE PATTERNS

The internal states can be split into two groups: (1) expert initiated and (2) technician initiated.

#### 5.5.3.1 EXPERT INITIATED

Typical expert initiated scenarios include: (1) the expert utters a procedural statement and provides additional explanation to emphasize the importance, (2) the expert provides advice or knowledge that can help the technician perform the task better. Both of these are present in the following transcript (Table 10).

Sometimes, the expert waits for the technician to acknowledge the procedural statement before providing an internal state, as seen in Table 11. In other cases, when the expert suspects that the technician is struggling or doing the wrong thing (e.g. they have not said anything in a while), the expert might ask how they are doing and then provide advice, as seen in the two examples on the next page (Table 12 and Table 13).

Descrip.	The expert provides additional guidance and reasoning to the technician
Trigger	The expert identifies a critical task
Excerpt	E: And then please put all the cables where they belong. Is best from top to bottom all cables through. ((advice)) T: Yes. E: and please pay attention to the position of the cables, because if they are not in their correct position there may be a short circuit. ((explanation))
Pattern	E: PROCEDURAL, PROCEDURAL / INTERNAL STATE T: ACKNOWLEDGEMENT E: PROCEDURAL, INTERNAL STATE
Source	Attempt 33, Line 134

Table 11 Expert identifies a critical task

Descrip.	The expert provides additional details and an explanation to illustrate the importance of the step.
Trigger	The expert identifies critical step
Excerpt	E: make sure that you insert the cables too short again too long T: yes E: and that the cables are stuck in the right place. T: yes E: otherwise there may be a short circuit. T: all clear T: OK, the lines are back on.
Pattern	E: PROCEDURAL / INTERNAL STATE T: ACKNOWLEDGEMENT E: PROCEDURAL / INTERNAL STATE T: ACKNOWLEDGEMENT E: INTERNAL STATE T: ACKNOWLEDGEMENT T: TASK STATUS
Source	Attempt 37, Line 202

Table 10 The expert identifies a critical step

Descrip.	The expert asks how it's going, the technician provides a task status upon which the expert responds with an explanation for the technician's observation.
Trigger	The expert checks in on the technician after not hearing anything for a while
Excerpt	E: does it work for you? T: I need to take another screw. The screws are quite tightly screwed. E: yes, the screws have been in use for a long time.
Pattern	E: INQUIRY T: TASK STATUS / INTERNAL STATE E: ACKNOWLEDGEMENT, INTERNAL STATE
Source	Attempt 38, Line 179

Table 12 Expert check in 1

Descrip.	The expert explains that the technician's actions will prevent him from being able to perform the subsequent task (plug at the bottom).
Trigger	Expert proactively asks the technician if he had completed a task. The technician's response reveals that he was performing the incorrect action.
Excerpt	E: are you through? T: yes, I tighten the screws straight at the top. [12] E: please do not do it yet T: o. k. E: because, first you need the plug that is at the bottom,... [8] T: yes E: ... Plug in again. [8] T: o. k.
Pattern	E: INQUIRY T: ACKNOWLEDGEMENT, TASK STATUS E: PROCEDURAL / INTERNAL STATE T: ACKNOWLEDGEMENT E: INTERNAL STATE T: ACKNOWLEDGMENT E: PROCEDURAL T: TASK STATUS
Source	Attempt 33, Line 106

Table 13 Expert check in 2

### 5.5.3.2 TECHNICIAN INITIATED

There are two types of technician initiated internal states: (1) the technician utters a “task status”, e.g. “the screws are very tight” or “this looks broken”, which implicitly requires a response from the expert and (2) the technician explicitly asks the expert a question or for advice.

For the first type, the technician is usually asking for reassurance (Table 15) or asking for advice on how to proceed (Table 14).

For the second type of technician initiated internal states, the technician is also usually looking for advice (Table 16) or reassurance (Table 17).

Descrip.	The expert responds with an explanation
Trigger	The technician observes that a component looks broken (“task status”)
Excerpt	T: this is a slot T: this looks a bit broken. E: yes the same amplifier has been used many times. T: all clear. E: After some time it just breaks T: I think so.
Pattern	E: REFENTIAL  T: TASK STATUS  E: ACKNOWLEDGEMENT, INTERNAL STATE  T: ACKNOWLEDGEMENT  E: INTERNAL STATE  T: ACKNOWLEDGEMENT / INTERNAL STATE
Source	Attempt 38, Line 45

Table 15 Technician observes component state

Descrip.	The expert responds by giving him a tip on how to solve the problem.
Trigger	The technician alerts the expert to an issue he is having.
Excerpt	T: there is a small problem, one of the orange plugs I still have to solve. Otherwise I won't get out. E: you can also get out without, you just need the servo amplifier a bit above it.
Pattern	T: TASK STATUS / INTERNAL STATE, INTERNAL STATE  E: PROCEDURAL / INTERNAL STATE
Source	Attempt 38, Line 142

Table 14 Technicians reports problem

Descrip.	The expert responds with a description of the best practice and then provides additional information concerning the safety. Based on the technician's initial question, he infers that the technician might be worried of getting electrocuted.
Trigger	The technician identifies that he should execute the task in a specific way
Excerpt	E: and then simply pull out all the cables. T: is there a certain order to consider? E: best to go from top to bottom T: o. k. E: is no more electricity on it, you don't need to worry
Pattern	E: PROCEDURAL T: INQUIRY E: INTERNAL STATE T: ACKNOWLEDGEMENT E: TASK STATUS, INTERNAL STATE
Source	Attempt 38, Line 88

Table 16 Technician asks for task guidance

Descrip.	The technician mentions a problem, the expert responds with an action he can take to solve it. Lastly, the expert mentions that the problem isn't unexpected considering the age.
Trigger	The technician encounters a problem during a task.
Excerpt	T: may it be that the right screw breaks the thread? E: if the thread is broken there is a replacement above. T: o. k. then I use. E: yes, is not so bad they have been in use for a long time.
Pattern	T: INQUIRY E: PROCEDURAL / INTERNAL STATE T: ACKNOWLEDGEMENT, TASK STATUS E: ACKNOWLEDGEMENT, INTERNAL STATE
Source	Attempt 38, Line 179

Table 17 Technician asks for clarification



---

#### 5.5.4 RELEVANCE FOR THE REPORTING USE CASE

The use of acknowledgements, e.g. “ok”, “all clear”, are widespread and serve to indicate that the other party has understood the utterance. If no acknowledgement is uttered, the other party might think that the utterance was not understood. This is especially true in the analysed example as one party is providing instructions to the other. Moore & Aral (2019) state that during storytelling, acknowledgements also indicate that the listener is interested in hearing more details. It should be noted that many of the words coded as acknowledgements were interjections (a more general term), e.g. “good”, “beautiful”, “ok” (in some cases). Literature suggests there are many types of interjections and they play an important role in human communication (Ameka, 1992; Norrick, 2009). Previous work has also indicated that when people interject comments at just the right moment with each other, they build rapport, feel more positively about each other and are more likely to be persuaded (Gratch et al., 2006).

Although, not present in the transcripts, research on the use of fillers like “hm” or “uh” by conversational agents show that it did not affect the perceived naturalness or satisfaction (Pfeifer & Bickmore, 2009).

Regarding the development of the CAIA prototype, the identified utterance patterns can be used to train the intent model, define stories and natural responses.

When trying to elicit a response from the user if they have been silent for a while, the CAIA could ask the user “how are you doing?”. Further research will be needed to determine

how disruptive this is and when is a good moment to do so.

To elicit further explanation or advice regarding a task, the CAIA could ask “what’s the best way to do this?”. However, further research is needed to determine how the CAIA can identify suitable opportunities.

---

#### 5.5.5 EVALUATION OF CONVERSATION ANALYSIS

The coding scheme provided a mechanism for quickly spotting utterance patterns, however, in many cases it did not capture the subtleties present in human language. For example, there were many sub-types of “internal states” and “acknowledgements” that was not differentiated from each other. For future research, I suggest splitting internal states into (1) explanations, (2) advice, (3) intentions and (4) feelings, and split “acknowledgements” into (1) “acknowledgements”, (2) approval/praise injections, and (3) preamble interjections.

The original transcripts were in German which were translated using the Microsoft Word Translate function. Based on my Dutch and basic knowledge of German, I was able to determine the meaning of the utterances, even when poorly translated. However, the sentence structure and phrasing may not be completely representative of English-language maintenance personnel.

Lastly, there is some subjectivity involved when applying the coding scheme, especially when the intention of the original statement is ambiguous. A few of the transcripts were coded by a colleague, however, I went through them afterwards to ensure consistency with the rest. There were only a few minor differences that I had to adjust.

## 5.6 SCIENTIFIC EXPERIMENT

The decision to focus on automating reporting for corrective maintenance tasks was supported by several hypotheses (see Section 5.6.1 below). To develop a scientific knowledge base that can inform future development, I performed user experiments to test these hypotheses.

As discussed in Section 4.2.3, maintenance personnel at Stryker and Whirlpool have reported that paperwork takes a lot of time and is sometimes of poor quality. At Stryker, paperwork was reported to be the biggest “pain”. Personnel at both facilities expressed interest in something being done to improve the reporting process. In both facilities, maintenance technicians rely significantly on internal explicit and tacit (implicit) knowledge when troubleshooting and fixing problems. At Whirlpool, many of the procedures for fixing problems (e.g. error codes) are not documented anywhere and exist only in the minds of the maintenance technicians. Therefore, this knowledge is not readily

accessible to other technicians and is lost when they leave.

### 5.6.1 HYPOTHESES

Hypothesis 1 (Primary): People performing maintenance work will provide higher-quality reports (number of relevant utterances and references) when talking to a conversational agent while they perform the work as opposed to writing reports down on paper after completing the work.

Hypothesis 2 (Secondary): People performing maintenance work will finish their task quicker when they can create a report by talking to a conversational agent during the task as opposed to writing report on paper afterwards.

Hypothesis 3 (Secondary): People performing maintenance work will perceive the workload (mental demand, physical demand, temporal demand, performance, effort, frustration) of their job as lower when they can create a report by talking to a conversational agent during the task as opposed to writing report on paper afterwards.

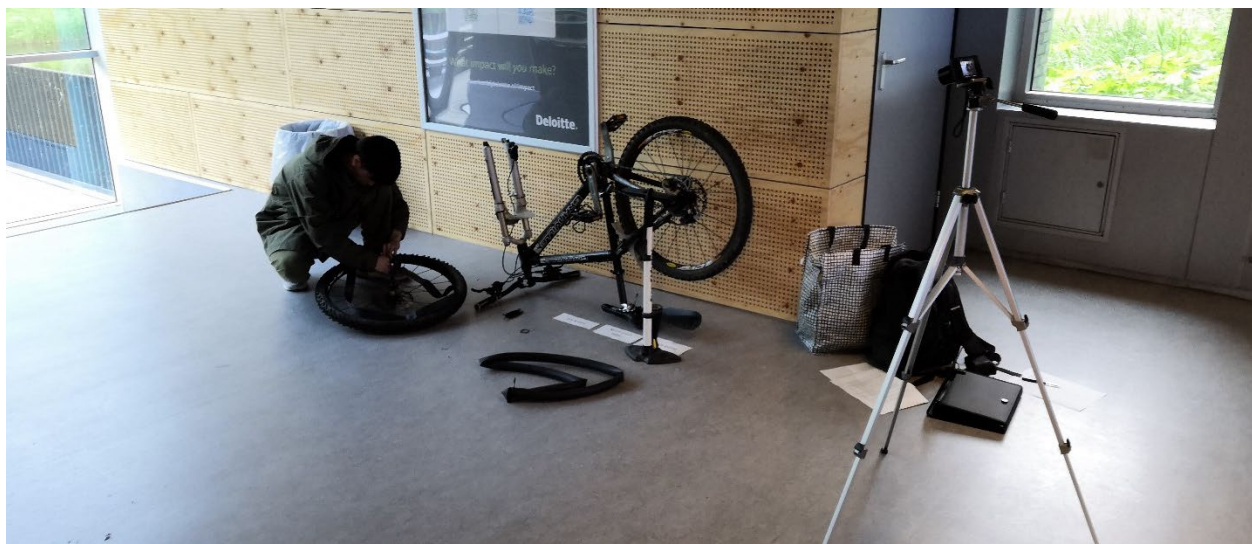


Figure 40 A “control” experiment participant

## 5.6.2 METHOD

To test the hypotheses, I created an experiment that involved participants replacing a bicycle inner tube (see Figure 41) and creating a report. One group was asked to write a maintenance report after completing the task, whereas the other group could converse with a conversational agent while they work, who automatically created the report for them. I acted as the conversation agent, i.e. Wizard of Oz, to avoid bugs and performance issues of the prototype affecting the results. I followed a set of rules and scripts that simulated the theoretical capabilities of existing conversational agents (see Appendix 9.4.13.3).

Due to the limited access to maintenance workers due to COVID-19 outbreak, I selected a maintenance task that many people living in the Netherlands know how to do. As a result, I could ask neighbours, colleagues and acquaintances to participate. The reason I chose to a task that many the general public

are familiar with as opposed to providing them with a novel problem to solve, was so that they could utilise internal tacit and explicit knowledge. I considered them as experts that might have useful knowledge to share.

One of the goals for the CAIA is to elicit valuable knowledge from experienced workers. Therefore, both participants groups were told that the purpose of the report was that a novice colleague could learn from their experience and understand what they did, how they did it and why they did it that way. I consciously deviated from the current practice of reporting at factories, a simple list of performed actions, as this does not facilitate the capture of internal knowledge.

I used a between-groups setup as I expected the learning effect to be significant and the experiment would take very long (>45 min) if participants were asked to perform the task twice.

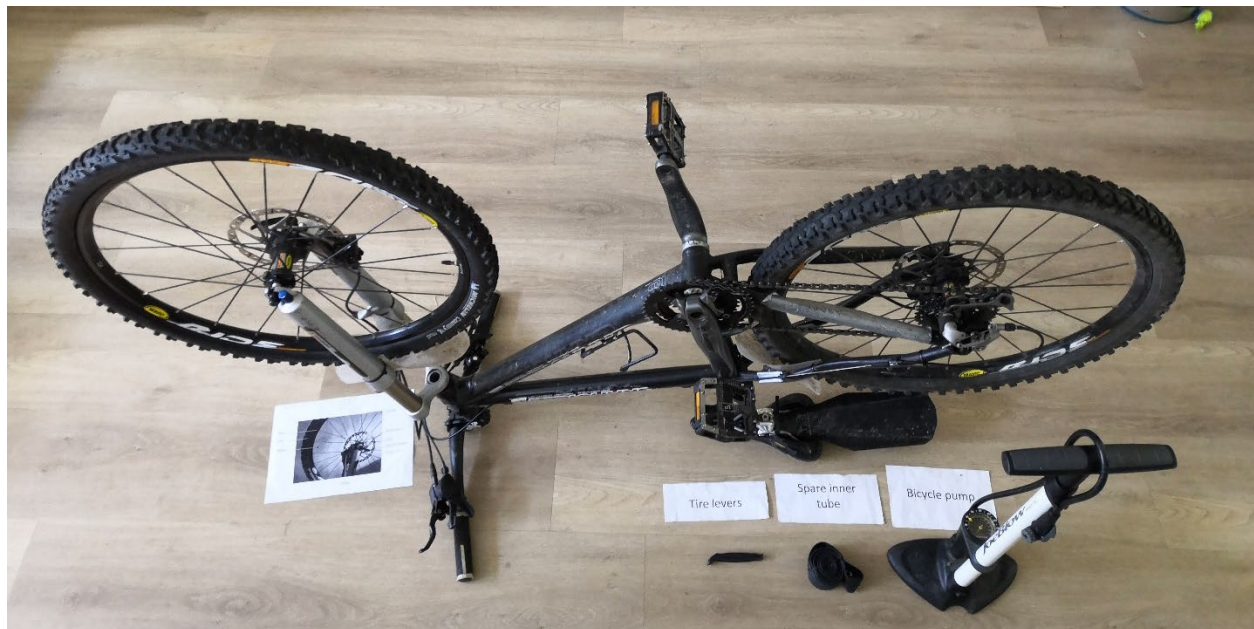


Figure 41 Experiment setup

---

### 5.6.3 DEPENDENT VARIABLES

1. Quality of reporting: number of relevant details (see Appendix 9.4.14).
2. Time to complete the entire task (in the case of the control group, the time to complete the report will also be measured). If needed, the video recordings can be used to compare the time per subtask.
3. Perceived workload: mental demand, physical demand, temporal demand, performance, effort, frustration (NASA-TLX as seen in Appendix 9.4.9).

---

### 5.6.4 INDEPENDENT VARIABLES

The reporting mechanism:

1. Talking to a conversational agent that uses natural dialogue techniques (e.g. status requests, acknowledgements, continuers, interjections and follow-up questions, see Appendix 9.4.3.2 for more details).
2. Writing the report down on paper.

Note the reporting mechanisms differ in two ways: the time of reporting (during the task versus afterwards) and the mode of reporting (writing on paper versus talking to a CAIA). Therefore, I will not be able to definitively infer how each aspect contributed to the results. However, if the results are promising, further research could be performed to understand the relationship better.

---

### 5.6.5 PARTICIPANTS

Participants were recruited if they had changed a bicycle tire at least once and were confident that they could do so again unassisted. Additionally, they had to be capable of communicating clearly through spoken and written English. To minimise the

statistical corrections needed due to skill, the participants were distributed equally over both groups based on their self-reported bicycle tyre-changing skill level.

---

### 5.6.6 RESULTS

In the following section, the statistical results of the experiment are summarised. The two groups, (1) replace the inner tube and report afterwards on paper and (2) replace the inner tube while conversing with a CAIA, are referred to as the control and experimental group respectively.

---

#### 5.6.6.1 STATISTICS AND VISUALISATION METHODS

The resulting data was tested for normality using the Shapiro-Wilks test. (Shapiro & Wilk, 1965). Seven of the 14 measured factors tested as following a normal distribution for at least one of the groups. Because the sample size is relatively small at 12 per group, it is unlikely that all factors would test as normal, therefore, I will assume that the rest also follow a normal distribution. As there are two groups, t-tests were used to test the population means. Furthermore, a Spearman's rank-order correlation test was used to check for correlations between the control variables and control variables. Lastly, t-tests were used to confirm that the means of the tyre changing-related control variables were not significantly different between the groups.

For each factor, the mean, median, significant rating and p-value of the t-tests are presented in tables. Additionally, boxplots from the Seaborn library were used (L. Waskom, 2020). These feature boxes that extend from the Q1 to Q3 quartiles of the data with a line on Q2, the median.



### 5.6.6.2 FACTORS

Participants from the experimental group uttered more unique pieces of information than the control group for all the measured utterance types, as seen in Figure 44 and Table 20 below. Whereas most control group participants provided a short description of the task, e.g. “pump up the bike”, experimental participants were more likely to provide additional information, e.g. “attach the grey side of the pump to the bike, flip the lever backwards, and start to pump. Expect to lose some air when you attach it”. The differences were most stark between “task status” and “internal state” utterances as the control group uttered almost none. There are no major differences between the mean and median values, indicating that the data was not affected by outliers or a skewed data distribution.

The total time to complete the exercise (change the tyre and create the report) was significantly longer for the control group (mean of 1120 seconds versus 540 seconds)(Table 19 and Figure 43). The results

concerning the time taken to change a tyre are less conclusive. There is no statistically significant difference between the control and the experimental group. Also, the direction of the relationship differs depending on whether the mean (553 versus 540 seconds in favour of the experimental group) or median (530 versus 580 seconds in favour of the control group) is considered. The time measurements appear to be affected by outliers or skewed data distribution, therefore, further data is needed to conclude.

The perceived physical demand was significantly lower in the experimental group (mean of 6.67 versus 11.4,  $p = 0.0167$ )(Figure 45 and Table 18). There is an indication that the experimental group experiences a lower mental demand (mean of 6.58 versus 8.58), frustration (mean of 6.33 versus 7.67), effort (mean of 8.5 versus 10.5) and higher performance (mean of 17.7 and 16.1), and temporal demand (mean of 13.1 and 10.8) but the results are all not statistically significant. The same is true for the total workload (mean of 46.7 versus 51.7).

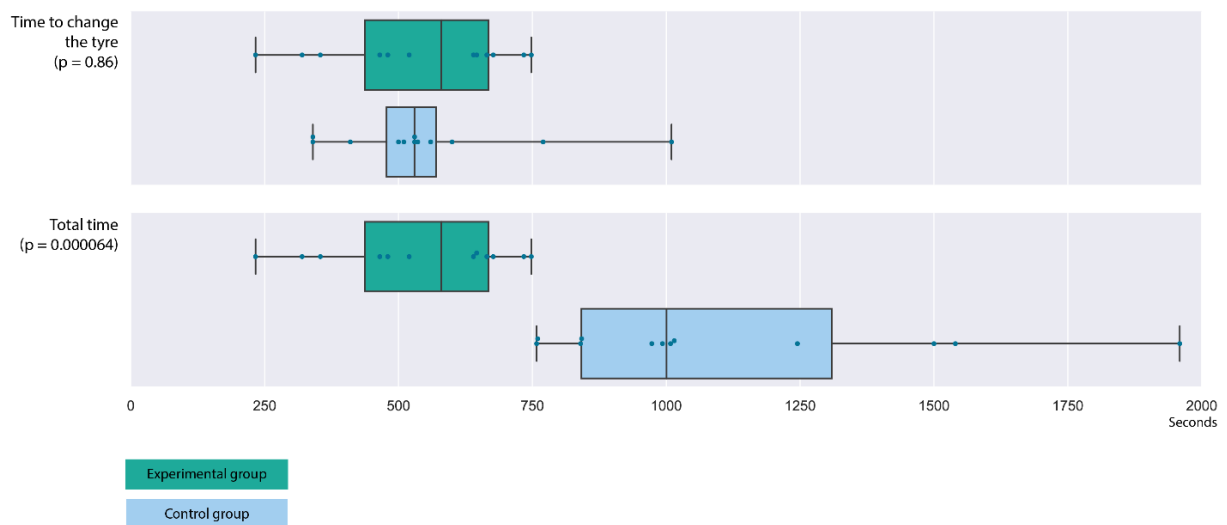


Figure 42 Time factors



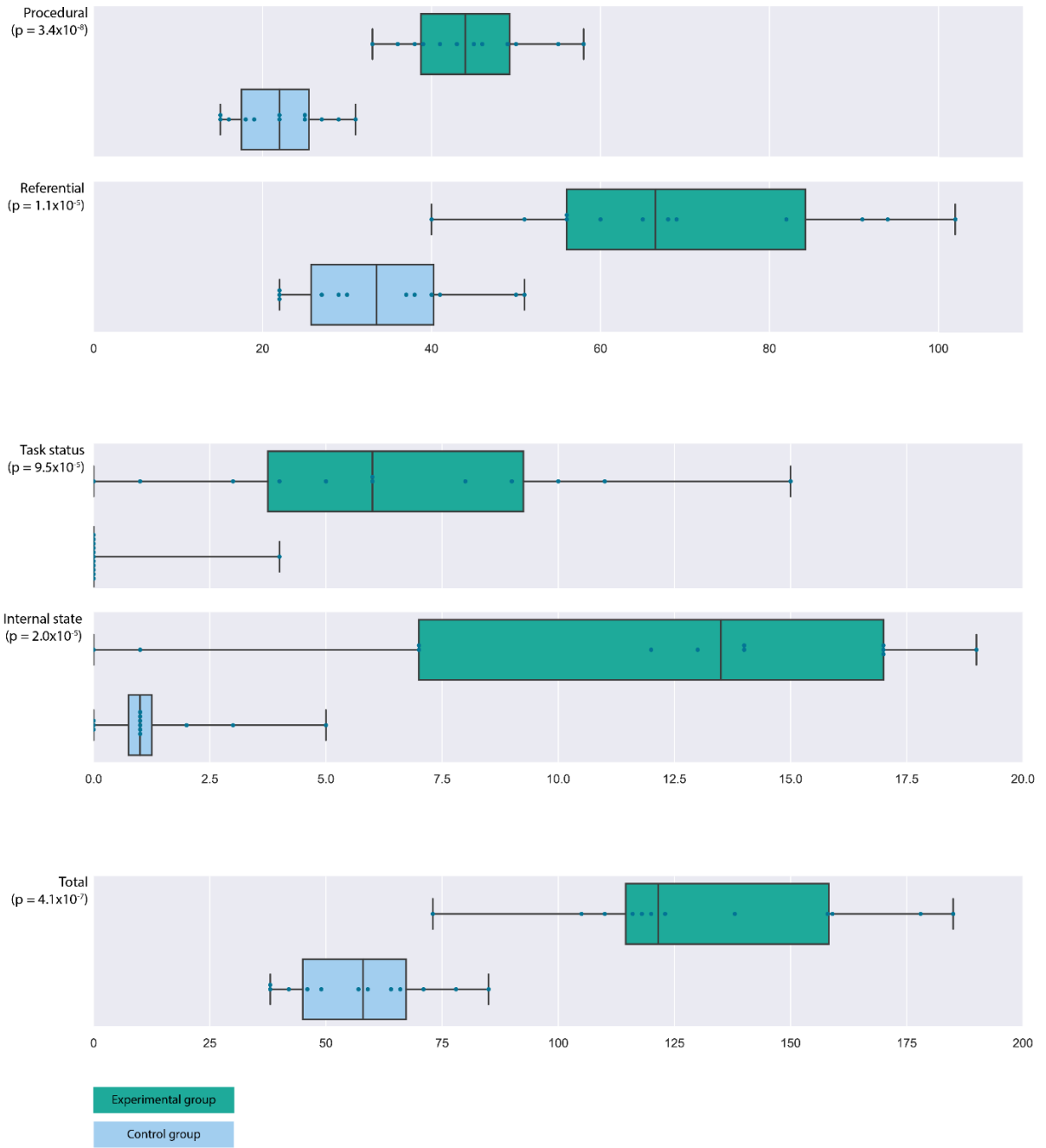


Figure 43 Relevant utterances factors (reporting quality)

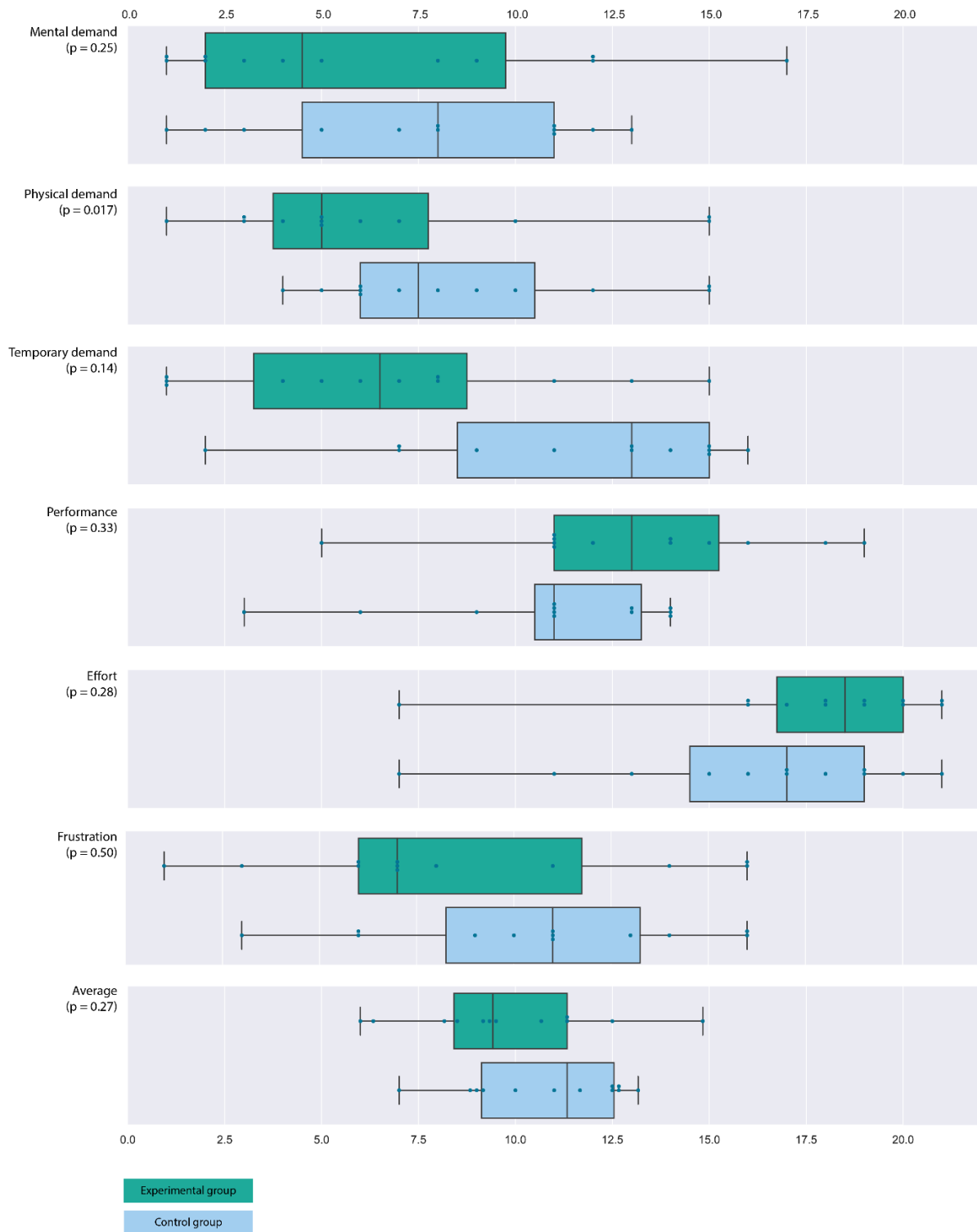


Figure 44 Workload factors

Variable	p-value	sig	Control mean (median)	Experimental mean (median)
Total score	0.274	ns	51.7 (54.0)	46.7 (44.8)
Mental demand	0.250	ns	8.58 (7.5)	6.58 (5)
Physical demand	0.0167	(*)	11.4 (13)	6.67 (6.5)
Temporal demand	0.134	ns	10.8 (11)	13.1 (13)
Performance	0.333	ns	16.1 (17)	17.7 (18.5)
Effort	0.282	ns	10.5 (11)	8.5 (7)
Frustration	0.495	ns	7.67 (8)	6.33 (4.5)

**Table 18 NASA-TLX perceived workload factors**

Variable	p-value	sig	Control mean (median)[sec]	Experimental mean (median)[sec]
Time to change a tyre	0.862	ns	553 (530)	540 (580)
Total time	6.64E-05	(**)	1120 (1000)	540 (580)

**Table 19 Time factors**

Utterance type	p-value	sig	Control mean (median)	Experimental mean (median)
Procedural	3.36E-08	(**)	22 (22)	44.4 (44)
Task status	9.47E-05	(**)	0.333 (0)	6.5 (6)
Referential	1.05E-05	(**)	34.1 (33.5)	69.5 (66.5)
Internal state	2.01E-05	(**)	1.33 (1)	11.5 (13.5)
Total	4.06E-07	(**)	57.8 (58)	132 (121.2)

**Table 20 Relevant utterances (report quality)**

---

## 5.6.7 DISCUSSION

---

### 5.6.7.1 REPORTING QUALITY

All four reporting quality indexes scored significantly in favour of the experimental group.

This result could be explained by several factors: (1) when thinking-aloud “live” as they conducted the task, the experimental group could easily recall and report more details, (2) because talking is far more efficient than writing on paper, participants are inclined to provide more details (3) the experimental participants were less concise in their descriptions and explanations, (4) experimental participants who took longer or had challenges performing the task inevitably talked more. Conversely, control participants, usually only reported the “correct steps” to complete the task, thereby excluding any exploratory steps or explanations.

---

### 5.6.7.2 TIME FACTORS

The results show that the experimental group takes significantly less time to complete the task. This could be attributed to the fact that the experimental group could effectively multitask by reporting as they performed the task. Furthermore, using speech to report takes less time than writing the report down on paper. It should also be noted that in terms of relevant utterances, the control group reports were roughly half the length of the experimental group reports.

There was an indication (not statistically significant) that the tyre changing part of the task took slightly longer for the experimental group. The additional time could be attributed to the fact that (1) participants sometimes paused to answer a question from the conversational agent, (2) they paused to

explain something in more depth or (3) thinking-aloud imposed a cognitive load that affected their ability to perform the tyre-changing task.

---

### 5.6.7.3 PERCEIVED WORKLOAD FACTORS

The perceived physical workload was the only index that resulted in a significant t-test. This was unexpected as the physical part of the task (changing the tyre) was the same for both groups. Possible explanations could be: (1) the total task time for the control group was significantly longer and possibly more mentally demanding, making them feel more tired. They then report the feeling of tiredness as physically demanding, (2) talking aloud while changing the tyre takes their attention away from the physical demands, (3) the process of writing a report can be tiring for your hand/arm. (This was reported by one participant only).

Participants reported that thinking aloud helped them with their reasoning process.

Control participants mentioned that having to remember what they did afterwards was cognitively demanding.

---

## 5.6.8 QUALITATIVE ANALYSIS OF THE UTTERANCE TYPES

Experimental participants frequently referred to components, tools, actions etc. with pronouns, as if their conversation partner could see what they were doing. In the last statement in the excerpt below (Figure 46), the participant is referring to the wheel, however, this is not evident from the previous conversation.

U: I can tighten the small silver ring to make it tight

U: put on the black ring to again

A: yeah

U: place it back. ((referring to the wheel))

Figure 45 Pronoun usage by participant 12

Information contained in “internal states”, e.g. an explanation on why the participant was doing something or providing a tip is a valuable source of information as seen in the excerpt below (Figure 47). These occurred far more frequently in the experimental group.

U: sometimes the inner tube gets caught between the outer tube and the rim

A: yeah

U: and that would cause problems or more often when it's flat.

Figure 46 Internal state by participant 12

Information contained in “task status” can provide information regarding the state of components – useful for maintenance planning. In the excerpt below (Figure 48), a

participant remarks that the tyre levers are bending.

U: maybe I put them here

A: okay

U: I might this thing is bending, the tyre lever is bending

Figure 47 Task status participant 5

### 5.6.9 QUALITATIVE ANALYSIS REGARDING THE DEVELOPMENT OF THE SOLUTION

Throughout the experiment, various qualitative observations were made that were relevant to the implementation of the solution. These are outlined below and in Table 21 and Table 22.

In many cases, the participants spoke at length without pausing more than a few seconds. Therefore, there were not many opportunities for the conversational agent to process the data and respond. For the final solution to work in these scenarios, the conversational agent will need to be proactive

User-behaviour related opportunities	Agent opportunity	Complexity
Users are not significantly slowed down when thinking aloud	Stimulate the user to talk more through “continuers”, sounding interested, asking follow-ups	Low
Users rarely needed to be reminded to describe what they were doing.	Less action required	Low
Users are good at identifying information that could be useful for colleagues/novices (this may be different for an expert who is less conscious of their thought process).	When a user indicates it has a useful tip but doesn't provide additional explanation, ask for more details. Possibly also show “your appreciation” for them sharing.	Medium

Table 21 Opportunities for the CAIA based on a qualitative analysis of experimental transcript

User-behaviour related challenges	Agent solution	Complexity
Incomplete utterances	Ask users to finish statement or ask for missing information	Medium/High
Pauses in the middle of utterances	Agent identifies statement as incomplete and waits for further information	Medium/High
Pausing and then repeating utterances	Agent will need to determine that the statement is a repeat and ignore it.	Low
Frequent use of pronouns	Agent can ask the user what they are referring to, e.g. "what's do you mean by this?"	Low
Referring to objects as if the agent could see what they are doing	Same as above	Low
Continuous speech with little room for the agent to ask questions	The agent will need to split long utterances up into sub-utterances. The agent will need to interrupt the user.	High
The gaps for the agent to interject are fleeting	Same as above	High
Not always clear when the user moves to the next "step" - step boundaries are sometimes fuzzy.	If the agent can't identify which step the user is referring to, it will need to ask for clarification.	Medium
Users may stop their work to answer a question that requires them to think	Minimise complex/cognitively intense questions unless the benefit is clear, e.g. providing explanation for a crucial step	Medium/High

Table 22 Challenges for the CAIA based on a qualitative analysis of experimental transcripts

and occasionally interrupt the user or save its queries and wait for the user to stop. The latter option is more technically feasible but less natural.

Utterances frequently didn't have a clearly defined beginning and end, which will make NLP challenging.

The procedural descriptions were sometimes poorly structured, therefore, one could argue that not every procedural utterance can be considered equal in quality.

The researcher rarely needed to ask "how are you doing?" to stimulate interaction (triggered when the participants were silent for >15 seconds).

## 5.6.10 CONCLUSIONS

In conclusion, using a CAIA for reporting saves a significant amount of time, results in reports of higher quality (i.e. they contained more utterances that were relevant for understanding or performing the assigned task), and a reduced physical workload. Furthermore, there are indications that other



workload indexes are reduced but more data is needed to confirm this.

---

#### 5.6.11 REFLECTION ON THE EXPERIMENT

For this experiment, it was assumed that the CAIA could automatically generate a completely accurate report. However, with the current technology available, users will need to proofread/check the generated report, resulting in additional time. Furthermore, the written reports contain more information than currently found in industry as the participants were asked to include explanations than might help novice colleagues. It should be pointed out that most participants in the experiment chose not to include any explanations, as evident in the number of internal state utterances.

The indexes used to score reporting the reporting quality relied on the researcher's subjective judgment to perform the coding correctly and award points to relevant utterances. Furthermore, it is arguable how well these indexes can be used to fairly compare the quality of the control and experimental results. As mentioned in the qualitative analysis, participants in the experimental group tended to report all their troubleshooting steps and thought processes, whereas the control group participants typically only reported on the steps they eventually used to complete the task (Both groups had been given the same instructions: report on what, how and why they were accomplishing the task). Depending on the use of the reports, either one could be considered more valuable. However, if the goal is to collect as much information as possible, it is clear that the experimental group is superior.

Lastly, this experiment focused entirely on using word-based natural language, whereas it might have been more efficient to make a drawing or take a picture of a step.

---

#### 5.6.12 RECOMMENDATIONS FOR FUTURE RESEARCH

This experiment was designed to validate the potential value of using a CAIA-based reporting system. The results and qualitative analysis of the experiment have highlighted various interesting areas for future research.

- 1) How can the information collected by the CAIA be effectively used? This research points to the following uses: to provide tips for maintenance technicians, as input for a predictive maintenance model, for training purposes, to capture and transfer tacit knowledge.
- 2) Which other ways can the CAIA proactively facilitate (rich) information capture? For example, identify steps for which the technician could provide an additional explanation?
- 3) Explore the value and feasibility of multi-modal reporting e.g. combining speech with videos, images or sketches.
- 4) How to use computer vision to track the context to supplement the speech-input.
- 5) Further data collection to isolate the independent variables: (1) written versus spoken reporting and (2) reporting during the task versus afterwards.

#### 5.7 SOLUTION REQUIREMENTS

Additional requirements specific to the chosen use case, automatic reporting of corrective maintenance work, are presented here.

The requirements listed below are based on the features demonstrated in the scientific experiment. They represent a minimal viable product that demonstrates the potential for the proposed solution. Multiple requirements stipulate that they should be achieved in 75% of the instances. This per cent is an arbitrary number that I specified as an indicator that the prototype is reliable enough for user testing.

---

### 5.7.1 FUNCTIONAL REQUIREMENTS

- 1) The CAIA should correctly extract at least 75% of the relevant entities (actions, components and tools) from user utterances.
- 2) The CAIA should be able to generate a step-by-step report based on the interaction with the user.
- 3) The CAIA should be able to discern between different user intents (e.g. description of task step, send the report by email, provide an answer to a question, etc) for at least 75% of the instances.
- 4) The CAIA should attempt to determine the true meaning of pronouns and ask for clarification when unsure for at least 75% of the instances (compared to a human listener).
- 5) The CAIA should utter a continuer or interjection when appropriate (e.g. when a user utters a procedural statement or provides an answer to a query).
- 6) The CAIA should be able to send the generated report to the user after their conversation.
- 7) Ask the user for an update after they had been silent for 15 seconds.

---

### 5.7.2 USABILITY REQUIREMENTS

- 1) The CAIA should be able to identify scenarios in which it should remain silent.

- 2) The CAIA should respond rapidly when indicating it has understood the user's utterance.

---

### 5.7.3 TECHNICAL REQUIREMENTS

- 1) The CAIA should be able to interface with the host company's reporting system

---

### 5.7.4 SCOPE OF DEVELOPMENT AND TESTING

The different functional requirements pose widely varying challenges. The ability to send a report by email is relatively straightforward as others have already developed similar solutions. Conversely, getting the CAIA to use the conversational context to determine whether or not to ask for clarification when the user uses a pronoun is a unique feature that requires significant development work. Therefore, I chose to implement (a simplified) version of this in the final prototype, namely, the agent will support respond to usages of "it" in procedural statements. This represents the most common pairing of statement type and pronoun.

The desired behaviour is that when the user uses "it" to describe a component within a procedural statement and it is not obvious what they are referring to from the context, the agent should ask for clarification. To achieve this, the prototype will need various supporting features, such as (1) the ability to extract and track components (including when they are referred to by pronouns), (2) classify the intent of user statements (e.g. procedural utterances, answers to clarifications and requests).

The usability requirements will not be tested as it would require redesigning how the existing utterance capturing mechanism

provided by Mycroft works. This would take a significant amount of time and is out-of-scope for this study.

The technical requirement regarding the interface to the host company is not being tested as I do not have access to it.

## 5.8 PROTOTYPING

The purpose of the prototype to evaluate the technical feasibility of the proposed solution by testing if it meets the functional requirements defined above. The requirements related to entity extraction and intent classification were tested using conversations from the scientific experiment.

Developing a CAIA for reporting is challenging as there are no existing solutions to learn from and existing frameworks are not built with this use case in mind. Therefore, I evaluated two different approaches to creating a

reporting CAIA (1) Using Rasa's FormPolicy, which provides a structured way for the CAIA to ask for missing slots and (2) a free-form approach. I tested the FormPolicy approach first because it was simpler and easier to test. Further explanation is provided below.

A third and final prototype was built based on the learnings from the previous iterations and scientific experiment.

As Rasa does not support a voice interface directly, the text-to-speech and speech-to-text functionality was handled by Mycroft (Figure 49).

---

### 5.8.1 KEY COMPONENTS

In Section 4.1.5, System architecture, the components that constitute a CAIA as outlined. These are revisited here with respect to the functional requirements outlined above (Section 5.7.1).



Figure 48 Mycroft setup on a Raspberry Pi (for the voice interface)

---

### 5.8.1.1 ENTITY EXTRACTION

The basic requirement for entity extraction is to provide labelled training data for each user intent (e.g. procedural utterance or answer to a question). A single training example for the phrase "I'm going to put the inner tube in the outer tyre" is shown in Figure 50 below. Rasa learns to extract the entities based on several features such as the position in the sentence and meaning of the word.

```
I'm going to [put>{"entity": "action", "group": "1"} the [inner tyre>{"entity": "component", "role": "primary"} in the [outer tyre>{"entity": "component", "role": "secondary"}
```

Figure 49 Labelled training data

Normally, Rasa would then automatically overwrite the respective slots (e.g. action, component or tool) with the extracted entity, however, for the reporting use case we would like to validate the data first and append it to slots as opposed to overwriting them. Therefore, additional custom code is required. The captured information can then be used to generate a report.

---

### 5.8.1.2 INTENT CLASSIFICATION AND STORIES

Modern CAIAs rely on an intent-based system to determine how to respond to user utterances. Basic intents include things like "greeting", "affirmative", "deny", and "goodbye". Also, the prototype may feature intents for "send the report by email", "show me the report", "procedural statement" and "answer question". Rasa determines which "story" to follow based on the classified intent, extracted entities and previous conversation steps. Stories are predefined conversation paths. A basic example is a greeting, e.g. if the user

intent is "greet", the agent should respond with a greeting. For the prototype, the requirements stipulate that the agent should respond with a continuer whenever the user provides a "procedural" utterance. Stories are also used to define when custom code should be executed, e.g. when the user intent is "procedural" utterance, the agent should check for any missing information and append the relevant entities to slots.

---

### 5.8.1.3 CONNECTOR TO EXTERNAL SERVICES

As per the fourth functional requirement, the agent should be able to send the generated report by email. This involves writing some custom code that utilises Gmail's API. A similar approach will be needed to read and write to the host company's reporting database.

The feature that asks the user for an update if they have been silent for 15 seconds requires using custom code that accesses the computer's clock and sets a Rasa "reminder".

---

### 5.8.2 FORMPOLICY PROTOTYPE

For the first iteration, I used an inbuilt Rasa feature called FormPolicy. In a scenario when the CAIA needs to fill a set of slots (e.g. machine name, start time, symptoms, tools used etc), the training "stories" can become lengthy and complex when the users frequently stray from the standard order of providing information and asking questions. A FormPolicy can be seen as a sub-thread of the main story and allows the CAIA to ask for the missing information. Resultantly, the main "story" is greatly simplified. A test conversation with the FormPolicy prototype is shown in Figure 51 below. Note that the agent

(A= agent, U = User) leads the conversation by asking for specific pieces of information.

U: hello  
A: Hello! I'm a maintenance assistant! How can I help?  
U: I'd like to make a report  
A: what device are you fixing?  
U: I'm fixing my computer  
A: what are the symptoms?  
U: it's getting very hot  
A: what caused the problem  
U: the dust caused the problem  
A: How did you fix the problem?  
U: I cleaned it out  
U: I restarted the computer and it was fixed  
A: which parts did you use?  
U: I replaced the battery  
A: Which tools did you use?  
U: I used a screwdriver  
A: I am going to create a maintenance report with the following formation:  
Machine: computer  
Symptoms: very hot  
Cause: dust  
Fix: used a  
Parts: battery  
Tools: screwdriver

Figure 50 FormPolicy prototype demonstration

The weaknesses of using a FormPolicy are:

- 1) The agent enforces a rigid structure to the conversation that cannot adapt to the situation or user (e.g. when the user is not ready to provide a certain piece of information).
- 2) It relies on a single, "catch-all", intent for all inputs. This is problematic for a few

reasons: (1) all information inputs are handled the same (i.e. it cannot discern between procedural utterances or internal states) which limits how intuitively the agent can respond, (2) the entity extractor is more likely to miss entities.

The positives of these characteristics are that the added structure makes it is easier to build a reliable prototype and it could potentially guide a user through a proven root-cause-analysis procedure. The fact that the FormPolicy uses a "catch-all" intent means that it cannot fulfil requirement 3. A FormPolicy might be useful to use at the beginning or end of a job to capture some basic details if the user had not mentioned them yet (e.g. the time, the machine number etc). For the main phase of a job, it makes more sense to have a CAIA that allows the user to take initiative in providing descriptions and explanation. This led me to develop the "flexible" prototype outlined below.

---

### 5.8.3 FREE-FORM PROTOTYPE

The second prototype is built on a more flexible basis because it can differentiate between different types of utterances (e.g. procedural, task status, internal state, inquiry).

This requires more planning and preparation up front. One way to bootstrap this process is to use interactive learning. When interactive learning is active, the agent determines a response but asks for confirmation or correction before continuing as seen in Figure 58 below. This allows the developer to build more conversation varieties without having to explicitly type them out.



## 5.8.4 FINAL PROTOTYPE

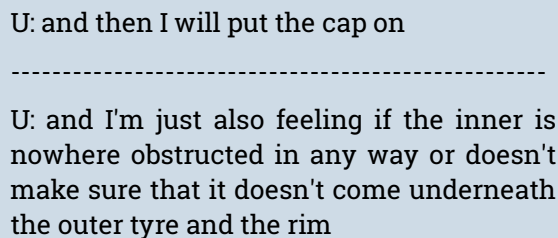
Intents and stories are relatively simple, human-understandable mechanisms that allow developers to design conversations. They work well for short interactions like asking for the weather, playing some music or for some simple troubleshooting help. However, they become a hindrance if conversations become longer or more complex. This shortcoming is fueling research to create end-to-end conversational agents. However, for the time being, building a CAIA outside of an experimental research lab still involves using intents and stories.

The first challenge for the final prototype is to decide what mechanisms to use to achieve the required functionality and translate this into intents, entities and custom code.

### 5.8.4.1 INTENTS AND STORIES

Intents for sending the report by email or greeting are straightforward as the interaction is short and the user utterance distinct (i.e. easy for the intent classifier to discern from other intents).

One of the challenges for classifying reporting intents is that the utterance structure, content and length of varies widely within a conversation and between users as demonstrated in Figure 52 below.

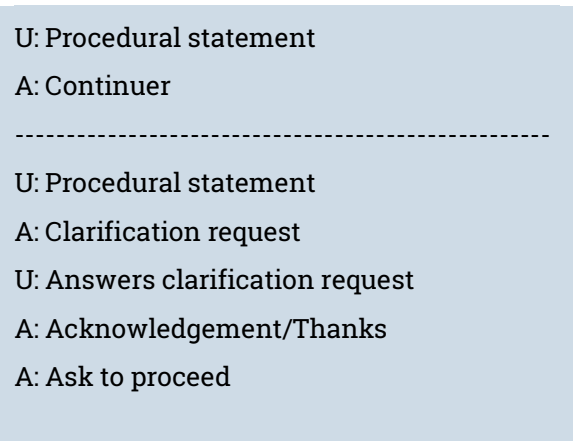


U: and then I will put the cap on  
-----  
U: and I'm just also feeling if the inner is nowhere obstructed in any way or doesn't make sure that it doesn't come underneath the outer tyre and the rim

Figure 51 Two different procedural statements

This makes it difficult for the intent classifier and entity extractor to spot patterns. Processing the long utterance that features multiple conjunctions are very difficult for current NLP models and beyond the scope of this project unless they are broken down into shorter utterances.

The prototype relies on two intents for processing reporting utterances (1) regular procedural statements and (2) when the user answers the agent's request for clarification. This allows the stories to be kept short, containing not more than four interaction pairs (see Figure 53).



U: Procedural statement  
A: Continuer  
-----  
U: Procedural statement  
A: Clarification request  
U: Answers clarification request  
A: Acknowledgement/Thanks  
A: Ask to proceed

Figure 52 Basic stories

The basic interaction consists of the user uttering a procedural statement followed by the agent extracting relevant information and then uttering a continuer or clarification request. This system results in simple stories and allows additional "logic" to be added to the custom code without needing to completely re-write a large number of complex stories. The support story paths are visualized in Figure 55 below (the blue boxes are user intents, the white boxes are agent responses and actions). The context tracking and logic regarding the usage of pronouns for components occurs in the



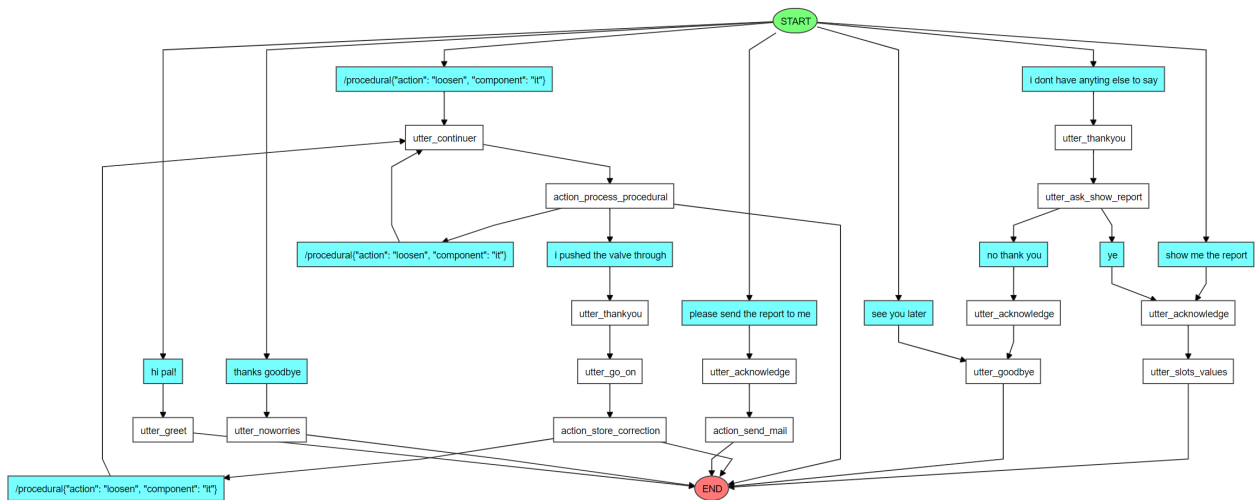


Figure 54 Visualisation of stories

“action\_process\_procedural” step, a Python class that runs on the “action server”.

#### 5.8.4.2 ENTITIES AND CONTEXT

The final prototype extracts and tracks three types of entities: components, actions and tools. It can discern between primary and secondary components, e.g. in steps when the user refers to an action they perform on one component (primary) relative to another (secondary), as seen in the Figure 54 below. The entity extractor also captures both parts of

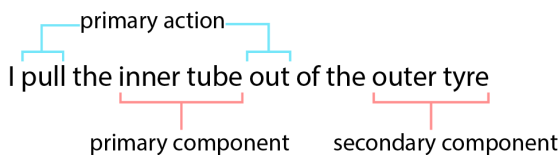


Figure 53 Components and actions

the primary action. It should be noted that this utilises an experimental feature of the Rasa framework and may be discontinued in future releases.

The following example (Figure 56) features a tool and a referral to the primary component by a pronoun. In this case, there is only one component that the action “inflate” could refer to, the inner tube. Therefore, the agent assumes that is what the user is referring to. When it is not obvious what the primary component is, the agent formulates a question based on the extracted action, e.g. “what did you *pull out?*”.

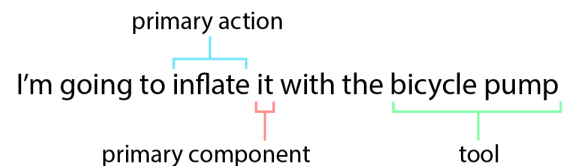


Figure 55 Entity extraction with a pronoun

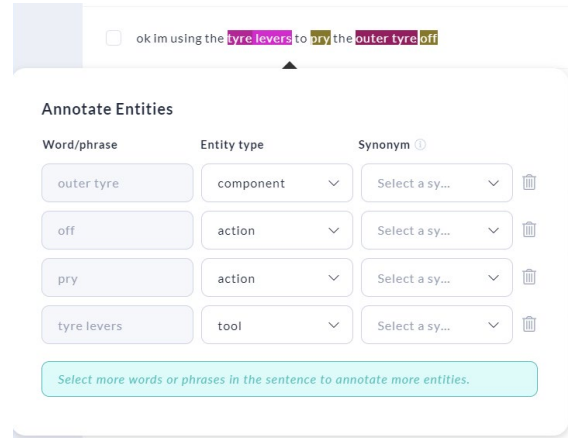
The final prototype was built feature-by-feature and improved over various iterations. The transcripts from the scientific experiment were used to train the intent classifier and the entity extractor. The development process is outlined below:

- 1) Plan the required intents, entities, stories and custom logic.
- 2) Use the transcripts from the experiment to provide labelled training examples for the intent classifier and entity extractor, test and iterate. This is a lengthy process as every entity must be manually labelled, as seen below.

```
I [push>{"entity":"action", "group":"1"} the
[outer tyre>{"entity":"component",
"role":"primary"} [over>{"entity":"action",
"group":"1"} the [tyre
rim>{"entity":"component",
"role":"secondary"}
```

**Figure 56 Training example for intent classification and entity extraction**

- 3) Define agent responses and stories, test and iterate.
- 4) Write code that handles the “pronoun clarification feature”, test and iterate.
- 5) Write code that handles email integration, test and iterate.
- 6) Write code that provides integration to Mycroft (open-source voice assistant framework) and create a skill in Mycroft that connects to the Rasa server, test and iterate.
- 7) Further testing and improvements including interactive learning as seen below in Figure 58. Interactive learning allows testing the assistant and directly correct intent classification and entity extraction.



**Figure 57 Interactive learning with Rasa X**

- 8) Add support for multiple components and actions per utterance, test and iterate.
- 9) Add timestamps to steps and a start and end time to the maintenance report, as seen below.

```
(1) 17:26:08 Action(s): remove, Component(s):
outer tyre, Tool(s): none
```

**Figure 58 Timestamp added to step description**

- 10) Add status request feature (e.g. “how’s it going?”) for when the user is silent for more than 15 seconds (as used in the user test). Note: this feature only works when running Rasa X.

The logic behind the handling of the pronouns and tracking the context is contained in two Python classes that run on the “custom action” server. Additionally, two classes are used for the status request feature (not included in the final prototype code). Any external connectors, like the functions to send the report by mail, also run on this server. See the code in Appendix 9.8.3. The key logic regarding asking for clarification for usages of “it” to describe a component is outlined in pseudo-code below (Figure 60).

```

IF the primary extract component is "it"
{
    IF primary action contains "pump"
    or "inflate"
        Primary component is
"inner tube"
    ELSE IF the previous primary
component is not "it" or "none"
        Primary component is
previous component
    ELSE IF previous primary
component is "if" or "none"
    {
        IF primary action exists
            Send message:
            "what did you
            [primary action]?"
        ELSE
            Send message:
            "what do you mean
            by it?"
    }
}
}

```

Figure 59 Pseudo-code pronoun clarification feature

An alternative mechanism to provide the same feature would be to use more complex "stories". The story flow could be informed based on the extracted entities. Thereby avoiding the need to rely nested if statements running on a separate server. Either way, both strategies are very time consuming to develop and require the developer to explicitly define what the agent should do for every possible pronoun and component combination. A more powerful solution would be to use machine learning to train a model to predict the meaning of a pronoun based on the rest of the utterance (e.g. using the BERT transformer model (Devlin, Chang, Lee, & Toutanova, 2019)).

By tracking the entities over time, the agent can create a step-by-step report of the task with timestamps. It also compiles the components, tools and actions taken in separate lists. Lastly, it keeps a record of the raw user utterances. An example report is shown in Figure 61. Note that it is generated by the CAIA and contains some errors, like extracting "tyre" as a tool instead of "tyre lever".

### 5.8.5 TEST OF THE FINAL PROTOTYPE

The final test serves to demonstrate the technical feasibility of the solution as specified by the functional requirements. As the voice interface provided by Mycroft would require significant tweaking to accommodate user testing beyond the scope of this project (e.g. it needs to "listen" for longer than typically necessary and process user utterances more quickly), the prototype was tested using the transcripts from the scientific experiment.

The testing consists of multiple parts, designed to evaluate the different functional requirements (Section 5.7.1) as objectively as possible.

Some of the requirements are straightforward to test as it is a matter of supporting a feature or not. These include requirement 2, 5, 6 and 7, namely: the ability to generate a step-by-step report, uttering a continuer when applicable, send the report by email and ask the user for an update by email. All of these have been demonstrated to work in the final prototype except for requirement 7, which worked temporarily but at the time of writing doesn't work due to a bug in the Rasa framework. The bug is related to the "reminder" feature and should be fixed in future Rasa updates.

Dear Maintenance Tech,

I've collected the following information from our conversation:

Start time: 19:33:06 on Wed, 01 Jul 2020

End time: 19:35:22

Extracted steps: (1) 19:33:06 Action(s): none, Component(s): quick release lever, Tool(s): none (2) 19:33:29 Action(s): remove, Component(s): wheel, Tool(s): none (3) 19:33:55 Action(s): use remove, Component(s): outer tyre, rim, Tool(s): tyre (4) 19:34:49 Action(s): unscrew, Component(s): valve cap, ring, Tool(s): none (5) 19:35:22 Action(s): pull out, Component(s): inner tube, outer tube, Tool(s): none

Raw user input: (1) i start by turning the quick release lever (2) i remove the wheel from the dropouts (3) i use the tyre levers to remove the outer tyre from the rim (4) then unscrew the valve cap and ring (5) pull the inner tube out of the outer tube

Tools: (3) tyre

Components: (1) quick release lever (2) wheel (3) outer tyre, rim (4) valve cap, ring (5) inner tube, outer tube

Actions: (2) remove (3) use remove (4) unscrew (5) pull out

Kind regards,

The maintenance bot

Figure 60 A report emailed by the CAIA

The other requirements (1, 3, 4) require a more rigorous approach to testing as it is highly unlikely that the CAIA will fulfil the requirements at all times. The requirements stipulate that they should exhibit the desired behaviour for at least 75% of the instances.

There are numerous examples from the experiment that can be used to test the entity extraction. However, during the experiment, the participants never had to request the report be sent by email and the clarification requests were worded differently by the agent.

Therefore, these requirements will be evaluated qualitatively using a combination of experiment transcript excerpts and synthesised utterances designed to demonstrate the capabilities and limitations.

---

#### 5.8.5.1 ENTITY EXTRACTION

The entity extraction performance according to the inbuilt cross-validation method is 84%. Additionally, 10 utterances, which are not in the training data were tested. These resulted in an accuracy rate of 81.5% (originally there

were 11 test utterances but the intent classifier misclassified one of them). It should be noted that accuracy varied greatly per entity as shown in Table 23 (the full table with the test utterances can be seen in Appendix 9.10.1). The poor performance of tool and secondary action entity extraction could be attributed to their infrequency of usage.

The utterances used to test the entity extractor were “clean” and relatively short (i.e. they did not contain any repetitions, figures of speeches or confusing sentence structure) and thus comparable to the training data in complexity.

### 5.8.5.2 INTENT CLASSIFICATION

The inbuilt Rasa cross-validation method reported an accuracy of 86%. The results are depicted in a confusion matrix (Appendix 9.10.3Figure 66). It shows that the most frequent error is when the CAIA incorrectly predicts clarification request answers (“answer\_question”) as a “procedural” utterance (“procedural”). This is unsurprising as they both contain component entities and a similar sentence structure.

### 5.8.5.3 COMPONENT CONTEXT AND CLARIFICATION

Eleven short excerpts were selected from the scientific experiment transcript where “it” was used to describe a component and used to test the prototype. The prototype was given a point for each time it correctly assumed the true meaning of “it” or correctly asked the user to provide clarification. The agent was awarded

half a point if it was partially correct. The resulting accuracy was 60%. The excerpts and scoring can be seen in Appendix 9.10.2 The prototype relies on a few “If” statements to determine how to handle the context, namely(1) if the primary action is “pump” or “inflate” it assumes that the primary component is “inner tube”, (2) if there is a non “it” component in an utterance, it assumes that it is the primary component, (3) if the primary component is “it” and the previous utterance contained a non-“it” primary component, it assumes that the new “it” is the same. As shown by the results, these relatively simple assumptions worked in 60% of the cases, however, it still fails at some simple scenarios that would be obvious to a human. The example below (Figure 62) shows a scenario in which a human would likely understand that the user is still referring to the valve when they say “I push it through”, however, the agent assumes that they meant the “ring” which is identified as the primary component in the previous utterance.

```

u: I'm going to put the inner tyre in the outer
  tyre and start with the valve again
a: yeah
u: you need to take the ring off
a: ok
u: I push it through
a: yeah (assumed "it" = "ring")
  
```

Figure 61 Incorrect component assumption

Based on these results, additional logic could be added to improve the performance for the

	components	components	Action	Action		correct		
Per cent	85% (10)	100% (4)	90% (10)	0% (1)	25%	81.5%	14.8%	16.7%
					(2)			

Table 23 Entity extractor accuracy

inner tube bike changing scenario (e.g. when “it” is pulled out of the outer tube, “it” = “inner tube” or when “tyre levers” are used, assume that “it” = “outer tyre” or to develop a mechanism to track component context over more than two utterances). However, the downside of this technique is that it does not translate well to other tasks or scenarios as the logic is very specific. As previously mentioned, future research could explore how to approach this problem by using machine learning.

## 5.9 ROADMAP FOR FUTURE DEVELOPMENT

The current CAIA prototype was built to demonstrate the technical feasibility of the features as specified by the functional requirements in Section 5.7.1.

The prototype was trained using data from the experiment that involved changing a bicycle inner tube (Section 5.6). As discussed in Section 5.8.5, Test of the final prototype, the prototype achieved all but one of the requirements. It should be noted that the prototype was trained and tested on “clean” user utterances (concise and well-structured sentences). To improve the overall performance and robustness of the prototype, a roadmap for future development is outlined below.

---

### 5.9.1 PERFORMANCE (SPEED OF THE PROCESSING)

In its current form, the prototype takes a few seconds to process and extract the information from each utterance. This is largely due to the Python code running on the separate “action server”. With more training data and more intent types (e.g. split the procedural intent into multiple (sub)intents),

the solution can be made less dependent on hard-coded logic on the “action server”. This should improve the flexibility of the CAIA and improve the processing time.

---

### 5.9.2 ACCURACY OF INTENT CLASSIFIER AND ENTITY EXTRACTION

The featurisers used in the prototype’s NLU pipeline are trained on general English vocabulary and utterances. If featurisers were training on industrial maintenance conversations, the accuracy of the intent classifier and entity extraction should significantly increase. This will require significant data collection.

As previously mentioned, the entity role and group feature that was used in the prototype has proven very useful for tracking the relationship between entities, however, it is an experimental feature that may be discontinued in future releases.

---

### 5.9.3 HANDLING “REAL” SPEECH

The prototype works reliably when the user speaks in well-structured, concise sentences. However, regular human speech is frequently poorly structured and fragmented, as evident in the transcripts of the scientific experiment. There are few strategies to consider here: (1) train the users to speak in a manner that is more understandable for the agent or (2) develop a custom preprocessor that “tidies” the text input up at the beginning of the NLU pipeline (3) train the intent classifier and entity extractor using a large volume of data so that it learns to deal with the “low quality” input.



---

#### 5.9.4 MORE ADVANCED CONTEXT TRACKING AND INFORMATION REQUESTS

During the experiment, the simulated CAIA asked for clarifications when the participant referred to components, actions or tools without providing sufficient detail or context for the CAIA to understand (e.g. using pronouns like “it” and “this” or referring to a component using a general term like “mechanism”). The final prototype incorporates a context tracking mechanism for components, actions and tools but it only asks for clarification when components are referred to be “it” and it cannot discern the true

meaning from the context. This scenario represents one of many in which the CAIA might need to ask for clarification. The current mechanism utilises hard-coded logic that is time-intensive to programme but reliable for a single scenario. However, as the solution grows in features, e.g. be able to recognise more scenarios in which I should ask for clarification, the hard-coded logic may become extremely complex and restrictive. Therefore, when more data is available, it is recommended to rely more on machine learning to determine if a response is necessary.

## 6 EVALUATION AND CONCLUSIONS

This project is an exploration of potential applications of Conversational AI Agents (CAIA) for industrial maintenance. It involves the scientific validation and development of a CAIA for a specific application: the capture of (rich) information for the automatic creation of maintenance reports during the execution of industrial maintenance.

The high-level requirements of this project were: (1) the solution should solve key challenges that the industry is facing, (2) the solution should leverage core affordances of CAIAs, (3) the solution fits the needs and interests of the end-users, and (4) the solution should be technically feasible. How these requirements are met is outlined in the following sections.

### 6.1 THE APPLICATION

The selection of the application, using a CAIA for maintenance reporting, was based on literature research, (in-situ) context analyses (Section 4.2.3) and a review of CAIA frameworks and design guidelines (Section 4.1). This research revealed that maintenance workers rely heavily on their own experience and intuition when solving problems but mechanisms for capturing and accessing this were non-existent. This knowledge is highly valuable and can represent a significant part of a company's worth. Once captured, it can be safeguarded and transferred to other employees. Furthermore, the maintenance technicians reported that the existing maintenance reporting mechanisms were inconvenient or a "pain". Lastly, audits had revealed that maintenance reports were frequently incomplete or of poor quality.

In turn, CAIAs (Conversational AI Agents), have various affordances that make them well-suited to the context of industrial maintenance. They are (1) hands and gaze-free, (2) highly efficient (for recording text compared to writing or typing and provided faster access to specific information), (3) they can adapt to the skill level of the user and (4) impose a minimal cognitive load (Section 4.1).

### 6.2 SCIENTIFIC EXPERIMENT

A between-subjects experiment with 24 participants was used to test three hypotheses regarding the potential value of the solution (Section 5.6). These were: by using a CAIA for reporting during task completion as opposed to writing the reports on paper afterwards would result in (1) reports of higher quality (more information relevant to understanding and learning about the task), (2) save time and (3) lower perceived workload (NASA TLX). T-tests confirmed that all three hypotheses were true. Critically, it demonstrated that the CAIA could facilitate the capture of valuable "expert knowledge" such as the inclusion of explanations and tips alongside a step-by-step description of the completed tasks.

### 6.3 PROTOTYPE

The CAIA features used in the experiment are demonstrated in a prototype built in the open-source Rasa and Mycroft frameworks (Section 5.8.4). These include (1) using continuers and interjections after user statements, (2) tracking context, (3) asking for clarifications when there is a lack of mutual understanding and (4) asking for status updates when the user is silent for a while (a full list of requirements can be found in Section 5.7.1). The prototype fulfilled all but one of the seven functional requirements, the ability to track component context and ask for clarification

when there was a lack of mutual understanding. The prototype achieved an accuracy of 60%, whereas, the requirement had set an arbitrary minimum of 75%. This is no surprise given that the prototype employs rudimentary logic that still has some flaws, as discussed in the testing (Section 5.8.5). Generally, the tests have demonstrated that using a CAIA for this application has huge potential but further work is necessary to uncover this potential fully. Because the collection of more conversational data is important for future developed, maintenance technicians could use a speech-to-text system to record their work, as seen in the medical industry. This would allow the capture of a high volume of data that could be used to develop the next generation of the CAIA.

#### 6.4 CONVERSATIONAL DESIGN METHOD

The Natural Conversation Framework (Moore & Aral, 2019) was used as a reference to understand the components and theory behind conversation design. However, the development process outlined by the framework is tailored towards typical conversational user experiences, e.g. for booking a flight or frequently asked questions. Ultimately, this project focused on the validation and implementation of a single, novel feature for a CAIA, automatic maintenance reporting. To do this, I strayed from the standard process by employing a more extensive conversation analysis and using a scientific experiment to validate the value and create new knowledge. Nevertheless, for a good foundation of knowledge about conversation design and conversation analytics I would still strongly recommend the book by Moore & Aral (2019), *Conversation UX Design*, for any designer developing a CAIA.

#### 6.5 RELEVANCE BEYOND THIS PROJECT

The knowledge generated through the scientific experiment (Section 5.6) has implications for other domains besides industrial maintenance. It indicates that applying think-aloud techniques with a speech-to-text system is an efficient way to elicit rich user knowledge. This could be relevant for other fields where expert knowledge plays an important role in troubleshooting such as in medicine, IT support and vehicle maintenance.

The techniques employed to facilitate a long interaction between the CAIA and user, generate a report from the entities and clarify misunderstandings are innovative applications for CAIAs. These include using the context from previous utterances to determine the meaning of pronouns and asking for clarification when unable to this (Section 5.8.4). The experiment showed that users frequently use pronouns when referring to components. This is also likely the case in other domains and applications, therefore, the techniques employed here could be relevant elsewhere.

#### 6.6 FUTURE WORK

Future research could explore integrating multi-modal information capture (e.g. through Smart glasses), additional uses for the captured data (e.g. for prescriptive maintenance or providing tips) and improving the existing solution on functionality and usability (e.g. tracking context over a longer period and improving the response time).

## 7 REFLECTION

For my graduation, I was keen to ceaselessly push my comfort zone, expand my areas of expertise and demonstrate my capabilities as an industrial design engineer. Reflecting on the activities undertaken and the accomplishments I have achieved, I believe I was successful in these respects. To name a few (1) I performed field research in two factories abroad as part of DIAMOND, an EIT project, (2) I performed a scientific experiment to evaluate the potential value of a new application for CAIAs, (3) I developed a CAIA prototype based on previous learnings to demonstrate the technical feasibility of the proposed solution. This was especially challenging as the scope of the project focused on the digital aspects of the solution, whereas, my area of expertise as an industrial design engineer is skewed towards physical product development. However, it did involve applying data analytics and machine learning, an area that I've been exploring in recent years.

There were two major challenges I faced throughout the project (1) narrowing the scope of the project to the point that I could generate new knowledge and (2) how to convey the knowledge regarding CAIAs. Numerous additional factors contributed to these challenges such as the travel limitations and social distancing required due to the COVID-19 pandemic, the fact that CAIAs is a relatively

novel topic for both myself and the Faculty and the explorative nature of the project itself. In hindsight, I could have narrowed the focus earlier in the project as this may have reduced the time pressure at the latter part of the project. If I had more time, I would have used it to develop the prototype further. I am especially interested in exploring how to improve the context tracking over multiple turns, using machine learning models to predict the meaning of pronouns and to improve the usability of the voice interface.

Ultimately, this project was primarily focused on evaluating the technical feasibility of some of the underlying features of the CAIA, whereas the practical usability of the prototype in the field is untested. Similarly, although the solution attempted to solve a problem that many maintenances faced, user acceptance still needs to be evaluated further. These are the logical next steps to fully demonstrate the potential value of the proposed solution.

In my project brief, I stated that I wanted to expand my expertise regarding complex intelligent systems, implemented innovative technologies and address the needs and expectations of the involved stakeholders. All things considered, I believe that I have achieved these goals. I hope that this project has set the groundwork for future research and development.

## 8 REFERENCES

- #PlacethatMatter: Łódź, Poland. It is written “Łódź” but it is... | by Whirlpool EMEA | Whirlpool Corporation | Medium. (n.d.). Retrieved July 16, 2020, from <https://medium.com/whirlpool-corporation/placethatmatter-łódź-poland-70243bb38cf4>
- Ahamed, S. I., Sharmin, M., Ahmed, S., Hague, M. M., & Khan, A. J. (2006, April). Design and implementation of a virtual assistant for Healthcare professionals using pervasive computing technologies. *Elektrotechnik Und Informationstechnik*. Springer. <https://doi.org/10.1007/s00502-006-0335>
- AirPods Pro Active Noise Cancellation and Transparency mode - Apple Support. (n.d.). Retrieved May 22, 2020, from <https://support.apple.com/en-us/HT210643>
- Albrecht, U. V., Von Jan, U., Kuebler, J., Zoeller, C., Lacher, M., Muensterer, O. J., ... Hagemeyer, L. (2014). Google glass for documentation of medical findings: Evaluation in forensic medicine. *Journal of Medical Internet Research*, 16(2), e53. <https://doi.org/10.2196/jmir.3225>
- Alesich, S., & Rigby, M. (2017). Gendered Robots: Implications for Our Humanoid Future. *IEEE Technology and Society Magazine*, 36(2), 50–59. <https://doi.org/10.1109/MTS.2017.2696598>
- Alexa Fund. (n.d.). Retrieved May 22, 2020, from <https://developer.amazon.com/en-US/alexa/alexa-fund>
- Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., ... Trippe, E.-B. D. (2017). Text Summarization Techniques: A Brief Survey. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Amazon Announces 6 New HIPAA Compliant Alexa Skills. (n.d.). Retrieved May 22, 2020, from <https://www.hipaajournal.com/hipaa-compliant-alexa-skills/>
- Ameka, F. (1992). Interjections: The universal yet neglected part of speech. *Journal of Pragmatics*, 18(2–3), 101–118. [https://doi.org/10.1016/0378-2166\(92\)90048-G](https://doi.org/10.1016/0378-2166(92)90048-G)
- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... Horvitz, E. (2019). Guidelines for human-AI interaction. In *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290605.3300233>
- Ansari, F., Erol, S., & Sihni, W. (2018). Rethinking Human-Machine Learning in Industry 4.0: How Does the Paradigm Shift Treat the Role of Human Learning? In *Procedia Manufacturing*. <https://doi.org/10.1016/j.promfg.2018.04.003>
- Aromaa, S., Heimonen, T., Vääänen, A., & Aaltonen, I. (2015). A model for gathering and sharing knowledge in maintenance work. <https://doi.org/10.1145/2788412.2788442>
- Aschenbrenner, D., Rojko, M., Leutert, F., Verlinden, J., Lukosch, S., Latoschik, M. E., & Schilling, K. (2018). Comparing Different Augmented Reality Support Applications for Cooperative Repair of an Industrial Robot. In *Adjunct Proceedings - 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2018*. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00036>
- “Athena, Transform Manufacturing”: Voice-Activated Machining Technology in Multiple Demos: Modern Machine Shop. (n.d.). Retrieved July 16, 2020, from <https://www.mmsonline.com/blog/post/athena-transform-manufacturing-voice-activated-machining-technology-in-multiple-demos>
- Automotive AI Assistants : BMW AI assistant. (n.d.). Retrieved May 30, 2020, from <https://www.trendhunter.com/trends/bmw-ai-assistant>
- Belciug, S., & Gorunescu, F. (2020). Intelligent Decision Support Systems - A Journey to Smarter Healthcare. In *Intelligent Decision Support Systems - A Journey to Smarter Healthcare* (pp. 71–102). <https://doi.org/10.1007/978-3-030-14354-1>
- Berg, M. M. (2015). *Modelling of Natural Dialogues in the Context of Speech-based Information and Control Systems*. [https://doi.org/10.1007/978-3-319-19581-0\\_12](https://doi.org/10.1007/978-3-319-19581-0_12)
- Buist, A. H., Kraaij, W., & Raaijmakers, S. (2004). Automatic Summarization of Meeting Data: A Feasibility Study.
- Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Calero Valdez, A. (2019). “Hey,

- Siri”, “Ok, Google”, “Alexa”. Acceptance-Relevant Factors of Virtual Voice-Assistants. In *IEEE International Professional Communication Conference*. <https://doi.org/10.1109/ProComm.2019.00025>
- Campbell, E. M., Sittig, D. F., Ash, J. S., Guappone, K. P., & Dykstra, R. H. (2006). Types of Unintended Consequences Related to Computerized Provider Order Entry. *Journal of the American Medical Informatics Association*. <https://doi.org/10.1197/jamia.M2042>
- Chen, H., Liu, X., Yin, D., & Tang, J. (2017). A Survey on Dialogue Systems: Recent Advances and New Frontiers, *19*(2), 25–35. Retrieved from <http://arxiv.org/abs/1711.01731>
- Conversation design process—Is conversation the right fit? (n.d.). Retrieved May 22, 2020, from <https://designguidelines.withgoogle.com/conversation/conversation-design/what-is-conversation-design.html>
- De Crescenzo, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., & Salti, S. (2011). Augmented reality for aircraft maintenance training and operations support. *IEEE Computer Graphics and Applications*, *31*(1), 96–101. <https://doi.org/10.1109/MCG.2011.4>
- Desouza, K. C. (2003). Facilitating tacit knowledge exchange. *Communications of the ACM*, *46*(6), 85–88. <https://doi.org/10.1145/777313.777317>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Retrieved from <https://github.com/tensorflow/tensor2tensor>
- Does Your Doctor Need a Voice Assistant? | WIRED. (n.d.). Retrieved May 22, 2020, from <https://www.wired.com/story/does-your-doctor-need-a-voice-assistant/>
- Dopico, M., Gomez, A., De la Fuente, D., García, N., Rosillo, R., & Puche, J. (2016). A vision of industry 4.0 from an artificial intelligence point of view. In *Proceedings of the 2016 International Conference on Artificial Intelligence, ICAI 2016 - WORLDCOMP 2016*.
- Du, W., & Black, A. W. (2019). Boosting Dialog Response Generation. <https://doi.org/10.18653/v1/p19-1005>
- Explore Focals - North. (n.d.). Retrieved May 30, 2020, from <https://www.bynorth.com/focals>
- Fabbri, D., & Ehrenfeld, J. M. (2016, October 1). Hands Free Automatic Clinical Care Documentation: Opportunities for Motion Sensors and Cameras. *Journal of Medical Systems*. Springer New York LLC. <https://doi.org/10.1007/s10916-016-0570-x>
- Falk, J., Poulakos, S., Kapadia, M., & Sumner, R. W. (2018). PICA: Proactive intelligent conversational agent for interactive narratives. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA 2018*. <https://doi.org/10.1145/3267851.3267892>
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender Bias in Chatbot Design. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11970 LNCS, pp. 79–93). Springer. [https://doi.org/10.1007/978-3-030-39540-7\\_6](https://doi.org/10.1007/978-3-030-39540-7_6)
- Fessler, L. (2017). We tested bots like Siri and Alexa to see who would stand up to sexual harassment. *Quartz*.
- Finley, G. P., Edwards, E., Robinson, A., Sadoughi, N., Fone, J., Miller, M., ... Axtmann, N. (2018). *An automated medical scribe for documenting clinical encounters* \*.
- Friedberg, M. W., Chen, P. G., Van Busum, K. R., Aunon, F., Pham, C., Caloyer, J., ... Tutty, M. (2014). Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. *Rand Health Quarterly*.
- Gao, X., Lee, S., Zhang, Y., Brockett, C., Galley, M., Gao, J., & Dolan, B. (2019). Jointly Optimizing Diversity and Relevance in Neural Response Generation. <https://doi.org/10.18653/v1/n19-1125>
- Ghosh, S., & Pherwani, J. (2015). Designing of a natural voice assistants for mobile through user centered design approach. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/978-3-319-20901-2\\_29](https://doi.org/10.1007/978-3-319-20901-2_29)
- Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., Werf, R. J. van der, & Morency, L.



- L.-P. (2006). *Intelligent Virtual Agents* (pp. 14–27). Springer.
- Gupta, I., Di Eugenio, B., Ziebart, B., Liu, B., Gerber, B., Sharp, L., ... Baiju, A. (2018). Towards building a virtual assistant health coach. In *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018* (pp. 419–421). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICHI.2018.00081>
- Hearables Market Size, Share & Growth | Trends & Analysis - 2026. (2019). Retrieved May 31, 2020, from <https://www.alliedmarketresearch.com/hearables-market>
- Henderson, S., & Feiner, S. (2011). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10), 1355–1368. <https://doi.org/10.1109/TVCG.2010.245>
- Hexagon Invests in New Smart Factory | Fabricating and Metalworking. (n.d.). Retrieved July 16, 2020, from <https://www.fabricatingandmetalworking.com/2017/06/hexagon-invests-in-new-smart-factory/>
- Highest Paying Diesel Mechanic Jobs - Diesel Mechanic Guide. (n.d.). Retrieved July 16, 2020, from <https://dieselmechanicguide.com/highest-paying-diesel-mechanic-jobs/>
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*. <https://doi.org/10.1126/science.aaa8685>
- How Alexa, Siri, and Google Assistant Will Make Money Off You | MIT Technology Review. (n.d.). Retrieved May 22, 2020, from <https://www.technologyreview.com/2016/05/31/8199/how-alex-siri-and-google-assistant-will-make-money-off-you/>
- How do I get started? - Conversation design process - Conversation design. (n.d.). Retrieved May 31, 2020, from <https://designguidelines.withgoogle.com/conversation/conversation-design-process/how-do-i-get-started.html>
- How to activate “Hey, Cortana!” in Windows 10 Laptop / PC. (n.d.). Retrieved July 16, 2020, from <https://sguru.org/activate-hey-cortana-windows-10-laptop-pc/>
- Huang, M., Zhu, X., & Gao, J. (2020). Challenges in Building Intelligent Open-domain Dialog Systems. *ACM Transactions on Information Systems*. <https://doi.org/10.1145/3383123>
- Initiatives - Mycroft. (n.d.). Retrieved May 22, 2020, from <https://mycroft.ai/initiatives/>
- Intelligent Virtual Assistant Market Size | Industry Report, 2027. (n.d.). Retrieved May 22, 2020, from <https://www.grandviewresearch.com/industry-analysis/intelligent-virtual-assistant-industry>
- Kääriä, A. (2017). Technology Acceptance of Voice Assistants: Anthropomorphism As a Factor. *University of Jyväskylä*. <https://doi.org/10.1177/1460458216660754>
- Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., ... Noh, S. Do. (2016). Smart Manufacturing: Past Research, Present Findings, and Future Directions. *INTERNATIONAL JOURNAL OF PRECISION ENGINEERING AND MANUFACTURING-GREEN TECHNOLOGY*, 3(1), 111–128. <https://doi.org/10.1007/s40684-016-0015-5>
- Kans, M. (2019). within the Swedish maintenance ecosystem Maintenance in the digital era - an interview study of challenges and opportunities within the Swedish maintenance ecosystem. In I. Arranz & M. C. Karim (Eds.) (pp. 68–73). Retrieved from <https://www.diva-portal.org/smash/get/diva2:1426298/FULLTEXT01.pdf>
- Kessler, S. K., & Martin, M. (2017). How do potential users perceive the adoption of new technologies within the field of Artificial Intelligence and Internet-of-Things? A revision of the UTAUT 2 model using Voice Assistants, (May), 1–90. Retrieved from <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8909840&fileId=8909844>
- Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*. [https://doi.org/10.1207/S15327051HCI1812\\_2](https://doi.org/10.1207/S15327051HCI1812_2)
- Krupitzer, C., Lesch, V., Züfle, M., Kounev, S.,

- Müller, S., Edinger, J., ... Lemken, A. (2020). A Survey on Human Machine Interaction. *ArXiv Preprint* *ArXiv:2002.01025*. <https://doi.org/10.1145/1122445.1122456>
- L. Waskom, M. (2020). Seaborn. Retrieved from <https://github.com/mwaskom/seaborn>
- L'Abbate, M., & Thiel, U. (2003). The use of contextual information in a proactivity model for conversational agents. In *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*. [https://doi.org/10.1007/3-540-44958-2\\_38](https://doi.org/10.1007/3-540-44958-2_38)
- L'Abbate, M., Thiel, U., & Kamps, T. (2005). Can proactive behavior turn chatterbots into conversational agents? In *Proceedings - 2005 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT'05*. <https://doi.org/10.1109/IAT.2005.49>
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., ... Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*. <https://doi.org/10.1093/jamia/ocy072>
- Large, D. R., Burnett, G., Anyasodo, B., & Skrypchuk, L. (2016). Assessing Cognitive Demand during Natural Language Interactions with a Digital Driving Assistant. <https://doi.org/10.1145/3003715.3005408>
- Lee, J., Lapira, E., Bagheri, B., & Kao, H. an. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*. <https://doi.org/10.1016/j.mfglet.2013.09.005>
- Leitão, P., Karnouskos, S., Ribeiro, L., Lee, J., Strasser, T., & Colombo, A. W. (2016). Smart Agents in Industrial Cyber-Physical Systems. *Proceedings of the IEEE*. <https://doi.org/10.1109/JPROC.2016.2521931>
- Liu, Y. C., & Ou, Y. K. (2011). Effects of Age and the Use of Hands-Free Cellular Phones on Driving Behavior and Task Performance. *Traffic Injury Prevention*, *12*(6), 550–558. <https://doi.org/10.1080/15389588.2011.607197>
- Lorenz, M., Rüßmann, M., Strack, R., Lueth, K. L., & Bolle, M. (2015). How Will The Technology Transform the Industrial Workforce Through 2025. *Man and Machine in Industry 4.0*.
- Lotterbach, S., & Peissner, M. (2005). Voice user interfaces in industrial environments. In *INFORMATIK 2005 - Informatik LIVE!, Beitrage der 35. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*.
- Lotti, G., Villani, V., Battilani, N., & Fantuzzi, C. (2018). Towards an integrated approach for supporting the workers in Industry 4.0. In *Proceedings - 2018 IEEE Industrial Cyber-Physical Systems, ICPS 2018*. <https://doi.org/10.1109/ICPHYS.2018.8390775>
- McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2019.05.009>
- Mitchell, W. J., Ho, C. C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? Explicit-implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2010.09.002>
- Moore, R. J., & Aral, R. (2019). *Conversational UX Design*. (M. Tamer Ozsu, Ed.). Association for Computing Machinery. <https://doi.org/10.1145/3304087>
- Nguyen, Q. N., Ta, A., & Prybutok, V. (2019). An Integrated Model of Voice-User Interface Continuance Intention: The Gender Effect. *International Journal of Human-Computer Interaction*, *35*(15), 1362–1377. <https://doi.org/10.1080/10447318.2018.1525023>
- Nichol, A. (2019). It's About Time We Get Rid of Intents. Retrieved May 24, 2020, from <https://blog.rasa.com/its-about-time-we-get-rid-of-intents/>
- Nihei, & Nakano. (2019). Exploring Methods for Predicting Important Utterances Contributing to Meeting Summarization. *Multimodal Technologies and Interaction*, *3*(3), 50. <https://doi.org/10.3390/mti3030050>
- Norman, D. (1988). The Design of Everyday Things (Originally published: The psychology of everyday things). In *The Psychology of Everyday Things*. <https://doi.org/10.2307/1423268>

- Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics*, 41(5), 866–891. <https://doi.org/10.1016/j.pragma.2008.08.005>
- Number of voice assistants in use worldwide 2019-2023. (2019). Retrieved from <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/>
- Nunes, L., & Recarte, M. A. (2002). Cognitive demands of hands-free-phone conversation while driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 133–144. [https://doi.org/10.1016/S1369-8478\(02\)00012-8](https://doi.org/10.1016/S1369-8478(02)00012-8)
- Pfeifer, L. M., & Bickmore, T. (2009). *Should Agents Speak Like, um, Humans? The Use of Conversational Fillers by Virtual Agents*. *LNAI* (Vol. 5773).
- Plattner, H. (2016). An Introduction to Design Thinking; Process Guide. *Institute of Design at Stanford*. [https://doi.org/10.1007/978-1-4302-6182-7\\_1](https://doi.org/10.1007/978-1-4302-6182-7_1)
- Polanyi, M. (1966). The tacit dimension, knowledge in organizations. *L. Prusak, Ed.*
- Ponathil, A., Ozkan, F., Welch, B., Bertrand, J., & Chalil Madathil, K. (2020). Family health history collected by virtual conversational agents: An empirical study to investigate the efficacy of this approach. *Journal of Genetic Counseling*, jgc4.1239. <https://doi.org/10.1002/jgc4.1239>
- Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3173574.3174214>
- Predictive Maintenance Template | Azure AI Gallery. (n.d.). Retrieved July 16, 2020, from <https://gallery.azure.ai/Collection/Predictive-Maintenance-Template-3>
- Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology: A scoping review. *Journal of Medical Internet Research*. <https://doi.org/10.2196/jmir.6553>
- Quint, F., & Loch, F. (2015). Using Smart Glasses to Document Maintenance Processes in: *Mensch und Computer 2015 – Workshopband. Mensch Und Computer 2015 – Workshopband*, 203–208. Retrieved from <https://www.degruyter.com/view/book/9783110443905/10.1515/9783110443905-030.xml>
- Quiroz, J. C., Laranjo, L., Kocaballi, A. B., Berkovsky, S., Rezazadegan, D., & Coiera, E. (2019). Challenges of developing a digital scribe to reduce clinical documentation burden. *Npj Digital Medicine*, 2(1), 1–6. <https://doi.org/10.1038/s41746-019-0190-1>
- Rateau, H., Clay, A., & Bottecchia, S. (2018). Combining digital pen, tactile, and AR for technical documentation during maintenance procedures, 2p. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Reason, J. T. (1997). *Managing the risks of organizational accidents*. Ashgate.
- Ring, L., Barry, B., Totzke, K., & Bickmore, T. (2013). Addressing loneliness and isolation in older adults: Proactive affective agents provide better support. In *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*. <https://doi.org/10.1109/ACII.2013.17>
- Romero, D., Bernus, P., Noran, O., Stahre, J., & Berglund, Å. F. (2016). The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *IFIP Advances in Information and Communication Technology* (Vol. 488, pp. 677–686). Springer New York LLC. [https://doi.org/10.1007/978-3-319-51133-7\\_80](https://doi.org/10.1007/978-3-319-51133-7_80)
- Sankar, C., Subramanian, S., Pal, C., Chandar, S., & Bengio, Y. (2019). Do Neural Dialog Systems Use the Conversation History Effectively? An Empirical Study. <https://doi.org/10.18653/v1/p19-1004>
- Setchi, R., & White, D. (2003). The development of a hypermedia maintenance manual for an advanced manufacturing company. *International Journal of Advanced Manufacturing Technology*. <https://doi.org/10.1007/s00170-002-1513-x>
- Shabbir Moiyadi, H., Desai, H., Pawar, D., Agrawal, G., MPatil, N., & Gandhi, R. (2016). NLP Based Text Summarization Using Semantic Analysis. *International Journal of Advanced Engineering*,

- Management and Science (IJAEMS)*, 2(10). Retrieved from [www.ijaers.com](http://www.ijaers.com)
- Shachak, A., Hadas-Dayagi, M., Ziv, A., & Reis, S. (2009). Primary care physicians' use of an electronic medical record system: A cognitive task analysis. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-008-0892-6>
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*. <https://doi.org/10.2307/2333709>
- Smart Earbuds for Hearing Enhancement | Nuheara. (n.d.). Retrieved May 22, 2020, from <https://www.nuheara.com/>
- Smart glasses in aid of maintenance. (n.d.). Retrieved July 16, 2020, from <https://www.theagilityeffect.com/en/article/smart-glasses-aid-maintenance/>
- Smart Speaker Market by IVA, Component. (2018). Retrieved May 22, 2020, from <https://www.marketsandmarkets.com/Market-Reports/smart-speaker-market-44984088.html>
- Smart Wearable Market | Growth, Trends, Forecast (2020-2025). (n.d.). Retrieved May 22, 2020, from <https://www.mordorintelligence.com/industry-reports/smart-wearables-market>
- Sun, Y., Yuan, N. J., Wang, Y., Xie, X., McDonald, K., & Zhang, R. (2016). Contextual intent tracking for personal assistants. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Aug, 273-282. <https://doi.org/10.1145/2939672.2939676>
- Useful artificial intelligence apps for your mobile phones. (n.d.). Retrieved May 30, 2020, from <https://bigdata-madesimple.com/useful-artificial-intelligence-apps-for-your-mobile-phones/>
- Valencell - The most accurate PPG Sensors for wearable technology. (n.d.). Retrieved May 22, 2020, from <https://valencell.com/>
- Voice Assistant Consumer Adoption Report*. (2018). Retrieved from <https://voicebot.ai/wp-content/uploads/2019/01/voice-assistant-consumer-adoption-report-2018-voicebot.pdf>
- Wachter, R., & Goldsmith, J. (2018). To Combat Physician Burnout and Improve Care, Fix the Electronic Health Record. *Harvard Business Review Digital Articles*.
- Weidauer, A. (2018). Your Guide to Five Levels of AI Assistants in Enterprise. Retrieved May 30, 2020, from <https://blog.rasa.com/conversational-ai-your-guide-to-five-levels-of-ai-assistants-in-enterprise/>
- Weizenbaum, J. (1966). ELIZA-A computer program for the study of natural language communication between man and machine. *Communications of the ACM*. <https://doi.org/10.1145/365153.365168>
- West, M., Kraut, R., & Chew, H. E. (2019). *I'd blush if I could: closing gender divides in digital skills through education. Unesco-Equals*.
- Wireless & Bluetooth Headphones | Bose. (n.d.). Retrieved May 22, 2020, from [https://www.bose.com/en\\_us/products/headphones.html](https://www.bose.com/en_us/products/headphones.html)
- Wu, C.-S., Madotto, A., Hosseini-Asl, E., Xiong, C., Socher, R., & Fung, P. (2019). Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems. <https://doi.org/10.18653/v1/p19-1078>
- Wu, W., Guo, Z., Zhou, X., Wu, H., Zhang, X., Lian, R., & Wang, H. (2020). Proactive human-machine conversation with explicit conversation goals. In *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.
- Xu, L. Da, He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*. <https://doi.org/10.1109/TII.2014.2300753>
- Zhang, L., Luo, Y., Tao, F., Li, B. H., Ren, L., Zhang, X., ... Liu, Y. (2014). Cloud manufacturing: a new manufacturing paradigm. *Enterprise Information Systems*. <https://doi.org/10.1080/17517575.2012.683812>
- Zhu, C., Xu, R., Zeng, M., & Huang, X. (2020). *End-to-End Abstractive Summarization for Meetings*.

## 9 APPENDICES

### 9.1 PROJECT BRIEF

DESIGN  
FOR OUR  
future

TU Delft

## IDE Master Graduation

### Project team, Procedural checks and personal Project brief

This document contains the agreements made between student and supervisory team about the student's IDE Master Graduation Project. This document can also include the involvement of an external organisation, however, it does not cover any legal employment relationship that the student and the client (might) agree upon. Next to that, this document facilitates the required procedural checks. In this document:

- The student defines the team, what he/she is going to do/deliver and how that will come about.
- SSC E&SA (Shared Service Center, Education & Student Affairs) reports on the student's registration and study progress.
- IDE's Board of Examiners confirms if the student is allowed to start the Graduation Project.

**! USE ADOBE ACROBAT READER TO OPEN, EDIT AND SAVE THIS DOCUMENT**  
Download again and reopen in case you tried other software, such as Preview (Mac) or a webbrowser.

**STUDENT DATA & MASTER PROGRAMME**  
Save this form according to the format "IDE Master Graduation Project Brief\_familyname\_firstname\_studentnumber\_dd-mm-yyyy". Complete all blue parts of the form and include the approved Project Brief in your Graduation Report as Appendix 1 !

<p>family name <u>Kernan Freire</u></p> <p>initials <u>S</u> given name <u>Samuel</u></p> <p>student number <u>4091787</u></p> <p>street &amp; no. _____</p> <p>zipcode &amp; city _____</p> <p>country _____</p> <p>phone _____</p> <p>email _____</p>	<p>Your master programme (only select the options that apply to you):</p> <p>IDE master(s): <input checked="" type="checkbox"/> IPD <input type="checkbox"/> Dfl <input type="checkbox"/> SPD</p> <p>2<sup>nd</sup> non-IDE master: _____</p> <p>individual programme: _____ (give date of approval)</p> <p>honours programme: <input type="checkbox"/> Honours Programme Master</p> <p>specialisation / annotation: <input type="checkbox"/> Medisign</p> <p><input type="checkbox"/> Tech. in Sustainable Design</p> <p><input type="checkbox"/> Entrepreneurship</p>
---	---

**SUPERVISORY TEAM \*\***  
Fill in the required data for the supervisory team members. Please check the instructions on the right !

<p>** chair <u>Zoltán Rusák</u> dept. / section: <u>SDE/IoT</u></p> <p>** mentor <u>Doris Aschenbrenner</u> dept. / section: <u>SDE/M&amp;M</u></p> <p>2<sup>nd</sup> mentor _____</p> <p>organisation: _____</p> <p>city: _____ country: _____</p> <p>comments (optional) _____</p>	<p>Chair should request the IDE Board of Examiners for approval of a non-IDE mentor, including a motivation letter and c.v.</p> <p><b>!</b> Second mentor only applies in case the assignment is hosted by an external organisation.</p> <p><b>!</b> Ensure a heterogeneous team. In case you wish to include two team members from the same section, please explain why.</p>
--	---

IDE TU Delft - E&SA Department /// Graduation project brief & study overview /// 2018-01 v30 Page 1 of 7

Procedural Checks - IDE Master Graduation

**APPROVAL PROJECT BRIEF**

To be filled in by the chair of the supervisory team.

chair Zoltán Rusák date 12.02.2020 signature [Signature]

**CHECK STUDY PROGRESS**

To be filled in by the SSC E&SA (Shared Service Center, Education & Student Affairs), after approval of the project brief by the Chair. The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total: 27 EC  
 Of which, taking the conditional requirements into account, can be part of the exam programme: 27 EC

YES all 1<sup>st</sup> year master courses passed

NO missing 1<sup>st</sup> year master courses are:

List of electives obtained before the third semester without approval of the BoE

[Empty box for list of electives]

[Empty box for missing 1<sup>st</sup> year master courses]

name \_\_\_\_\_ date 24.2.2020 signature CB

**FORMAL APPROVAL GRADUATION PROJECT**

To be filled in by the Board of Examiners of IDE TU Delft. Please check the supervisory team and study the parts of the brief marked \*\*. Next, please assess, (dis)approve and sign this Project Brief, by using the criteria below.

- Does the project fit within the (MSc)-programme of the student (taking into account, if described, the activities done next to the obligatory MSc specific courses)?
- Is the level of the project challenging enough for a MSc IDE graduating student?
- Is the project expected to be doable within 100 working days/20 weeks?
- Does the composition of the supervisory team comply with the regulations and fit the assignment?

Content:  APPROVED  NOT APPROVED

Procedure:  APPROVED  NOT APPROVED

[Empty box for comments]

comments

name Mv Mergin date 2-3-20 signature [Signature]



A digital intelligence assistant for predictive maintenance operators project title

Please state the title of your graduation project (above) and the start date and end date (below). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

start date 03 - 02 - 2020 19 - 06 - 2020 end date

### INTRODUCTION \*\*

Please describe, the context of your project, and address the main stakeholders (interests) within this context in a concise yet complete manner. Who are involved, what do they value and how do they currently operate within the given context? What are the main opportunities and limitations you are currently aware of (cultural- and social norms, resources (time, money,...), technology, ...).

The DIAMOND project aims to develop a DIA (Digital Intelligent Assistant) to facilitate predictive maintenance at industrial facilities (see figure 1). Predictive maintenance involves monitoring the status of a machine such that failures can be predicted and preemptively fixed. It involves the processing of lots of data using machine learning algorithms to create predictive models. Modern industrial machines contain countless sensors which can be used to collect the data (see figure 2). Many companies and industries stand to benefit from the implementation of predictive maintenance as it can help avoid sudden failures and can make scheduled maintenance procedures more efficient.

DIAMOND is a collaboration of academic, industry and government stakeholders. EIT (European Institute for Innovation and Technology), TU Delft, BIBA (Bremen Institute for production and logistics), Whirlpool (a home appliances manufacturer) and Stryker (a medical equipment manufacturer).

The TU Delft will explore the various use scenarios for the DIA, the dialogue design and user interface of the information retrieval platform. Upon completion of the project, TU Delft can apply the knowledge in their education programme and spread it to other interested parties. The TU Delft is interested in the research and educational value through the DIAMOND project and the potential to foster new corporate ties. BIBA (Bremen Institute for Production and Logistics) will be coordinating the project and responsible for the development of the DIA. Like the TU Delft, they will also be interested in the lessons learned and new partners. Whirlpool and Stryker stand to benefit from the DIA if the goals regarding cost reductions and efficiency are met. It is in the interest of each company for DIAMOND to work on use scenarios that are most relevant to them. Furthermore, both companies operate in different industries and likely have different company cultures. EIT (European Institute for Innovation and Technology) provides the funds for the project and sees it as an investment for the EU. If successful, the resulting knowledge could help improve the efficiency of manufacturing plants across the EU. Other companies stand to benefit from the implementation of the development results. They will be appreciative of the open-source nature of the project.

#### Opportunities

- I will have access to at least two large companies, Whirlpool and Stryker, for research and testing purposes.
- The potential to have a significant impact through the consortium's network, open-source results and creation of a start-up
- The timing of various technologies (open-source voice-assistants, predictive maintenance, augmented reality, digital twins) that allow the development of the DIA.

#### Limitations

- A tight budget imposed on the project by EIT may pose problems if it limits the trips we can make to the factories or to meet with partners.
- The clashing needs of the various stakeholders may result in slower and turbulent development.

space available for images / figures on next page

Personal Project Brief - IDE Master Graduation

introduction (continued): space for images

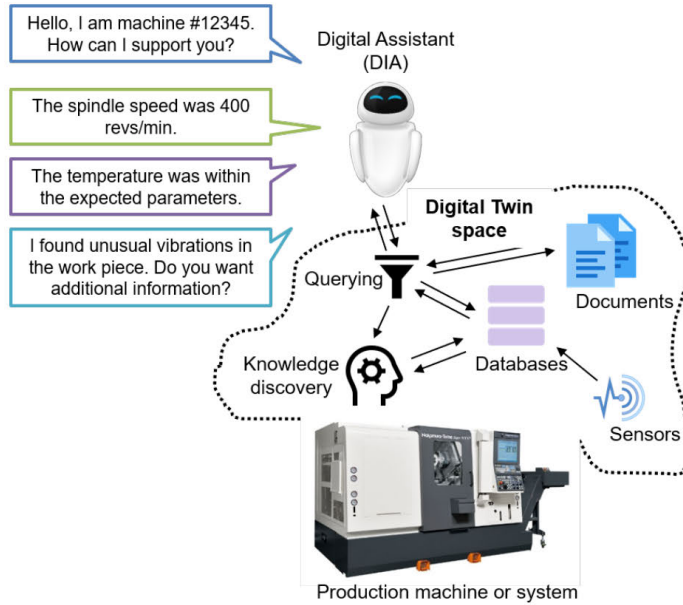


image / figure 1: Envisioned digital assistant (Stefan Wellsandt, BIBA)

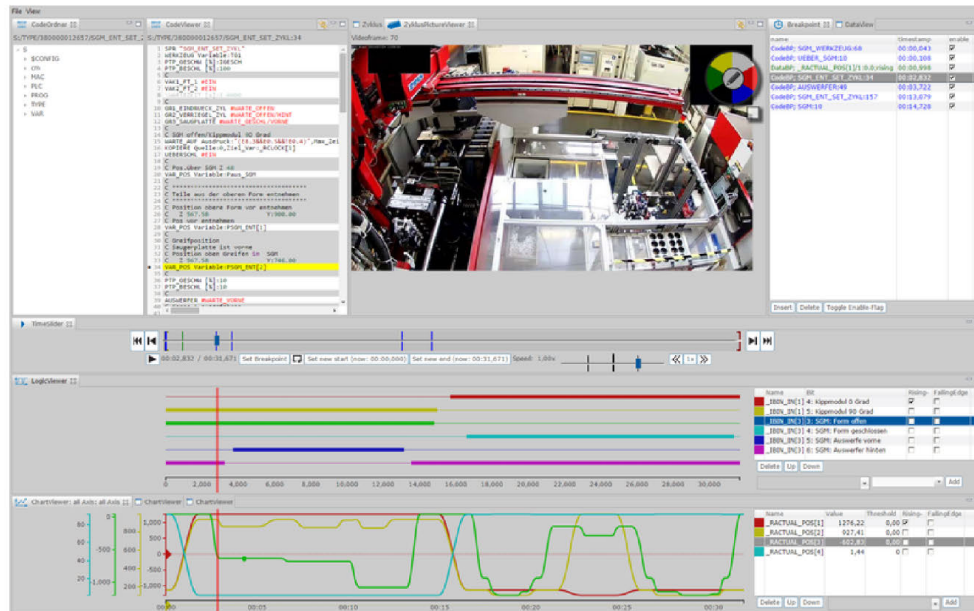


image / figure 2: Visualisation of data from an industrial machine (Doris Aschenbrenner, TU Delft)

**PROBLEM DEFINITION \*\***

Limit and define the scope and solution space of your project to one that is manageable within one Master Graduation Project of 30 EC (= 20 full time weeks or 100 working days) and clearly indicate what issue(s) should be addressed in this project.

Predictive maintenance systems can output a huge volume of data, most of which is unlikely to be relevant or understandable for regular machine operators. The primary challenge is to define what the operators need to know and how best to communicate it. This challenge can be divided further into three themes.

Technical

- How will the digital assistant identify the relevant information?
- How to support and guide a user on the various types of maintenance tasks (e.g. cleaning, inspection, lubrication) ?
- A voice assistant may have trouble understanding non-native speakers and operating in a noisy environment.

Human-centered

- Informational ergonomics and ease-of-use: as the operators are used to following paper check-lists and analog displays, they may struggle to interpret digital interactions and instructions.
- Work culture: the status-quo is to do fixed maintenance tasks at regular intervals as requested by your manager. Receiving customised instructions from a digital assistant may disrupt the current company culture.
- How to gather information about the user (e.g. location, job, availability) in an unobtrusive and simple manner.

Business

- Flexibility: each company may use different information systems, procedures and types of machines. It will be a challenge to create a product-service-system that can be readily implemented independent of these factors.
- How can the assistant be phased into the current workflow with minimal disruption?

**ASSIGNMENT \*\***

State in 2 or 3 sentences what you are going to research, design, create and / or generate, that will solve (part of) the issue(s) pointed out in "problem definition". Then illustrate this assignment by indicating what kind of solution you expect and / or aim to deliver, for instance: a product, a product-service combination, a strategy illustrated through product or product-service combination ideas, ... . In case of a Specialisation and/or Annotation, make sure the assignment reflects this/these.

In the context of predictive maintenance, research what the machine operators need to know and how best to communicate it. Based on this research, design and prototype a digital assistant capable of effectively supporting the machine operators.

The goal is to create a prototype of the digital assistant such that aspects of its functionality can be tested. These tests will be used to iteratively improve the design and act as a proof-of-concept. This will likely involve simulating its intelligence by programming some rule-based interactions. The digital assistant will be designed such that it can be integrated into an overarching product-service-system. For my project, I will consider the predictive maintenance model as a black-box that outputs current and predicted machine health information. Furthermore, I expect the system to be used as part of a machine maintenance service.

**PLANNING AND APPROACH \*\***

Include a Gantt Chart (replace the example below - more examples can be found in Manual 2) that shows the different phases of your project, deliverables you have in mind, meetings, and how you plan to spend your time. Please note that all activities should fit within the given net time of 30 EC = 20 full time weeks or 100 working days, and your planning should include a kick-off meeting, mid-term meeting, green light meeting and graduation ceremony. Illustrate your Gantt Chart by, for instance, explaining your approach, and please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any, for instance because of holidays or parallel activities.

start date 3 - 2 - 2020 end date 19 - 6 - 2020

	Fulltime during semester 2																			
Calendar week	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Project week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kick-off meeting																				
Reflect on progress and adjust plan if necessary																				
Presentation/reporting material																				
Research stakeholders and relevant technologies																				
Initiate contact with experts and other parties																				
Analyse and draw conclusions from research																				
Create simple prototypes to test at Stryker																				
Prep, execute and evaluate Stryker visit																				
Analyse and draw conclusion from Stryker visit																				
Create simple prototypes to test at Whirlpool																				
Prep, execute and evaluate Whirlpool visit																				
Test simple prototypes at Whirlpool																				
Analyse and draw conclusions on Whirlpool visit																				
Draw overall conclusions of the research phase																				
Synthesis research and test results into concepts																				
Create and test prototypes locally																				
Mid-term evaluation																				
Develop concepts further																				
Formalise ideas/concepts into a single concept																				
Green light meeting																				
Finalise design and recommendations																				
Finalise report and presentation																				
Graduation day																				
Project week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Through the project, I will use an iterative approach to rapidly research, prototype, test and evaluate my work. I will use the initial weeks to perform research, approach experts and perform context analyses. This will involve researching state of the art examples of predictive maintenance, digital assistants and human-machine-interactions and how these tackle the issues defined in the problem statement. Additionally, I will investigate what the current practices are at Whirlpool, Stryker and other representative companies, such that I can define their use scenarios; their specific needs, opportunities and risks. The results of this phase will be used to develop recommendations, requirements and wishes regarding the solutions, for example, the type of digital assistant, potential features and best practices. Throughout the process I will use prototypes and user tests to validate these findings.

The visits to Whirlpool (Poland) and Stryker (Italy) offer an ideal opportunity to test my designs and analyse the target environment. To facilitate faster development loops, I will also do research and test my solutions at local companies, in or near Delft. Furthermore, I will regularly reflect on my progress and correct my approach accordingly. Throughout the project I will prepare and present my results such that they can be used to support the IDE group within the DIAMOND consortium. Ultimately, I aim to have a well-supported design accompanied by a prototype.

I am planning on working full-time on my graduation project from February 3rd till June 19th.

**MOTIVATION AND PERSONAL AMBITIONS**

Explain why you set up this project, what competences you want to prove and learn. For example: acquired competences from your MSc programme, the elective semester, extra-curricular activities (etc.) and point out the competences you have yet developed. Optionally, describe which personal learning ambitions you explicitly want to address in this project, on top of the learning objectives of the Graduation Project, such as: in depth knowledge a on specific subject, broadening your competences or experimenting with a specific tool and/or methodology, ... . Stick to no more than five ambitions.

I am setting this graduation project up because I am interested in developing complex intelligent systems. I've recently completed an elective on Machine learning for intelligent products and would like to prove my worth in the development of an intelligent product.

As an IPD student, most of my curricular experience so far has been focused on the development of physical products, therefore, I would like to use this opportunity to expand my competencies to include cyber-physical product-service-systems. Furthermore, I would like to prove that I can effectively incorporate innovative technologies, such as voice-assistants and augmented reality into new products.

This project involves various stakeholders with different needs and expectations, I would like to show that I can develop a concept that balances these well and perform an effective context analysis. Lastly, I believe that the knowledge gained through this project has the potential to have a positive impact for all parties involved, whether it be knowledge, improved production or supporting personnel.

**FINAL COMMENTS**

In case your project brief needs final comments, please add any information you think is relevant.

## 9.2 USE CASE SELECTION

Use-case	D1: Fit with conversation design according to the guidelines of Google, IBM and Rasa	D2: Frequency of occurrence	D3: The positive impact of each usage (short and long-term value)	D4: Solves something that the users perceive as a problem	D5: How easily can the required technology and infrastructure be implemented	R1: The goal of my project: evaluate the potential for a conversational AI agent for maintenance tasks in the	R2: The DIAMOND project: currently focused on getting a working MVP to test basic features at Whirlpool and	R3: The proposed follow-up DIAMOND project on tacit knowledge capture	D-score (max. 10)	R-score (max. 10)	Total (max. 10)
Weight	0.4	0.1	0.1	0.2	0.2	0.6	0.3	0.1	0.5	0.5	1
Create maintenance reports	9	10	9	9	8	10	9	9	8.9	9.6	9.25
Check spare part availability	7	7	9	8	9	7	8	5	7.8	7.1	7.45
Check machine status	7	6	6	4	7	8	8	7	6.2	7.9	7.05
Retrieve error code meaning	9	7	9	9	9	7	9	6	8.8	7.5	8.15
Call/message/email colleagues	7	10	8	5	9	8	8	6	7.4	7.8	7.6
Guide user through preventative maintenance procedures	8	7	8	4	7	9	6	9	6.9	8.1	7.5
Check changes to machine parameters	7	5	8	9	7	7	8	7	7.3	7.3	7.3
Assign tasks to mechanics	6	9	7	4	6	6	5	3	6	5.4	5.7
Digital authorization/approval of maintenance tasks	5	10	8	5	6	6	5	3	6	5.4	5.7
Retrieve availability of technicians	6	8	7	5	6	6	5	3	6.1	5.4	5.75



## 9.3 UTTERANCE PAIR BRAINSTORMING

### 9.3.1 IN SCOPE AND ON TOPIC

Define utterance pairs based on use case, outcome statement and agent persona.

U: I don't understand what you need from me?

A: I need you to explain your actions and thoughts step-by-step.

U: [Procedural statement]?

A: okay / I see / Go on / Interesting

A: That was it?

U: Yes, I'm done. / No, I'm still fastening the bolts.

A: Please keep me updated on your progress

U: Sorry, I was just cleaning up the oil that I split.

U: I don't understand what you need from me?

A: I need you to explain your actions and thoughts step-by-step.

U: [Procedural statement]?

A: okay / I see / Go on / Interesting

U: Okay, that was everything

A: Thank you, I'll create the report for you now.

A: Can you take me through the steps you took?

U: First, I checked if the oil levels were high enough.

### 9.3.2 OUT OF SCOPE AND ON TOPIC

A: What time did you start?

U: 10 52 am / I just started / Just now / 10 minutes ago

A: Could you repeat that?

U: [user repeats the last statement]

U: [Procedural statement that mentioned the problem or machine]?

A: okay / Uh-huh / Okay, if I understand you correctly, you're having a problem with the [machine]. / Okay, that's unfortunate.

A: What is the problem?

U: The machine keeps going into an error state

A: Which tools did you use?

U: I used a flat-head screwdriver

A: Shall I submit the maintenance report?

U: Yes, please.

U: What can you do?

A: I can automatically create a maintenance report for you based on our conversation.

A: Which materials did you use?

U: I needed to replace temperature sensor and top up with lubricant oil.

U: Has this problem occurred before?

A: Yes, this machine had the same error code two weeks ago. Would you like to hear how they fixed it?

U: What does error code 1772 mean?

A: It means the main spindle axis is out of alignment

U: How can I solve this error code?

A: The machine manual says that you need to replace the timing belt.

A: Did you find the cause of the problem?

U: Not yet. / It seems like the dust covered the sensor.

A: could you explain that in more detail?

U: [user explains step in more detail]

---

### 9.3.3 IN SCOPE AND OFF TOPIC

U: Are you human?

A: No, I'm a virtual assistant that's specialized in maintenance reporting

U: You're a bitch

A: That's very rude. Please refrain from using that language.

---

### 9.3.4 OUT OF SCOPE AND OFF TOPIC

U: Tell me a joke?

A: [Tells joke]

## 9.4 EXPERIMENT PLAN

### 9.4.1 PROBLEM

Maintenance personnel at Stryker and Whirlpool have reported that paperwork takes a lot of time and is sometimes of poor quality. At Stryker, paperwork was reported to be the biggest “pain”. Personnel at both facilities expressed interest in something being done to improve the reporting process. In both facilities, maintenance technicians rely significantly on internal explicit and tacit (implicit) knowledge when troubleshooting and fixing problems. At Whirlpool, many of the procedures for fixing problems (e.g. error codes) are not documented anywhere and exist only in the minds of the maintenance technicians. Therefore, this knowledge is not readily accessible to other technicians and is lost when they leave.

### 9.4.2 HYPOTHESES

#### 9.4.2.1 HYPOTHESIS 1 (PRIMARY)

People performing maintenance work will provide higher quality reports (number of relevant details) when talking to a conversational agent while they perform the work as opposed to writing it down on paper after completing the work.

#### 9.4.2.2 HYPOTHESIS 2 (SECONDARY)

People performing maintenance work will finish their task quicker when they can create a report by talking to a conversational agent during the task as opposed to writing it down on paper afterwards.

#### 9.4.2.3 HYPOTHESIS 3 (SECONDARY)

People performing maintenance work will perceive the workload (mental demand,

physical demand, temporal demand, performance, effort, frustration) of their job as lower when they can create a report by talking to a conversational agent during the task as opposed to writing it down on paper afterwards because the conversational agent will guide them through the process while they work.

### 9.4.3 VARIABLES

#### 9.4.3.1 DEPENDENT VARIABLES

D1. Quality of reporting: number of relevant details (see Appendix 9.4.14).

D2. Time to complete the entire task (in the case of the control group, the time to complete the report will also be measured).

D3. Perceived workload: mental demand, physical demand, temporal demand, performance, effort, frustration (NASA-TLX).

#### 9.4.3.2 INDEPENDENT VARIABLES

The reporting mechanism:

3. Talking to a conversational agent that uses natural dialogue techniques (e.g. status requests, acknowledgements, continuers, interjections and follow-up questions, see Appendix 9.4.13.3 for more details).
4. Writing the report down on paper.

#### 9.4.3.3 CONTROL VARIABLES

Task

C1. All participants will be asked to replace the inner tube of the front tyre on a mountain bike.

C2. The same mountain bike will be used.

C3. The same tools will be available.

- C4. The same instructions will be given.
- C5. The tyre pressure will be the same at the start of the test (empty).
- C6. The position of the bike will be the same at the start of the test (turned upside down).

#### Participant

- C6. All participants must know how to replace an inner tube and have done so at least once before.
- C7. All participants must be physically capable of replacing a tyre without pain/discomfort.

#### Conversational agent

- C8. The researcher, who will Wizard of Oz the conversational agent, will not be watching the progress of the test participant, only listening.
- C9. The researcher, who will Wizard of Oz, will follow a few basic rules that best simulate the current capabilities of conversational agents (Appendix 9.4.13.3).

#### Environment

- C9. The test will occur in a quiet environment, without loud noises, disruptions or risk of rain/high winds/heat.

---

#### 9.4.4 TEST SETUP

The researcher will recruit participants who are willing to participate in the study, know how to replace a bicycle inner tube and have done so at least once before, are physically capable of replacing an inner tube without discomfort/pain, live within a 30 minute drive (by car) of the researchers home (Mexicohof, Delft), are willing to receive the researcher at their home (or travel to the researchers home

or another location) and agree to the conditions as defined in Appendix B. The target is to recruit 20 participants, the minimum is 10.

The researcher's personal mountain bike, tools and spare inner tube will be used during the experiment to guarantee consistency. The test will begin with the mountain bike upside down (see below in Figure 63)) and with a flat front tyre.



Figure 62 Orientation of bicycle

The camera will be placed such that it is possible to see what the participants are doing with their hands. It will be placed next to the rear tyre, pointing toward the front tyre and participant (see placement below).



Figure 63 Camera position

---

#### 9.4.5 MATERIALS

- 1x Mountain bike
- 1x Spare 26" inner tube
- 2x Tyre levers
- 1x Consent form
- 1x Preliminary questionnaire (Appendix 9.4.11)
- 1x Workload questionnaire (Appendix 9.4.9)
- 1x Unlined page for the report (Appendix 9.4.12)
- 1x Labels for tools and materials
- 1x Reference page for component names (Appendix 9.4.15)
- 1x Disinfectant spray
- 1x Disposable paper towel
- 1x Disposable gloves

---

##### 9.4.5.1 RECORDED DATA

1. Video recording of the participant performing the task.
2. Audio recording

- a. Experimental condition: full task and debrief.
  - b. Control condition: debrief only.
3. Time to perform task by stopwatch.
  4. Online questionnaire (includes consent form, preliminary questions, workload survey (NASA-TLX)).

---

##### 9.4.5.2 PILOT

A pilot test will be performed for each condition (control and experimental) to evaluate the procedure.

---

#### 9.4.6 TEST PLAN

1. Fill out consent + preliminary questionnaire
2. Standing next to the upturned mountain bicycle with a flat front tyre, tools and spare inner tube, the researcher will explain the task to the participant and presents the available tools and materials (see script in Appendix 9.4.13).
3. For the experimental condition, the research will explain the concept of the think-aloud method and give a demonstration. The researcher will then ask the participant to practice the think-aloud method for a simple task: check if the front brakes are working.
4. The researcher starts the camera, audio recording, starts a stopwatch, and indicates that the participant can start.
5. The participant performs the task.
  - a. For the control scenario, the researcher does not interact with the participant and waits for them to indicate that they are done.

- b. For the conversational agent scenario, the researcher listens to the participant, acknowledges their statements (using acknowledgements, continuers and interjections), and when deemed applicable, ask follow-up questions. The conversational agent does not answer any questions regarding the task.
6. If a tool breaks during the experiment, the researcher will pause the time and provide a replacement tool. If the participant gets stuck on the same task for 3 minutes or says that they are stuck, the researcher will provide a tip. If the participant punctures the inner tube, the researcher will instruct the participant to continue as if the tyre wasn't punctured.
7. When the participant indicates they are finished, the stopwatch and camera are stopped.
8. The researcher will check if the task has been completed successfully: the wheel rotates freely, and the tyre has been pumped up sufficiently to ride on.
9. The researcher asks the participant to fill in a NASA-TLX perceived workload survey (Appendix 9.4.9).
10. The researcher asks the participant about their experience. The goal is to gather qualitative results about challenges they had during the test and what their impressions were of the reporting process. The researcher will ask: How do you feel it went? What are your impressions of this reporting process? Anything else you wish to share?

11. The audio recording is stopped.

---

#### 9.4.7 CORONA MEASURES

- 1.5m separation will always be maintained.
- The bicycle, materials and tools will be disinfected after every experiment.
- The experiment won't be performed in an area where third parties may walk by in close proximity.

---

#### 9.4.8 ANALYSIS PLAN

1. The duration of the entire task and subtasks will be compared and tested for significance between groups.
2. The overall and individual workload scores (mental demand, physical demand, temporal demand, performance, effort, frustration) will be compared and tested for significance between groups.
3. The recordings of the tasks will be transcribed, coded and scored on reporting quality according to the process outlined in Appendix 9.4.14. Results will be compared and tested for significance between groups.
4. A qualitative analysis will be performed on the video recordings and feedback from the participants. The key factors that will be observed are how the interactions with the conversational agent affect the task progression, for example, do the participants stop their work when talking to the conversational agent. Additionally, the videos will be used to identify if the users used suboptimal techniques or if they got stuck on a subtask.

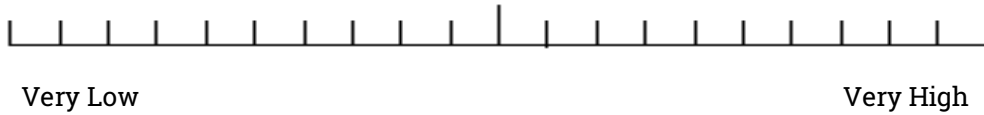


---

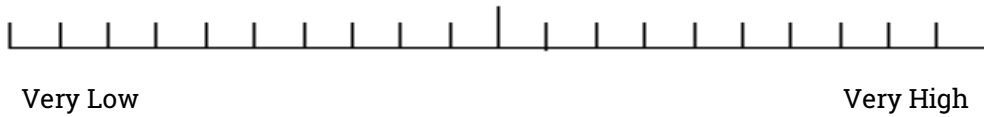
#### 9.4.9 NASA-TLX

Please answer the following questions for the entire task (repair and reporting). Indicate your answer by circling one of the vertical lines.

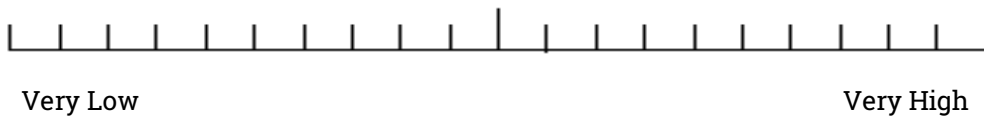
How mentally demanding was the task?



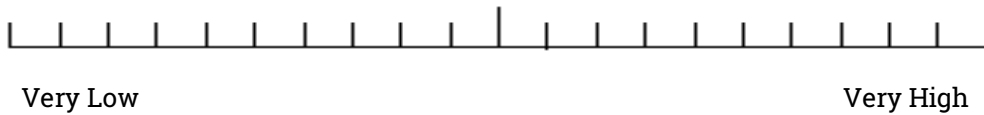
How physically demanding was the task?



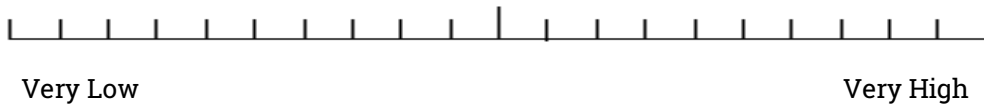
How hurried or rushed was the pace of the task?



How successful were you in accomplishing what you were asked to do?



How hard did you have to work to accomplish your level of performance?



How insecure, discouraged, irritated, stressed and annoyed were you?



---

#### 9.4.10 CONSENT FORM

In order to contribute you must agree the consent form below:

Consent for participation in the evaluation experiment of 'Bicycle repair and reporting'.

1. I volunteer to participate in a product evaluation experiment conducted by Delft University of Technology Delft, Faculty of Industrial Design Engineering.
2. My participation in this evaluation experiment is voluntary. I understand that I will not be paid for my participation. I may withdraw and discontinue participation at any time without penalty.
3. Participation involves answering personal background questions, replacing a bicycle inner tube, creating a report and a debrief. This process will take around 20 minutes.
4. I understand that the results and gathered (anonymous) data (video, audio, questionnaire responses and feedback) from the experiment will be saved and stored for scientific processing.
5. I understand that no specific participant related performance measurement is transferred to any external source.
6. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.

For further information or questions, please contact:

Samuel Kernan Freire ([s.kernanfreire@student.tudelft.nl](mailto:s.kernanfreire@student.tudelft.nl))

Date        \_\_\_-\_\_\_-\_\_\_\_\_

Name        \_\_\_\_\_

Signature    \_\_\_\_\_

---

#### 9.4.11 PRELIMINARY QUESTIONS

Q1. What is your age?

18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65+

Q2. What is your gender?

Male | Female | Non-binary | I prefer not to say

Q3. How many times have you changed a bicycle inner tube in your life?

1-2 | 3-5 | 6-10 | 11-20 | 21+

Q4. How long ago was the last time you changed a bicycle inner tube?

0-4 weeks | 1-3 months | 3-6 months | 6-12 months | 1-2 years | 2 years +

Q5. I can change a bicycle inner tube without any problem.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
-------------------	----------	-------------------	---------	----------------	-------	----------------

Q6. I can think aloud while performing a task.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
-------------------	----------	-------------------	---------	----------------	-------	----------------

Q7. I can write a clear/understandable report on a task I recently completed.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
-------------------	----------	-------------------	---------	----------------	-------	----------------

Q8. I can communicate effectively through spoken English.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
-------------------	----------	-------------------	---------	----------------	-------	----------------

Q9. I can communicate effectively through written English.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
-------------------	----------	-------------------	---------	----------------	-------	----------------

Q10. I am wordy/verbose (I use more words than necessary to convey something).

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
-------------------	----------	-------------------	---------	----------------	-------	----------------

---

#### 9.4.12 MAINTENANCE REPORT

Description of actions (what, how and why)

--

---

## 9.4.13 SCRIPTS

---

### 9.4.13.1 START CONTROL

---

The researcher asks the participant to fill in the consent form and preliminary questionnaire.

"After the experiment, I need you to fill in a short questionnaire. The experiment will be filmed and recorded for scientific research. Do you have any questions now?"

"This bike's front tyre is flat, I need you to replace inner tube, inflate it and mount the wheel on the bike again. You don't need to flip the bike back up. Afterwards, I need you to write a report of your work. The purpose of this report is so that a novice colleague can learn from your experience and understand what you did, how you did it and why you did it that way. \*Researcher places report on table\* (appendix D). These are the tools you need: tyre levers, a bicycle pump and a spare inner tube. The researcher asks if the participant is familiar with the type of bicycle pump, valve type, quick-release and disc brakes on the bike. If not, the researcher will explain how they work. "

"There is no time limit but please complete the tasks in an efficient manner. Once the experiment starts, I will not be able to answer any of your questions. Unless, of course, you wish to stop or are stuck. Do you have any questions now?"

The researcher asks the participant to fill in consent form and the preliminary questionnaire.

The researcher repeats the task instructions and checks if the participant has any more questions.

---

### 9.4.13.2 START EXPERIMENTAL

---

The researcher asks the participant to fill in the consent form and preliminary questionnaire.

"After the experiment, I need you to fill in a short questionnaire. The experiment will be filmed and recorded for scientific research. Do you have any questions now?"

"This bike's front tyre is flat, I need you to replace inner tube, inflate it and mount the wheel on the bike again. You don't need to flip the bike back up. I also need you to create a report of your work. If you describe what you are doing while you work, by thinking aloud, I will automatically create the report for you. I may also ask you some questions, but I won't be watching you work, only listening. The purpose of this report is so that a novice colleague can learn from your experience and understand what you did, how you did it and why you did it that way. These are the tools you need: tyre levers, a bicycle pump and a spare inner tube."

The researcher asks if the participant is familiar with the type of bicycle pump, valve type, quick-release and disc brakes on the bike. If not, the researcher will explain how they work.

"There is no time limit but please complete the tasks in an efficient manner. Once the experiment starts, I will not be able to answer any of your questions. Unless, of course, you wish to stop or are stuck. Do you have any questions now?"

The researcher asks if the participant is familiar with the think aloud method. If not, the researcher demonstrates the think aloud method for “what type of bike is this?” and asks the participant to practice the think-aloud method by checking if the front brakes work.

The researcher repeats the task instructions and checks if the participant has any more questions.

---

#### 9.4.13.3 CONVERSATIONAL AGENT SCRIPT

- Respond whenever the participant finishes an utterance (e.g. Okay, nice, mhm, uh-huh).
- Respond with interjections or acknowledgements, if the participant indicates that they have completed a subtask (e.g. Okay, cool, nice, great).
- If the participant has been silent for 15 – 20 seconds, ask them: “how are you doing?” or “how is it going?”. If they are silent again within the next minute, wait 25-30 seconds before asking again.
- If they participant uses a pronoun, but it is not clear which object they are referring to, the agent will ask them which object they are referring to.
- If the participant indicates that they’ve completed a subtask but hasn’t mentioned how or with what tool, the agent will ask for further details (e.g. did you use a tool? How did you do that?).



#### 9.4.14 QUALITY OF REPORTING MEASURE

To define the number of relevant details in each report, they will be coded using the scheme below.

(sub-)Utterance type	Definition
PROCEDURAL	Descriptions of tasks (e.g. "next, remove the cover")
TASK STATE	Diagnosis of object or situation (e.g. "the tyre is flat", "the quick release is very tight").
REFERENTIAL	Identifying or localising task objects (e.g. "the big disc in the middle")
INTERNAL STATE or INTERNAL STATE	Explanation, reasoning, knowledge, etc. (e.g. "because, it might be slippery")

The sub-utterances are then awarded points based on the following criteria:

- Each unique "PROCEDURAL" is awarded 1 point.
- Each unique "INTERNAL STATE" is awarded 1 point.
- Each unique "TASK STATE" is awarded 1 point.
- Each unique "REFERENTIAL" within the same "PROCEDURAL", "INTERNAL STATE" or "TASK STATE" is awarded 1 point.
- A point is only awarded if the sub-utterance is relevant to understanding how the task was completed, why the participant completed it in that way or is a piece of knowledge that could help someone else perform the task.

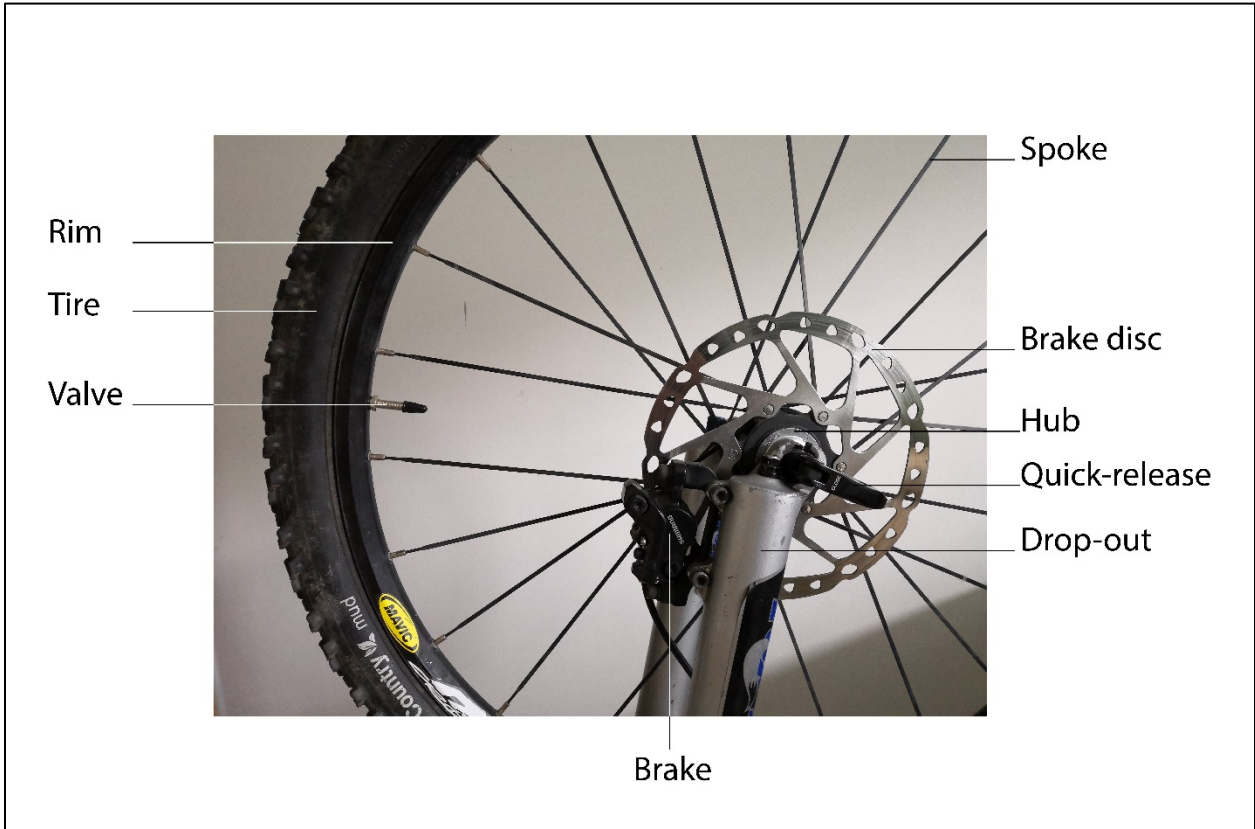
Example coding from one of the pilots (control 1):

Unique (sub)utterance	
Problem: flat inner tube	2
Solution: replace it with a new inner tube	3
With the help of tyre levers, the out tire was removed.	3
Then the entire wheel was taken off	2
so that the inner tube could be removed.	2
For this the valve has to be loosened.	2
A new inner tube is placed within the outer tire.	3
The valve is placed in the correct hole.	3
Outer tire is put back on with the tire levers.	3
Wheel is set back on with the tire levers.	3
Wheel is set back on the bicycle.	3
Made sure that the brake is set back correctly.	2

Set everything fixed.	1
Used bicycle pump to get the inner tube inflated again at a pressure of a bit over 2 bar.	4
Wheel is spinning without difficulties.	2
Inner tube is still inflated	2
Total	40

Based on the total scores, the reports will be ranked as good (top 33.3%), satisfactory (middle 33.3%) and poor (bottom 33.3%).

9.4.15 REFERENCE IMAGE FOR COMPONENT NAMES



## 9.5 DIAMOND SURVEY FOR WHIRLPOOL AND STRYKER

Titles:

1. Maintenance survey
2. Introduction
3. General questions

Introduction:

The purpose of this survey is to learn more about the maintenance situation at your factory and potential areas for improvement. It includes maintenance-related scenarios that we observed during the factory visits. Depending on your role within the factory, a different set of questions will be asked (don't be alarmed if you seem to skip some questions, this is normal). It should take 5 - 10 minutes. The results are anonymous and will only be used within the context of this project.

Question descriptions:

1. Please indicate how strongly you agree with the following statements. You may elaborate your answer using the comment section.
2. Please rank the scenarios by the order in which you would like them to be fixed (drag and drop or use the arrows to rearrange the statements into the desired order)

Questions:

1. Where do you work?
2. What is your role?
3. Sometimes, the meanings of error codes are not immediately obvious
4. When I consult old maintenance reports, they are often poorly documented
5. Filling out paperwork for maintenance tasks is time consuming
6. When performing maintenance work, I can't access all the necessary information (e.g. troubleshooting steps) while at the machine
7. I spend a lot of time waiting for my work to be approved/authorized
8. I go to the spare-parts warehouse only to find out that what I needed was not in stock
9. When fixing a problem, sometimes I get stuck and need tips
10. When performing planned maintenance tasks, I sometimes forget some steps
11. Sometimes I forget my planned maintenance tasks
12. It takes a long time to get a response from quality control
13. It is difficult to determine why someone else changed the machine parameters
14. Digitizing maintenance reports is time consuming

15. When making the maintenance schedule, I miss information regarding machine status'
16. When making the maintenance schedule, I need to gather information from many different sources
17. When assigning maintenance tasks, it takes time to figure out who is available
18. When assigning maintenance tasks, I need to physically find the maintenance technician
19. When performing planned maintenance tasks, I can't access all the necessary information (e.g. troubleshooting steps) while at the machine
20. Do you have any other comments or problematic scenarios you wish to mention?

Answers:

1. Other (please specify)
2. Maintenance technician
3. Maintenance coordinator
4. Machine operator
5. Setter
6. Line manager / shift leader
7. Strongly agree
8. Agree
9. Neither agree nor disagree
10. Disagree
11. Strongly disagree
12. Not applicable
13. Comments

## DIAMOND Umfrage für Stryker

Titel:

1. Umfrage zur Wartung
2. Einleitung
3. Allgemeine Fragen

Einleitung:

Diese Umfrage dient den DIAMOND Projektpartnern dazu die Wartungsaufgaben in Ihrer Einrichtung besser zu verstehen. Zunächst bitten wir Sie darum anzugeben in welcher Rolle Sie in Ihrer Einrichtung arbeiten. Abhängig von dieser Angabe folgen dann weitere Fragen. Die Umfrage dauert 5-10 Minuten. Die Ergebnisse sind anonym (wir fragen nicht nach Ihrem Namen) und wird nur im Zusammenhang mit dem DIAMOND Projekt verwendet.

Beschreibung der Fragen:

1. Bitte geben Sie an, wie stark sie den folgenden Aussagen zustimmen. Sie können zusätzlich das Feld für Kommentare nutzen, um Ihre Angaben genauer zu erklären.
2. Bitte sortieren sie die folgenden Szenarien durch Ziehen mit der Maus. Die Reihenfolge gibt an wie wichtig die Lösung der Szenarien aus Ihrer Sicht ist (je weiter oben, desto wichtiger).

Fragen:

1.	Wo arbeiten Sie?
2.	Welche Rolle haben Sie in Ihrer Einrichtung?
3.	Manchmal ist die Bedeutung von Fehlercodes nicht klar.
4.	Wenn ich abgeschlossene Wartungsprotokolle lese, sind diese oft schlecht dokumentiert.
5.	Das Ausfüllen der Papiere zur Dokumentation von Wartungsaufgaben kostet viel Zeit.
6.	Wenn ich eine Maschine warte habe ich nicht auf alle benötigten Informationen (z.B. Anleitungen zur Fehlerbehebung) zugriff.
7.	Ich muss lange warten bis meine Arbeit freigegeben wurde.
8.	Oft finde ich erst im Ersatzteillager heraus, dass ein benötigtes Ersatzteil nicht auf Lager ist.
9.	Manchmal komme ich beim Lösen eines Problems nicht weiter und brauche Tipps.
10.	Manchmal vergesse ich Arbeitsschritte bei einer geplanten Wartungsaufgabe.
11.	Manchmal vergesse ich meine geplanten Wartungsaufgaben.



12.	Es dauert sehr lange bis ich eine Rückmeldung von der Qualitätssicherung bekomme.
13.	Es ist schwer zu verstehen warum jemand anderes die Einstellungen an einer Maschine verändert hat.
14.	Wartungsberichte zu digitalisieren kostet viel Zeit.
15.	Wenn ich den Wartungsplan erstelle fehlen mir Informationen zum Zustand der Maschinen.
16.	Wenn ich den Wartungsplan erstelle muss ich Informationen aus vielen verschiedenen Quellen sammeln.
17.	Wenn ich Wartungsaufgaben zuweise dauert es lange bis ich weiß welcher Mitarbeiter verfügbar ist.
18.	Wenn ich Wartungsaufgaben zuweise muss ich zum Wartungstechniker laufen.
19.	Wenn ich an einer Maschine geplante Wartungsaufgaben erledige, kann ich nicht auf alle notwendigen Informationen (z.B. Anleitungen zu Fehlerbehebung) zugreifen.
20.	Gibt es noch weitere Kommentare oder Situationen, die Sie uns mitteilen möchten?

Antworten:

1. Andere (bitte angeben)
2. Wartungstechniker
3. Wartungskoordinator
4. Maschinenführer
5. Einrichter
6. Linienmanager / Schichtleiter

Stimme absolut zu

Stimme zu

Teils / teils

Stimme nicht zu

Stimme absolut nicht zu

Nicht anwendbar

Kommentare

## 9.6 EVALUATION OF MYCROFT

Mycroft install on ubuntu server 18.04 virtual machine. Followed the standard instructions [here](#).

The audio worked straight away but the mic failed after registering the Mycroft wake command. Plugged in my C270 webcam and restarted the client. It works now. Also used alsamixer to adjust volumes (don't know if that had any effect).

---

### 9.6.1 FIRST IMPRESSIONS

- There is a noticeable pause between the wake phrase and the beginning of recording. Perhaps this is due to the speed of the virtual machine or is a setting. After being used to Alexa, where you can say Alexa, "query", without a pause. If you want to be sure that Mycroft hears your entire query, you need to wait until the beep/feedback that he is listening. Similarly, there is a delay between the mycroft's text response and his speech response.
- The speech-to-text is working ok, misunderstands more frequently than Alexa.
  - Mycroft misunderstands on for off very frequently.
- Mycroft's voice is quite robotic and doesn't pronounce things very well.
- Nice that you can install skills with your voice.
- It sometimes stops listening half-way through a query

#### Thoughts

- A robotic voice/mispronunciation may not be an issue if operators aren't familiar with better systems (Alexa etc).
- The S2T component is close to the speed of Alexa, perhaps an auditory response is not necessary. Use the advantage of S2T and a screen to leverage the best of both worlds. S2T is useful to skip endless menus and the intelligent component can package the info, add extra things and suggestions.
- It can take a very long time to communicate basic information, especially numbers. More room for misunderstanding.

#### Request speed versus Alexa (measured from end of command)

- "Play music by Ennio Morricone"
  - Alexa responds in 1.5 seconds and starts the music at 4 seconds
  - Mycroft text-to-speech takes 1.5 seconds but it takes 3 seconds before he says he will look into it, 11 seconds for him to confirm query and 13 seconds to start playing
- "What's 2 x 5?"
  - Alexa: 1.5 to respond
  - Mycroft: S2T in 2-3 seconds, response at 5.5 sec.
- "what's the temperature now?"
  - Alexa: 1.5 to respond
  - Mycroft: 2 for S2T, 6 for response (misunderstood first time)
- "What's the capital of the Democratic People's republic of Korea"?

- Alexa: 1.5 secs
- Mycroft: (misunderstand first time, second time only got half the message, third time even less, fourth also)

---

## 9.6.2 PICROFT

Picroft is a Raspberry Pi operating system with Mycroft preinstalled and configured.


- Using American male voice (default)
- The voice sounds more realistic, less robotic
- When respond with a long phrase, MyCroft speeds up his speech -> nice feature
- The quality of the 3.5mm output on the Pi is poor -> ordered a USB soundcard from [Amazon.de](#)
- [Kickstarter for Mark II](#)
- Need better microphone array for realistic performance in noisy environment?


## 9.7 RASA EVALUATION


### Talk to your bot (Interactive Learning)


action\_session\_start

The chat interface shows a sequence of messages between a user and a bot. The bot's messages are in light blue bubbles, and the user's messages are in dark blue bubbles. The bot's messages include a greeting, a request for user ID, a request for technician name, a request for job description, and a confirmation of the user ID, technician name, and job description. The user's messages include a greeting, a request to create a workorder, the user ID '8080', the technician name 'hugh', and the job description 'not urgent'.


hi  greet


 Hello! I am a workorder assistant! How can I help?

i want to create a workorder  create\_workorder


 what is your user id?


slot["requested\_slot": "user\_id"]

8080  inform["user\_id": "8080"]


 Who would you like to assign the work order to?


slot["requested\_slot": "assigned\_technician"]

to hugh  inform["assigned\_technician": "hugh"]

 How urgent is this job?

slot["requested\_slot": "urgency"]

not urgent  inform["urgency": "urgent"]

 please provide a description of the job

slot["requested\_slot": "job\_description"]

Figure 64 Initial testing with Rasa X and interactive learning

# Deployment

## Level 1: Run locally

- Simplest to setup
- Does not support interactive learning (conversations need to be exported later)
- Less stable / slower
- You can use Ngrok to expose your assistant to the web and allow for user testing.

- 
- Installed locally but Rasa X GUI was buggy. The rest worked.

## Level 2: Run on server/docker

- Supports interactive learning
- More complex to setup
- Suitable for testing
- Not suitable for production (not very scalable)

- 
- Installed in docker on windows but haven't got it working yet.

- Installed in docker on an Ubuntu 18.04 server. The various components startup but Rasa X doesn't function.

## Level 3: Run on server/Kubernetes

- Rasa has "one-line deploy script"
- Suitable for production (scalable)

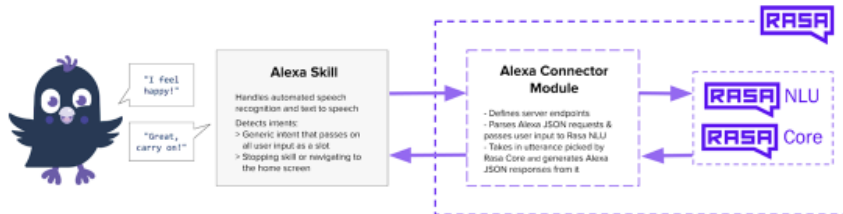
- 
- Tried the one-line deploy script on a Ubuntu 18.04 server but it never finishes "deploying".

# Rasa and user testing

Rasa supports connectors to common chat platforms like Slack, mattermost, Facebook messenger etc.

There is **no official support for voice platforms** like Google Assistant or Alexa but there are unofficial guides. The guide for Google is two years old and may no longer work well. The Alexa guide is new.

In both cases, users could test the prototype by downloading the respective Alexa or Google assistant app and installing a "skill". Alternatively, they could use an Amazon Echo or Google home device.



## Voice vs text user testing

	Voice	Text
Pros	<ul style="list-style-type: none"><li>- More representative</li><li>- Evaluate the interplay of STT and NLU</li></ul>	<ul style="list-style-type: none"><li>- Provide insight for conversation flow, potential questions, the logic behind the assistant</li><li>- Bypass limitations of STT</li><li>- Easier to implement and test</li></ul>
Cons	<ul style="list-style-type: none"><li>- More complex to setup</li><li>- Less stable (not official supported)</li><li>- User testing will be more involved</li></ul>	<ul style="list-style-type: none"><li>- Users may communicate differently by text versus speech</li><li>- Poor for simulating "handsfree" situations</li><li>- The validity of the results may be questionable</li></ul>



## 9.8 CODE

Full code here: [https://github.com/skernanfreire/rasa\\_voice\\_assistant](https://github.com/skernanfreire/rasa_voice_assistant)

---

### 9.8.1 DEPENDENCIES

```
from typing import Dict, Text, Any, List, Union, Optional
```

```
from rasa_sdk import Action
```

```
from rasa_sdk.events import SlotSet
```

```
from time import localtime, strftime
```

```
from rasa_sdk.executor import CollectingDispatcher
```

```
from rasa_sdk import Tracker
```

```
import smtplib
```

```
from decouple import config
```

---

### 9.8.2 CLASSES FOR HANDLING PROCEDURAL STATEMENTS AND CONTEXT

```
class ActionStoreFullDescription(Action):
```

```
    def name(self):
```

```
        return "action_store_full_description"
```

```
    def run(self, dispatcher, tracker, domain):
```

```
        prev_number = tracker.get_slot('step_number')
```

```
        step_number = prev_number + 1
```

```

step_num = str(step_number)

print(step_num)

message = tracker.latest_message['text'] #get the entire raw utterance from the user

time = strftime("%H:%M:%S")

if step_num=='1':
    date = strftime("%a, %d %b %Y")
    start_time = time+" on "+date
else:
    start_time = tracker.get_slot('start_time')

try: #to catch exceptions if no entity is found
    primary_components = list(tracker.get_latest_entity_values(entity_type='component',entity_
role='primary'))
    if len(primary_components) != 0:
        for x in primary_components: #searches for a non-
it primary component, otherwise primary component = it
            if x != "it":
                latest_component = x
                break
            else:
                latest_component = 'it'
        else:
            latest_component = 'none'
        print("no primary component found")
except Exception:
    latest_component = 'none'

```

```

print("no primary component found")

try:
    secondary_components=next(tracker.get_latest_entity_values(entity_type='component',entity_role='secondary'))
except Exception:
    print("no secondary component found")

try:
    actions_group1 = list(tracker.get_latest_entity_values(entity_type='action',entity_group='1'))
    latest_action = actions_group1[0]
    if len(actions_group1)==2: #if action group 1 has two parts, add it
        latest_action = latest_action+" "+actions_group1[1]
    primary_action = latest_action
    if len(actions_group1)==0:
        print("no action found")
except Exception:
    print("no action found")

try:
    actions_group2 = list(tracker.get_latest_entity_values(entity_type='action',entity_group='2'))
    if latest_action is None:
        latest_action = ""
    for x in actions_group2:
        latest_action = latest_action + " "+x
    if len(actions_group2)==0:

```

```

    print("no action group 2 found")

except Exception:
    print("no action group 2 found")

#latest_component = next(tracker.get_latest_entity_values("component"),None)

#latest_action = next(tracker.get_latest_entity_values("action"),None)

latest_tool = next(tracker.get_latest_entity_values("tool"),None)

#curraction = tracker.get_slot('action')

#currtool = tracker.get_slot('tool')

#currcomponent = tracker.get_slot('component')

prevmessage = tracker.get_slot('full_description')

prevsteps = tracker.get_slot('steps_list')

prevaction = tracker.get_slot('action_list')

prevtool = tracker.get_slot('tool_list')

prevcomponent = tracker.get_slot('component_list')

last_pcomp = tracker.get_slot('prev_component')

flag = False #reset the 'it' flag (for conversation flow predictions)

print(last_pcomp)

if 'latest_action' in locals():
    actions = prevaction+" (" +step_num+" )"+latest_action
else:
    actions = prevaction

    latest_action = "none"

```

```

if latest_component != "none":
    components = prevcomponent+" (+step_num+)"
    new_components = ','.join(primary_components)

    if 'secondary_components' in locals():
        new_components = new_components+", "+secondary_components

    components = components+" "+new_components
else:
    components = prevcomponent
    new_components = 'none'

if latest_tool is not None:
    tools = prevtool+" (+step_num+)" +latest_tool+" "
else:
    tools = prevtool
    latest_tool = "none"
    print("no tool found")

if ('latest_component' in locals()) and (latest_component == "it"):
    if 'primary_action' in locals():
        if ('pump' in primary_action) or ('inflate' in primary_action):
            latest_component = 'inner tube'
            new_components = new_components+" (assumed that 'it': inner tube)"
    elif (last_pcomp!='it') and (last_pcomp!='none'):

```

```

        new_components = new_components+" (assumed that 'it': "+last_pcomp+")"
elif (last_pcomp == "it") or (last_pcomp == "none"):
    if 'primary_action' in locals():
        check_message = "what did you "+primary_action+"?"
    else:
        check_message = "what do you mean by it?"
    dispatcher.utter_message(check_message)
    flag = True

print((" "+step_num+" ") +time+" Action(s): "+latest_action+", Component(s): "\
    +new_components+", Tool(s): "+latest_tool)

if message is not None:
    message = prevmessage+" (" +step_num+" ") +message
    steps = prevsteps+" (" +step_num+" ") +time+" Action(s): "+latest_action+", Component(s): "\
        +new_components+", Tool(s): "+latest_tool

return [SlotSet('full_description', message),SlotSet('pronoun_flag', flag),\
        SlotSet('step_number', step_number),SlotSet('steps_list', steps),SlotSet('action_list', actions), \
        SlotSet('component_list', components),SlotSet('tool_list', tools),SlotSet('prev_component', la
test_component),\
        SlotSet('start_time',start_time), SlotSet('current_time',time)]

class ActionStoreCorrection(Action):

    def name(self):

```



```

return "action_store_correction"

def run(self dispatcher, tracker, domain):

    latest_component = next(tracker.get_latest_entity_values("component"),None)

    prevmessage = tracker.get_slot('full_description')

    prevsteps = tracker.get_slot('steps_list')

    prevcomponent = tracker.get_slot('component_list')

    message = tracker.latest_message['text']

    flag = False

    if latest_component is not None:

        components = prevcomponent+" "+latest_component

    else:

        components = prevcomponent+" "+message

        latest_component = 'none'

    if message is not None:

        message = prevmessage+" (correction: "+message+" )"

        steps = prevsteps+" (it = "+ latest_component+" )"

    return [SlotSet('pronoun_flag', flag), SlotSet('full_description', message),\

            SlotSet('steps_list', steps),SlotSet('component_list', components),SlotSet('prev_component', la
test_component)]

```

---

### 9.8.3 CLASSES FOR SENDING THE REPORT BY EMAIL

```
class ActionSendmail(Action):
```

```
    def name(self) -> Text:
```

```
        return "action_send_mail"
```

```
    def run(self, dispatcher: CollectingDispatcher,
```

```
            tracker: Tracker,
```

```
            domain: Dict[Text, Any]) -> List[Dict[Text, Any]]:
```

```
        SENDER_EMAIL = config('EMAIL') #get sensitive info from local environment variables (in .env
file)
```

```
        SENDER_PWD = config('PWD')
```

```
        RECIPIENT_EMAIL = config('RECIPIENT')
```

```
        #Subject = tracker.get_slot('Subject')
```

```
        subject = 'Generated maintenance report'
```

```
        full_description = tracker.get_slot('full_description')
```

```
        steps = tracker.get_slot('steps_list')
```

```
        tools = tracker.get_slot('tool_list')
```

```
        actions = tracker.get_slot('action_list')
```

```
        components = tracker.get_slot('component_list')
```

```
        start_time = tracker.get_slot('start_time')
```

```
        time = tracker.get_slot('current_time')
```

```
        preamble = "Dear Maintenance Tech, \n\n I've collected the following information from our con
versation."
```

```

#Recipient = tracker.get_slot('Recipient') #if you want the user to specify the recipient

server = smtplib.SMTP_SSL('smtp.gmail.com', 465) #connect to smtp server
server.login(SENDER_EMAIL, SENDER_PWD)

msg = "Subject: {} \n\n{} \n\nStart time: {} \nEnd time: {} \n\nExtracted steps{} \n\nRaw user input
{} \n\nTools{} \
\n\nComponents{} \n\nActions{} \n\nKind regards,\n\n The maintenance bot" \
.format(subject,preamble,start_time,time,steps,full_description,tools,components,actions) #c
reating the message

server.sendmail(          #send the email
SENDER_EMAIL,
RECIPIENT_EMAIL,
msg)
server.quit()

dispatcher.utter_message(" Email sent! ")

return []

```

---

#### 9.8.4 CLASSES FOR STATUS REQUEST FEATURE

```
class ActionSetReminder(Action):
```

```
    """Schedules a reminder"""
```

```
    def name(self) -> Text:
```

```
        return "action_set_reminder"
```

```
    async def run(
```

```
        self,
```

```
        dispatcher: CollectingDispatcher,
```

```
        tracker: Tracker,
```

```
        domain: Dict[Text, Any],
```

```
    ) -> List[Dict[Text, Any]]:
```

```
        date = datetime.datetime.now() + datetime.timedelta(seconds=20)
```

```
        reminder = ReminderScheduled(
```

```
            "EXTERNAL_reminder",
```

```
            trigger_date_time=date,
```

```
            name="my_reminder",
```

```
            kill_on_user_message=False,
```

```
        )
```

```
        return [reminder]
```

```
class ForgetReminders(Action):  
    """Cancels all reminders."""  
  
    def name(self) -> Text:  
        return "action_forget_reminders"  
  
    async def run(  
        self, dispatcher, tracker: Tracker, domain: Dict[Text, Any]  
    ) -> List[Dict[Text, Any]]:  
  
        # Cancel all reminders  
        return [ReminderCancelled()]
```

## 9.9 EXPERIMENT RESULTS

Variable	p-value	significance	Mean control group	Mean experimental group
Time to change a tyre	0.861608	ns	553	540.1667
Total time	6.64E-05	(**)	1119.417	540.1667
NASA-TLX total score	0.273543	ns	51.65344	46.69312
Mental demand	0.249772	ns	8.583333	6.583333
Physical demand	0.016669	(*)	11.41667	6.666667
Temporary demand	0.139685	ns	10.83333	13.08333
Performance	0.332902	ns	16.08333	17.66667
Effort	0.281951	ns	10.5	8.5
Frustration	0.494992	ns	7.666667	6.333333
Total (sub)utterances	4.06E-07	(**)	57.75	131.9167
Procedural (sub)utterances	3.36E-08	(**)	22	44.41667
Task status (sub)utterances	9.47E-05	(**)	0.333333	6.5
Referential (sub)utterances	1.05E-05	(**)	34.08333	69.5
Internal state (sub)utterances	2.01E-05	(**)	1.333333	11.5

**Table 24 Dependent variables by group (p<0.01 = (\*\*), p<0.05 = (\*), p<0.1 = (+))**



	age	gender	last_tyre	tyres_in_life	tyre_no_prob	think_aloud	report_skill	spoken_eng	written_eng	verbose	
tyre_time	0.69	1	0.822	0.006	0.036	0.065	0.929	0.963	0.521	0.413	0.439
time	0.17	4	0.325	0.179	0.055	0.070	0.792	0.270	0.535	0.481	0.831
tlx_total	0.15	4	0.707	0.204	0.311	0.034	0.223	0.569	0.667	0.613	0.686
tlx_ment	0.31	9	0.544	0.251	0.283	0.007	0.646	0.921	0.656	0.535	0.984
tlx_phys	0.20	1	0.763	0.037	0.028	0.023	0.580	0.445	0.732	0.973	0.424
tlx_pace	0.72	9	0.081	0.473	0.454	0.360	0.239	0.608	0.161	0.232	0.594
tlx_succes	0.97	4	0.165	0.877	0.890	0.139	0.179	0.133	0.482	0.789	0.607
tlx_work	0.56	6	0.425	0.974	0.693	0.699	0.497	0.596	0.360	0.445	0.628
tlx_discomfort	0.70	5	0.763	0.156	0.673	0.083	0.014	0.964	0.793	0.737	0.409
report_pro	0.30	3	0.222	0.363	0.754	0.504	0.737	0.054	0.707	0.974	0.849
report_task	0.30	6	0.904	0.116	0.866	0.553	0.599	0.235	0.425	0.236	0.568
report_ref	0.44	7	0.096	0.462	0.931	0.453	0.777	0.126	0.754	0.515	0.869
report_int	0.46	7	0.263	0.286	0.601	0.941	0.809	0.507	0.828	0.885	0.897
report_raw	0.39	7	0.168	0.285	0.839	0.599	0.987	0.104	0.856	0.579	0.838

Table 25 Correlations between factors

## 9.10 PROTOTYPE TESTS

### 9.10.1 ENTITY EXTRACTION

Input	Primary components	Secondary components	Primary Action	Secondary Action	Tools	Correct (max)	False	
and then I use the tyre levers to just release it from the entire rim	It	Entire rim Release	Use		Tyre levers	3.5(4)	1	
I'm just going to pop out the outer tyre over the rim here	Outer tyre	Rim	Pop out over			3(3)		
just turn it a little bit loose	It		Turn a little bit loose			2(2)		
releasing the quick release	Quick release		Releasing			NA	NA	
turn the lever to tighten it	Lever it		Turn to tighten		lever	2(3)	1	
and grab the black cap.	Black cap		Grab black			1(2)	1	
Rotate it into place.	It		Rotate into place			1.5(2)		
you can loosen the valve	Valve		Loosen			2(2)		
check that you didn't snag the inner tube anywhere	Inner tube	Perimeter	Check			3(3)		

along the perimeter								
you put it back as before	It		Put back			2(2)		
get the pump and place it near the bike		Bike it	get	Place near	pump	2(4)	1	
Totals	8.5(10)	4(4)	9(10)	0(1)	0.5(2)	22(27)	4(27)	
Per cent	85% (10)	100% (4)	90% (10)	0% (1)	25% (2)	81.5%	14.8%	

Table 26 Entity extraction test scores

## 9.10.2 "IT" COMPONENT TRACKING AND CLARIFICATION

### Participant 23

u: releasing the quick-release yeah
a: Thanks, please carry on [incorrect intent]
u: just turn it a little bit loose and then the other side as well.
a: ok (components: "it", "side")
u: Turn it a little bit loose otherwise it won't come free
a: ok (assumed the "it" = "side")

The agent assumed that "side" was the primary component in the second utterance. This is not entirely wrong as it was one of two components referenced, however, it is still missing the true meaning (quick release nut).

u: yeah, that's always a bit tricky because you don't want to put them too far away but you kind of have to because it needs to pop out at one point so ...
a: ok
u: so... and then I use the tyre levers to ... to just release it from the entire rim
a: ok (components: "it", "entire rim", "release")

A human might have asked what "it" was, however, the agent assumed that the "entire rim" or "release" was the primary component, thereby missing that it was the outer tyre.

### P15

u: at which point I can push the outer tube over the rim
a: yeah
u: I push it all the way over so I don't need any tools
a: ok (assumed "it" = "outer tube")

u: I put the pump on the on the valve
a: yeah
u: and I put a little bit of air in there
a: yeah
u: make sure it's not twisted
a: ok, what did you make sure?

### Participant 14

u: make sure there's no twists or bends
a: yeah
u: there we go... give it one more pump so that it sits inside
a: Yeah (components: it, it, inside)
u: then start putting the outer tyre back
a: ok

A human could have guessed that "it" was the inner tube based on the "pump" action, however, the agent assumed that "inside" was the primary component.

u: put these aside
a: yeah
u: find the valve
a: yeah
u: and you pull it out
a: yeah (assumed "it" = "valve")
u: and then I'll pull it out of the tyre
a: yeah (components = it, tyre)

A human may have guessed or asked what "it" was ("inner tube").

### P11

01:38 u: now that the tube is kind of loose you can take the valve off
01:41 a: yeah

01:43 u: so take the cap off and the holding nut
01:47 a: ok
01:48 u: place it aside so that I can still see where I put it
01:51 a: ok (assumed that "it" = "cap")

It tracked half of the primary components (cap and holding nut) correctly.

u: and we have our tyre spinning again and now we just have to pump it up
a: ok (assumed "it" = "inner tube")
u: pump it up... take the seal off again.
a: Ok (assumed "it" = "inner tube")
u: check the state of our valve
a: yeah

P12

u: I don't put the black one yet, I can already release the valve a bit
a: yeah
u: I take the bike pump
a: yeah
u: and I put the small connector on it
a: yeah, what did you put on?

P8

u: and again, I take off the cap and the ring that's on there.
a: yeah
u: then I will push the new valve through the hole in the rim
a: ok
u: where it's supposed to go
a: yeah (assumed that "it" = "valve")

P5

u: then I screw it, tighten it
--------------------------------

138

A: ok (components = "screw it")
u: I tightened it too much
a: ok (assumed that "it" = "screw it")

The agent incorrectly identifies screw it as a component and therefore, assumes that "it" in the next utterance means "screw it". Had the experiment participant uttered "then I tighten it", the agent would have correctly queried what they meant by it, as demonstrated below.

P5 Modified:

u: then I tighten it
a: ok
u: I tightened it too much
a: ok, what did you tightened?

u: mmm, then i grab the spare inner tube
a: yeah
u: and i wrap it, and i make sure I know where the vent is
a: ok (assumed that "it" = "spare inner tube")

P1

u: I'm going to put the inner tyre in the outer tyre and start with the valve again
a: yeah
u: you need to take the ring off
a: ok
u: I push it through
a: yeah (assumed "it" = "ring")

60% correct

### 9.10.3 INTENT CLASSIFICATION

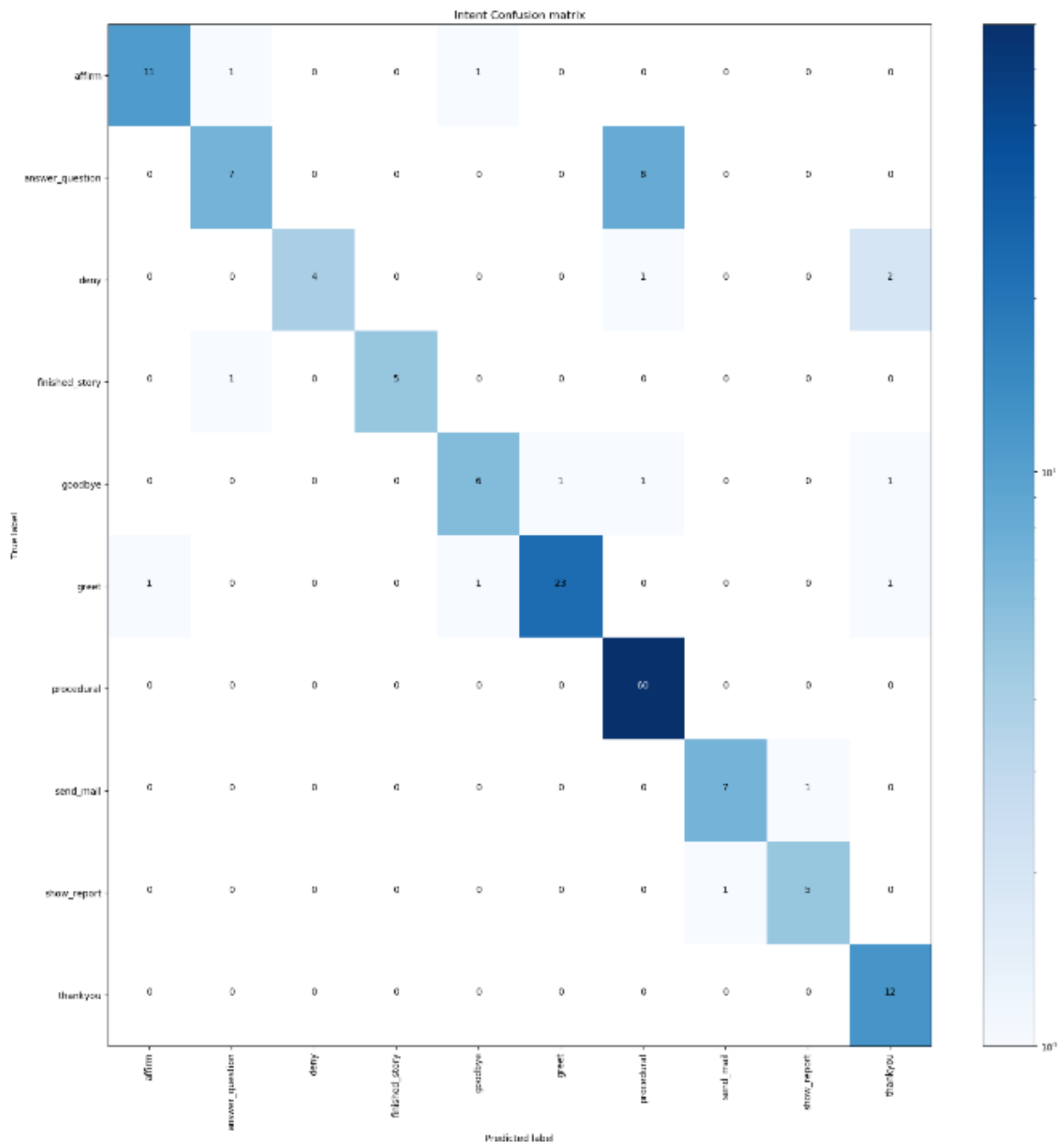


Figure 65 Confusion matrix for intent classification